FORENSIC EPIDEMIOLOGY

Principles and Practice

Edited by

Michael D. Freeman and Maurice P. Zeegers

FORENSIC EPIDEMIOLOGY

This page intentionally left blank

FORENSIC EPIDEMIOLOGY

PRINCIPLES AND PRACTICE

Edited by

MICHAEL D. FREEMAN
MAURICE P. ZEEGERS

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
50 Hampshire Street, 5th Floor, Cambridge, MA, 02139, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-404584-2

Publisher: Sara Tenney
Acquisitions Editor: Elizabeth Brown
Editorial Project Manager: Joslyn Chaiprasert-Paguio
Production Project Manager: Lisa Jones
Designer: Matthew Limbert

Typeset by TNQ Books and Journals
www.tnq.co.in

Printed and bound in China

Dedication

Our goal in writing this book was to contribute to the balanced delivery of justice by improving the accuracy of causal determinations presented in legal settings. We dedicate this work to those who strive toward the same goal and who further the advancement of forensic epidemiology.

This page intentionally left blank

Contents

7. Death Investigation

S.A. KOEHLER

8. Injury Biomechanics

D. KIESER, D. CARR, M. JERMY, A. MABBOTT, AND J. KIESER

9. Biomechanical, Epidemiologic, and Forensic Considerations of Pediatric Head Injuries

W.E. LEE III, AND J.D. LLOYD

10. Survival Analysis

H.D. TOLLEY, J.M. BARNES, AND M.D. FREEMAN

III

APPLICATIONS OF FORENSIC EPIDEMIOLOGY

15. Criminal Investigation

M.D. FREEMAN, AND F. FRANKLIN

This page intentionally left blank

List of Contributors

J.M. Barnes Brigham Young University, Provo, UT, United States

M.J.L. Bours Maastricht University, Maastricht, The Netherlands

A. Broadbent University of Johannesburg, Johannesburg, South Africa

D. Carr Cranfield University at the Defence Academy of the United Kingdom, Shrivenham, United Kingdom

A. Eriksson Umeå University, Umeå, Sweden

M. Faure Erasmus University Rotterdam, Rotterdam, The Netherlands; Maastricht University, Maastricht, The Netherlands

F. Franklin Oregon Health & Science University-Portland State University, School of Public Health, Portland, OR, United States; Morehouse School of Medicine, Atlanta, GA, United States; Thomas R. Kline School of Law, Drexel University, Philadelphia, PA, United States

M.D. Freeman Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

Alani Golanski Weitz & Luxenberg, P.C., New York, NY, United States

S.C. Gold Rutgers University-Newark, Newark, NJ, United States

M.D. Green Wake Forest University, Winston-Salem, NC, United States

M. Jermy University of Canterbury, Christchurch, New Zealand

D. Kieser University of Otago, Christchurch, Canterbury, New Zealand

J. Kieser University of Otago, Dunedin, Otago, New Zealand

S.A. Koehler Forensic Medical Investigations, Pittsburgh, PA, United States

W.E. Lee III University of South Florida, Tampa, FL, United States

J.D. Lloyd Brains, Inc., San Antonio, FL, United States

A. Mabbott Cranfield University at the Defence Academy of the United Kingdom, Shrivenham, United Kingdom

J. Sanders University of Houston, Houston, TX, United States

H.D. Tolley Brigham Young University, Provo, UT, United States

L. Visscher Erasmus University Rotterdam, Rotterdam, The Netherlands

M.P. Zeegers Maastricht University, Maastricht, The Netherlands

This page intentionally left blank

Introduction

M.D. Freeman, M.P. Zeegers

And now here is my secret, a very simple secret: It is only with the heart that one can see rightly; what is essential is invisible to the eye. ***Antoine de Saint-Exupéry (1900–1944)***

In the Sign of Four, Sherlock Holmes famously commented to Dr. Watson that "Once you have eliminated the impossible, whatever remains, no matter how improbable, must be the truth." He was adamant about this, repeating it two more times in the same story, and then again in The Adventure of the Bruce-Partington Plans and The Adventure of the Blanched Soldier. Arthur Conan Doyle wrote the Sign of Four in 1890, 42 years after John Snow first employed epidemiologic methods to investigate the cause of a cholera outbreak in London in 1848, noting the fact that the disease was distributed in the same pattern as the common water supply to the homes made it likely that the source of the disease was the water. Snow was criticized at the time for suggesting that the cause of the outbreak was contaminated water, as there was no evidence that there was anything in the water that was producing the disease (such evidence was discovered 6 years later, in 1854 by Filipo Pacini). Snow's theory regarding the cause of the cholera outbreak was based on an inference drawn from an observed association, rather than a direct observation; no one could "see" that the water was the cause of the disease.

More than 100 years after Snow's discovery, Austin Bradford Hill gave a lecture to the Royal Society of Medicine in 1965 in which he outlined nine viewpoints by which an observed association could be evaluated for causality. In this famous lecture, immortalized now as the "Hill criteria" (a characterization that has persisted to the present day despite Hill's protestations that he did not want his viewpoints turned into a checklist) he made it clear that, despite advances in science and medicine, we are still vexed by questions of how best to approach investigations of causality. But he also noted that a strong association is usually the best evidence of a causal relationship, using as an example the fact that chimney sweeps sustain scrotal cancer 200 times more often than other occupations stands as powerful evidence of a causal relationship between sweeping chimneys and cancer of the scrotum.

Holmes, Snow, and Hill were all describing the same fundamental truth; that a cause cannot be seen, and for this reason it must be inferred.

It is widely accepted that unreliable evidence is a significant problem in the forensic sciences generally and in forensic medicine specifically. One of the explanations for this phenomenon is the lack of validated and reliable standards and methods for common tasks performed in a forensic setting. Determination of the cause of injury or disease is a pivotal issue in virtually all criminal and civil actions, and one that is often vigorously contested. Despite this fact, there are no published standards regarding what constitutes scientifically valid evidence of causation, nor a systematic means of quantifying and weighing evidence of

causation. The single largest explanation for this state of affairs, as noted above, is the fact that *causation cannot be observed*, and thus conclusions of causation are not observations but rather inferences based on a presumed degree of association between an exposure and injury. The lack of a generally accepted systematic approach to what is essentially an exercise in probabilistic reasoning results in the reliance by lay fact finders (ie, judge and jury) on what is often speculative and unreliable evidence regarding causation.

Outside of a forensic or legal setting, causal evaluations are most commonly performed in a medical setting by physicians. This is because the determination of the diagnosis of the condition for which the cause is sought is the responsibility of the physician, rather than because clinicians are routinely trained in causal methodology (they are not).

Courts expect clinicians to be able to "see" a cause as readily as they can "see" a diagnosis, despite the fact that the process of arriving at a diagnosis is entirely different than determining a cause. A child can see a broken femur on an X-ray and make the diagnosis of a fracture, a fact that anyone who sees the X-ray would have to agree with. We do not qualify the diagnosis of a fracture as being present on a "more likely than not basis," the standard for most expert testimony, because it is undeniable. If, however, we are informed that the individual with the broken leg was in two car crashes that happened one right after the other; the first one involving a far side crash and the second a frontal impact, how do we determine which crash was the *cause* of the fracture? Such a determination is not based on observation, nor is it based on diagnosis. The clinician might have sufficient experience or knowledge to know that femur fractures are much more *common* in frontal crashes than side impacts, but this is an inference rather than an observation.

Determination of causality is an important part of the practice of forensic pathology, where the primary purpose of the postmortem examination is to determine the manner and cause of death. In this setting, when there is a high degree of association between the diagnosis and the cause of the death (for example, a gunshot wound to the head), the determination of causation is easily made as a matter of common sense. This is because the strength of the association, like Hill's example of the chimney sweeps, tends to rule out competing causes. In the example of a gunshot wound to the head, causation is obvious because such injuries are nearly always fatal, and the probability of an alternative cause of death coinciding with the time of the gunshot wound is exceedingly low in most circumstances. In contrast, the cause of death in a patient with pneumonia, an 80% blockage of the left coronary artery, and who received an intravenous injection of a narcotic 30 min before going into respiratory arrest, cannot be determined as a matter of common sense. In such a circumstance the only causal analysis that can yield valid and repeatable results is the assessment and comparison of the *risk* of death associated with each of the plausible causes.

Risk is a population-based metric defined as the probability or chance that an event will occur in the future, based on what has happened in the past. The field of study from which risk is estimated is epidemiology. Epidemiology is broadly described as the branch of medicine dedicated to the study of the cause of disease and injury in populations. Epidemiologic study examines the relationships between exposures and outcomes (and vice versa) and describes the results in terms of frequencies, rates, and probabilities. Epidemiologists use standardized methods to describe disease and injury occurrence in specified populations in order to identify populations that are at higher risk than others and to evaluate factors that may account for the risk differences. In assessing causes, epidemiologists consider

components of cause both individually and collectively, as well as which components are necessary (required) for causation, and the components that are sufficient for causation.

Although a primary function of epidemiology is to understand the causes of disease and injury in populations, epidemiology is largely silent about methods for investigating the cause of disease and injury in individuals (specific causation). Despite this fact, when there is a low degree of association between an injury observed in an individual and a suspected cause, or there are multiple competing causes, an evidence-based causal assessment requires the quantification and comparison of risks acting on the individual at the time of the injury.

The discipline of forensic epidemiology (FE), essentially a hybrid of principles and practices common to both forensic medicine and epidemiology, is directed at filling the gap between clinical judgment and epidemiologic data and methods in the evaluation of both general and specific causation in civil and criminal matters. The purpose of an FE causal analysis is to provide an evidence-based foundation for an opinion regarding the probability of causation, suitable for presentation in a medicolegal setting.

As questions pertaining to risk and causality are pervasive in virtually all aspects of civil and criminal litigation, the applications of FE methods are potentially quite broad. In this book we have endeavored to give the reader an overview of concepts and methods of FE and provide illustrations of the methods with case studies and examples. We have not attempted to describe all applications of FE, nor have we written a primer on epidemiologic or biostatistical methods, as there are many well-written texts that do this already. The goal of this book is to introduce the reader to FE, rather than make the novice an expert in the field.

This book is organized into three major sections, with Chapters 1 through 5 describing the principles of FE practice, including a historical perspective on how epidemiologic evidence has been used in courts, and methods used in FE investigations. In Chapters 6 through 10 are descriptions of non-epidemiologic forensic disciplines that in some cases are incorporated in an FE investigation. In Chapters 11 through 15 we provide the reader with examples of how FE methods have been applied in a wide variety of circumstances as a means of assessing causal relationships.

We hope that the reader finds as much intrigue and enjoyment in this text as we experienced in putting it together.

This page intentionally left blank

PART I

PRINCIPLES OF FORENSIC EPIDEMIOLOGY

This page intentionally left blank

CHAPTER

1

Legal Considerations of Forensic Applications of Epidemiology in the United States

Alani Golanski

Weitz & Luxenberg, P.C., New York, NY, United States

OUTLINE

HISTORICAL CONTEXT OF THE *FRYE* STANDARD

By the final decades of the 20th century, the two-page opinion in *Frye v. United States*,[1] issued in 1923, largely defined the way in which state and federal courts treated the question of the admissibility of scientific and technical proof. The specific issue in that second-degree murder case was whether exculpatory expert testimony about the result of *Frye*'s "systolic

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00001-X

blood pressure deception" test (precursor to the polygraph machine) ought to have been allowed. In ruling it inadmissible, the Court of Appeals for the District of Columbia Circuit set forth its famous standard:

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.[2]

Before *Frye*, most courts handled the problem of whether to trust experts by asking only about the expert's qualifications and whether the subject matter of the testimony was beyond the range of knowledge of the average juror. As the New Hampshire high court in *Jones v. Tucker* stated in 1860: "When a witness is offered as an expert, three questions necessarily arise: (1) Is the subject concerning which he is to testify, one upon which the opinion of an expert can be received? (2) What are the qualifications necessary to entitle a witness to testify as an expert? (3) Has the witness those qualifications?"[3]

There were no reported decisions involving epidemiological testimony prior to *Frye*. Indeed, in the late 19th century and well into the 20th century, technical, nonscientific testimony, more so than the scientific, tended to generate admissibility controversies. The expert in *Jones v. Tucker* was a surveyor, and the court ruled his testimony admissible on the issue of whether "the marks upon the corners" of a property were "ancient."[4] Discussing the range of experts whose "superior skill in relation to" various subjects qualified them to testify, the court referred to handwriting experts and "house joiners" (in modern times, construction contractors).

Soon after the Civil War, in a strict liability action arising from injuries caused by a "mischievous deer" that defendant permitted to roam on his property, the US Supreme Court ruled that expert testimony was not needed on the issue of the offending animal's character or disposition, which was "common knowledge."[5] The court's survey of the range of expert testimony with which it was then familiar was telling:

> Medical men, for example, may give their opinions not only as to the state of a patient they may have visited, or as to the cause of the death of a person whose body they have examined, or as to the nature of the instruments which caused the wounds they have examined, but also in cases where they have not themselves seen the patient, and have only heard the symptoms and particulars of his state detailed by other witnesses at the trial. . . . It must appear, of course, that the witness is qualified to speak to the point of inquiry, whether it respects a patented invention, a question in chemistry, insurance, shipping, seamanship, foreign law, or of the habits of animals, whether feroe naturoe or domestic.[6]

That pre-*Frye* controversies, under the two-part special knowledge/qualifications test, would tend to involve witnesses offering technical rather than scientific testimony is understandable. On the one hand, scientific knowledge is more clearly beyond the ken of the average juror; technical understanding fell closer to the divide between the ordinary citizen's common knowledge and the specialist's enhanced awareness. On the other hand, the marketplace could more readily attest to the expertise of scientific specialists, since they would be cloaked with an established professional status; their success in their profession corroborated their expertise.[7]

Even in the case of scientific experts, however, the issue of the expert's qualifications was seen in relation to the nature and quality of competing expertise in the case. Courts viewed

even scientific expert testimony with some degree of suspicion, and required that any such testimony pass through the prism of necessity. For example, in the medical malpractice case of *Martin v. Courtney*,[8] the Minnesota Supreme Court overturned the jury's verdict in favor of the plaintiff, unfavorably comparing the qualifications of the plaintiff's expert, a surgeon who had graduated from medical school less than 5 years before trial, to those of the defendant physician, who had 15 years of experience and had been chief surgeon at his hospital.[9]

In general, however, courts in the pre-*Frye* period attempted to answer the question posed in *Jones v. Tucker* as did the Supreme Court of Errors of Connecticut in *Taylor v. Town of Monroe*[10]:

> The true test of the admissibility of such testimony is not whether the subject matter is common or uncommon, or whether many persons or few have some knowledge of it, but whether the witnesses offered as experts have any peculiar knowledge or experience, not common to the world, which renders their opinions founded on such knowledge or experience any aid to the court or jury in determining the questions at issue.[11]

Such was law's outlook in the decades preceding *Frye*. Nor have these twin requisites to the admission of expert testimony—(1) the expert's qualifications to render testimony in an area of particular expertise and (2) the need for any such testimony to assist the fact finder—lost their urgency in modern times. The explanation for this begins with the constitutional right to a trial by jury possessed by criminal defendants and civil litigants. The Sixth Amendment provides, inter alia, that criminal defendants have the right to "an impartial jury of the state and district wherein the crime shall have been committed…. ."[12] Similarly, the Seventh Amendment guarantees trial by a jury of one's peers in "suits at common law, where the value in controversy shall exceed twenty dollars…."[13] State constitutions afford local litigants analogous rights.[14]

Against the backdrop of the jury trial rights afforded by the US Constitution as well as parallel state constitutional provisions, the legal system has always recognized the importance of adopting procedures and evidentiary doctrines that avoid intruding upon, or usurping, the jury's role. Witnesses presented to the jury as "experts" come cloaked in auras of authority[15] and reliability.[16] Therefore, expert testimony in areas accessible to the ordinary juror's common sense and everyday experience would tend to muscle out the jury's fact-finding function. For a similar reason, an expert witness's proffer of a legal conclusion has traditionally been deemed impermissible.[17]

PRELUDE TO THE FEDERAL RULES OF EVIDENCE

The federal court's decision in *Frye* went largely unnoticed for several decades. When courts cited to *Frye*, they tended to do so on the narrow issue in that case, namely, the admissibility of "lie detector" evidence,[18] or of evidence concerning the findings associated with similar mechanical devices.[19]

Also during the decades following *Frye*, various attempts at drafting a comprehensive set of evidentiary rules were unsuccessful. These included efforts by the Commonwealth Fund Committee in 1927 and by the American Law Institute in 1942.[20] Nevertheless, in 1961, the Judicial Conference of the United States approved the creation of an Advisory

Committee to report on the feasibility of creating uniform rules of evidence for the federal courts, and Chief Justice Earl Warren appointed the Committee's eight members. Ultimately, Congress adopted the Federal Rules of Evidence in 1975.

As the Committee worked to formulate the uniform rules, some of the *dicta* in the case law suggested *Frye*'s limited application in practice and revealed questionable judicial treatment of expert testimony and methodologies notwithstanding *Frye*. In *United States v. McNeil*,[21] for instance, a psychiatrist testified that the child abuse defendant had "partially recovered" his sanity and would not be dangerous upon release from St. Elizabeth's Hospital if he refrained from consuming alcohol.[22] In his separate, concurring opinion, the Chief Judge on the appellate court, David L. Bazelon, construed the trial judge's improper expressions of hostility and skepticism toward the expert witness to reflect the judge's "personal feeling of revulsion" about the underlying misconduct.[23]

Two years later, in *United States v. Leazer*,[24] Judge Bazelon wrote a similar concurring opinion, noting that, "[i]n the course of his questioning, the trial judge made a number of pointed comments and observations which could have been taken by the jury to reflect some skepticism of the insanity defense and the manner in which it was presented."[25] Significantly, however, in seeking to inject a heightened level of objectivity into judicial oversight and review of expert proofs, even Judge Bazelon did not then take *Frye* to set down the standard for examining experts or their methodologies, but referenced *Frye* as exemplary of a validity test.

Although by the early 1970s *Frye* had not weaned the judiciary from its reliance upon subjective, folk psychological intuitions in certain circumstances, courts showed significant respect for epidemiology.[26] Epidemiology had become a well-established science in the study of infectious diseases and a significant institutional component of public health programs at prestigious universities. Following World War II, and certainly by the 1970s, the courts were primed to welcome epidemiology's paradigm shift to risk factor causation assessment and its attendant role in toxic product-related litigation arising from cancers and other chronic diseases.[27]

The *Frye* standard, however, was viewed as conservative and restrictive, for the most part. For example, the general acceptance approach usually resulted in the continuing preclusion from evidence of often-exonerative lie detector tests.[28] In 1976, the California Supreme Court noted that "[s]ome criticism has been directed at the *Frye* standard, primarily on the ground that the test is too conservative, often resulting in the prevention of the admission of relevant evidence…. [W]e are satisfied that there is ample justification for the exercise of considerable judicial caution in the acceptance of evidence developed by new scientific techniques."[29]

Further criticisms of *Frye* flowed from the case's language and application. Whether a principle or discovery qualified as "scientific" in the first instance was not always self-evident.[30] Even if clearly "scientific," *Frye* appeared to impose "a protracted waiting period that valid scientific evidence and techniques must endure before gaining legal acceptance."[31] Nor is it typically clear, however, when a scientific principle has, in fact, been "generally accepted."[32] Inconsistencies also characterized the resolution of the issue of what precisely must be generally accepted to pass muster under *Frye*, the theory informing the expert's analysis, the technique applying that theory, or both?[33] Moreover, because general acceptance "in the particular field" is at issue, admissibility determinations may well skew in favor of less rigorous fields.[34]

ENTER THE FEDERAL RULES OF EVIDENCE

When the Federal Rules of Evidence became effective in 1975, Rule 702 provided:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

This Rule therefore patterned the two-pronged pre-*Frye* standard for admitting expert testimony into evidence, asking whether the witness was "qualified" and, if so, whether the testimony would "assist" the trier of fact.

Under the new Federal Rules of Evidence, however, Rule 702 did not constitute a stand-alone admissibility rule, but part of a larger scheme. Rule 703 introduced the idea of a modern "reasonable reliance" restriction, providing:

> The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to him at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence.

Although Rule 703 departed from the "general acceptance" standard of *Frye*, the Advisory Committee notes explained that "the court is called upon to reject testimony that is based upon premises lacking any significant support and acceptance within the scientific community…."[35] Nevertheless, leading commentators, most notably Professor Charles McCormick, viewed the expert testimony rules as repudiating, or the least superseding, the *Frye* standard.[36]

In addition to Rules 702 and 703, the general evidentiary principle that evidence may be precluded if it carries a danger of unfair prejudice that outweighs its probative value, embodied at Rule 403, also applied to expert testimony as an additional safeguard. This was a somewhat redundant exercise, because the inadequacy of an expert witness's qualifications, or the unlikelihood that such testimony would assist the fact finder, is typically what would render it unduly prejudicial.[37] In all events, continuing a long-standing standard of review, rulings under Rules 403 and 702 were deemed to fall within the sound discretion of the trial court, hence not to be reversed unless an appellate court determined the outcome was "manifestly erroneous."[38]

THE JUDICIAL DIVIDE INTERPRETING THE FEDERAL RULES OF EVIDENCE

Courts divided over the nature and extent of the judicial scrutiny of proffered expert testimony warranted by the Federal Rules of Evidence. One "liberal" line of cases held that, under the new rules, expert opinion testimony was admissible if the expert was generally qualified in her field, if there was some factual basis for her opinion, and if the facts or data underlying the expert's methodology met a threshold criterion of reliability. Another "restrictive" line of cases stood for the proposition that trial judges should more rigorously

scrutinize the expert's proffer and independently assess the quality and appropriateness of the data, methodology, and conclusions, to determine whether these are actually reliable.[39]

In his opinion in the *Agent Orange* litigation, Judge Jack B. Weinstein explained the divergent approaches that had developed as of the late 1980s in a slightly different way:

> Courts have adopted two general approaches to Rule 703: one restrictive, one liberal…. The more restrictive view requires the trial court to determine not only whether the data are of a type reasonably relied upon by experts in the field, but also whether the underlying data are untrustworthy for hearsay or other reasons. The more liberal view… allows the expert to base an opinion on data of the type reasonably relied upon by experts in the field without separately determining the trustworthiness of the particular data involved.[40]

The District of Columbia Circuit Court of Appeal's decision in *Ferebee v. Chevron Chemical Co.*,[41] was frequently cited as illustrative of the liberal interpretation of the new rules governing the admissibility of expert testimony.[42] Richard Ferebee was an agricultural worker employed by the US Department of Agriculture, whose estate alleged that he had contracted pulmonary fibrosis as a result of long-term skin exposure to dilute solutions of paraquat, a herbicide distributed in the United States solely by Chevron.

At the trial in *Ferebee*, Chevron maintained that paraquat is only acutely toxic, and hence that any injuries caused by the herbicide would occur within a very short time of exposure. In the case, however, Ferebee had not experienced any of the symptoms of pulmonary fibrosis until 10 months after last spraying the product.

The District of Columbia Circuit instructed that appellate review of a jury's verdict must be "very limited" and free from assessing witness credibility. The court explained that "[j]udges, both trial and appellate, have no special competence to resolve the complex and refractory causal issues raised by the attempt to link low-level exposure to toxic chemicals with human disease. On questions such as these, which stand at the frontier of current medical and epidemiological inquiry, if experts are willing to testify that such a link exists, it is for the jury to decide whether to credit such testimony."[43]

The restrictive approach, by contrast, arises from a double-edged skepticism: that concerning the validity of unfamiliar scientific methods or conclusions; and that reflecting a lack of confidence in the jury's capabilities to sift through technical proofs and testimony. Judge Weinstein's rulings in the *Agent Orange* litigation played a central role in articulating and implementing this restrictive approach during the pre-*Daubert* period.

The plaintiffs in *Agent Orange* were veterans who alleged that exposure to the herbicide during the Vietnam War had caused their cancers, respiratory and skin disorders, and other injuries. Judge Weinstein engaged in detailed scrutiny of the experts' analyses to determine whether there was a legally sustainable link between the veterans' dioxin exposures and their injuries. After all, he said, "[w]e are not dealing here with exposure of workers in a factory or laboratory to dioxin in concentrated amounts where the probative force of the evidence on causality may be substantial."[44] So he carefully examined, inter alia, "[t]he most intensive Agent Orange study of effects on veterans published to date," the Air Force-sponsored *An Epidemiologic Investigation of Health Effects in Air Force Personnel Following Exposure to Herbicides*, issued in 1984. The court "reevaluat[ed]" the study, noting, for

example, "that Air Force personnel who generally have clean clothes and showers available at the end of their missions are in a far different situation from a marine or soldier in the jungle who may be drinking contaminated water and living under primitive conditions in sprayed areas."[45]

In another *Agent Orange* opinion issued the following year, Judge Weinstein rested on both the established status of epidemiology and the traditional skepticism of other areas of expertise that might inform a causal analysis. He stated, for instance, that "[a] number of sound epidemiological studies have been conducted on the health effects of exposure to Agent Orange. These are the only useful studies having any bearing on causation. All the other data supplied by the parties rests on surmise and inapposite extrapolations from animal studies and industrial accidents."[46]

In this vein, for instance, Judge Weinstein rejected the proffered testimony of Dr. Barry M. Singer, a physician board certified in internal medicine, hematology, and oncology. In his expert affidavit, Dr. Singer had first noted that the principal chemical agents contained in Agent Orange, 2,4-D, 2,4,5-T, and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin "are potent and toxic agents capable of inducing a wide variety of adverse effects both in animals and in man."[47] Dr. Singer asserted that the dioxin compounds were capable of producing marked alteration in hepatic architecture and function, and that 2,4,5-T causes liver enzyme abnormalities, liver swelling, and centrilobular necrosis; he concluded that the liver abnormalities plaintiffs alleged were "consistent with" and "clearly compatible with" the known effects of polychlorinated herbicides.[48]

Singer proceeded in a similar manner to link numerous additional symptoms complained of by the injured veterans to the reported effects of exposure to Agent Orange. Ultimately, Singer opined:

> "Assuming the truth of the [veterans'] affidavits submitted, and absent any evidence of preexisting, intervening, or superseding causes for the symptoms and diseases complained of in these affidavits, it is my opinion to a reasonable degree of medical probability (that is, more likely than not) that the medical difficulties described by the affiants were proximately caused by exposure to Agent Orange."[49]

In reaction against Dr. Singer's analysis, Judge Weinstein stated that "one need hardly be a doctor of medicine to make the statement that if X is a possible cause of Y, and if there is no other possible cause of Y, X must have caused Y."[50]

Turning to a legal discussion of the standards governing the admissibility of expert opinion, Judge Weinstein noted that, under the federal rules, the assessment of novel testimony involves a balancing of the relevance, reliability, and helpfulness of the evidence against the likelihood of waste of time, confusion, and prejudice. Hence, the court stated, "when either the expert's qualifications or his testimony lie at the periphery of what the scientific community considers acceptable, special care should be exercised in evaluating the reliability and probative worth of the proffered testimony under Rules 703 and 403."[51] Retreating to a skeptical mode that he himself had suggested was contrary to the spirit of the new federal rules,[52] the judge stressed that such "'rigorous examination' is especially important in the mass toxic tort context where presentation to the trier of theories of causation depends almost entirely on expert testimony."[53]

Daubert

Three decisions issued by the US Supreme Court in the 1990s sought to decisively construe the admissibility standard governing expert testimony embodied in the Federal Rules of Evidence. By most accounts, these decisions, led by *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,[54] interpreted the rules as prescribing the restrictive approach involving substantial judicial scrutiny of the proffered testimony. Subsequently, in a dialectical turn, as shown in the next section, the rules were amended to conform to the decisions.

Some commentators have expressed the view that the political climate likely played a role in the skeptical gloss ultimately placed on the rules that concerned expert testimony. Professors Michael D. Green and Carl F. Cranor, for example, have noted that the Bendectin litigation from which *Daubert* arose occurred at a time when influential critics of the tort system perceived a crisis in tort law. Those critics alleged an unwarranted expansion of liability for product manufacturers, as well as over-litigiousness and increasingly high compensatory and punitive damage awards resulting in "overdeterrence" of innovative, if risky, technologies and products.[55]

Against the backdrop of the tort crisis rhetoric and in the context of the pre-*Daubert* tort regime, plaintiffs alleging harm caused by toxic substances faced an uphill battle. All specific causation—which links a specific exposure or set of exposures to *this plaintiff's* harm—passes through the gate of general causation, the causal capabilities of the toxic substance at issue. What proof is there, for instance, that silicone exposure increases the presence of injurious antinuclear antibodies?[56] How can a plaintiff establish that Bendectin ingestion causes birth defects?[57] More than 100,000 substances or their derivatives are registered for use in commercial applications, but researchers have studied the health implications of only a small portion of these.[58] This is the difficulty to which commentators seeking plaintiff-friendly tort reform responded in suggesting that traditional causal evidentiary requirements discouraged corporations from researching—and from thereby rendering accessible to litigants—potential health hazards that may be associated with their products.[59]

The *Daubert* plaintiffs were children born with birth defects allegedly caused by the antinausea drug Bendectin taken by their mothers during pregnancy. The district court had granted the manufacturer's motion for summary judgment because its single expert showed that none of the published epidemiological research—more than 30 studies—had found Bendectin to be a human teratogen. Eight experts had testified on behalf of the plaintiffs, each concluding that Bendectin can cause birth defects based on the results of in vitro and in vivo animal studies, pharmacological studies of Bendectin's chemical structure, and the "reanalysis" of previously published epidemiological studies.[60]

In the plaintiffs' initial appeal, the US Court of Appeals for the Ninth Circuit discounted the animal and chemical studies because other courts had already ruled that these were insufficient to establish a link between Bendectin and birth defects. The court also discounted the plaintiffs' reliance on reanalyses under the *Frye* standard, holding that "the reanalysis of epidemiological studies is generally accepted by the scientific community only when it is subjected to verification and scrutiny by others in the field."[61]

The Supreme Court in *Daubert* vacated the Ninth Circuit's judgment, commenting, somewhat ironically in retrospect, that "a rigid 'general acceptance' requirement would be at odds with the 'liberal thrust' of the Federal Rules and their 'general approach of relaxing the

traditional barriers to "opinion" testimony.'"[62] The court stated that *Frye*'s displacement by the Federal Rules of Evidence in the federal courts did not mean that there were no limits on the admissibility of purportedly scientific evidence. On the contrary, under the Rules, "the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."[63] The court then provided its epistemological analysis of Rule 702:

> The subject of an expert's testimony must be "scientific…knowledge." The adjective "scientific" implies a grounding in the methods and procedures of science. Similarly, the word "knowledge" connotes more than subjective belief or unsupported speculation. The term "applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds." …Of course, it would be unreasonable to conclude that the subject of scientific testimony must be "known" to a certainty; arguably, there are no certainties in science…. But, in order to qualify as "scientific knowledge," an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—ie, "good grounds," based on what is known. In short, the requirement that an expert's testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability.[64]

Now the court had to assure that judges themselves would be epistemically qualified to fulfill the "gatekeeping role"[65] of screening proffered scientific testimony for evidentiary reliability. The court stated that it was "confident that federal judges possess the capacity to" assess "whether the reasoning or methodology underlying the testimony is scientifically valid…."[66] According to the court, the factors bearing on the inquiry should typically include:

1. whether the scientific theory or technique at issue could be or had been tested, and thus whether it could be "falsified";
2. whether the theory or technique had been subjected to peer review and publication;
3. the known or potential rate of error;
4. the existence of "standards controlling the technique's operation"; and
5. the extent to which the relevant scientific community accepted the theory or technique.[67]

Note that the court derived the fifth factor from *Frye* as one of several factors by which a nonexpert district court judge may ordinarily appraise scientific validity. The court then stressed that the inquiry should be "a flexible one," and that "the focus, of course, must be solely on principles and methodology, not on the conclusions that they generate."[68]

Nor did the *Daubert* opinion explain the nature of certain of the factors it summoned, or manifest an appreciation of their interrelationships. For instance, the opinion spoke of a "rate of error" without regard to whether this may be Type I or Type II error, and without suggesting what might be an acceptable error rate, level of confidence, or level of statistical significance. And the court failed to elucidate the relationship between the falsifiability criterion embodied in the first factor and the error rate evaluation set down as the third factor, although a scientific rejection of an hypothesis in the absence of a stated confidence level for evaluating the hypothesis would be considered meaningless.

Importantly, and perhaps of the most lasting significance, the *Daubert* court emphasized that Rule 702's admissibility requisite—that the evidence or testimony "assist the trier of fact to understand the evidence or to determine a fact in issue"—was a relevancy condition best described as one of "fit." The *Daubert* opinion said that "'[f]it' is not always obvious, and

scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes."[69]

Responding to the suggestion that the screening role for judges would sanction "a stifling and repressive scientific orthodoxy and will be inimical to the search for truth," Justice Blackman countered that "there are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly."[70] In his dissenting opinion, Chief Justice Rehnquist, joined by Justice Stevens, nevertheless objected that "definitions of scientific knowledge, scientific method, scientific validity, and peer review [were] matters far afield from the expertise of judges."[71]

Joiner

The second decision in the *Daubert* trilogy was the Supreme Court's opinion in *General Electric Co. v. Joiner*.[72] The immediate question addressed in *Joiner* was whether the appellate court, reviewing a trial court's decision to admit or preclude expert testimony under *Daubert*, should apply an "abuse of discretion" standard of review or one that takes a fresh look at the proffered testimony and pays less deference to the trial judge's determination. On this technical legal point, the court noted that a trial court's evidentiary rulings are typically reviewed merely to determine whether the court abused its discretion. By that standard, the appellate court will not reverse the ruling unless it is deemed to have been manifestly erroneous.[73] The *Joiner* court held that the usual standard for reviewing evidentiary decisions—abuse of discretion—should similarly apply with respect to rulings concerning expert testimony. The trial court engages in a somewhat rigorous gatekeeping exercise, and appellate courts should show deference to its work.

Other aspects of the decision in *Joiner*, however, signaled a tightening of the judicial gate through which expert testimony is required to pass, and concomitantly an expansion of the range of scrutiny expected from the trial judge. The case arose from Robert Joiner's exposures to polychlorinated biphenyls (PCBs) while working as an electrician around electrical transformers. When Joiner was diagnosed with small-cell lung cancer, he sued the manufacturers of the PCBs, transformers, and the dielectric fluid used in the transformers and contaminated with the PCBs.[74]

The federal district court precluded the proffered testimony of plaintiff's experts under *Daubert*, concluding that "the opinions of Plaintiffs' experts do not fit the facts in this case because the opinions are inextricably bound up with the experts' assumption that Joiner was exposed to furans and dioxins," the latter being PCB derivatives.[75] In the later appeal, the Supreme Court agreed because the infant mice in the animal studies on which the experts relied had had massive doses of PCBs, in a highly concentrated form, injected directly into their peritoneums or stomachs, whereas Joiner was an adult human claiming a lesser level of exposure. Further, the mice developed alveologenic adenomas, whereas Joiner had become afflicted with small-cell carcinomas. And one of the experts acknowledged that no study had then demonstrated that PCBs lead to cancer in any other species.[76]

The Supreme Court in *Joiner* then offered a significantly tweaked view of the judicial gatekeeping role, stating:

> conclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the ipse dixit of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.[77]

The court's extension of the range of scrutiny to be expected from federal trial judges in their *Daubert* assessments, now reaching the experts' conclusions, and its initiation of an evaluation of whether "there is simply too great an analytical gap" between the underlying data and any such conclusion bolstered *Daubert*'s restrictive tendency. To the extent that lay judges tended to *perceive* analytic gaps, and with lay appellate panels reluctant to deem such rulings manifestly erroneous, *Joiner*'s gloss naturally accentuated *Daubert*'s pull disfavoring the admission of controversial, or difficult, expert testimony.

Kumho

The *Daubert* trilogy was completed in 1999 with the Supreme Court's decision in *Kumho Tire Co. v. Carmichael*.[78] *Kumho Tire* arose from a minivan accident and involved the testimony of an expert in tire failure analysis. The decision held that *Daubert* scrutiny applies not solely to scientific testimony, but also to that of engineers and other experts who are not scientists, and whose testimony is based on "technical" and "other specialized" knowledge. Because of the variety and nature of nonscientific, technical expertise, *Kumho Tire* reasserted *Daubert*'s openness to a "flexible" test of reliability, whereby "*Daubert*'s list of specific factors neither necessarily nor exclusively applies to all experts or in every case."[79]

THE AMENDED FEDERAL RULES OF EVIDENCE

In light of the *Daubert* trilogy's interpretation of the evidentiary rules relating to expert testimony, the Supreme Court approved amendments to the relevant rules codifying the principles announced in those cases. Rule 702, Federal Rules of Evidence, as amended in 2000 and then restyled a bit in 2011, now reads as follows:

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> (b) the testimony is based on sufficient facts or data;
>
> (c) the testimony is the product of reliable principles and methods; and
>
> (d) the expert has reliably applied the principles and methods to the facts of the case.[80]

Rule 703 was amended to clarify the use of an expert's underlying data, in the event that the data would "otherwise be inadmissible." Rule 703 presently provides:

> An expert may base an opinion on facts or data in the case that the expert has been made aware of or personally observed. If experts in the particular field would reasonably rely on those kinds of facts or data in forming an opinion on the subject, they need not be admissible for the opinion to be admitted. But if the facts or data would otherwise be inadmissible, the proponent of the opinion may disclose them to the jury only if their probative value in helping the jury evaluate the opinion substantially outweighs their prejudicial effect.[81]

There are circumstances in which the party offering the expert may not want to disclose the facts or data underlying that witness's opinion. The facts or data may embody confidential patient information, for example,[82] or more generally could contain information that would enable the opposing litigant to impeach the expert's opinion testimony.[83] In that event, Rule 705 of the Federal Rules of Evidence prescribes:

> Unless the court orders otherwise, an expert may state an opinion — and give the reasons for it — without first testifying to the underlying facts or data. But the expert may be required to disclose those facts or data on cross-examination.[84]

DAUBERT JURISPRUDENCE HAS IMPACTED THE *FRYE* ANALYSIS

Most of the states have patterned their rules of evidence on the federal scheme.[85] Some, however, remain committed to the *Frye* general acceptance standard.[86] Even for the latter group, the *Daubert* gatekeeping guidelines, as they have evolved, have strongly influenced the *Frye* analysis, resulting in what has become, in effect, a hybrid evidentiary approach.

In the *Neurontin Product Liability Litigation*, for example, the New York state court and the federal district court held a joint *Frye/Daubert* hearing to address the defendant pharmaceutical companies' motion to preclude the testimony of the plaintiffs' experts on the issue of general causation—that is, the issue of whether Neurontin is capable of causing suicide-related injuries.[87] Neurontin (generically known as gabapentin) was originally intended to treat epilepsy as well as neuropathic pain. The plaintiffs in the litigation alleged defendants' wrongful marketing of the drug for off-label uses such as the treatment of bipolar disorder and claimed that Neurontin caused suicide-related thoughts and events.

The state court judge in *Neurontin* adopted the findings of the federal court, which after an exhaustive *Daubert* analysis had affirmed the reliability of plaintiffs' experts' methodology and conclusions on the general causation issue.[88] The state court then undertook to evaluate the proposed testimony based on the state *Frye* standard.

The New York state court in *Neurontin* began by emphasizing the interplay between the *Frye* and *Daubert* standards. Because "general acceptance" is one of the suggested *Daubert* factors, said the state judge, the *Frye* standard can inform the federal reliability inquiry. Conversely, "a *Daubert*-type analysis of the plaintiff's expert's methodology is relevant where the scientific issue is not whether a novel scientific technique should be held admissible, but whether the methodology employed by the plaintiff's expert leads to a reliable theory or opinion on causation."[89]

The *Neurontin* state court emphasized that *Frye* does not intend that "novel" theories be excluded simply because they have neither been conclusively established in the scientific literature nor unanimously supported by the scientific authorities. "Thus, 'general acceptance' does not necessarily mean that a majority of the scientists involved subscribe to the conclusion."[90] However, in a *Frye* regime, said the court, "if the methodology is not novel and the issue is whether the methodology leads to a reliable theory of causation, the theory should arguably be scrutinized not under *Frye* for general acceptance, but under foundational principles for reliability."[91]

State courts still ostensibly applying the *Frye* standard have also gone so far as to summon the "analytical gap" test articulated in the restrictive *Joiner* opinion. Thus, in *Ratner v. McNeil-PPC, Inc.*,[92] for example, the New York appellate court looked skeptically upon plaintiff's experts' conclusions that, where acetaminophen (as contained in Tylenol) has been deemed to be a proven hepatotoxin associated with liver failure in certain cases of massive overdose, it was more probable than not that plaintiff's chronic ingestion of acetaminophen over several years, at maximum allowed doses, had similarly caused her cirrhosis. After citing directly to *Joiner*, the appellate court said, "the analytical gap between the plaintiff's scientific data and her experts' theory of causation is widened by the contrary scientific articles submitted by the defendant which, among other things, concluded that acetaminophen is safe in therapeutic doses, even for individuals suffering from liver disease."[93]

Although courts may evaluate whether an expert's theory, methodology, or conclusions have been generally accepted on an empirical basis and without themselves engaging in scientific or technical analysis, the same is not ordinarily true with respect to the "analytical gap" determination. In *Ratner*, it was arguable that "the thing from which the deduction was made"—the theory that acetaminophen overdoses cause liver damage—was generally accepted in the fields of toxicology, pharmacology, and hepatopathology. One sustainable option, applying a liberal approach, would have been to now let the jury evaluate the weight of the inference of a causal connection between the plaintiff's chronic, maximum-dose ingestion and her liver injury. The evolving *Frye/Daubert* hybrid enlists courts of law in a more probing and substantive extra-legal analysis.

THE EVOLVING SET OF *DAUBERT* FACTORS

Remarks in the *Daubert* and then *Kumho* opinions counseling the "flexibility" of the reliability determination opened the door to the use of an array of considerations beyond the original five factors detailed in *Daubert*. As shown, in addition to requiring that courts determine whether the expert proofs are sufficiently relevant to, and thereby "fit," the facts of the case, *Daubert* recommended that courts scrutinize proffered expert testimony with respect to the following reliability factors: (1) whether the theory or method can be/has been tested; (2) the known or potential error rate of the method; (3) whether it has been subject to peer review and publication; (4) whether it is generally accepted in the scientific community; and (5) the existence of standards controlling the method's operation.[94]

Apart from their awareness that they might range more broadly into other areas of inquiry, depending on the circumstances and the type of expertise at issue, judges remained cognizant that *Daubert*'s goal was not only to permit lay fact finders to weigh competing

scientific testimony without having a scientist's "understanding" of the subject matter, toward the end "that truth may be ascertained,"[95] but was also to enable lay judges to screen such testimony at the starting gate without themselves having such a sophisticated scientific understanding. The pull for those judges, therefore, was in the direction of developing evaluative criteria that were more accessible to the lay intellect.[96]

One such criterion is rooted, quite simply, in the expert's writing style, and arises from the impression made upon the court by the clarity and coherence of the expert's explanation of her methods, procedures, and theory. If the court discerns a logical flow in the expert's reasoning, it is more likely to uphold the admissibility of the testimony.

For example, in the litigation resulting from the 1979 "blow out" within the reactor at the Three Mile Island nuclear power facility, the plaintiffs sought to prove that high quantities of radioactive noble gases were forced into the atmosphere, only then to contaminate certain land areas throughout the area surrounding Three Mile Island.[97] Dr. Ignaz Vergeiner was a meteorologist proffered as an expert in boundary level meteorology in alpine regions. Highlighting the court's rationale for excluding the expert's testimony was its opening statement that "[t]he logical starting point for an analysis of Dr. Vergeiner's methodology is his own somewhat convoluted discussion of that methodology."[98]

With specific regard to opinions that draw inferences of general causation (ie, the general capability of the substance or mechanism to cause the harm), several courts implementing *Daubert*'s gatekeeping assignment have attempted to apply the "Bradford Hill" criteria.[99] Those courts have acknowledged that scientists are guided by the various Bradford Hill factors when determining whether an observed association between a substance or chemical and a disease is causal. These factors include: (1) strength of association (ie, whether the association is strong and statistically significant); (2) temporality (ie, whether the timing of the exposure and the onset of disease is consistent with the disease's latency period); (3) consistency (ie, whether the results of multiple scientific studies have themselves been sufficiently consistent to support the causal inference); (4) biological plausibility (ie, whether there exists a biologically plausible mechanism by which the agent could cause the disease); (5) consideration of alternative explanations (ie, whether the association could be accounted for by other factors); (6) specificity (ie, how closely the specific substance is associated with a specific population or occupational group and with a particular disease); (7) dose–response relationship (ie, whether an increase in exposure yields an increase in risk); and (8) analogy (ie, whether the effects of a similar substance or chemical has been established such that we may draw a causal inference regarding the substance at issue).[100]

Hill himself cautioned, however, against an overly rigid application of the criteria. Regarding strength of association, for instance, he stated that "we must not be too ready to dismiss a cause-and-effect hypothesis merely on the grounds that the observed association appears to be slight." Further, counseling against weighting consistency too heavily, he noted, "there will be occasions when repetition is absent or impossible and yet we should not hesitate to draw conclusions." Hill also emphasized that plausibility "is a feature I am convinced we cannot demand. What is biologically plausible depends upon the biological knowledge of the day. The association we observe may be one new to science or medicine and we must not dismiss it too light-heartedly as just too odd." In general, except for temporality, Hill did not view his "criteria" to be "necessary conditions," and he warned against laying down "hard and fast rules of evidence that must be obeyed before we accept cause and effect."[101]

While courts may take a flexible view of the extent to which the Bradford Hill factors must be satisfied before permitting causation testimony, they have been less welcoming of use of those criteria in the absence of an epidemiological foundation.[102] On the other hand, when experts have relied upon epidemiological data coupled with a Bradford Hill analysis, courts have readily excused other perceived weaknesses in the testimony.

For instance, in *In* re *Gadolinium-Based Contrast Agents Products Liability Litig.*,[103] defendants manufactured gadolinium-based contrast agents used in magnetic resonance scans, exposure to which plaintiffs alleged had caused their nephrogenic systemic fibrosis ("NSF"), a progressive fibrotic kidney disease affecting tissues and organs with no known cure. The court admitted the testimony of Joachim H. Ix, a nephrologist who engaged in substantial epidemiological research, and deemed him capable of conducting a meta-analysis of studies involving NSF, as well as of applying the Bradford Hill criteria. Given Dr. Ix's expertise in analyzing epidemiological and Bradford Hill-related data, the court did not place undue weight on the fact that his opinions were developed solely for litigation, that he may not have followed procedures he used in his own independent research, that he did not consult with other experts as to which studies to include in his meta-analysis, that he used retrospective, nonrandomized studies for his meta-analysis, or that his testimony was "based upon studies with wide confidence intervals."[104]

The purpose for which the research underlying an expert's testimony was conducted, however, has become one of the several available post-*Daubert* factors. Even in *Gadolinium-Based Contrast Agents*, the district court explained that, because Dr. Ix's opinions were developed solely for litigation, his testimony would be examined with greater scrutiny than if his opinions had been developed independent of the litigation.[105] As stated by the US Court of Appeals for the Eighth Circuit, cases subsequent to *Daubert* "have proposed additional factors, including, whether the expertise was developed for litigation or naturally flowed from the expert's research; whether the proposed expert ruled out other alternative explanations; and whether the proposed expert sufficiently connected the proposed testimony with the facts of the case."[106]

Other cases in the post-*Daubert* period have considered the breadth of the underlying data or studies upon which the proffered expert has based her conclusions. In *Boyd v. CSX Transp., Inc.*,[107] for example, the plaintiff was a train engineer who sued his employer under the Federal Employers' Liability Act, alleging that he suffered back injuries due to excessive whole-body vibrations. Rejecting the defendants' argument that plaintiff's expert, an orthopedic surgeon, would not offer reliable causation testimony, the court stated that the expert's review of multiple sources, including several studies, publications, and journal articles, ensured that the expert was "not relying on 'junk science' in rendering his opinion in a given case."[108]

Some legal rulings have also considered the importance of whether testing and data has derived from real-world, as opposed to solely laboratory, settings. The district court judge in *United States v. Ramirez*,[109] for example, precluded polygraph testimony in large part because, "[w]hile it is important to consider the known or potential rate of error…, which for polygraphs in laboratory research has been shown to be very low, the rate of error in 'real life' situations is not known to a reasonable degree of scientific certainty."[110]

Significantly, however, legal decision-making is normally rooted in precedent, and the doctrine of *stare decisis* informs judicial habit and custom when tasked with the need for a

ruling. The impact of persuasive prior rulings regarding similar expert proffers should not be underestimated. In addition to the usual goals motivating an adjudicative system built on respect for precedent in the articulation of legal rules and principles—including coherence, stability, consistency, predictability, certainty, fairness, equality, and legitimacy—prior *Daubert* gatekeeping determinations involving the same or a similar expert afford courts in subsequent cases an opportunity to realize substantial efficiencies. Although a prior ruling in an unrelated litigation does not relieve the present court of its gatekeeping task, the busy judge will ordinarily welcome guidance that opens the way to an efficient shortcut either in the *in limine* proceedings or during a later evidentiary analysis.[111]

Finally, judges face a choice not only of which *Daubert* criteria to weigh most heavily in their admissibility analyses, but also of how to engage in that weighing—how, in other words, to define the gate through which the testimony must pass. Some judges, applying a highly restrictive, "corpuscular" approach, examine evidence or factors in isolation, determining whether some factor may disqualify the testimony from being deemed reliable.[112] Others examine the cumulative weight of the evidence in the way the scientific community reaches a consensus of opinion, not condemning studies based on an isolated flaw or lack of perfect clarity, but rather assessing the "collective meaning" of the group of studies.[113]

FURTHER LEGAL APPROACHES TO FORENSIC EPIDEMIOLOGY

Over the past decades, courts have struggled to define epidemiology's forensic role, and in particular the use that may be made of relative risk findings in satisfying the burden of persuasion on the issue of actual causation, the factual aspect of the larger proximate causation inquiry. Perhaps the earliest judicial decision addressing the appropriateness of using epidemiological relative risk findings for determining actual causation was *Cook v. United States*.[114] The plaintiffs in *Cook* claimed that their rare neurological disorder, Guillain–Barre Syndrome ("GBS"), had been caused by their federally sponsored swine flu vaccinations. By holding that the plaintiffs' epidemiological evidence was *legally insufficient* to prove causation because it failed to show a relative risk greater than twice the upper limit of the baseline risk, the court sanctioned the significance of such a two-fold finding.[115] The *Cook* court elaborated, "Once the relative risk rises above two, it becomes more probable than not that a given case was caused by the vaccine."[116]

Other courts have departed from a bright-line relative risk requirement. In an oft-cited hybrid admissibility/sufficiency ruling, in *Grassis v. Johns-Manville Corporation*,[117] New Jersey's intermediate appellate court reasoned that imposing a 2.0 threshold correlation

> proves too much. Assuming a large group of potential plaintiffs, a causative factor of 1.99 and significant evidence eliminating other known causes, defendants' proposition would still exclude the epidemiological proof. Even though the physical problems of just under one-half of the plaintiffs (without reference to the additional causative proof) would have been statistically "caused" by the factor being studied, none could recover. Yet, if a new study raised the risk factor to 2.01, all of the plaintiffs could use the study to collect damages, although for nearly one-half of the group, the risk factor was not an actual cause of the condition. This makes little sense, scientifically or legally.[118]

Although the appeal of a bright-line standard nevertheless remains significant for jurists charged with providing guidance both to the litigants before them and to courts in future cases, courts have inclined toward finding a conceptually more satisfying way to reconcile epidemiology's traditional unit of study—being not individuals but population aggregates—with its potential usefulness in specific causation deliberations in toxic tort litigations. In the latter circumstance, when the epidemiological study is the only evidence grounding the plaintiff's causal claim, a direct inference of specific causation, even rooted in relative risk ratios greater than 2.0, would appear unstable hence incapable of sustaining the plaintiff's burden. When, however, there is some other case-specific factor—perhaps the plaintiff's high exposure, perhaps the presence of the offending substance near the tumor site, perhaps a "differential diagnosis" (more accurately: "differential etiology") eliminating certain alternative causal possibilities in the plaintiff's life—then the causal hypothesis should take on enhanced confirmational value.

In *Zandi v. Wyeth*,[119] for example, the plaintiff claimed that the defendant drug manufacturers' hormone replacement therapy ("HRT") drugs had caused her hormone-dependent breast cancer. The plaintiff's experts, a pathologist and an oncologist, relied on a combination of epidemiological studies and differential diagnosis in opining that HRT was the likely cause of Zandi's breast cancer. While endorsing this approach to the specific causation inquiry, the court concluded that, in this case, the experts' differential diagnosis analysis was unsustainable and lacked "foundational reliability," because "the scientific community has not accepted that breast cancer has a limited number of discrete and recognized possible causes such that ruling out one cause would implicate another."[120]

Interestingly, while courts in their *sufficiency* analyses have often denied that the bright-line 2.0 relative risk threshold aligns too strictly with the preponderance-of-the-evidence burden of proof, this ratio has had some significant play in judicial assessments of reliability in the context of *Daubert admissibility* determinations. In this regard, some jurists have taken the position that the 2.0 relative risk threshold is relevant in apprising the reliability of the testifying expert's methodology, as "one piece of evidence, among others,… which the Court is to consider in its evaluation."[121] This perspective effectively modifies the analytical gap inquiry *Joiner* assigned to the judicial gatekeeping task, more liberally applying the "collective meaning" approach by considering the epidemiological and other scientific evidence as a whole.

CONCLUSION

Legal adjudication simultaneously resists and embraces extra-legal expertise. Opinion testimony rooted in expertise beyond the ken of the average lay juror carries with it an aura of authoritativeness in tension with the jury's fact-finding mission. The legal system's standards for determining the courtroom status of expert testimony, and judicial attitudes construing those standards, have therefore always tended to embody some level of skepticism concerning the admissibility of such evidence.

Before *Frye*, courts somewhat begrudgingly permitted expert testimony to reach juries, tolerating this potential usurpation of the role of the litigants' lay peers "only from necessity."[122] *Frye*'s general acceptance test suffered from perceived vagueness and an apparent

bias against cutting-edge scientific theories or discoveries, but was also inevitably deemed overly simplistic in the era of complex, toxic tort litigation. The Federal Rules of Evidence, adopted in 1975 and amended in 2000, incorporated an admissibility threshold that both preserved the traditional standard—requiring the expert to possess adequate qualifications, and her testimony to assist the fact finder—but also enhanced the judiciary's role in engaging in some level of evaluation of the proffered testimony for a base level of reliability.

The *Daubert* trilogy ultimately aligned with the "restrictive" interpretation of the Rules, requiring not only reasonable reliance upon the methodology and type of data at issue by experts in the field, but also a fairly rigorous substantive scrutiny of the proffered testimony to determine both fit and reliability. The competing "liberal" approach had placed more confidence in the ability of the litigants to expose flaws and weaknesses in their adversary's evidence by means of rigorous cross-examination.

Evolving criteria for engaging in judicial scrutiny of proffered expert testimony, however, have tended to shift the terrain back toward the sort of review that is more accessible to the lay judicial intellect. Courts will nevertheless remain divided in their selection from the growing list of *Daubert* factors available to inform their gatekeeping task, in their approach to applying those factors, and over the liberality with which they permit the testimony—and, in litigations in which the admissibility of expert testimony is outcome determinative, the cases themselves—to reach the jury. This division will, of course, similarly characterize the judicial assessment of epidemiological proofs.

ENDNOTES

1. 293 F. 1013 (D.C. Cir. 1923).
2. *Id*., at 1014.
3. 41 N.H. 546, 547 (1860).
4. *Id*., at 548.
5. *Spring Co. v. Edgar*, 99 U.S. 645, 658 (1878).
6. *Id*., at 657.
7. *See* David L. Faigman, et al., Modern Scientific Evidence: The Law and Science of Expert Testimony § 1-2.1, at p. 4 (St. Paul, MN., West Publishing Co., 2002).
8. 77 N.W. 813 (Minn. 1899).
9. *Id*.
10. 43 Conn. 36 (1875).
11. *Id*., at 44.
12. U.S. Const., Amend. VI.
13. U.S. Const., Amend. VII.
14. For example, N.Y. Const., Art. I, § 2 (2013).
15. *Siring v. Oregon State Bd. of Higher Educ*., 927 F. Supp. 2d 1069, 1079 (D. Or. 2013).
16. *United States v. Rodriguez-Berrios*, 445 F. Supp. 2d 190, 192 (D. P.R. 2006).
17. *United States v. Bilzerian*, 926 F.2d 1285, 1294 (2d Cir. 1991); *Travelers' Ins. Co. v. Thornton*, 46 S.E. 678, 678 (Ga. 1904) ("The expert may aid the jury, but he cannot act as a member of the jury; nor, while on the stand, can he transcend the functions of a witness and, under the guise of giving testimony, state a legal conclusion").
18. *State v. Bonner*, 246 N.W. 314, 317 (Wis. 1933); *People v. Forte*, 4 N.Y.S.2d 913, 915 (Kings County Ct.), *aff'd*, 18 N.E.2d 31 (N.Y. 1938); *Marks v. United States*, 260 F.2d 377, 382 (10th Cir. 1958).
19. For example, *Medley v. United States*, 155 F.2d 857 (D.C. Cir. 1946) (sustained the admission in evidence of the results of a "spectroscopy" finding); *United States v. Yates*, 16C.M.R. 629, 633-34 (A.F.C.M.R. 1954) (applying *Frye* standard to affirm guilty finding in court-martial based on chromatography analysis of urine sample showing Air Force airman's use of morphine).

20. Irving Lehman, Review: *The Law of Evidence: Report of the Commonwealth Fund Committee*, 27 Columbia L. Rev. 890 (1927); Robert P. Mosteller, *Evidence History, the New Trace Evidence, and Rumblings in the Future of Proof*, 3 Ohio St. J. Crim. L. 523, 524 (2006).
21. 434 F.2d 502 (D.C. Cir. 1970).
22. *Id*., at 504 n.1.
23. *Id*., at 512 (Bazelon, C.J., concurring). Taking an ostensive approach, Judge Bazelon quoted at length from the trial judge's intrusive questioning. For example, "Do you think that is normal? Can you pick out a woman's genitalia in the Rorschach test?" *Id*., at 506-07.
24. 460 F.2d 864 (D.C. Cir. 1972).
25. *Id*., at 868 (Bazelon, C.J., concurring); *see also United States v. Alexander*, 471 F.2d 923, 955 (Bazelon, C.J., dissenting) (noting the then-general acceptance of psychological tests as diagnostic tools, saying "it is troublesome to see counsel or the court attempting to discredit them in a particular case by ridicule, rather than exploring their acknowledged strengths and weaknesses").
26. For example, *Wyeth Labs., Div'n of American Home Products Corp. v. Reyes*, 419 U.S. 1096 (1974) (granting motion of American Academy of Pediatrics and Conference of State and Territorial Epidemiologists for leave to file briefs as *amici curiae*).
27. For example, *Reserve Mining Co. v. United States*, 498 F.2d 1073, 1078 (8th Cir. 1974) ("A study by Dr. Selikoff of workers at a New Jersey asbestos manufacturing plant demonstrated that occupational exposure to amosite asbestos poses a hazard of increased incidence of asbestosis and various forms of cancer").
28. For example, *United States v. Betham*, 348 F. Supp. 1377, 1378 (S.D. Cal.) ("In support of his contention that he lacked knowledge of the heroin in the car, the defendant proffered as evidence the results of certain polygraph, or 'lie detector,' examinations to which he had submitted himself. If admitted, they would presumably indicate that the defendant was not attempting to deceive the polygraph examiner"), *aff'd*, 470 F.2d 1367 (9th Cir. 1972).
29. *People v. Kelly*, 549 P.2d 1240, 1244 (Cal. 1976).
30. 1 David L. Faigman, et al., Modern Scientific Evidence: The Law and Science of Expert Testimony § 1-2.1, at p. 5 n.8 (West Publishing Co., St. Paul, Minn. 2002) ("The Frye test also cannot distinguish between science and pseudoscience; after all, astrological forecasts are 'generally accepted' in the 'particular field' of astrology").
31. *Id*., at § 1-2.4, at p. 8.
32. J. Alexander Tanford, et al., Novel Scientific Evidence of Intoxication: Acoustic Analysis of Voice Recordings from the Exxon Valdez, 82 J. Crim. L. & Criminology 579, 595 (1991).
33. *See* Carlton Bailey, *The Admissibility of "Novel Scientific Evidence" in Arkansas: Does* Frye *Matter*?, 52 Ark. L. Rev. 671, 685 (1999).
34. Faigman et al., Modern Scientific Evidence, *supra* note 30, at § 1-2.4, at p. 10.
35. Preliminary Draft of Proposed Amendments to Federal Rules of Civil Procedure and Federal Rules of Evidence, 137 F R.D. 53, 157 (1991).
36. Charles McCormick, Evidence, at § 203, at p. 607 (West Publishing Co., St. Paul, Minn., 3rd ed. 1984) ("[p]lainly, 'reasonable reliance' is not synonymous with general acceptance").
37. *See generally United States v. 87.98 Acres*, 530 F.3d 899, 905 (9th Cir. 2008) ("We need not decide whether Rule 403 or Rule 703 governs in this case because we conclude that exclusion of the evidence was not an abuse of discretion under either rule").
38. *United States v. Everett*, 825 F.2d 658, 662 (2d Cir. 1987).
39. *See* Alani Golanski, *Judicial Scrutiny of Expert Testimony in Environmental Tort Litigation*, 9 Pace Envtl L. Rev. 399, 400 (1992).
40. *In re "Agent Orange" Product Liability Litig.: Gibbs v. Dow,* 611 F. Supp. 1223, 1243-44 (E.D.N.Y. 1985), *aff'd*, 818 F.2d 187 (2d Cir. 1987).
41. 736 F.2d 1529 (D.C. Cir. 1984).
42. David E. Bernstein, *The Misbegotten Judicial Resistance to the* Daubert *Revolution*, 89 Notre Dame L. Rev. 27, 36 (2013).
43. *Ferebee*, 736 F.2d 1529, 1534; *see also Barefoot v. Estelle,* 463 U.S. 880, 898-99 (1983) (addressing expert psychiatric testimony, stating "the rules of evidence generally extant at the federal and state levels anticipate that relevant, unprivileged evidence should be admitted and its weight left to the fact finder, who would have the benefit of cross-examination and contrary evidence by the opposing party").
44. *In re "Agent Orange" Product Liability Litig.,* 597 F. Supp. 740, 783 (E.D.N.Y. 1984).

45. *Id*., at 788.
46. *In re "Agent Orange" Product Liability Litig.: Gibbs v. Dow, supra,* 611 F. Supp. 1223, 1231.
47. *Id*., at 1236.
48. *Id*.
49. *Quoted at id*., at 1237.
50. *Id*., at 1238.
51. *Id*.
52. *Id*., at 1242 ("Courts abandoning *Frye* stress Rule 702's liberal attitude toward the admissibility of relevant expert testimony whenever it would be helpful to the trier").
53. *Id*., at 1244.
54. 509 U.S. 579 (1993).
55. Carl F. Cranor, Toxic Torts: Science, Law, and the Possibility of Justice, pp. 46-47 (Cambridge Univ. Press, New York, N.Y., 2006); Michael D. Green, Bendectin and Birth Defects: The Challenges of Mass Toxic Substances Litigation, pp. 19-20 (Univ. of Pennsylvania Press, Philadelphia PA., 1996); *see also* Arthur R. Miller, The Pretrial Rush to Judgment: Are the "Litigation Explosion," "Liability Crisis," and Efficiency Cliches Eroding Our Day in Court and Jury Trial Commitments?, 78 N.Y.U. L. Rev. 982, 1000 (2003).
56. *See Allison v. McGhan Med. Corp*., 184 F.3d 1300 (11th Cir. 1999).
57. *See, eg*, *Daubert v. Merrell Dow Pharm*., 43 F.3d 1311 (9th Cir. 1995); *DeLuca v. Merrell Dow Pharm., Inc*., 911 F.2d 941 (3d Cir. 1990).
58. *See* Carl F. Cranor & David A. Eastmond, Scientific Ignorance and Reliable Patterns of Evidence in Toxic Tort Causation: Is There a Need for Liability Reform?, 64 Law & Contemp. Probs. 5, 11 (2001); James Huff & David Hoel, Perspective and Overview of the Concepts and Value of Hazard Identification as the Initial Phase of Risk Assessment for Cancer and Human Health, 18 Scand. J. Work Env't & Health 83, 85 (1992) (estimating that there are as many as 100,000 chemicals in the marketplace); *see also* Office of Tech. Assessment, U.S. Congress, Screening and Testing Chemicals in Commerce 1 (1995) (estimating approximately 70,000 chemicals in commerce).
59. Margaret A. Berger, *Eliminating General Causation: Notes Towards a New Theory of Justice and Toxic Torts*, 97 Colum. L. Rev. 2117, 2134 (1997); *see also* Alani Golanski, *General Causation at a Crossroads in Toxic Tort Cases*, 108 Penn State L. Rev. 479, 500 (2003).
60. *Daubert*, 509 U.S. 579, 583.
61. *Daubert v. Merrell Dow Pharmaceuticals, Inc*., 951 F.2d 1128, 1131 (9th Cir. 1991) (citing Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact*, 7 Harv. Envtl. L. Rev. 429, 438-39 (1983)).
62. 509 U.S. at 588 (quoting *Beech Aircraft Corp. v. Rainey*, 488 U.S. 153, 169 (1988)).
63. 509 U.S. at 588, 589.
64. *Id*., at 589-90 (citations omitted).
65. *Id*., at 597.
66. *Id*., at 592-93.
67. *Id*., at 593-94. In suggesting the five factors that characterize the scientific methodology, and that inform the judicial gatekeeping exercise, the court cited for support both Carl Hempel, *Philosophy of Natural Science* (1966), and Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (5th ed., 1989). Hempel argued against Popper, however, that the line separating science from pseudoscience cannot be clearly demarcated. *See* Adina Schwartz, *A "Dogma of Empiricism" Revisited:* Daubert v. Merrell Dow Pharmaceuticals, Inc., *and the Need to Resurrect the Philosophic Insight of* Frye v. United States, 10 Harv. J. Law & Tech. 149, 168 (1997); Carl G. Hempel, Philosophy of Natural Science 22-32 (Prentice Hall, Englewood Cliffs, N.J., 1966). Nor did Popper, unlike Hempel, conclude that confirmation of a scientific theory, or of its reliability, was ultimately possible, but rather that every theory can, at best, only be held contingently. *See* Thomas S. Ulen, A Nobel Prize in Legal Science: Theory, Empirical Work, and the Scientific Method in the Study of Law, 2002 U. Ill. L. Rev. 875, 882-83. For Hempel, theories could be deemed confirmed in varying degrees, depending on the evidence. Carl G. Hempel, *Studies in the Logic of Confirmation*, *in* Aspects of Scientific Explanation and Other Essays in the Philosophy of Science 3, 3-46 (The Free Press, New York, N.Y., 1965).
68. 509 U.S., at 594-95.
69. *Id*., at 591.

70. *Id*., at 596-97.
71. *Id*., at 599-600 (Rehnquist, C.J., concurring in part, dissenting in part). On the Karl Popper influence, Rehnquist added, "I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its 'falsifiability,' and I suspect some of them will be, too." *Id.; cf.* Alani Golanski, *Why Legal Scholars Get Daubert Wrong: A Contextualist Explanation of Law's Epistemology*, 22 Whittier L. Rev. 653, 658-61 (2001).
72. 522 U.S. 136 (1997).
73. *Id*., at 141-42.
74. *Id*., at 139.
75. *Joiner v. General Elec. Co*., 864 F. Supp. 1310, 1320 (N.D. Ga. 1994).
76. 522 U.S. 136, 144.
77. *Id*., at 146.
78. 526 U.S. 137 (1999).
79. *Id*., at 141. The Supreme Court held that the district court had not abused its discretion in excluding the expert's testimony. The district court accepted the expert's qualifications, which included a masters degree in mechanical engineering, a decade of work at Michelin America, Inc., and testimony as a tire failure consultant in other tort cases. Rather, it excluded the testimony because, despite those qualifications, the expert's methodology was unreliable in assessing the cause of the tire's separation from its steel-belted carcass resulting in the crash.
80. Fed. Rule Evid. 702 (as amended Apr. 17, 2000, eff. Dec. 1, 2000; and Apr. 26, 2011, eff. Dec. 1, 2011).
81. Fed. Rule Evid. 703 (as amended Apr. 17, 2000, eff. Dec. 1, 2000; and Apr. 26, 2011, eff. Dec. 1, 2011).
82. For example, *Tibbs v. Adams*, Nº CIV S-05-2334, 2008 WL 2633233 (E.D. Cal., June 25, 2008).
83. *Curt Bullock Builders, Inc. v. H.S.S. Development, Inc*., 586 N E.2 d 1284, 1295 (Ill. App. 1992) (Lund, J., concurring).
84. Fed. Rule Evid. 705 (as amended Apr. 26, 2011, eff. Dec. 1, 2011).
85. Gil Seinfeld, *The Federal Courts as a Franchise: Rethinking the Justifications for Federal Question Jurisdiction*, 97 Calif. L. Rev. 95, 134 n. 120 (2009).
86. For example, *People v. LeGrand*, 867 N E.2d 374, 379 (N.Y. 2007).
87. *In re Neurontin Product Liability Litig: Delaney v. Pfizer Inc*., 897 N.Y.S.2d 671, 2009 N.Y. Misc. LEXIS 1777 (N.Y. Sup. Ct., 2009); *In Re Neurontin Marketing, Sales Practices, and Products Liability Litig*., 612 F. Supp. 2d 116 (D. Mass. 2009).
88. 897 N.Y.S.2d 671, 2009 N.Y. Misc. LEXIS 1777, at *2.
89. *Id*., at *6.
90. *Id*., at *9.
91. *Id*., at *8. The state court judge attributed this turn under New York law toward a reliability analysis to the state high court's decision in *Parker v. Mobil Oil Corp*., 857 N E.2d 434 (N.Y. 2006), wherein the New York Court of Appeals stated that "[t]he introduction of novel scientific evidence calls for a determination of its reliability. Thus, the *Frye* test asks 'whether the accepted techniques, when properly performed, generate results accepted as reliable within the scientific community generally.'" 857 N E.2d 434, 446 (quoting *People v. Wesley*, 633 N E.2d 451, 454 (1994)).
92. 933 N.Y.S.2d 323 (N.Y. App. Div. 2011).
93. *Id*., at 333; *see also Blackwell v. Wyeth d/b/a Wyeth, Inc*., 971 A.2d 235, 255 (Md. 2009) ("Generally accepted methodology, therefore, must be coupled with generally accepted analysis in order to avoid the pitfalls of an 'analytical gap'"); *Goeb v. Tharaldson*, 615 N W.2d 800, 816 (Minn. 2000) ("Given these concerns with Dr. Kilburn's methodology, the district court concluded that Dr. Kilburn made too great a leap to get from 'mere exposure of an unquantified amount of Dursban' to his conclusions about appellants' illnesses"); *Betz v. Pneumo-Abex, LLC*, 44 A.3d 27, 33 (Pa. 2012) ("the court expressed concern with an 'analytical gap' between the scientific proofs and the pathologist's conclusion…. A *Frye* hearing ensued, which was supplemented by other testimonial and documentary evidence").
94. *Daubert*, 509 U.S. 579, 593-94.
95. Fed. Rule Evid. 102.
96. *See* Lloyd Dixon & Brian Gill, *RAND Institute for Civil Justice: Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision* 38 (2001).
97. *In* re *TMI Litig. Consolidated II*, 911 F. Supp. 775 (M.D. Pa. 1996).

98. *Id*., at 793; *see also Chicago Title Ins. Corp. v. Magnuson*, Nº 2:03-CV-368, 2004 WL 5499517, at * (S.D. Ohio, Dec. 30, 2004) (admitting testimony of accounting expert William Minadeo on the issue of lost profits, concluding that "to regard Minadeo as excludable under *Daubert* would impermissibly predicate reliability on error-free analysis. Minadeo apparently employed *a rough but coherent methodology*, calculating lost profits based on projections, with expenses deducted based on the information he had") (emphasis added).
99. Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 Proceedings of the Royal Society of Medicine 205, 295-300 (1965).
100. *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 592-93 (D.N.J. 2002), *aff'd*, 68 Fed. Appx. 356 (3d Cir. 2003); *Gannon v. United States*, 292 Fed. Appx. 170, 172-73 n.1 (3d Cir. 2008).
101. Cranor, Toxic Torts, *supra*, at 102-05 (quoting Hill, *The Environment and Disease*, at pp. 15-19).
102. *In re Fosamax Prods. Liab. Litig.*, 645 F. Supp. 2d 164, 188 (S.D.N.Y. 2009) ("Several courts that have considered the question have held that it is not proper methodology for an epidemiologist to apply the Bradford Hill factors without data from controlled studies showing an association").
103. Nº 1:08 GD 50000, MDL Nº 1909, 2010 U.S. Dist. LEXIS 43444 (N.D. Ohio May 4, 2010).
104. 2010 U.S. Dist. LEXIS 43444, at *86–89.
105. *Id*., 2010 U.S. Dist. LEXIS 43444, at *86.
106. *Polski v. Quigley Corp*., 538 F.3d 836, 839 (8th Cir. 2008).
107. Nº 2:08-CV-108-TLS, 2011 U.S. Dist. LEXIS 24052 (N.D. In., Mar. 7, 2011).
108. 2011 U.S. Dist. LEXIS 24052, at *9.
109. Nº H-93-252, 1995 WL 918083 (S.D. Tex. Nov. 17, 1995).
110. 1995 WL 918083, at *2.
111. *See*, eg, *Security Nat'l Bank of Sioux City v. Abbott Labs*., Nº C 11-4017-MWB, 2013 U.S. Dist. LEXIS 77331, at *27-28 (N.D. Iowa, June 3, 2013) ("In this case, I am informed – if not actually aided – by three prior district court decisions considering whether or not to exclude essentially the same opinions from the very same experts based on essentially the same methodologies"); *Woodard v. Stryker Corp*., Nº 11-CV-36-F, 2012 U.S. Dist. LEXIS 119096, at *3 (D. Wy., July 16, 2012) ("Litigation regarding the use of pain pumps and chondrolysis is prevalent in jurisdictions all over the country and many cases use the same experts present in this case").
112. *See* Lisa Heinzerling, *Doubting Daubert*, 14 J. L. & Pol'y 6,5, 69 (2005) (criticizing Chief Justice Rehnquist's opinion in *Joiner* for "looking at each study in isolation from the other… [and] condoning the exclusion of the entire group of studies purporting to show a link between PCBs and cancer").
113. *Gadolinium-Based Contrast Agents*, 2010 U.S. Dist. LEXIS 43444, at *86-89.
114. 545 F. Supp. 306 (N.D. Cal. 1982).
115. *Id*., at 316.
116. *Id*., at 308 n.1.
117. 591 A.2d 671 (N J. App., 1991).
118. *Id*., at 676.
119. Nº A08-1455, 2009 WL 2151141 (Minn. Ct. App., July 21, 2009).
120. *Id*., 2009 WL 2151141, at *7; *cf. Norris v. Baxter Healthcare Corp*., 397 F.3d 878, 885 (10th Cir. 2005) (stating that experts' "reliance on differential diagnosis without supporting epidemiological evidence is misplaced and demonstrates the unreliable nature of the testimony").
121. *Pritchard v. Dow Agro Sciences*, 705 F. Supp. 2d 471, 486 (W.D. Pa. 2010); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp.2d 584, 606 (D.N.J. 2002).
122. *Martin*, 77 N W. 813, 815.

CHAPTER

2

Epidemiologic Evidence in Toxic Torts

S.C. Gold

Rutgers University-Newark, Newark, NJ, United States

M.D. Green

Wake Forest University, Winston-Salem, NC, United States

J. Sanders

University of Houston, Houston, TX, United States

OUTLINE

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00002-1

INTRODUCTION

The widespread use of epidemiologic evidence is a relatively recent phenomenon in the courts of the United States and elsewhere. Today, this type of evidence is introduced in a number of legal contexts, including in medical malpractice and other tort cases as well as in certain criminal cases, as described throughout the remainder of this book. However, the great majority of cases in which epidemiology has been used involves toxic torts. This chapter focuses on the use of epidemiology in these types of cases.[1]

Generally, the law of torts addresses claims for redress (damages) of a variety of injuries caused by a wide range of wrongful behavior. Subject to applicable legal requirements, tort law compensates injured people and other legal entities for harm to the body, property, emotional state, economic well-being, liberty, reputation, and privacy, for example.[2] Depending on the harm caused and the conduct or activity engaged in, tort liability may result from conduct that is intentional, reckless, negligent, or (in cases of "strict liability") without fault. A tort claim is specified by the combination of an interest protected by tort law and wrongful behavior injuring that interest. For example, the tort of defamation protects one's interest in reputation against another's communicating false information to third persons; the tort of negligence protects one's interest in freedom from bodily harm, property damage and sometimes emotional harm against another's behavior that fails to take reasonable care to avoid such injuries. Tort cases are civil cases initiated and prosecuted by the injured party, as contrasted with criminal cases, which are prosecuted by the government to vindicate societal interests (although sometimes the same behavior may result in both tort liability and criminal liability).

Toxic torts are a subspecies of torts. The primary distinction from other torts that protect bodily integrity is that the harm suffered is usually a disease rather than a traumatic injury, and the disease is allegedly caused by exposure to an environmental agent. Examples include asbestos, tobacco, and Vioxx. Although this category of cases can present a variety of challenges for the legal system, the most prominent one is determining causation.

Civil Litigation Generally

A basic outline of the American civil legal system will help the reader understand the special features of toxic tort cases and the role of epidemiology in these cases. In the United States, an injured party (the plaintiff) commences a lawsuit (often called a civil action) by filing a complaint—a formal document alleging that one or more other parties (the defendants) are legally liable for the plaintiff's injury.[3] The complaint must describe at least one valid legal theory that would entitle the plaintiff to relief against a defendant and must allege a set of facts that, if true, would be sufficient to meet the requirements of that legal theory.[4] If the complaint does state a valid claim for relief (or cause of action), the defendant may defend by denying some or all of the plaintiff's factual allegations or by alleging facts that would satisfy the requirements of one or more affirmative defenses, which are legal theories that allow a defendant to avoid liability.

Lawsuits then enter a phase in which the parties develop and exchange relevant evidence. Depending on the circumstances, evidence may consist of tangible objects, documents, or testimony of witnesses with firsthand knowledge of the events at issue. Witnesses may include experts who explain matters beyond lay understanding and who are permitted to state opinions based on their expert knowledge and analysis.[5]

At various points during the pendency of a civil action, procedural devices allow the court to terminate the case if the court concludes that a trial is unnecessary, such as when a party who is required to do so cannot produce reliable expert testimony, when all of a plaintiff's legal theories are deficient or when the factual evidence permits only one legal conclusion. If the case is not decided on one of these grounds and is not resolved by a settlement, the case proceeds to a trial for determination of disputed facts. In the United States, unlike most other countries, the finder

of fact is usually a jury. Evidence is introduced on disputed issues of fact, and the jury must resolve those issues based on the standard of proof required, which in civil cases is a preponderance of the evidence or, its equivalent, more likely than not. For convenience, this chapter uses the term "jury" to refer to the finder of facts in a legal proceeding, and "judge" or "court" to refer to the judicial officer(s) responsible for determining issues of law.

Whether by trial or by use of a pretrial procedural mechanism, a case is concluded by entry of a judgment that states whether or not the plaintiff is entitled to relief from each defendant, and if so, what that relief shall be. A dissatisfied party who believes the judgment resulted from some error committed in the trial court may then appeal the judgment to an appellate court. Appeals may continue through multiple levels of appellate courts until the court of last resort for a particular case has decided the case or declined to grant review.

In the federal system of government prescribed by the US Constitution, lawmaking is divided between state and federal government. The federal legislature, the Congress, has certain specified authority and the remaining lawmaking is left to states. Thus, the procedural and substantive rules that govern tort law in general, and toxic torts in particular, have several sources. Much tort law is state law and thus may vary depending on which state's law applies; even federal courts hearing tort claims generally use state tort law, unless the claim falls within one of a few types for which federal law applies.[6] Although tort law began as judge-made common law and even today tort rules to a large extent are crafted by judges, state legislatures and the US Congress have enacted many statutes that affect various aspects of tort litigation. Finally, the requirements of the US Constitution (and, in the states, each state constitution) are binding on all courts in the United States and may not be contravened by tort law.

Features of Toxic Torts

The most fundamental common feature of toxic torts is that they allege that injury (typically, a disease, rather than trauma) was caused by exposure to a toxic agent. Most toxic tort cases involve claims of bodily harm. Some toxic tort cases involve claims for property damage (eg, claims that spilled or leaked toxic chemicals migrated through soil or groundwater to other properties that lost value as a result). A few cases seek medical monitoring or other preventive action after exposure but before manifestation of disease.

A lawsuit on behalf of a person carelessly administered a fatal overdose of a drug or carelessly exposed to a lethal dose of pesticide could be considered a toxic tort. Such cases of acute poisoning, however, although of vital importance to the people involved, are not particularly novel in the broader context of tort law and can be handled readily by traditional legal doctrine.

The toxic torts that have engendered the most difficulty and consequent controversy among lawyers, judges, and legal scholars, by contrast, involve one or more of a group of distinctive characteristics that present unusual challenges for the application of conventional legal doctrine. These characteristics include chronic rather than acute alleged toxic effect; a period of induction and/or latency, often lengthy, between exposure and manifestation of disease; corresponding difficulties in proving the sources of entities responsible for, and extent of, exposure; unknown or opaque mechanisms of toxicity or questions about whether any toxic mechanism exists at all; existence of competing causes of the disease in question

and/or significant incidence of disease with no known cause; and the existence of instances of exposure that do not lead to disease. Viewed from this perspective, toxic tort claims include not only claims related to exposure to physical–chemical toxicants (whether intentional as with pharmaceuticals, unintentional as with environmental releases, or occupational as with asbestos fibers), but also claims of harm caused by exposure to radiation, biological agents (such as vaccines or pathogens), surgical implants, or other items that may cause disease in ways that exhibit some of these features (such as railroad ballast alleged to cause damage to the joints of railroad workers).

Toxic tort claims may arise from exposures that occur in many ways—environmental pollution, industrial or transportation accidents, occupational uses, use of consumer products containing or contaminated by toxic substances, and use of pharmaceuticals, for example. By no means do all toxic torts display all of the features described above. In the case of thalidomide, for example, none of these issues presented great difficulties: the time between exposure to the drug and manifestation of effects was relatively short, the link between the drug and birth defects was clear, and there was only one manufacturer. In claims for asbestos-related mesothelioma, the focus usually is not on the causal relation between asbestos fibers and disease, but on the difficulty of identifying the manufacturers to whose products a plaintiff was exposed. Asbestosis claims present this issue as well, and because of the cumulative nature of the disease, also involve questions about how much each defendant contributed to the severity of a plaintiff's illness. In other cases, such as those involving brand-name pharmaceuticals, the source of the exposure is unquestioned but the capacity of the agent to cause the disease in question may be hotly contested. Or the agent's capacity to cause the plaintiff's disease may be accepted but not the conclusion that the plaintiff's individual case of disease arose from exposure to the agent rather than from some other source. And some cases may present all of these issues, as when a plaintiff alleges that a variety of volatile organic compounds dumped in a landfill reached the plaintiff's childhood home and caused a particular subtype of leukemia diagnosed many years later.

The History of Epidemiologic Evidence in Courts

Until well into the latter part of the 20th century, courts frequently dismissed efforts to prove a proposition using statistical evidence. For example, the court in *Smith v. Rapid Transit*[7] refused to permit liability to be imposed based on statistical evidence that could have supported a finding by a preponderance of the evidence (the requisite standard of proof). Quoting an earlier case, the court said that it is:

> not enough that mathematically the chances somewhat favor a proposition to be proved; for example, the fact that colored automobiles made in the current year outnumber black ones would not warrant a finding that an undescribed automobile of the current year is colored and not black, nor would the fact that only a minority of men die of cancer warrant a finding that a particular man did not die of cancer…[A] proposition is proved by a preponderance of the evidence if it is made to appear more likely or probable in the sense that actual belief in its truth, derived from the evidence, exists in the mind or minds of the tribunal notwithstanding any doubts that may still linger there.[8]

Instead, disputed propositions of fact had to be proved with evidence specific to the case at hand.

Although there continues to be considerable controversy among commentators about the nature of proof and the value of statistical evidence, courts in toxic tort cases have accepted biostatistical evidence regarding the matter of causation for several decades. Indeed, in one of the earliest toxic tort cases, *Stubbs v. City of Rochester*,[9] the plaintiff contracted typhoid fever, allegedly as the result of the city's having intermingled drinking water and unsanitary water used for firefighting. Plaintiff employed rudimentary statistical evidence that addressed the increased number of typhoid cases during the year in which the intermingling occurred and the excess number that occurred during the months when the intermingling existed as compared to the rest of the year. The court relied on that evidence, in part, to affirm the plaintiff's verdict.

Atomic energy provided one of the first opportunities to confront the need for epidemiologic evidence in court proceedings. In the post–World War II era, as the nascent industry of atomic energy power generation emerged, Sam Estep, a professor of law at the University of Michigan, raised the matter of human injury and the importance of determining "biological causation" for those exposed to radiation. He observed:

> Of the many types of injuries which may result from irradiation of human beings, the greatest difficulties will be presented by those which as yet can be related scientifically to radiation only by an increased incidence among an exposed population. When the onset of the disease or injury is latent (delayed), predictions of future incidence are based on statistical possibilities. When, in addition, the biological causal relationship also is nonspecific (it may be caused by radiation but also arises among unexposed groups and no differentiation between those cases caused by radiation and those caused otherwise is possible), the legal problems, difficult before, become unmanageable under existing rules.[10]

Although epidemiology has long roots dating back to the Enlightenment,[11] modern epidemiologic methods were developed in the same postwar period[12] with several important prospective studies undertaken by public health officials, including the Framingham cardiovascular health study and the Salk vaccine trial.[13] In the ensuing years, epidemiologists uncovered causal relationships that have played an important role in toxic tort litigation, including smoking and lung cancer, swine flu vaccine and Guillain-Barré syndrome, asbestos and mesothelioma, lung cancer, and asbestosis, and diethylstilbesterol (DES) and vaginal adenocarcinoma.[14]

Today, when a causal issue arises in a toxic tort case, courts not only accept—but prefer—epidemiologic evidence.[15] Of course, part of the reason for that preference is that, as compared with traumatic injury torts, biological mechanisms are less well understood and alternative causes are often plausible explanations for the disease in question. Many claimants today find that without epidemiologic evidence to support their causal allegations, courts dismiss their cases after ruling that their expert witnesses may not testify about causation.[16]

LEGAL ISSUES ARISING IN TOXIC TORTS

The Elements of Legal Theories Employed in Toxic Tort Claims

All legal claims have "elements" that must be established in order for a plaintiff to prevail and recover relief from the defendant. Proving a tort claim requires proving all of its "prima facie elements."

The plaintiff bears the burden of proof on the prima facie elements of a tort claim. A plaintiff must introduce evidence that is sufficient for the jury to find that each element exists. The burden of proof in civil cases, unlike criminal cases, is the preponderance of the evidence—that is a jury must find that each of the elements is more likely than not to exist in order to find for the plaintiff.

"Negligence," a type of unintentional wrongdoing, is the most common type of tort claim and one frequently employed in toxic tort litigation. There are five elements to a negligence claim: (1) the existence of a duty of care by the defendant; (2) breach of that duty of care; (3) factual causation; (4) harm within the defendant's scope of liability (also referred to as "proximate cause"); and (5) legally cognizable harm. Breach of the ordinary duty of reasonable care is frequently referred to as "negligence."[17]

In addition to negligence, US courts developed in the 1960s and 1970s, a theory of liability for products that does not require a showing of unreasonable conduct but merely requires proof that the product is defective.[18] This theory was, in part, the product of difficulties faced by claimants in obtaining the requisite proof of unreasonable conduct by the product manufacturer and the growing societal concern with consumer protection. The theory is widely used in toxic torts today when the allegedly harmful agent has been sold as a product (eg, cigarettes, drugs, and heart valves), which includes most toxic torts save for those that involve hazardous waste storage and disposal. Products liability claims are widely used today in almost all states. Most often the basis for a claim that a product is defective in the toxic tort context is that the warning accompanying the product failed to inform or inadequately informed the consumer of the dangers posed by the product.

The sale of a product, a contract between the seller and buyer, creates warranties.[19] For our purposes the most important warranty is the implied warranty of merchantability, which requires that a product be fit for the ordinary purposes for which it is used.[20] While there is considerable overlap between what warranty law and products liability require for the safety of products, often there are differences that make one claim or the other more advantageous to the plaintiff[21] who may choose to pursue either or both claims.

Strict liability (in the sense that wrongful conduct is not required) has been applied in the area of abnormally dangerous activities. This theory depends on a judicial determination that the defendant is engaged in especially dangerous activity and therefore should bear the costs of that activity when it results in injuries to others. This theory has limited applicability to most toxic torts as the manufacture and sale of legal products is not a basis for this theory. The predominant area in which this theory has emerged in toxic tort litigation is in the storage and disposal of hazardous waste.[22]

Two other torts theories that are sometimes brought to bear in the hazardous waste line of toxic torts are nuisance and trespass, which protect interests in the possession and enjoyment of real property—they are the torts that make one's home her castle. Trespass is also a no-fault tort and if hazardous waste is disposed and deposited in or on another's property, trespass would provide a remedy to the owner of the contaminated property. Nuisance is the branch of tort law protecting one's enjoyment of real property and is often invoked when hazardous waste affects an owner, regardless of whether that waste crosses onto the owner's property.[23]

It deserves emphasis that, unlike some other areas of law, tort law requires proof that an individual defendant was responsible for an individual plaintiff's harm. Of course, a victim may show that each of several different wrongdoers was a cause of the victim's injury or disease, as when a physician negligently prescribes a contraindicated drug for a patient and the manufacturer negligently designed the drug. But tort law is said to be "private law," which requires that the plaintiff show that the defendant's wrongdoing caused harm to the plaintiff in order for the plaintiff to recover damages. Thus, factual causation is an element of every one of the theories of liability described above. A plaintiff, to be successful, must establish that the defendant's tortious conduct (or breach of warranty) was a necessary condition[24] for the harm suffered by plaintiff. Epidemiologic evidence plays a role in these cases primarily as a method of proving factual causation, and thus this chapter focuses most of its attention on this element.

Even if a plaintiff establishes the prima facie elements of a tort claim, there are a number of affirmative defenses that the defendant may raise and must prove if liability is to be defeated. Some affirmative defenses result in denying the plaintiff any recovery of damages from the defendant, while others reduce the amount of damages the defendant must pay. Each affirmative defense also has required elements. The defendant bears the burden of proving these elements. We discuss the defenses that are most significant in toxic tort cases in the section on "Defenses."

If a plaintiff is successful, the jury will award the plaintiff damages. The theory of tort damages is to return plaintiffs, as best as can be done, to the position they were in before the defendant's tort. Of course, with personal injury and disease, monetary damages are an imperfect device for restoring a victim to preinjury status. Damages are awarded for what are termed pecuniary or economic harm: these are losses for which there is a market value, such as lost wages, medical expenses, or destruction of property. By contrast, nonpecuniary damages are awarded for losses for which there is no market equivalent: pain and suffering or the lost opportunity for enjoyment of life due to, say, losing a limb. The determination of such damages in tort law is done on an individual basis based on evidence specific to the plaintiff's losses in each of these categories.[25] In addition to compensatory damages, punitive damages are sometimes awarded when the behavior of the defendant has been so outrageous and odious as to justify punishment in the form of an extra award of damages. Punitive damages have been considerably restricted in the past quarter century both by US Supreme Court decisions and state law reforms.

Special Problems Posed by Toxic Torts

Tort claims of any type may create difficulties for courts attempting to resolve them. Often there are uncertain facts even in routine cases like automobile accidents: did the defendant run a red light as some witnesses claim or proceed only after the light turned green, as other witnesses state? Sometimes there are difficulties in applying the law to facts: did the plaintiff's crossing unguarded railroad tracks without stopping constitute a failure to exercise reasonable care? Toxic torts can pose any of these difficulties but also present two particularly difficult problems: causal proof and apportionment.

Factual Causation

Plaintiff must establish that defendant's wrongdoing was necessary in causing the disease (or harm) for which plaintiff seeks to recover. Toxic torts often pose very difficult problems of proof for this element. Unlike many traumatic injury cases, the biological mechanisms involved in causing disease are often not well understood. Epidemiologic (or toxicologic) studies are often used to fill the proof gap, but they create additional problems that are described later in this chapter. Because many of the diseases involved in toxic torts have multi-decade latency periods, plaintiff may have been exposed to different defendants' agents as well as nonenvironmental risk factors during that period. Determining which exposure(s) played a causal role in the development or severity of plaintiff's disease may be impossible because of the lack of understanding about the biology of the initiation and development of the disease.[26] Or, plaintiff may have been exposed to only one defendant's toxic agent, but because of the passage of time, plaintiff cannot establish which one, as in the cases of daughters who developed cancer because of their mothers' ingestion of DES during pregnancy.[27] Thus, depending on the toxic tort, there may be one or several of these difficult aspects of factual causation presented. Factual causation in toxic torts has proved to be the most serious and often the most intractable problem for the legal system.

Apportioning Harm Among Defendants

Some harm is distinct or readily divisible. When a hiker is struck in an arm and a leg by bullets simultaneously fired by two separate hunters, a jury can (with the help of ballistics experts) identify which hunter shot the hiker's arm and which hunter shot the hiker's leg. In other cases, the harm may be theoretically divisible but evidence does not permit causal apportionment. Two cows from two different ranches escape and plunder a neighboring farm's corn crop. No evidence is available to assess how much of the destruction was due to each cow. Other harm—ones that are a result of multiple causes—are not even theoretically divisible. Thus, if two automobiles crash, resulting in a wheel coming off of one of the automobiles and hitting a nearby pedestrian, there is no basis to apportion the harm among the three drivers on the basis of causation—all that can be said is that each is a but-for cause of the harm. Apportionment is possible on the alternative basis of the comparative fault of the three drivers, a matter discussed in the section on "Contributory and Comparative Fault and Assumption of Risk."[28]

The biology of some diseases is comparable to the plundered corn crop, above. Thus, asbestosis is a progressive disease whose severity is a function of the magnitude of the dose. If an asbestosis victim was exposed to different manufacturers' asbestos products, each may have caused some incremental harm to the victim. However, like the plundered corn crop, the available evidence does not permit a determination of what increments were caused by which exposures.[29] Courts in the United States initially ruled that each manufacturer is liable for the full extent of the asbestotic harm. However, as asbestos litigation reached maturity and state laws were enacted that encouraged apportionment, some courts employed risk, rather than causation, as the basis of determining an asbestos defendant's liability.[30] Exposure is used as a rough proxy for risk, thereby providing a feasible means to assign liability to multiple toxic tort defendants who each cause only a portion of the victim's disease. Other courts use a very different metric, one unrelated to causation, the relative culpability of the parties, to divide the harm among multiple defendants.

By contrast with asbestosis, some diseases are nonprogressive and, once caused by a toxic exposure, do not become more severe with increased exposure. Mesothelioma and lung cancer, two other asbestotic diseases, are examples. Because the likelihood of contracting these diseases is related to the individual's total dose, each defendant who contributed to the threshold causal dose is a factual cause of the entirety of the harm (even if a "one-hit" genetic mechanism is involved), just as in the two-automobile examples above. A few courts have approved of apportionment based on risk contribution, ie, a defendant's share of the total dosage, in such cases.[31]

The Role of Epidemiologists

Epidemiologists and their research often inform legal decision-makers on the issue of factual causation in toxic tort cases. Lay witnesses cannot testify to causation in toxic tort cases in the way that they can when they see an apparently healthy pedestrian fall and break a kneecap after slipping on ice. Thus, expert witnesses are required to provide testimony about factual causation in toxic tort cases. Unlike lay witnesses who are limited to testifying to what they perceive with their five senses (putting aside the philosophical objection that causation is never directly observed and requires inference), expert witnesses are permitted to testify to their opinions based on their expertise and reliable methodology employed to reach that opinion. Epidemiologists (and others, including toxicologists, molecular biologists, and physicians) may serve as expert witnesses in a case, providing their opinions about the relationship between an exposure and a harm. In such instances, epidemiologists may testify about their own research, the research of others, or the results of research evaluations performed by public health entities such as the International Agency for Research on Cancer or the Agency for Toxic Substances and Disease Registry.

Epidemiologic evidence on a relevant question is often supplemented with toxicologic evidence on biological mechanisms of a disease or the implications of in vivo or in vitro research for factual causation. In the absence of epidemiologic evidence, plaintiffs may attempt to prove causation exclusively through these means. Neither epidemiologic nor toxicologic evidence directly answers the matter of whether a specific individual's disease was caused by the suspected agent, but both may be relevant to that issue. We return to the proof of specific causation for a given plaintiff below.[32]

APPLYING THE LAW OF FACTUAL CAUSATION IN TOXIC TORT CASES

The *Sine Qua Non*, But-For, or Necessary Element Test for Factual Causation

The factual cause element requires that plaintiff prove that a defendant's wrongdoing was a *sine qua non* (or, in plainer terms, a but-for) element in the disease for which plaintiff seeks recovery. Thus, in *Stubbs v. City of Rochester*,[33] there were several potential sources of typhoid bacteria that could have been the cause of the plaintiff's disease. The plaintiff had to prove that it was more likely than not that he contracted typhoid as a result of drinking water from the defendant City's water system that had been contaminated with water from the

City's sewage system. Assessing causation necessarily requires a counterfactual inquiry: What would have happened if the agent or conduct of interest had not occurred?

This account of factual causation means that there are multiple (perhaps an infinite number of) causes of an event. The causal role of the City's intermingling of drinking and sewage water required, among other conditions, that: (1) Stubbs had decided to live and work in Rochester; (2) he remained in the City during the period of contamination, rather than taking vacation at that time; (3) he drank public water rather than purchasing bottled water; and so on. That there are other factual causes does not undermine the fact that the City was *a* cause (rather than *the* cause) of the plaintiff's typhoid. Other elements of a tort claim eliminate innocent causes such as those mentioned above from consideration for liability. Thus, one can say that there are always multiple causes of legally cognizable harms.

Before closing this explanation of factual causation in the legal realm, multiple sufficient causes, which are quite different from the idea of multiple causes discussed above, must be addressed. Two or more causes are multiply sufficient if each, in the absence of the other, would be sufficient to cause the harm in question.[34] The classic example is two brush fires that are independently started, either of which would spread and burn down plaintiff's house. The two fires combine into one and plaintiff's house is burned to the ground. Neither of the two fires is a but-for cause of the destruction of the house because, in one fire's absence, the other fire would have destroyed the house. When multiple sufficient causes concur to bring about harm, each one is treated as a factual cause of the harm. To take a slightly more complex example, if a plaintiff receives six different doses of a toxin from different sources and the threshold for causing disease is five doses, each of the six doses is a factual cause of plaintiff's disease even though removing any one of the six would not change the outcome.

Legal terminology often employs the term "proximate cause." Proximate cause is used to mean different things, which often creates confusion. Sometimes it is used as a synonym for factual cause. However, its distinct usage addresses an element of a tort case that is quite different from factual cause. This usage imposes limits on a defendant's liability when an extended chain of events results in harm that is quite different from the harm risked by the defendant's misconduct that subjects the defendant to liability. Thus, if a person negligently stores unlabeled rat poison in a refrigerator and the jar of poison falls on another's toe when the door is opened, the negligent storer would not be liable for the toe injury because the risk of poisoning, not the risk of falling jars, was the basis for finding the storer negligent. This "proximate cause" limitation on liability has nothing to do with causation and little to do with proximity. The influential third Restatement of Torts adopts a different term, "scope of liability," to replace proximate cause.[35]

The Different Aspects of Factual Causation in a Toxic Tort Case

The causal question in a toxic tort case is whether defendant's tortious conduct caused plaintiff's harm. In some cases, however, the harm that was caused was not the existence of disease but that defendant accelerated plaintiff's contracting the disease. In such cases, the harm is the presence of disease during the period of time that plaintiff would have been disease-free without defendant's wrongdoing.[36] Similarly, in a case in which a plaintiff claims exposure to a toxic agent causes her to fear that she will contract a disease in the future, the harm is not the disease but the emotional harm suffered by the plaintiff. Another

such instance of a defendant causing something other than the entirety of the disease is when the plaintiff has already contracted a progressive disease and defendant's conduct enhances the degree of the disease.

Once the harm and defendant's wrongful conduct have been clearly specified, the various aspects of the causal inquiry that must be resolved to determine whether factual cause is satisfied can be identified.

Agent-Disease Causation

Often the most problematic aspect of causation in a toxic tort case is the issue of whether plaintiff's exposure to an agent caused plaintiff's disease. Such a determination involves three sub-issues:

- Exposure[37]: For plaintiff's disease to be caused by a toxic agent, plaintiff must have been exposed (through any of the known routes, absorption, skin contact, ingestion, inhalation, implantation, irradiation, or injection) to that agent. In some types of toxic tort cases, typically those involving the use of a drug or a chemical in a consumer product, exposure is straightforward. However, in occupational settings, exposure may not be so straightforward. In hazardous waste litigation, there may be a question of whether the waste reached the location where plaintiff resided or worked. Exacerbating proof problems is that the dose of exposure is often critical, and little evidence of dose over the period of time plaintiff was exposed occupationally or residentially is available, especially with the passage of decades due to latency periods. Further complications occur because of the need to tie defendant's conduct to plaintiff's exposure and to determine the portion of any exposure for which defendant is responsible. Thus, in asbestos litigation, an industrial worker may have been exposed to dozens of different asbestos products over an entire career. Determining which manufacturers provided those products and the extent of exposure to each defendant's products presents difficult, nearly impossible, proof problems.
- General causation[38]: Tort law has developed two additional aspects for the agent-disease causal inquiry. One, general causation, asks the question of whether the agent in question (or in a more refined form, the agent at the dosage in question) is capable of causing the disease in the human population. Courts employ general causation because the source of evidence about agent-disease causation often is provided by group studies of the agent, either epidemiologic or toxicologic. Because epidemiology inquires into the effect of an agent on a group and whether there is an increased incidence of disease in the exposed population, it does not directly address the matter of whether any individual's disease was caused by exposure to an agent. However, if general causation does not exist, then agent-disease causation does not exist, and consideration of the case can be truncated because of the absence of factual causation.
- Specific causation[39]: If general causation exists, then a further inquiry is required: did the defendant's agent cause this plaintiff's disease? Before an injured plaintiff can recover from a defendant, the plaintiff must show that the defendant caused that plaintiff's harm.

Many toxic agents cause diseases that have other independently sufficient causes. These other causes are called "competing causes" because only one was involved in the actual

causal set of factors that caused plaintiff's disease. The other competing causes are also capable of causing the disease even though in this instance they did not. Thus radon and cigarette smoking are competing causes of lung cancer. Asbestos and smoking have a more complicated relationship with regard to lung cancer. In some instances they are competing causes. But they also have a synergistic effect on risk, so they may in other instances be multiple causes of the disease. In competing cause situations, the plaintiff must establish that defendant's toxic agent was more likely the cause of disease as contrasted with all other competing causes.[40] In the section on "Epidemiology and Proof of Specific Causation," we review how courts have used relative risk information in addressing this issue.

Other evidence, in some cases, may supplement evidence from group studies to support or negate causation. Thus challenge/dechallenge/rechallenge evidence in acute disease cases (discussed in the section on "Epidemiology and Proof of Specific Causation") or evidence about biologic mechanism may affect the probability of causation. In some cases, especially when the agent is pathognomonic, evidence from group studies may not be required. For example, thalidomide was identified as a teratogen without any epidemiologic study, and its teratogenicity was only later supported by toxicologic inquiry.[41]

Defendant and Defendant's Misconduct

Agent-disease causation, while necessary, is not sufficient to satisfy the requirement of factual causation. The causal link to the defendant and the defendant's misconduct must be established as well. *Zuchowitz v. United States*[42] illustrates the latter issue. Defendant's negligence in prescribing or dispensing a drug resulted in the patient receiving twice the authorized dose of the drug. The patient developed primary pulmonary hypertension (PPH). Defendant's negligence only applied to the double dose; prescribing and providing the proper dose was not negligent. Thus, the court had to determine if it was only the excess dose that caused the patient's disease or whether the proper dose, alone, would have caused the disease, in which case the defendant would not have been liable.

Another aspect of demonstrating that it was defendant's misconduct that caused the harm, rather than merely defendant's product, arises when the alleged misconduct is a failure to warn or an inadequate warning of the risks posed by the product sold by the defendant. (Sellers of products are required to provide reasonable warnings of those risks of which the seller is aware or should have been aware.) If there would have been no difference in plaintiff's using or being exposed the product even if an adequate warning had been provided, then the warning failure made no difference in the plaintiff's contracting the disease and causation is absent. In the prescription drug arena, some plaintiffs are unsuccessful because the prescribing physician testifies that the decision to prescribe would not have been affected by the additional information that an adequate warning would have provided.[43]

A different difficulty in connecting defendant's tortious conduct with plaintiff's harm arose in litigation over DES, a drug prescribed to pregnant women to prevent miscarriage. During the time the drug was on the market, several hundred manufacturers produced and sold the drug. A generation later, the female offspring who were exposed in utero developed several adverse effects, most of which were pathognomonic. Yet, because of the passage of time, most were unable to establish which DES manufacturer provided the drug that her mother ingested. Because of the generic risk posed by DES, about half of the states confronted

with this situation employed a new "market share" theory of liability that enabled victims to recover from DES manufacturers based on the manufacturers' shares of the DES market[44] while about an equal number declined to modify their requirement that causation requires identification of the actor responsible for plaintiff's harm. However, courts have rejected a market share theory of liability in virtually every setting that does not involve DES.[45]

USING EPIDEMIOLOGY TO PROVE CAUSATION

Testimony by epidemiologists, or testimony based on the results of epidemiologic studies, figures prominently in toxic torts. Prior to the 1970s, however, there were very few references to epidemiology in federal or state cases. As toxic tort claims became more common in the 1970s and 1980s, initially some courts—adhering to the traditional judicial view of statistical evidence—held that it was improper to draw conclusions of causation in individual cases based on population-based data showing that the frequency of a disease was higher in a group of persons exposed to a toxic agent. At the same time, however, courts were suspicious of the extrapolations required to infer causation from the results of traditional in vivo and in vitro toxicological studies.[46] As courts gained experience with toxic torts and with epidemiology, judicial treatment of epidemiology became more complex.

One of the earliest widespread uses of epidemiology in court occurred in cases alleging that the swine flu vaccine caused those who received it to contract Guillain-Barre syndrome.[47] Since then, the role of epidemiology has grown remarkably. Epidemiologic evidence has played a central role in many toxic torts, including those alleging injury due to asbestos, electromagnetic radiation, IUDs, silicone implants, tobacco products, chlorine gas, benzene, herbicides, pesticides, solvents, vinyl chloride, polychlorinated biphenyls (PCBs), and many pharmaceuticals.

Epidemiology is a two-edged sword for toxic tort plaintiffs. On the one hand, this evidence sometimes provides irrefutable evidence linking exposure to injury. For example, our first knowledge of the enormous health dangers caused by tobacco consumption was largely the result of epidemiologic research.[48] On the other hand, the absence of epidemiologic evidence tying an exposure to an injury often acts as a barrier to recovery.

The essence of epidemiologic research is to look for statistical associations between disease incidence and exposure and to assess whether observed associations are causal rather than coincidental or spurious. In drug testing, randomized double-blind clinical trials provide quasi-experimental conditions for the drug being investigated. In most toxic tort contexts, however, if epidemiologic research is available it consists of one or more observational studies using a case-control, cohort (prospective or retrospective), or cross-sectional design.[49]

Epidemiology and Proof of General Causation

Recognizing Epidemiology's Advantages

Many courts have emphasized epidemiology's central advantage over toxicological evidence: epidemiologic research studies people, eliminating the need for interspecies extrapolation or inference from simpler experimental systems. In significant part this emphasis arose because of noteworthy cases or case clusters in which repeated, large, well-designed

epidemiologic studies failed to find statistically significant increased risk of disease associated with exposure despite the existence of animal or other toxicological studies that suggested the possibility of causation. These included the claims of women who had received silicone gel breast implants that their implants caused connective tissue diseases and the claims that the morning sickness drug Bendectin had caused major birth defects.[50] A number of courts that rejected plaintiffs' causation evidence in these cases stated or implied that epidemiologic proof of causation was required or at least that epidemiologic studies constituted the best evidence.[51]

Courts also recognize, however, that epidemiologic data often is not available.[52] Courts have acknowledged that epidemiologic studies are expensive and time-consuming and that it may be impracticable for epidemiologic studies to detect effects that are subtle, rare, or associated with rare exposures. As a matter of legal principle, therefore, most courts have rejected a blanket rule that a plaintiff must produce statistically significant epidemiologic results to support a claim of causation. Nevertheless, only very rarely have courts in mass exposure cases determined that a toxic tort plaintiff lacking epidemiologic proof nevertheless has produced evidence of general causation sufficient for the causation issue to be presented to a jury are quite rare. Courts are more lenient when the plaintiff has been exposed to a localized release of a substance about which there is no substantial body of epidemiologic evidence. And it is very difficult for a plaintiff to overcome a substantial body of epidemiologic studies showing no statistically significant association between the plaintiff's exposure and the plaintiff's disease, even if some epidemiologic studies have found an association or other evidence suggesting agent-disease causation exists.

Recognizing Epidemiology's Limitations: Methodological Issues

At the same time that courts have emphasized the importance of epidemiologic observations of increased disease incidence in exposed human populations, they have also recognized the limitations of epidemiologic research. The most fundamental limitation is that an association alone does not establish causation. In general, courts have looked to a number of considerations to assess whether an observed association reflects a causal relation. The most widely cited set of considerations are the well-known Bradford Hill factors.[53] Some courts have at times appeared reluctant to accept the fact that application of these considerations in a particular case requires an informed exercise of scientific judgment rather than completion of a mandatory checklist.

When epidemiologic evidence does exist, the courts are confronted with the question of whether it is admissible to prove general causation. Testimony based on well-conducted studies that are precisely on point is universally admitted. Beyond this, the legal rules vary from jurisdiction to jurisdiction and sometimes from claim to claim within a jurisdiction, reflecting varying degrees of judicial tolerance for reliance on epidemiologic studies that are not exactly on point or are methodologically imperfect.

Some order can be brought this complex and sometimes contradictory mosaic, however, by focusing on several factors that influence admissibility. Epidemiologic evidence is most clearly admissible when it is based on a number of well-designed, large studies that indicate a strong and statistically significant relationship between the exact substance to which the plaintiff was exposed and the exact injury the plaintiff has suffered at a dose identical to that the plaintiff is known to have experienced. As each of these factors (design of studies,

number of studies, strength of the relationship, statistical significance of the relationship, substance similarity, injury similarity, dose similarity) is removed, the value of the epidemiological research is weakened and the admissibility of the testimony becomes more problematical. None of these factors by itself inevitably leads to exclusion. For example, courts are quite reticent, but not entirely unwilling to accept testimony based on epidemiologic results that lack statistical significance.

Some courts are less cognizant of other methodological issues in epidemiologic research—potential problems of bias, measurement error, and uncontrolled confounding—that if not addressed can produce spurious associations or distort any real association that exists.[54] In the area of injuries due to drugs, randomized clinical trials may sidestep many of these problems but unfortunately the incidence of some diseases that may be caused by drugs is frequently so low that even large-scale clinical trials often have insufficient power to detect all of the adverse side effects in premarketing testing. Amendments to the Federal Food, Drug, and Cosmetic Act recognize this fact and call for heavy use of observational epidemiologic studies as a source of regulatory and scientific evidence for postmarketing drug safety regulation.[55]

Recognizing Epidemiology's Limitations: Power and Significance Testing

Courts have been relatively insensitive to the low power of some clinical trials and observational studies to detect effects that actually exist. The power of a test is usually expressed as β and the probability of making a Type II error is $1 - \beta$. The power of a study is a function of a study's sample size, the size of the effect one wishes to detect, and the significance level (usually expressed as α) used to guard against Type I error. Because power is a function of, among other things, the significance level used to guard against Type I errors, all things being equal, minimizing the probability of one type of error can be done only by increasing the probability of making the other.

In most toxic tort contexts, the defendant would prefer to minimize Type I error while the plaintiff would prefer to minimize Type II error. The power of any test is reduced as the incidence of exposure in cohort studies or the incidence of an effect in case-control studies decreases. Type II threats to causal conclusions are particularly relevant with respect to rare events. Plaintiffs make a fair criticism of randomized trials or epidemiologic cohort studies when they note that in these situations studies may have a low probability of detecting a relationship. In situations of low exposure rates, case-control studies can be particularly valuable because of their relatively greater power but at the risk of magnifying the effect of any biases in the study.

Given the importance of power in assessing epidemiologic evidence, surprisingly few appellate opinions discuss this issue. *DeLuca by DeLuca v. Merrell Dow Pharmaceuticals, Inc.*[56] contains one of the better court discussions of this subject. The opinion discusses the two types of error and suggests that courts should be concerned about both. *Merrell Dow Pharmaceuticals, Inc. v. Havner*[57] recognizes that there is a trade-off between making Type I and Type II errors in any given study and that scientists may wish to vary alpha and beta depending on the costs of making either a Type I or a Type II error. Nevertheless, *Havner* selected the 95% statistical significance level as a legal threshold for reliability and hence admissibility of scientific expert testimony.[58]

Because the burden of persuasion in civil cases is a "preponderance of the evidence," ie, more probable than not, legal commentators and others, including scientists, sometimes incorrectly attempt to equate this legal test to tests of statistical significance and power in epidemiological research. The most egregious and yet common examples of this intuition comes from cases and commentators arguing that in civil cases courts should simply disregard the typical p value of tests of significance in the legal setting, 0.05, because this is far higher than the preponderance of evidence burden.[59] Such statements are based on a misunderstanding of what tests of significance accomplish. Significance testing states the probability that the study outcome (or an even more extreme outcome) would occur if the null hypothesis is correct. It does not permit us to conclude that the null hypothesis has only a 5% chance of being correct. The p value tells us what is likely to happen when the null hypothesis is correct; it cannot tell us the probability that the hypothesis is true. Nor is it correct to say that a 0.50 significance level is similar to a preponderance of the evidence legal standard of proof. Were we to adopt this standard, the result would be to greatly increase the ratio of false positive (Type I) errors to false negative (Type II) errors.

A more nuanced proposal is to choose a test of significance that equalizes the chance of a false positive and a false negative on the ground that this corresponds to the "more-probable-than-not" burden of proof. It is not the case, however, that α and $1 - \beta$ ordinarily give the probabilities of the null and alternative hypotheses.

Epidemiology and Proof of Specific Causation

Epidemiology studies the factors that influence disease incidence in populations rather than the cause of disease in any individual, so by its nature epidemiology does not directly address the legal issue of specific causation. Thus epidemiologic research can discern that a disease occurs more frequently in a population of exposed individuals but cannot distinguish which individuals in that exposed population would have become sick even in the absence of the exposure. This characteristic of epidemiology has been much discussed in court decisions and fueled substantial early skepticism of the value of epidemiologic research, standing alone, as proof of causation in toxic tort cases. To some extent that skepticism endures, as seen in a decision of the United Kingdom Supreme Court, *Sienkiewicz v. Greif (UK) Ltd.*,[60] as well as in some decisions of courts in the United States. The population-based data of epidemiology have sometimes been contrasted with preferred "particularistic" proof that would link an individual case of disease to an individual exposure, as, for example, mechanism evidence might.

The problem with a judicial desire for particularistic proof is that such proof rarely exists. To be sure, the inference linking exposure to disease may be very strong in some cases. Pathognomonic or signature diseases are strongly and uniquely linked to one exposure as are some illnesses that appear after large acute exposures. In addition, a challenge/dechallenge/rechallenge ("CDR") process may be employed to assess the adverse effects of various substances on a particular individual because a positive response to rechallenge reduces the probability that some competing cause is responsible for the reaction. Unfortunately, CDR is only available for acute responses that are reversible upon removal of the agent, rendering it useless in toxic tort cases involving chronic toxicity. It is unavailable for many unintentional exposures

and cannot be used when rechallenge would be unethical. Consequently, CDR's usefulness as a method of assessing specific causation is quite limited. Thus for a wide range of toxic tort cases that involve a disease that occurs both with and without exposure and an exposure that occurs both with and without disease, a legal requirement of solid particularistic proof would be tantamount to a rule that many plaintiffs could never prevail.

Recognizing that reality, some courts have found ways to harmonize the legal system's focus on specific causation with the group-based output of epidemiologic studies. A common, though far from universal, approach has invoked the probabilistic frame of legal burdens of proof. As described in the section on "The Elements of Legal Theories Employed in Toxic Tort Claims," a plaintiff in a civil case has the burden of proving causation by a preponderance of the evidence, which is generally defined as evidence that causes the jury to believe that a disputed proposition of fact is more likely than not true. Beginning in the early 1980s, some courts reasoned that the attributable fraction of disease incidence in an exposed population, which can be computed from epidemiologic results, could be treated as a measure of the probability that an individual case occurred because of the exposure. This new approach raises two related questions. First, what is the legal significance of a relative risk greater than 2.0 and second, what is the legal significance of a relative risk less than 2.0.[61]

Relative Risk Greater Than 2.0

With respect to relative risks greater than 2.0, some courts reason that an attributable fraction exceeding 50% (equivalent to a relative risk greater than 2 in a cohort study) is evidence permitting a finding that specific causation has been established to the "preponderance of the evidence" standard. Such reasoning has been criticized in various ways by some legal writers, epidemiologists, and courts. Critics have argued, for example, that: specific causation should never be inferred from epidemiologic data; relative risk values slightly above 2.0 are too likely to be artifacts of methodological bias, random error, differences between the plaintiff and the subjects in the study, or confounding; focus on a threshold relative risk value incorrectly ignores the inherent variability of epidemiologic results; epidemiologic relative risk systematically understates the individual probability of causation. Despite these criticisms, there appears to be a trend toward permitting plaintiffs to use a doubling of the risk as evidence of specific causation.

Relative Risk Less Than 2.0

The second question posed by this line of reasoning is the legal significance of a relative risk less than 2.0. This issue was discussed by the US Court of Appeals for the Ninth Circuit in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[62]:

> While plaintiffs' epidemiologists make vague assertions that there is a statistically significant relationship between Bendectin and birth defects, none states that the relative risk is greater than two. These studies thus would not be helpful, and indeed would only serve to confuse the jury, if offered to prove rather than refute causation. A relative risk of less than two may suggest teratogenicity, but it actually tends to disprove legal causation, as it shows that Bendectin does not double the likelihood of birth defects.[63]

The opinion went on to note, however, that in some situations a plaintiff might be able, despite relative risks under 2.0, to show it is more likely than not that the alleged cause is responsible for a particular plaintiff's injury. For example, if a study finds a relative risk of

a particular birth defect of 1.8 among children born of all mothers who took a particular drug, a given plaintiff might successfully disaggregate this data between mothers who smoke and those who do not and demonstrate that among children born to nonsmoking mothers, the relative risk is over 2.0.[64] Some courts have also agreed that a plaintiff should be able to produce other types of evidence, including animal research, to bolster their argument that the suspect substance is more likely than not the cause of their disease.[65]

Note that the *Daubert* opinion quoted above did not mandate that a plaintiff must show a relative risk greater than 2.0 in order to reach a jury. This appears to be the opinion of a number of courts that have considered the issue with others taking the opposite position.[66] However, in all jurisdictions, a general assertion of an expert that his or her assessment of the relevant literature resulted in the conclusion that an injury was caused by the exposure in question will likely not be sufficient to overcome the adverse epidemiologic evidence.[67] See Chapter 1, *Legal Considerations of Forensic Applications of Epidemiology in the United States* for additional discussion of this issue.

Differential Diagnosis and Specific Causation

As noted in the section on "Agent-Disease Causation," because of the distinction courts have drawn between general causation and specific causation, many plaintiffs have attempted to introduce evidence of specific causation apart from their evidence of general causation. Very frequently this attempt takes the form of a physician or scientist testifying to an opinion of specific causation by a reasoning process commonly known in legal texts as "differential diagnosis," although in toxic tort cases it might better be called "differential etiology" as the objective is not to diagnose a disease but to infer its cause.

The reasoning process is simple enough to describe: an expert testifies that competing causes of the plaintiff's disease can be ruled out definitively, concluding that the toxic exposure, as the only potential cause not ruled out, must have caused the illness. Of course such testimony may appear in many gradations, depending on the extent to which the causes of plaintiff's disease have been identified and the confidence with which competing causes can be excluded. Despite the conceptual simplicity of the approach, attempts to introduce differential diagnosis testimony frequently have been challenged. In many cases these challenges have succeeded due to deficiencies in the challenged testimony.

Courts have uniformly agreed that differential diagnosis cannot, as a logical matter, establish general causation: even if all known causes of a plaintiff's disease can be ruled out, absent some proof of agent-disease causation it does not follow that exposure to the accused agent caused the plaintiff's illness. Thus courts have routinely refused to admit differential diagnosis testimony without sufficient admissible evidence of general causation as a predicate.[68]

Even when adequate evidence of general causation accompanies differential diagnosis testimony, the admissibility and sufficiency of the differential diagnosis may be attacked. As early as *Stubbs v. City of Rochester* (see the section on "The *Sine Qua Non*, But-For, or Necessary Element Test for Factual Causation"), courts understood that a plaintiff need not rule out every possible alternative cause—an impossibility if any cases of the disease in question are idiopathic. Conversely, if too high a proportion of a disease's incidence is unexplained, courts have been unwilling to accept that a differential diagnosis ruling out the limited known causes means much of anything. How much of the incidence of a disease must be explained, and how much of the explained incidence must be ruled out, have been the fault

lines on which battles over differential diagnosis testimony have typically been fought. Most difficult are those situations where a large percentage of injuries are from idiopathic causes. In these situations, even when an expert successfully rules out all other known causes, it still may be substantially more likely than not that an individual's injury was caused by something other than the agent involved in the lawsuit.[69]

Genetic Epidemiology

Genetic epidemiology uses biochemical markers—and in particular, genomic and epigenomic information—to increase the resolving power of traditional epidemiologic methods.[70] Thus a study sample might be divided into exposure groups based on measurements of biomarkers of exposure rather than work histories or recollection. Or a study sample might be divided into groups sharing certain genotypes to determine whether disease risk varies across genotypes (either with or independently of exposure). The collection of genetic information in epidemiologic studies has increased and is likely to accelerate as gene sequencing becomes faster and less expensive. As of this writing, very few court decisions in toxic tort cases have considered genetic epidemiologic evidence in any detail. Genetic epidemiology, together with toxicogenomics, may well present the next frontier in the development of toxic tort law, which we explore in the section on "Scientific Advances: Genetic Epidemiology and the 'Omics'."

Judicial Scrutiny of Expert Testimony

As described in the section on "The Role of Epidemiologists," the agent-disease causation questions that are typically critical in a toxic tort case cannot be resolved by lay inference based on common experience. Therefore courts routinely hold that expert testimony is essential to support an inference of agent-disease causation. Epidemiologists are frequently called upon to offer such testimony. Additional discussion of the role of the epidemiologist as an expert is provided in Chapter 5, *The Role of the Expert Witness*.

The judge presiding over a case determines whether or not evidence offered by a party may be admitted (presented to the jury). In recent decades courts have taken a more active screening role before admitting testimony of expert witnesses.

For 70 years, scientific expert testimony was assessed predominantly using the "*Frye* test," named after *Frye v. United States*,[71] a decision that rejected a criminal defendant's attempt to have an expert testify that the pattern of the defendant's systolic blood pressure while being questioned about the crime indicated that the defendant answered truthfully. The *Frye* test requires that for scientific expert testimony to be admitted, "the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field to which it belongs." Many state courts adopted a similar rule, but in point of fact rarely applied it in civil cases.

In a toxic tort case decided 70 years later, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,[72] the US Supreme Court held that the Federal Rules of Evidence, adopted in 1975, did not incorporate the *Frye* "general acceptance" test. In a series of three decisions beginning with *Daubert*, the Court articulated and refined a "gatekeeping" role for federal trial courts asked to admit scientific or technical expert testimony.[73] These decisions, and many lower court

decisions interpreting and applying them, allow a court to admit expert testimony if the opinion is "reliable" and "fits" the facts of the case. Most courts interpret the reliability assessment to entail an inquiry into the scientific validity of the data, method, or theory upon which the testimony is based, and the "fit" assessment to entail an inquiry into whether the expert has made reasonable inferences from the underlying data, method, or theory to the conclusion offered in a particular case.

In theory, these inquiries bear only on an issue of the law of evidence—whether the expert testimony is admissible—and not on whether the evidence is legally sufficient to prove the alleged causal relation. In practice, however, in deciding whether an expert will be permitted to testify, many courts explicitly or implicitly make a substantive legal decision of what type of proof is needed to permit an inference of toxic causation. For example, a plaintiff might offer only one expert witness on general causation, a toxicologist who would testify that a certain agent is capable of causing the plaintiff's disease, based on in vivo studies showing a statistically significant increased incidence of the disease in animal models given high doses of the agent. Whether that testimony is admitted could depend, in part, on whether the court concluded that such evidence alone is sufficient to permit the jury to draw an inference of causation.

The Supreme Court's "*Daubert* trilogy" is binding on all federal courts and has been very influential in the states. Roughly two-thirds of the states adopted the *Daubert* approach to expert testimony admissibility or decided to apply the reliability factors discussed in *Daubert* in analyses nominally using the *Frye* test.[74] Most of the remainder adhered to some version of *Frye*, albeit with many variations among "*Frye*" jurisdictions. Over time, there has been a slow but steady movement of states from the *Frye* column to the *Daubert* column.[75] Regardless of which test courts have applied after *Daubert*, courts seem to have become substantially more willing to exclude expert testimony of causation in toxic tort cases than they had been before *Daubert*.[76] Because the plaintiff bears the burden of proving that the defendant caused the plaintiff's harm and proof of causation depends on expert testimony, exclusion of all of the plaintiff's causation expert(s) effectively decides the case in favor of the defendant(s). Of course, even though courts have increased the stringency of their screening of expert witnesses, courts do decide to admit much expert testimony. The vast majority of cases end in settlements after decisions to admit the expert's testimony; a few are tried to jury verdicts.

JUDICIAL TREATMENT OF NONEPIDEMIOLOGIC CAUSATION EVIDENCE

Testimony by epidemiologists, or testimony based on the results of epidemiologic studies has figured prominently in toxic torts. It remains important, however, to understand how this testimony is situated within the context of all types of expert evidence introduced to prove causation in toxic tort cases.

In Vivo Animal Toxicity Experiments

In vivo animal testing, long a staple of toxicological research, has had a decidedly mixed reception in the courts. The most obvious issues are the validity of extrapolation from the test

species to humans and the validity of extrapolation from test doses to the estimated actual exposure of the plaintiff.

A few courts have stated or implied that extrapolation from animal studies is simply too suspect to support admissible expert testimony on disease causation in humans. More typically, expert opinions based on animal studies showing a causal effect are relegated to inadmissible status when the animal research conflicts with a substantial body of epidemiologic research that does not observe an increased incidence of disease associated with exposure in human populations. Expert opinions based on animal studies are also frequently excluded if the opinion depends on an additional extrapolation beyond the interspecies and dose extrapolations mentioned above. For example, courts have rejected opinions based on animal studies that showed agents caused diseases other than that suffered by the plaintiff or based on animal studies that involved a different route of exposure from the plaintiff's.[77]

On the other hand, opinions based on animal studies have been admitted despite the need for interspecies and dose extrapolations. Such testimony is more likely to be admitted if the expert can explain why human studies are unavailable or impractical or why extrapolation is appropriate. Such testimony is also more likely to be admitted if the results of animal studies are consistent with other evidence, particularly epidemiologic data.[78]

In Vitro Toxicity Experiments, Mechanistic Evidence, and Toxicogenomics

An inference of causation from in vitro toxicity testing requires extrapolation from the tested embryo, tissue, cells, or cellular components to whole living organisms, and may require dose or interspecies extrapolation as well. Courts have shown decided reluctance to accept causal inferences based on in vitro toxicity testing alone.[79] A complicating factor often has been present in court decisions excluding expert testimony based on in vitro testing, however. In many of these cases, the expert attempted to rely on in vitro experiments in which either the agent tested or the effect found was different from the agent to which plaintiff was exposed or the effect experienced by the plaintiff. Thus these cases may reflect judicial doubts about the analogy from agent to agent or from effect to disease rather than judicial reluctance to accept in vitro studies per se as a basis for expert opinion.

Evidence of a toxic agent's (or its metabolites') mechanism of action may be derived from in vitro studies or from research observing biochemical or cytological changes in vivo. Depending on the nature of the research, varying degrees of extrapolation to the plaintiff's exposure and condition may be required. As with in vitro testing generally, mechanism evidence alone has rarely been held adequate to support an admissible expert opinion of causation. More commonly, mechanism evidence is accepted as a basis for confirming that observed epidemiologic associations are biologically plausible and consistent with biological knowledge. Courts have admitted testimony based only on mechanistic evidence, however, when the testifying expert was able to explain the absence of other types of evidence and the basis for inferring causation from the mechanistic evidence. For example, an expert was permitted to testify that an overdose of the endometriosis drug Danocrine caused PPH based on studies showing that the drug affected the levels of hormones known to be involved in vasoconstriction.[80] In another case, an expert was permitted to testify that occupational exposure to benzene can cause acute promyelocytic leukemia (APL) based in large part on studies

showing that benzene is clastogenic and that chromosomal abnormalities are strongly associated with APL.[81]

The relative importance of in vivo and in vitro evidence depends on the quantity and quality of epidemiologic evidence relevant to the issue being litigated.[82] In the absence of epidemiologic evidence, cases may turn on the quality of these other types of evidence.

The emergence of toxicogenomics has created new opportunities for the observation of agents' effects on chromosome structure and gene sequences as well as epigenetic factors affecting gene expression such as methylation and gene switches within noncoding DNA regions. As of this writing, genomic and genetic information has played a role in only a few court decisions in toxic tort cases, but its use in litigation is increasing. We explore the implications for toxic torts of genomics, toxicogenomics, and other "omics" technologies in the section on "Scientific Advances: Genetic Epidemiology and the 'Omics'."

"WEIGHT-OF-THE-EVIDENCE"

The preceding sections on "Using Epidemiology to Prove Causation" and "Judicial Treatment of Nonepidemiologic Causation Evidence" describe the standards that courts have employed for assessing whether the output of diverse types of scientific research can support the admissibility of expert testimony. Plaintiffs whose diseases are not pathognomonic commonly have assembled a constellation of research results in support of their causation claims. They have frequently acknowledged that each result is subject to some of the methodological or inferential problems described above but have argued that taken together the scientific evidence is mutually reinforcing and sufficient to allow a testifying expert's causation conclusion to be submitted to a jury. Plaintiffs have argued that assembling imperfect scientific evidence in this way is consistent with the "weight-of-the-evidence" approach that regulatory agencies and their scientific advisors employ when assessing toxicity, for example, in classifying agents for carcinogenicity. Defendants have argued that regulatory agencies take a precautionary approach that is inappropriate for imposing ex post liability upon defendants for harm allegedly caused to individual plaintiffs. The result of these disputes depends heavily on judicial attitudes regarding the nature of science, the appropriateness of applying inference and judgment to scientific questions, and the epistemology of assembling information to reach a conclusion.

One approach to these issues was exemplified by the US Supreme Court in *General Electric Co. v. Joiner*.[83] The plaintiff, a smoker who suffered from lung cancer, alleged that his occupational exposure to PCBs had promoted his cancer. His experts relied on several animal and epidemiologic studies to support their opinions. The trial court dismissed the case after excluding all of the plaintiff's expert causation testimony, and the Supreme Court affirmed. The Supreme Court's decision examined each study individually and found that each had a methodological problem, a statistically insignificant result, or a way in which it was not directly comparable to plaintiff's situation (eg, a different amount and route of exposure for an epidemiologic study of workers in a different industry; a different dose, route of exposure, and histological cancer type for an animal study). Therefore, the Court held, the trial court had a valid reason to conclude that each study did not support an inference of causation in the plaintiff's case and thus it was within the trial court's discretion to exclude plaintiff's experts' testimony.

The Supreme Court did not hold that lower courts were required to apply this atomistic analysis, but the approach came to dominate toxic tort decisions in the federal courts and many state courts as well. Examples of courts taking a holistic approach to assemblages of scientific evidence have been few. A 2011 decision of the US Court of Appeals for the First Circuit, *Milward v. Acuity Specialty Products Group*,[84] explicitly endorsed the "weight-of-the-evidence" approach, reversing the judgment that the trial court entered in defendants' favor after excluding the testimony of the plaintiffs' experts on general causation. It remains to be seen whether that opinion presages greater judicial acceptance of "weight-of-the-evidence methodology" and, if it does, how courts will distinguish acceptable expert testimony based upon the weight of individually insufficient scientific studies from unacceptable expert testimony based on evidence that even collectively cannot support a causal inference.[85]

DEFENSES

The law provides a number of affirmative defenses that, if proven by the defendant, would partly or completely defeat a plaintiff's toxic tort claim even if the plaintiff proved all elements of the prima facie case. The four affirmative defenses most often encountered in toxic tort litigation are discussed below.

Statutes of Limitations

Statutes of limitations prescribe a period within which a victim must sue on penalty of losing the right to bring a claim. Different jurisdictions impose their own time periods for statutes of limitations, and the time provided may be different for different tort claims. Most tort statutes of limitations require suit no later than 1−3 years after some specified event that begins the limitations period.

The triggering event for some statutes of limitations is the defendant's tortious act—ie, the wrongful toxic exposure (or, analogously for warranty claims, when a product is delivered to its purchaser). When the time period begins to run in this fashion, a toxic tort suit involving a lengthy latency period may be barred before the victim suffers harm. Because the plaintiff's harm is an element of every tort, statutes of limitations so defined create a Catch-22, barring a claim before it ever came into existence. For example, when DES emerged as the cause of certain diseases a generation after victims' in utero exposure, the New York statute of limitations already barred virtually all the victims' claims. The state legislature modified the statute of limitations retroactively to "revive" these claims.[86]

The more common time to begin the statute of limitations clock running is when the victim suffers harm, which avoids the Catch-22 of a "tortious act" trigger but presents another problem: a victim might suffer a common disease and be completely unaware that its source implicates a toxic environmental agent for which someone is legally responsible. Because of this difficulty, courts adopted a "discovery" rule that delays triggering the statute of limitations until the time when the victim discovers certain relevant facts or when a reasonable person would have discovered them. The definitions of relevant facts that start the statute of limitations clock under discovery rules vary from jurisdiction to jurisdiction, ranging from

knowledge of the injury alone to knowledge of the injury and its connection to both the toxin and the defendant's tortious conduct that exposed the plaintiff to the toxin.[87]

Another adjustment made to statutes of limitations because of the unique characteristics of toxic torts applies to agents that cause multiple diseases with different latency periods. Asbestos is the paradigm with asbestosis having the shortest latency period, mesothelioma the longest period (sometimes extending to 50 years), and lung cancer with a latency period somewhere in between. An adjunct to statutes of limitations, the single judgment rule, requires that a plaintiff who has multiple claims against another person that arise from the same transaction or course of conduct must sue for all such claims at the same time; claims not asserted are lost. Thus, an asbestos victim who suffers asbestosis must bring suit, within the statute of limitations period, for all claims that arise from exposure to defendant's asbestos products. But, at that point, plaintiff does not know whether another asbestotic disease will appear in the future. In response to this problem, numerous courts have crafted a "separate disease" exception to the single judgment rule, applying the statute of limitations independently to each separate disease, so the expiration of the statutory period for one disease does not affect claims for other diseases.[88]

Epidemiology may be relevant to statute of limitation issues in some toxic tort cases, particularly when application of the discovery rule is in issue. Litigants have sometimes used the existence or absence of epidemiologic studies linking an exposure to a disease, or the gradual accretion of epidemiologic evidence, as indicators of the time when a reasonable plaintiff would have discovered the connection between the plaintiff's disease and the defendant's tortious act.[89]

Statutes of Repose

Statutes of repose were enacted in a number of states during one of the cycles of "tort reform" that have occurred approximately every 10 years since the 1970s. Like statutes of limitations, they bar a claim (or impose a less-severe sanction) for delay in filing suit, but they employ a very different trigger to start the clock. Statutes of repose are triggered by an act of the defendant, such as (for a products liability claim) the time of initial retail sale of a product that later causes harm. Typically the specified repose period is between 6 and 15 years after the triggering event. These statutes have been applied to limit suits against architects and builders of residential construction, health-care providers, product sellers, and government entities.[90]

Federal Preemption

The Supremacy Clause in the United States Constitution provides that federal law is supreme to state law. Thus, so long as the US Congress acts within its constitutionally delegated powers, it may "preempt" state law that would interfere with federal legislation and interests. Federal preemption has played an important role in toxic tort cases since it was first invoked in tobacco litigation in the 1990s.[91] For example, the Supreme Court held that state products liability claims involving medical devices subject to the most rigorous FDA premarketing scrutiny were preempted by a federal statute that negates any state requirement "which is different from, or in addition to, any requirement" under the federal

statute.[92] The Court reasoned that tort liability could impose a "requirement" to change the design of or warnings with a medical device. The Supreme Court has also found state tort claims preempted, at least in some respects, for cigarettes, generic drugs, vaccines and insecticides, fungicides and rodenticides.[93] By contrast, the Court found that suits against branded prescription drug manufacturers were not preempted, although it left open whether there could be some limited preemption in future cases that present different facts.[94]

Contributory and Comparative Fault and Assumption of Risk

These three related defenses are based on the plaintiff's conduct. Beginning in the early 19th century, courts denied recovery to a plaintiff who failed to exercise reasonable care for his or her own safety.[95] This rule of "contributory negligence" was frequently criticized as unfair for making the victim bear all of the loss occasioned by the fault of both the victim and the defendant. In the latter part of the 20th century, opposition to contributory negligence coalesced in courts and legislatures, and today 46 states have replaced contributory negligence with comparative fault.[96] As its name suggests, comparative fault apportions liability for an injury to the parties involved in proportion to their fault in causing the harm. Apportioning liability is a function assigned to the jury.

In some states, comparative fault applies to many or all of the different tort claims that might be asserted. Other states apply comparative fault to negligence claims only. Other differences exist among the states' comparative fault systems as well.

Assumption of risk also addresses plaintiff's conduct, but this affirmative defense is distinct because it requires that the plaintiff knowingly and voluntarily confront the risk that results in plaintiff's harm. The voluntariness requirement leaves some wiggle room for juries and eventually resulted in a doctrine that requires the plaintiff's conduct be unreasonable as well. Thus, it might appear that a person who chose to drive a car despite knowing that the car had just been recalled because of a risk of engine fires had assumed the risk of getting burned. But if the fire occurred while she was driving her child with a medical emergency to the hospital and she had no other way of getting help for the child, assumption of risk would not bar recovery because use of the car in those circumstances was not unreasonable. Today, assumption of risk rarely completely bars recovery because it is treated as an aspect of comparative fault.

These defenses based on plaintiff's conduct have played a limited role in toxic tort cases. Consumers of pharmaceuticals rarely act unreasonably in doing so. Those who are involuntarily and unknowingly exposed to toxic agents also do not act unreasonably. Yet, in the first asbestos case that resulted in liability, the jury found that the plaintiff had been contributorily negligent, likely because he did not use respirators that were made available to industrial insulation workers like him.[97] That finding barred recovery on plaintiff's negligence claim, but he recovered on his products liability claim because, at the time, contributory negligence was not a defense to a products liability claim in the jurisdiction.

Another area where plaintiff's conduct has played a significant role is the use of tobacco products. Not only may the plaintiff be held wholly or partly responsible for injuries caused by tobacco, the plaintiff's tobacco use may become relevant when a disease can be caused by smoking and by other toxic substances. One of the most frequently litigated issues of this type is the role of smoking versus asbestos exposure in causing an individual's lung cancer. If a

court does determine that a plaintiff's smoking and the defendant's asbestos both played a role in the development of the plaintiff's injury, it is then confronted with the difficult question of how to apportion liability for the injury. Cases often rely on epidemiologic evidence to make this apportionment.[98]

Epidemiology may in various ways affect the defenses that are based on plaintiff's conduct. Epidemiology may be relevant to the issue of whether plaintiff's allegedly unreasonable conduct was a cause of plaintiff's illness. The amount of publicity given to epidemiologic studies that associate plaintiff's allegedly unreasonable conduct with plaintiff's illness may be relevant to the issue of whether plaintiff's conduct was, in fact, unreasonable.

SPECIAL TYPES OF TOXIC TORT LITIGATION

Claims Involving Pharmaceuticals

Because pharmaceutical drugs are intended to be biologically active, their therapeutic effects are often inseparable from undesirable side effects. Products liability law has long recognized that because drugs cannot for the most part be redesigned to remove risks while retaining their benefits, design defect claims are largely unavailable for prescription drugs.[99] Thus, improper warning, rather than defective design, has been the primary basis for products liability claims involving pharmaceuticals. There have been a few exceptions, including combination drugs where one active ingredient does not contribute to efficacy and drugs with dangerous inactive ingredients. The latter situation was highlighted in the 1937 sulfanilamide tragedy. The manufacturer suspended an antibiotic in liquid to facilitate administration to children. The liquid selected, diethylene glycol, resulted in some 100 deaths and additional injuries.[100]

Claims Covered by Workers' Compensation and Federal Employers Liability Statutes

Occupational injuries and diseases are subject to a special compensation scheme adopted in the early 20th century. In the latter part of the 19th century the industrial revolution took its toll on workers but a variety of judge-made tort law doctrines precluded successful claims by workers against their employers. Workers' compensation was developed to reverse this poor treatment of occupationally-injured workers. A grand bargain was forged: employees injured in the course of their employment could recover regardless of employer (or employee) fault, but the damages they could recover were more limited than those provided by tort law. Only medical expenses, a portion of lost wages, and payment for temporary and permanent disability could be recovered. The bargain also provided that workers' compensation was the "exclusive remedy" for an injured worker, which prevented the injured worker from suing the employer in tort, even if the employer had acted negligently in injuring the employee.[101]

The exclusive remedy provision, however, does not prevent an injured worker from suing third parties whose tort was a cause of the employee's harm. Thus, the manufacturer of an industrial machine without adequate safeguards could be sued when an employee was injured by the machine in the course of employment. Similarly, those exposed to asbestos

products at work who developed asbestotic disease could sue the manufacturers of those products as well as recover workers' compensation from the employer.[102] A system for adjusting losses among employee, employer, and third party prevents the employee from receiving double recovery and often reimburses the employer for its compensation payments.

The Federal Employers Liability Act (FELA)[103] was approved over a century ago, at the beginning of the states' adoption of no-fault workers' compensation statutes. It was motivated by the injury toll that railroading had on those employed in the industry and was designed to make compensation more available to those who were harmed. The FELA is a hybrid between a fault-centered tort system and workers' compensation. It retained a "lite" version of tort law, with more lenient standards for plaintiffs with respect to issues of fault, causation, and defenses that had posed severe obstacles to recovery. As the Supreme Court said in a leading case in 1957, all a plaintiff need do to satisfy the causation element is to show the railroad's negligence "played any part, even the slightest, in producing the injury."[104]

Workers' compensation initially had difficulty accommodating occupational disease because the model injury for the system was a sudden traumatic injury. However, that difficulty has been resolved and today the primary difficulty is determining whether the employee's disease was a product of workplace exposure or a nonoccupational cause. Epidemiology is brought to bear on that issue much as it is applied to the causation issue in tort claims.[105]

Claims Against Government Entities

Government agencies may be involved with toxic materials, such as radiation or radioactive material released from a government site. Based on the principle that the King can do no wrong, suits against the government were barred by "sovereign immunity," a rule that crossed the Atlantic when US settlers first established a legal system based largely on British law. This broad immunity has been eliminated or limited, usually by statute, in nearly all American jurisdictions.[106]

At the federal level, the Federal Tort Claims Act ("FTCA")[107] permits the government to be sued in the same fashion a nongovernmental entity could be, but carves out a number of exceptions that result in immunity being retained. For example, under the FTCA, the federal government remains immune to claims based on strict liability theories and to claims based on a government employee's exercise of a "discretionary function" (one which requires social, economic, or political determinations).[108] As interpreted by the Supreme Court, the FTCA also bars claims against the federal government that arise in the course of military service.[109] Thus, in one of the earliest cases to highlight the importance of epidemiology to claims of toxic causation, service members who were exposed to the defoliant Agent Orange in Vietnam could pursue claims against the manufacturers but not against the federal government.[110]

Claims Under the Childhood Vaccine Act

The Childhood Vaccine Injury Act of 1986 is designed to provide compensation to those who suffer adverse vaccine reactions but also to reduce the costs resulting from high tort

damage awards. This Act establishes a no-fault system of limited compensation funded through a tax paid on each dose of vaccine sold by vaccine manufacturers.[111]

Vaccine victims do not have to prove negligence or another tort by the vaccine manufacturer but proof of factual causation is required. Certain adverse effects that have been scientifically established and connected to particular vaccines are contained in a table and causation in such instances is presumed. However, the victim bears the burden to prove causation for any adverse effects not contained in the table.[112] Although the courts have held that epidemiologic evidence is not required to prove such an allegation,[113] epidemiologic evidence (or its absence) often plays an important role in decisions of the special court that hears vaccine claims.[114]

Claims Resulting in Bankruptcy or Against Bankrupt Entities

Individuals and corporations with debts that are greater than their assets or whose cash flow does not permit timely payment of financial obligations may choose (or be forced) to resolve their financial difficulties in bankruptcy. Bankruptcy has been employed, predominantly by asbestos product manufacturers, to resolve tort claims that the companies were unable to resolve in the ordinary course of their business. Since Johns-Manville filed for bankruptcy in 1982, well over 100 companies have employed bankruptcy to resolve the asbestos claims against them. When it filed for bankruptcy, Johns-Manville estimated that its current and future asbestos-tort obligations were $2 billion. The bankruptcy proceedings resulted in a reorganization plan that freed Johns-Manville from asbestos claims but created a trust to resolve the future claims of those who were exposed to Johns-Manville asbestos products. The total value of the trust was around $2.5 billion. It provided compensation to all existing and future claimants against Johns-Manville based on schedules that employed several criteria, such as the type of disease and age of the victim. Other asbestos product manufacturers who have filed for bankruptcy have followed, more or less, the template created in the Johns-Manville bankruptcy.[115]

Claims Seeking Compensation for Increased Risk of Disease

Some exposed to a toxic agent that increases the risk of contracting disease in the future have brought suit seeking to recover for the increased risk of harm—presumably instead of suing at some later date when the disease actually develops.[116] From an actuarial standpoint, the defendant's overall liability should not be affected by paying a greater number of discounted damage awards today as opposed to a lesser number of larger awards in the future. In effect, the defendant is providing funds by which the plaintiff insures against contracting the disease in the future. A practical concern about such a system, however, is the administrative costs required to resolve a larger number of suits. These administrative costs for example, attorney's fees, expert fees, and court costs are already quite high, consuming approximately 50 cents of every dollar paid by a defendant. Perhaps to avoid unnecessary transaction costs or because courts resist the use of tort law as a mechanism for providing insurance coverage, courts have not been sympathetic to suits by plaintiffs who have not yet suffered harm but are seeking to recover for the risk of future harm.

In addition to the risk of future disease, an exposed person may suffer the distinct injury of emotional harm at the prospect of contracting that disease in the future.[117] This is often labeled "pure emotional harm," because it is not a consequence of physical injury. American courts have been cautious about permitting recovery for pure emotional harm because it is widespread and not so clearly bounded as physical harm (consider the pure emotional harm that resulted from the events of 9/11, even in contrast to physical harm that includes alleged toxic injuries of neighbors and first responders). That caution has been modestly relaxed in recent years with courts recognizing limited circumstances (for example, a false diagnosis of HIV infection[118]) in which recovery is permitted. Nevertheless, a strong majority of courts have not extended that exception to fear of future disease, and the US Supreme Court adopted the traditional approach to pure emotional harm. Ruling in an FELA case applicable to railroad workers, the Court held that workers exposed to asbestos who were nonsymptomatic could not recover for their fear of future asbestotic disease.[119]

Claims Seeking Medical Monitoring

Toxic tort plaintiffs sometimes seek to recover the cost of tests for latent diseases that may occur in the future as a result of a toxic exposure. These costs are generally known as medical monitoring or medical surveillance damages. Medical monitoring damages are not controversial when they relate to future management of an exposure-caused disease or condition that the plaintiff has at the time of suit. In such a case, the cost of monitoring or diagnostic tests that are reasonably needed is simply a component of future medical expenses, which courts have long recognized are an element of recoverable damages. Courts have not reached consensus, however, on the correct treatment of claims for medical monitoring when the plaintiff is not suffering from clinical disease but alleges that the enhanced risk of future illness caused by the exposure warrants surveillance for which the defendant should pay. These "medical monitoring claims" are distinct from claims for damages for exposure-caused emotional distress or "cancerphobia" or for the increased risk of disease per se.

In the absence of manifest disease allegedly caused by the exposure, courts generally treat medical monitoring claims as attempts to recover for a distinct tort in which the injury, sometimes characterized as economic rather than physical harm, is the need to submit to and pay for additional medical testing. The response of US jurisdictions to medical monitoring claims falls into three roughly equal groups. In about one-third of the states, courts have allowed plaintiffs to pursue medical monitoring claims.[120] In about one-third of the states, courts have refused to allow medical monitoring claims.[121] And in about one-third of the states there are no reported decisions at the time of this writing.

Courts that have allowed medical monitoring claims have defined the elements of those claims in a variety of ways but with certain similarities. Most have required the plaintiff to allege a toxic exposure that is in some way greater than the toxic exposures daily experienced by the population at large. Most have also required the plaintiff to demonstrate that the exposure caused increased risk of one or more diseases such that the plaintiff requires medical monitoring or screening above and beyond what would normally be counseled for a similarly situated plaintiff who had not been exposed. The exact formulation of the degree of exposure

and risk that will satisfy the legal standard varies from state to state. Finally, some courts also have required the plaintiff to demonstrate the utility of the monitoring, ie, that monitoring will likely lead to earlier detection of the disease in question and will improve the plaintiff's prognosis, symptoms, or treatment.

Toxic exposures that lead to medical monitoring claims sometimes affect numerous people, as when an aquifer used for drinking water has been contaminated or a spill or accident has affected a workplace or a neighborhood. Plaintiffs often seek to pursue such cases as a class action on behalf of all persons exposed. Medical monitoring class actions are controversial and have often been rejected, even in states that otherwise allow medical monitoring claims, on grounds that the individual exposures and risks are too variable to be litigated in common.

THE FUTURE OF EPIDEMIOLOGY IN TOXIC TORTS

Scientific Advances: Genetic Epidemiology and the "Omics"

As noted above in the sections on "Genetic Epidemiology" and "In Vitro Toxicity Experiments, Mechanistic Evidence, and Toxicogenomics," toxicogenomics, other "omics" methodologies, and genetic epidemiology today provide unprecedented tools for elucidating chronic toxicity. This section discusses the potential applicability of these sciences to toxic tort claims, the limited judicial attention they have received to date, and the possible legal issues their use might raise.

Implications for General Causation

Toxicogenomics and related techniques study how exposure to suspected toxins interacts with variable genetic material to produce, or not produce, toxic effects.[122] Exposure may result in direct alteration of coding DNA sequences, in alteration of epigenetic factors, or in alteration of gene expression. Such alterations may serve as biomarkers of exposure or, if the alterations indicate or accompany clinical manifestations, as biomarkers of effect.

Valid biomarkers of exposure or effect could importantly supplement other evidence of general causation in toxic tort cases. Conceivably, as with other evidence of a toxic agent's mechanism of action, toxicogenomic evidence could suffice to support a finding of general causation even in the absence of epidemiologic support, particularly if limitations of power or other reasons explain the lack of epidemiologic evidence.

Alternatively, toxicogenomics could identify biomarkers of susceptibility—genetic or epigenetic variations that alter an individual's risk of developing disease after toxic exposure. Such variations could also be detected by genetic epidemiology studies, which extend the methods of classical epidemiology by considering genetic variability or using biomarkers to measure at least some properties of the sampled population.[123] The ability to take into account variations across genotypes in susceptibility to the toxic effects of exposure may be among the most important implications of these techniques to toxic tort cases. Discerning such differences could provide vital information in a case of an agent-disease

link that yields no association or a small association when investigated without accounting for genetic variability. Some studies have shown, for example, that the degree to which exposure to tobacco smoke increases breast cancer risk varies substantially across genotypes of the *NAT2* gene.[124]

Of course, if toxicogenomic studies observe no effect of an exposure, or genetic epidemiology studies find no association between exposure and disease independent of genotype, defendants would use the research to argue against a finding of general causation. Regardless of whether a plaintiff or defendant introduces expert testimony based on toxicogenomic or genetic epidemiologic studies, however, all involved—testifying experts, attorneys, and judges—should appreciate that such studies will not necessarily eliminate uncertainty and controversy concerning general causation. Questions may still be raised about the inferences needed from experimental toxicogenomic systems to actual people and real-world exposures, and about inadequate power, bias, confounding, and other methodological difficulties that may afflict genetic epidemiologic studies no less than traditional epidemiologic research.

Implications for Specific Causation

At least in theory, biomarkers identified through toxicogenomic and genetic epidemiologic research could provide the evidence that has been most strikingly unattainable in toxic tort cases, even when general causation has been established: convincing proof distinguishing whether a particular exposed plaintiff's disease was caused by the exposure or merely coincident with the exposure. To serve this purpose, the ideal biomarker would be both perfectly sensitive and perfectly specific; that is, it would always appear in cases of disease caused by the exposure in question and would never appear in other cases. The frequency with which this ideal will be obtained remains to be seen.[125]

But even if ideal biomarkers turn out to be scarce, toxicogenomics and genetic epidemiology may provide new types of evidence relevant to specific causation. Valid quantitative biomarkers of exposure (and to some extent even qualitative biomarkers of exposure) could reduce the uncertainty surrounding dosage that is encountered in many types of toxic tort claims. Biomarkers of susceptibility would also be useful: an exposed plaintiff with a genotype conferring heightened susceptibility to toxic effect would be more likely to succeed in proving specific causation, and vice versa. Thus the presence or absence of even imperfectly sensitive and imperfectly specific biomarkers of effect could nevertheless support inferences about specific causation.

The question confronting all these potential uses of toxicogenomics and genetic epidemiology in toxic torts will be how strongly the scientific research supports the desired inference about specific causation. How precise and reliable is a marker of exposure? How much more or less susceptibility to the toxin does the plaintiff's genotype confer? How sensitive and specific is a marker of effect, and what inferences can be drawn from its presence or absence? Is a marker of exposure or effect durable enough to be observable at the time of suit, often many years after the exposure? As courts confront these issues they will need to develop rules of admissibility and sufficiency of evidence, much as they have done with respect to toxicological and epidemiologic evidence to date.

Accumulating research in toxicogenomics and genetic epidemiology inevitably will find its way into toxic tort courtrooms. To date, however, relatively few litigants have relied on

such research. Ironically, in light of the difficulty specific causation can pose for plaintiffs and the potential for genetic-scale data to overcome that difficulty, to date defendants have made more use of toxicogenomics and related fields.[126] For example, in claims involving benzene exposure and various lymphomas or leukemias, defendants have argued that the testimony of plaintiffs' causation experts should be excluded if the plaintiff's cells do not display particular chromosomal aberrations statistically associated with benzene exposure.[127] In at least one case with a reported decision, an expert for a cigarette manufacturer testified that the plaintiff's lack of mutations associated with tobacco smoke in the *p53* and *k-ras* cancer suppression genes was evidence that the tobacco smoke did not cause the plaintiff's lung cancer.[128] Defendants in another case successfully introduced expert testimony that the gene expression pattern in a plaintiff's tumor was consistent with sporadic rather than radiation-induced thyroid cancer.[129]

Defense efforts to invoke the plaintiff's own genes as a competing cause of disease also seem to be increasing in frequency. The US government has been the most prominent proponent of this argument. In many cases under the vaccine compensation program (see the section on "Claims Under the Childhood Vaccine Act"), the government has asserted that the vaccine recipient suffered from an illness apparently caused by a high-penetrance variant allele, sometimes identified only after a compensation claim was filed.[130] More nuanced assertions that genetic risk factors or "predispositions" are the actual cause of plaintiff's disease have also begun to appear.[131] These arguments are much more apt to succeed if they are based on known, specific genetic traits of the plaintiff than on the general existence of genetic risk factors.

Given the popular mystique surrounding all things genetic, it is all too easy to imagine that a judge ruling on admissibility or a jury finding facts would accept uncritically an expert witness's assertion that genetic-level analysis has provided a definitive scientific answer to the specific causation question, yea or nay. Cases may arise in which all parties would accept such a conclusion. But attorneys for both plaintiffs and defendants, and their retained experts, will likely learn to question such testimony. Despite the potential of toxicogenomics and genetic epidemiology to provide more fine-grained information, their use in toxic tort cases is likely to bring its own controversies.

It is not yet clear whether plaintiffs could be required to submit to genetic or gene expression testing upon defendants' demand or what inferences courts will permit juries to draw from the fact of a plaintiff's refusal.[132] Even when test results are available, the parties to toxic tort cases may dispute whether studies of genetic contributions to disease risk reflect independent risk factors or toxic susceptibility; genomic studies rarely control for exposure to particular toxins or to toxins generally.[133] The parties may dispute whether mutation or gene expression patterns actually reflect sufficiently sensitive and specific "signatures" that allow discrimination between disease etiologies.[134] They may also dispute whether genetic epidemiology studies have adequately characterized the pertinent genetic differences or controlled for interactions among genetic, epigenetic, and environmental factors. Such disputes could be significant for the many toxic susceptibilities that are likely influenced by multiple genes, noncoding DNA gene switches, other epigenetic factors, and other environmental conditions. Finally, litigants will likely raise issues of power, bias, and confounding, just as they have with respect to classical epidemiology.

A case that originated in the massive class action against the manufacturers of certain diet drugs gives a taste of the issues that may arise. A settlement between members of the plaintiff class and one of the manufacturers resolved most claims but allowed suit by any class member who had taken the drugs, had developed PPH, and had ruled out familial causes of the PPH. One particular plaintiff had been adopted and had no information about the incidence of PPH among his biological relatives. The defendant argued that the plaintiff should undergo testing for a particular genetic mutation that is present in 70% of individuals diagnosed with familial PPH. The plaintiff argued that testing would mean little because 80–90% of persons with the mutation nevertheless do not develop PPH.[135]

Surely, the results of genetic testing on the plaintiff with PPH would have been relevant to the causation issue in his case. But, assuming the statistics given by the parties were both accurate, the results would hardly have been determinative. A judge would have had to decide whether the results were admissible, and if so, a jury would have had to decide their significance. If the difficulties that would have faced judge and jury sound much like the difficulties that judges and juries have faced in the era before genomics, that is because in this instance they are.

Continuing and New Legal Issues

Although courts in the United States have gotten past their reluctance to employ epidemiologic research for fact-finding in individual cases, the tension courts feel between the law's need to resolve individual cases and epidemiology's inherent focus on statistical properties persists. This section considers some of the ways in which that tension manifests and will continue to manifest itself.

Relative Risk Thresholds, Statistical Significance, and Power

As noted in the section on "Relative Risk Less Than 2.0," courts have sometimes equated a greater than doubling of risk in a sampled exposed population with a more likely than not probability of causation in an individual case. Despite scholarly criticism, this equation has persisted and slowly spread,[136] with less rigid application in some jurisdictions than in others, and with flexibility sometimes more apparent than real.[137]

In quite a number of jurisdictions, the courts have not yet spoken clearly on this issue. Even among those courts that have done so, new cases in different contexts may spark reappraisals.[138] Also, as toxicogenomics and genetic epidemiology produce more fine-grained risk information, courts may need to reconsider the appropriateness of relative risk thresholds and of the correct baseline incidence to which a plaintiff's risk should be compared.[139]

Epidemiology, "Fit" and Accounting for the Individual

The group nature of epidemiologic data necessarily forces courts to decide how similar a plaintiff must be to the members of a group that was the subject of epidemiologic study. As discussed in the section on "Judicial Scrutiny of Expert Testimony," courts answer this question in various legal contexts: when deciding whether a study "fits" the plaintiff well enough to allow an expert to rely on it as a basis for an admissible opinion, when deciding whether the proof of causation is sufficient to support a jury verdict that causation exists, or when making both decisions simultaneously.

The problem may present itself in at least two ways. Epidemiology may find a magnitude of risk that exceeds a jurisdiction's standard of proof, yet the plaintiff's exposure or other characteristics may suggest that the plaintiff was at lower risk than the exposed group in the epidemiologic research. Or a plaintiff's characteristics may suggest that the plaintiff was at higher risk than the exposed group in an epidemiologic study with results that did not exceed the jurisdiction's standard.

In the former scenario, courts have been leery of admitting challenged testimony or allowing the case to reach a jury. For the latter scenario, the Third Restatement of Torts recognized the possibility that the likelihood of specific causation could be "refined," beyond epidemiologic results, based on evidence of facts particular to an individual plaintiff.[140] Courts have been slow to embrace this possibility,[141] but they have begun to acknowledge that genetic variability may affect individual toxic susceptibilities,[142] and at least one court has allowed an expert to testify to a "specific odds ratio" based on genetics and other individual factors of the plaintiff.[143] The continued development of toxicogenomics and genetic epidemiology is likely to bring more attention to "personalized" measures of relative risk. It remains to be seen how completely and persuasively these sciences will allow such measurements.[144]

Thresholds, Single Hits, and "Any Exposure"

Courts have long understood well the toxicological truism that "the dose makes the poison."[145] But for certain types of toxicity, no dose low enough to have no adverse effects has yet been identified. Thus, for example, for some carcinogens government regulators use risk assessment models that implicitly assume that no safe dose threshold exists.[146]

Assertions of "no safe threshold" have featured prominently in recent mesothelioma cases involving defendants who manufactured products to which a plaintiff was casually or intermittently exposed, such as drywall mud or brake linings. In such cases, plaintiffs' experts typically testify in some way—the exact formulations vary—that every exposure to asbestos above background concentrations was a cause of, or contributed to the risk of, mesothelioma. The results have been somewhat inconsistent, but in general such testimony has not fared well in court.[147]

In low-dose cases involving multiple exposures, a few courts have used language that seems to reject the concept of linear no-threshold models of carcinogenesis.[148] Pending further scientific development of dose-response curves, the judicial treatment of low-dose exposures in mesothelioma and other cases will reflect a legal, rather than a scientific, determination. In many jurisdictions that development depends on judicial interpretation of the "substantial factor" test of causation, discussed in the section on "'Substantial Factor' Causation."

"Substantial Factor" Causation

The first Restatement of Torts, published in 1934, defined a "legal cause of harm" as conduct that "is a substantial factor in bringing about the harm."[149] The substantial factor formulation arose in response to the problem of overdetermined outcomes, as in the case of the combining fires described in the section on "The *Sine Qua Non*, But-For, or Necessary Element Test for Factual Causation." The traditional legal test of "but for" causation does not work well in such cases, because each cause independently is sufficient to bring about the

plaintiff's harm and therefore it could not be said that "but for" either cause plaintiff would not have been harmed. Many courts adopted the substantial factor formulation, which was retained in the Second Restatement of Torts in 1965.[150]

In routine cases the substantial factor test is indistinguishable from the *sine qua non* or "but for" description of causation.[151] But in difficult cases, including toxic torts, courts have sometimes treated the substantial factor test as an alternative to *sine qua non* causation.

A relatively small number of courts have invoked "substantial factor" to make it possible for plaintiffs to establish the factual causation element of their cases despite scientific indeterminacy that made it impossible to prove that but for defendant's conduct plaintiff would not have been injured.[152] To overcome the "irreducible uncertainty" of determining which of several asbestos products caused a plaintiff's cancer, for example, the California Supreme Court held that each product that "was a substantial factor contributing to" the *risk* of developing cancer would be considered a "substantial factor in causing or bringing about the disease."[153] This type of change in causation doctrine need not necessarily depend on the "substantial factor" formulation.

More often—as in many of the "every exposure" asbestos cases—courts have treated "substantiality" as an additional requirement, barring recovery against a particular defendant whose causal contribution to plaintiff's harm is not sufficiently large to satisfy the court. In an asbestosis case, for example, the Texas Supreme Court noted that popular notions of responsibility, and not only philosophical notions of "cause," lurk in the term "substantial."[154] Used in this way, the substantial factor test reflects more of a policy judgment than a method of deciding causation-in-fact.

Legal scholars have criticized courts' inconsistent and confusing use of the substantial factor test.[155] The Third Restatement of Torts eschews the term entirely.[156] Nevertheless, "substantial factor" is a phrase deeply ingrained in the American judicial tradition, and so far courts have been reluctant to abandon it.[157]

"Reasonable Medical Certainty": Square Scientific Pegs in Round Legal Holes

Expert witnesses (including epidemiologists), as described in the sections on "The Role of Epidemiologists" and "Judicial Scrutiny of Expert Testimony," above, uniquely are permitted to testify to their opinions about issues such as causation rather than being limited to testimony about their factual knowledge of the case. To admit such opinions, while giving some credence to the understanding that experts rarely testify about incontrovertible scientific truths, many courts require scientific or medical experts to state that opinions are held to a "reasonable degree of medical [or scientific] certainty [or probability]."[158]

This formulation has heuristic appeal to lay judges and juries, but as the Third Restatement of Torts explains, it is problematic for testifying scientists and physicians as well as for the legal system.[159] A "reasonable degree of scientific certainty" has no particular meaning to a scientist.[160] Moreover, this standard seems to mesh poorly with the "preponderance of the evidence" standard of proof, which would be challenging to apply to toxic tort causation issues even without the added confusion contributed by "reasonable degree of certainty." Most courts insist that the required level of certainty is no different from the preponderance standard,[161] but the two standards do not sound the same. And if a court were to construe "reasonable scientific certainty" to require greater certainty than "preponderance of the

evidence," the anomalous result would be to exclude from evidentiary consideration a range of opinions that were nevertheless held to a greater degree of certainty than required by the ultimate standard of proof.

To avoid these problems, the Third Restatement of Torts abjures the phrase "reasonable degree of medical [or scientific] certainty [or probability]," just as it does the phrase "substantial factor." Instead, the Third Restatement adopts "preponderance of the evidence" as the single standard for admissibility of expert causation testimony and the ultimate burden of persuasion.[162] Nevertheless, as with substantial factor, courts have been slow to relinquish a long-standing traditional verbal formula.

CONCLUSION

The emergence of toxic torts in the latter part of the 20th century created many challenges for the legal system and tort law. Traditional doctrines applied to traumatic injury cases have required reassessment in dealing with long-latency diseases whose causal mechanisms are not well understood. These cases have required courts to become more sophisticated about the sciences, including epidemiology, that provide evidence about causation—a central issue in private tort litigation. The flexibility of the common law—judge-made law on a case-by-case basis—has already produced several modifications of traditional doctrine for this new class of torts. Yet challenges in providing fair treatment of both sides in toxic tort cases remain and are not likely to end in the foreseeable future.

ENDNOTES

1. Portions of this chapter are drawn from Steve C. Gold & Michael D. Green, *Toxic Torts*, *in* 4 Encyclopedia of Toxicology 770 (3d ed. Philip Wexler ed. 2014).
2. The best contemporary treatment of tort law is Dan B. Dobbs et al., The Law of Torts (2nd ed. 2011).
3. For additional elaboration on the procedure in a civil case, see Jack H. Friedenthal et al., Civil Procedure Cases and Materials 4–20 (10th ed. 2009).
4. There are a variety of legal theories that are employed in toxic torts. We provide a brief explanation of them *infra* text accompanying notes 18–23.
5. For example, claims by veterans who alleged that they were sickened by exposure to Agent Orange required expert witnesses to explain that phenoxy herbicides sprayed during the Vietnam War were contaminated with dioxins as a result of the manufacturing process and to opine as to whether dioxin-contaminated herbicides caused the diseases about which military personnel complained. *See* In re "Agent Orange" Product Liability Litigation, 611 F. Supp. 1223 (E.D.N.Y. 1985). Expert testimony at trial is subject to judicial approval as explained in the section on "Judicial Scrutiny of Expert Testimony."
6. For example, pursuant to its authority over interstate commerce, Congress enacted the Federal Employers Liability Act, which addresses injuries to employees of interstate railroads in the course of their employment. Thus, in Metro-North Commuter R. Co. v. Buckley, 521 U.S. 424 (1997), the US Supreme Court decided that railroad workers who were exposed to asbestos but had not yet developed any asbestotic disease could not recover for their emotional harm due to the future risk they would develop such diseases. Admiralty is another subject specifically assigned to Congress and liability for injuries in that realm is governed by federal law. *See, eg*, Exxon Co., USA v. Sofec, Inc., 517 U.S. 830 (1996) (suit in admiralty by owner of oil tanker that ran aground right after its "breakout" from mooring facility against facility owner and operator).
7. 58 N.E. 2d 754 (Mass. 1945).

8. *Id*. at 755 (quoting *Sargent v. Massachusetts Accident Co.*, 29 N.E. 2d 825, 827 (Mass 1940)).
9. 124 N.E. 137 (NY 1919).
10. Samuel D. Estep, *Radiation Injuries and Statistics: The Need for a New Approach to Injury Litigation*, 59 Mich. L. Rev. 259, 262 (1960). Estep's article contains a prescient cataloging of legal issues that courts have confronted and resolved in the following half century.
11. *See* David E. Lillienfeld & Paul D. Stolley, Foundations of Epidemiology 22–33 (3d ed. 1994).
12. *See* 3 Modern Scientific Evidence § 23.1, at 224 (David L. Faigman et al. eds, 2013–14). ("The rise of epidemiology has accompanied the rise of toxic torts.")
13. Ken Rothman, in the first edition of his text on epidemiology published in 1986, wrote "a systemized body of epidemiologic principles by which to design and judge such studies has begun to form only the in the last two decades." Kenneth J. Rothman, Modern Epidemiology 1 (1986).
14. Interestingly, however, no epidemiologic study examined the connection between thalidomide and birth defects. *See* Michael D. Green, Bendectin and Birth Defects 72 (1996).
15. *See* 3 Modern Scientific Evidence § 23.3, at 230 (David L. Faigman et al. eds, 2013–14) ("epidemiology usually has enjoyed a privileged position in toxic tort cases"); Michael D. Green et al., *Reference Guide on Epidemiology* 551, 551 n. 2, *in* Federal Judicial Center & National Research Council, Reference Manual on Scientific Evidence (3d ed. 2011) ("Epidemiologic studies have been well received by courts deciding cases involving toxic substances."); Andrew S. Lipton, *Proving Toxic Harm: Getting Past Slice and Dice Tactics*, 45 McGeorge L. Rev. 707, 737 (2014). ("Epidemiology is often essential to proving a toxic tort case.")
16. *See* the section on "Judicial Scrutiny of Expert Testimony." The most extreme such example is a Bendectin case in which the court demanded statistically significant epidemiologic evidence connecting limb reduction birth defects with maternal ingestion of Bendectin. Brock v. Merrell Dow Pharms, Inc., 884 F.2d 166, 167 (5th Cir. 1989).
17. Thus, "negligence" has two different meanings: (1) the name of a type of tort involving unintentional wrongdoing; and (2) one of the elements of such a tort claim: failing to exercise reasonable care to avoid harm to others.
18. *See* Restatement (Second) of Torts § 402A (1965); John W. Wade, *On the Nature of Strict Tort Liability for Products*, 44 Miss. L.J. 825, 829 (1973).
19. We omit here the complication that there may be a series of sales from the manufacturer of the product through the distribution chain and ultimately to the consumer. For elaboration, see Dobbs et al., *supra* note 2, at § 450, at 892–894; U.C.C. § 2-318.
20. U.C.C. § 2-314.
21. *See, eg*, Cipollone v. Liggett Group, Inc., 893 F.2d 541 (3d Cir. 1990) (plaintiff awarded a $400,000 judgment based on breach of warranty while tort claim based on failure to warn was unsuccessful because of affirmative defense based on smoker's conduct), *aff'd in part and rev'd in part on other grounds*, 505 U.S. 504 (1992).
22. One of the first cases to employ abnormally dangerous activity in the toxic tort context is *State Dept. of Environmental Protection v. Ventron Corp*., 468 A.2d 150 (N.J. 1983). *See also* Restatement (Third) of Torts: Liability for Physical and Emotional Harm, rptrs. note to § 20 cmt. k (2010) (reporting "that the abnormally dangerous doctrine has a significant application in the context of environmental harms").
23. Most invocations of nuisance in toxic tort litigation entail private nuisance. But another strand, public nuisance, has been employed in suits against lead pigment manufacturers to require them to contribute to cleanup efforts. By and large, these efforts have been to no avail. *See* State v. Lead Industries Ass'n, 951 A.2d 428 (R.I. 2008).
24. One exception to the necessary condition requirement is when there are two or more causal chains each of which is sufficient to cause the plaintiff's harm. *See infra* text accompanying note 34.
25. By contrast, awards in other compensation systems, such as worker's compensation, may be more standardized, based on schedules of damages for various types of injuries.
26. Because of the lack of understanding of the biology of asbestotic disease, proof of which defendant caused or contributed to plaintiff's disease is virtually impossible when there are exposures to multiple defendants' asbestos products. The California Supreme Court adopted a "risk contribution" substitute for factual causation to address this problem of proof. *See* Rutherford v. Owens-Ill., Inc., 941 P.2d 1203 (Cal. 1997).
27. *See, eg*, Sindell v. Abbott Labs., 607 P.2d 924 (Cal. 1980) (crafting a theory of "market share" liability to overcome the evidentiary difficulties faced by plaintiff in proving which DES manufacturer caused plaintiff's disease).

28. On apportionment based on causal principles and fault principles, see Restatement (Third) of Torts: Apportionment of Liability § 26 (2000).
29. *See* Michael D. Green, *Second Thoughts on Asbestos Apportionment*, 37 Sw. U. L.J. 531 (2008).
30. *See, eg*, Rutherford v. Owens-Ill., Inc., 941 P.2d 1203 (Cal. 1997).
31. *See, eg*, Dafler v. Raymark Indus., Inc., 611 A.2d 136 (N.J. Super. Ct. App. Div. 1992).
32. *See* the sections on "Agent-Disease Causation" and "Epidemiology and Proof of Specific Causation."
33. *See supra* text accompanying note 9.
34. *See generally* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 27 (2010).
35. *Id*. § 29 (Special Note on Proximate Cause).
36. Thus, all "wrongful death" suits entail shortening the lifespan of the deceased.
37. *See* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(2) (2010).
38. *See Id*. § 28 cmt. c(3) (2010).
39. *See Id*. § 28 cmt. c(4) (2010).
40. This is unlike the multiple causes or multiple sufficient causes explained in the section on "The *Sine Qua Non*, But-For, or Necessary Element Test for Factual Causation."
41. The Sunday Times of London, Suffer the Children: The Story of Thalidomide 86–111 (1979).
42. 140 F.3d 381 (2d Cir. 1998).
43. *See, eg*, Seley v. G.D. Searle & Co., 423 N.E.2d 831 (Ohio 1981) (physician testified that he would have prescribed birth control pills for plaintiff even if warning had provided omitted information that made existing warning inadequate).
44. *See, eg*, Hymowitz v. Eli Lilly & Co., 539 N.E.2d (N.Y. 1989) (adopting a national market-share basis for liability of DES manufacturers).
45. *See* Shackil v. Lederle Laboratories, 561 A.2d 511 (N.J. 1989).
46. *See* the section on "Judicial Treatment of Non-epidemiologic Causation Evidence."
47. *See* Schonberger et al., *Guillain-Barre Syndrome Following Vaccination in the National Influenza Immunization Program, 1976–1977*, 110 Am. J. Epidemiology 105 (1979).
48. The first epidemiologic study connecting smoking with lung cancer was conducted by Sir Richard Doll and Bradford Hill. Richard Doll & A. Bradford Hill, *Lung Cancer and Other Causes of Death in Relation to Smoking: A Second Report on the Mortality of British Doctors*, 2 (5001) Brit. Med. J. 1071 (1956); Richard Doll & A. Bradford Hill, *The Mortality of Doctors in Relation to Their Smoking Habits: A Preliminary Report*, 1 (4877) Brit. Med. J. 1451 (1954).
49. *See* Michael D. Green et al., *Reference Guide on Epidemiology*, in Federal Judicial Center & National Research Council Reference, Manual on Scientific Evidence 549 (3d ed. 2011).
50. *See* Richardson by Richardson v. Richardson-Merrell, Inc., 857 F.2d 823, 832 (D.C. Cir. 1988).
51. *See* Brock v. Merrell Dow Pharm., Inc., 874 F.2d 307, *modified on reh'g*, 884 F.2d 166 (5th Cir. 1989).
52. *See* Callahan v. Cardinal Glennon Hosp., 863 S.W.2d 852, 864 (Mo. 1993). ("[T]he occurrence of vaccine-induced poliomyelitis has been extremely rare and there are no definitive studies indicating that a bacterial infection will or will not cause a person's immune system to be suppressed. It is not uncommon for areas of medicine involving rare diseases or causes to receive limited funding for research and studies.")
53. Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 Proc. Royal Soc'ty Med. 295 (1965).
54. *See* Margaret Kovera & Bradley McAuliff, *The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers?*, 85 J. Applied Psychol. 574 (2000).
55. Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, 121 Stat. 823 (2007).
56. 911 F.2d 941, 948 (3d Cir. 1990).
57. 953 S.W.2d 706, 722–24 (Tex. 1997).
58. *Id*. at 724 ("we should not widen the boundaries at which courts will acknowledge a statistically significant association beyond the 95% level"). The Texas Supreme Court has repeatedly reaffirmed this threshold. *See, eg*, Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332, 358-59 (Tex. 2014); Merck & Co., Inc. v. Garza, 347 S.W.3d 256, 264-66 (Tex. 2011). Some other courts have applied similar legal rules, and some have not. *Compare, eg*, Tumlinson v. Advanced Micro Devices, No. CV 08C-07-106 FSS, 2013 WL 7084888, *9 (Del. Super. Oct. 15, 2013) ("requiring results at the 95% confidence level is appropriate"); Estate of George v. Vt. League of Cities and Towns, 993 A.2d 367, 379 (Vt. 2010) (affirming exclusion of plaintiff's expert testimony when results of six

of eight studies were not statistically significant at $p < 0.05$) *with* Harris v. CSX Transp., Inc., 753 S.E.2d 275, 305 (W. Va. 2013) (reversing exclusion of plaintiff's expert's testimony because, among other errors, trial judge treated statistically insignificant differences as no differences); King v. Burlington N. & Santa Fe Ry. Co., 762 N.W.2d 24, 46–47 (Neb. 2009) (rejecting significance at $p < 0.05$ as threshold for admissibility). For discussions of scientific and judicial reliance on statistical significance at particular p values, see Kenneth J. Rothman, *Curbing Type I and Type II Errors*, 25 Eur. J. Epidemiology 223 (2010); Erica Beecher-Monas, *The Heuristics of Intellectual Due Process: A Primer for Triers of Science*, 75 N.Y.U.L. Rev. 1563, 1601–04 (2000); Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After* Daubert v. Merrell Dow Pharmaceuticals, Inc., 16 Va. Envtl. L.J. 1, 33–37 (1996); *Developments in the Law: Confronting the New Challenges of Scientific Evidence*, 108 Harv. L. Rev. 1481, 1535–56 (1995); Joseph Sanders, *From Science to Evidence: The Testimony on Causation in the Bendectin Cases*, 46 Stan. L. Rev. 1, 14–16 (1993); Brief of Amici Curiae Professors Kenneth Rothman, Noel Weiss, James Robbins, Raymond Neutra, and Steven Stellman in Support of Petitioners, Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993) (No. 92–102).

59. *See* Ethyl Corp. v. United States Envt'l Protection Agency, 541 F.2d 1, 28 n. 58 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976) (stating that "scientific fact is at least 95% certain" which may equate to "beyond a reasonable doubt" but is more stringent than "preponderance of the evidence"); Capizzano v. Sec'y of Health & Human Servs., 440 F.3d 1317 (Fed. Cir. 2006) (contrasting the medical research standard of "very near certainty–perhaps 95% probability" to the standard applied in civil cases, the preponderance standard); In re Ephedra Products Liability Litigation, 393 F. Supp. 2d 181 (S.D.N.Y. 2005) (stating that statistical significance at $p < 0.05$ "increase[es] the burden of proof imposed by substantive law"); Marmo v. IBP, Inc., 360 F. Supp. 2d 1019 (D. Neb. 2005) (expert toxicologist stated that science requires proof with 95 percent certainty while expressing his understanding that the legal standard requires more probable than not); Exxon Corp. v. Makofski, 116 S.W.3d 176, 187 (Tex. Ct. App. 2003) (expert testified that while science requires proof to a 95 percent certainty, appropriate standard for testifying in court is 51 percent); Carl Cranor, Regulating Toxic Substances: A Philosophy of Science and the Law (1993); Richard Goldberg, Causation and Risk in the Law of Torts: Scientific Evidence and Medicinal Product Liability 105 (1999); Larry Laudan, Truth, Error, and Criminal Law: An Essay in Legal Epistemology 64–65 (2006); K.S. Shrader-Frechette, Risk and Rationality: Philosophical Foundations for Populist Reforms 132–34 (1991); Ronald J. Allen, *Expertise and the* Daubert *Decision*, 84 J. Crim. L. & Criminology 1157 (1994); Margaret A. Berger, *What Has a Decade of* Daubert *Wrought?*, 95 Am. J. Pub. Health S59, S62 (2005); Margaret A. Berger & Aaron D. Twerski, *Uncertainty and Informed Choice: Unmasking Daubert*, 104 Mich. L. Rev. 257 (2005); Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. Rev. 385 (1985); Mark P. Denbeaux & D. Michael Risinger, Kumho Tire *and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 Seton Hall L. Rev. 15, 46–47 (2003); Edward J. Imwinkelried, *The Admissibility of Expert Testimony in* Christophersen v. Allied-Signal Corp.*: The Neglected Issue of the Validity of Nonscientific Reasoning by Scientific Witnesses*, 70 Denv. U. L. Rev. 473, 478 (1993); Harvey S. Frey, Letter, *When Scientific Data Become Legal Evidence*, 324 Sci. 335 (April 17, 2009); James E. Hullverson, Jr., *Reasonable Degree of Medical Certainty: A Tort et a Travers*, 31 St. Louis U. L.J. 577, 590 (1987); Jeff L. Lewin, *The Genesis and Evolution of Legal Uncertainty about "Reasonable Medical Certainty"*, 57 Md. L. Rev. 380, 400 (1998); Andrew A. Marino & Lawrence E. Marino, *The Scientific Basis of Causality in Toxic Tort Cases*, 21 U. Dayton L. Rev. 1, 23-24 & n.57 (1995); Paul R. Rice, *The Quagmire of Scientific Expert Testimony: Crumping the Supreme Court's Style*, 68 Mo. L. Rev. 53, 58–60 (2003); Leslie J. Sheffield & Ron Batagol, *The Creation of Therapeutic Orphans—or, What Have We Learned from the Debendox Fiasco*, 143 Med. J. Australia 143, 146 (1985); Steven R. Weller, *Book Review: Regulating Toxic Substances: A Philosophy of Science and Law*, 6 Harv. J. L. & Tech. 435, 436, 437–38 (1993); Wayne Roth-Nelson & Kathey Verdeal, *Risk Evidence in Toxic Torts*, 2 Envt'l Law. 405, 415–16 (1996); Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 Temple L. Rev. 55, 71 n.110 (1998); William M. Sage, *Lessons from Breast Implant Litigation*, 15 Health Affairs 206, 209 (1996); Cornelia Dean, *When Questions of Science Come to a Courtroom, Truth Has Many Faces*, N.Y. Times, December 5, 2006, at § F; William Glaberson, *The Courts vs. Scientific Certainty*, N.Y. Times, June 27, 1999, at § 4, p. 5. Even the Carnegie Commission has made this error. *See* Carnegie Commission on Science, Technology, and Government, Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges 28 (1993) ("But judicial decisions that appear to be based on 'bad science' may actually reflect the reality that the law requires a burden of proof or confidence level, other than the 95 percent confidence level that is often used by scientist to reject the possibility that chance alone accounts for observed differences.").

60. [2011] 2 AC 229.
61. In this discussion, we have used the term "relative risk" in a general sense. With respect to case-control studies, results are stated in terms of odds ratios because it is impossible to calculate a relative risk.
62. 43 F.3d 1311 (9th Cir. 1995). The court of appeals rendered this decision after the case was remanded to it by the Supreme Court in the decision discussed in the section on "Judicial Scrutiny of Expert Testimony."
63. *Id*. at 1321.
64. As this example suggests, the closer a relative risk is to 2.0, the easier it may be to produce evidence suggesting that in the case of a particular plaintiff the relative risk understates the probability of causation for that plaintiff.
65. *See, eg*, Landrigan v. Celotex Corp. 605 A. 2d 1079 (N.J. 1992).
66. Russelyn S. Carruth & Bernard D. Goldstein, *Relative Risk Greater than Two in Proof of Causation in Toxic Tort Litigation*, 41 Jurimetrics J. 195 (2001) (about half of judicial decisions that considered the issue between 1982 and 1999 held that proof of relative risk greater than two was required); Steve C. Gold, *The More We Know, the Less Intelligent We Are? How Genomic Information Should, and Should Not, Change Toxic Tort Causation Doctrine,* 34 Harv. Envtl. L. Rev. 369, 377 n. 33 (2010) (slightly fewer than half of judicial decisions that considered the issue between 1999 and 2009, exclusive of vaccine cases, held that proof of relative risk greater than two was required).
67. *See, eg*, Estate of George v. Vermont League of Cities and Towns, 993 A. 2d 367 (Vt. 2010).
68. *See* Cavallo v. Star Enter., 892 F. Supp. 756 (E.D. Va. 1995) ("It is also important to recognize that a fundamental assumption underlying this method is that the final, suspected "cause" remaining after this process of elimination must actually be capable of causing the injury. That is, the expert must "rule in" the suspected cause as well as "rule out" other possible causes."), *aff'd in relevant part*, 100 F.3d 1150 (4th Cir. 1996).
69. *See* Doe v. Ortho-Clinical Diagnostics, Inc. 440 F. Supp. 2d 465 (M.D.N.C. 2006); Perry v. Novartis Pharmaceuticals Corp. 564 F. Supp. 2d 452 (E.D. Pa. 2008); Milward v. Acuity Specialty Products Group, Inc., 969 F. Supp. 2d 101 (D. Mass. 2013).
70. For a general introduction to molecular (genetic) epidemiology and its early challenges, see Molecular Epidemiology (Paul A. Schulte & Frederica P. Perera eds., 1993). For more recent discussions, see Paolo Vineis & Frederica Perera, *Molecular Epidemiology and Biomarkers in Etiologic Cancer Research: The New in Light of the Old*, 16 Cancer Epidem., Biomarkers & Prev. 1954 (2007); Margaret R. Spitz & Melissa L. Bondy, *The Evolving Discipline of Molecular Epidemiology of Cancer*, 31 Carcinogenesis 127 (2010).
71. 293 F. 1013 (D.C. Cir. 1923).
72. 509 U.S. 579 (1993).
73. *Daubert*, *id*.; Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997) (also a toxic tort case); Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
74. *See* Heather G. Hamilton, Comment, *The Movement From* Frye *to* Daubert*: Where do the States Stand?*, 38 Jurimetrics J. 201, 208–09 (1998) (reporting that 33 states had reformed their test for determining the admissibility of expert witnesses while 17 retain the older *Frye* standard).
75. Between 2011 and 2014 four former *Frye* states, Alabama, Arizona, Florida and Wisconsin moved to the *Daubert* test. *See* Ala. R. Evid. Rule 702; Arizona Rules of Evidence, Rule 702; Fl. St. 90.702; Wis. Stat § 907.02.
76. *See generally* Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the* Daubert *Decision*, 8 Psychology, Pub. Pol'y & L. 251 (2002) (finding that in all types of civil cases, *Daubert* resulted in tightening of admissibility standards).
77. *See* Modern Scientific Evidence, *supra* note 12, at §§ 22.1–22.42.
78. *See, eg*, Marsee v. U.S. Tobacco Co., 639 F. Supp. 466, 470 (W.D. Okla. 1986) ("[T]he Court found evidence based on experiments with animals particularly valuable and important in this litigation since such experiments with humans are impossible. Under all these circumstances, the Court found this evidence probative on the issue of causation."), *aff'd in part*, 866 F.2d 319 (10th Cir. 1989).
79. *See* Wade-Greaux v. Whitehall Labs., Inc., 874 F. Supp. 1441, 1457 (D.V.I. 1994) ("The two chick studies cited by Dr. Gilbert are not helpful in determining effects in humans because of the principles of species specificity and "sledgehammer" teratology, and because the chick embryo model is so vastly different from the human experience."), *aff'd*, 46 F.3d 1120 (3d Cir. 1994).
80. *See* Zuchowicz v. United States, 140 F.3d 381, 386 (2d Cir. 1998).
81. *See* Milward v. Acuity Specialty Products Group, Inc., 639 F.3d 11 (1st Cir. 2011).
82. *See* Modern Scientific Evidence, *supra* note 12, at § 22.2 ("However, where a substantial body of

epidemiological evidence points in one direction, many courts will not permit an expert to express a contrary conclusion based on toxicological evidence.").

83. 522 U.S. 136 (1997).
84. 639 F.3d 11 (1st Cir. 2011).
85. For one discussion of this issue, *see* Reeps v. BMW of North America, LLC, 39 Misc. 3d 1234(A) (N.Y. Sup. Ct. 2013):

> The weight of evidence method is used in medical literature either in a rigorous scientific or metaphorical sense. It is used as methodology "where WOE points to established interpretative methodologies (e.g., systematic narrative review, meta-analysis, causal criteria, and/or quality criteria for toxicological studies) or where WOE means that all rather than some subset of the evidence is examined, or rarely, where WOE points to methods using quantitative weights for evidence." The metaphorical use of the term is, if nothing else, "a colorful way to say the body of evidence we have examined and judged using a method we have not described but could be more or less inferred from a careful between-the-lines reading of our paper."

Quoting Douglas L. Weed, 2005. Weight of evidence: a review of concept and methods. Risk Analysis 25, 1545, 1546–1547.

86. *See* Hymowitz v. Eli Lilly & Co., 539 N.E.2d 1069, 1073 (N.Y. 1989) (explaining the legislative revival of barred claims).
87. For a general discussion of the difficulties that toxic tort cases present for traditional statutes of limitations, *see* Michael D. Green, *The Paradox of Statutes of Limitations in Toxic Substances Litigation*, 76 Cal. L. Rev. 965, 973–974 (1988).
88. *See, eg*, Daley v. A.W. Chesterton, Inc., 37 A.3d 1175 (Pa. 2012).
89. *See, eg*, Annunziato v. City of New York, 624 N.Y.S.2d 544 (Sup. Ct. 1995), *modified*, 224 A.D.2d 31, 647 N.Y.S.2d 850 (App. Div. 1996); *In re* Mirapex Prods. Liab. Litig., 735 F. Supp.2d 1113 (D. Minn. 2010); Avila v. Willits Envt'l Remediation Trust, 633 F.3d 828 (9th Cir. 2011). Defendants' insurers have also sometimes argued that the publication of epidemiologic studies triggered their insureds' duty to notify the insurers of claims. *See, eg*, U.S. Fire Ins. Co. v. Vanderbilt Univ., 82 F. Supp. 2d 788 (M.D. Tenn. 2000).
90. Some statutes of repose contain provisions that exempt or limit their application to toxic torts. For example, the Texas statute of repose provides that if one is exposed before the end of the repose period the cause of action is not extinguished even if the injury resulting from the exposure does not manifest itself until after the end of the repose period. *See* Tex. Civ. Prac. & Rem. Code § 16.012(d).
91. Cipollone v. Liggett Group, 505 U.S. 504 (1992).
92. Riegel v. Medtronic, Inc., 552 U.S. (2008).
93. Cipollone v. Liggett Group, 505 U.S. 504 (1992) (cigarettes); PLIVA, Inc. v. Mensing, 131 S. Ct. 2657 (2011) (generic drugs); Bruesewitz v. Wyeth LLC, 131 S. Ct. 1068 (2011) (vaccines); Bates v. Dow Agrosciences LLC, 544 U.S. 431 (2005) (insecticides, fungicides and rodenticides).
94. Wyeth v. Levine, 555 U.S. 555 (2009).
95. Butterfield v. Forrester, 11 East. 60, 103 Eng. Rep. 926 (1809).
96. Restatement (Third) of Torts: Apportionment of Liability § 7 rprtrs. note to cmt. a (2000).
97. Borel v. Fibreboard Paper Products Corp., 493 F.2d 1076 (5th Cir. 1973).
98. Daffler v. Raymark Industries, Inc., 611 A.2d 136 (N.J. App. 1992).
99. *See* Restatement (Second) of Torts § 402A (1965) (listing drugs and vaccines as examples of "unavoidably unsafe products" that cannot be made entirely or even reasonably safe and therefore are not subject to products liability if properly manufactured and bearing proper warnings); Restatement (Third) of Torts: Products Liability § 6 (1998) (stating that prescription drugs subject their manufacturers to products liability if defectively manufactured, not reasonably safe due to inadequate warnings or instructions, or defectively designed in that foreseeable risks outweigh foreseeable therapeutic benefits so a reasonable health-care provider would not prescribe the drug for any class of patients).
100. *See* Michael D. Green, *Prescription Drugs, Alternative Designs, and the Restatement (Third): Preliminary Reflections*, 30 Seton Hall L. Rev. 207 (1999).
101. *See* Orin Kramer & Richard Briffault, Workers' Compensation: Strengthening the Social Compact (1991).
102. The first successful asbestos claim, and many thousands that followed it, were of this type. *See* Borel v. Fibreboard Paper Products Corp., 493 F.2d 1076 (5th Cir. 1973); Deborah R. Hensler, *Asbestos Litigation in the United States: Triumph and Failure of the Civil Justice System*, 12 Conn. Ins. L.J. 255 (2005–2006).

103. 45 U.S.C. §§ 51–60.
104. Rogers v. Mo. Pac. R. Co., 352 U.S. 500, 506 (1957). Such a standard for causation is extremely problematic as evidenced by lower courts' struggling to sort out what it means. *See* Michael D. Green, *The Federal Employers' Liability Act: Sense and Nonsense about Causation*, 61 DePaul L. Rev. 503 (2012).
105. For an example, see Estate of George v. Vt. League of Cities and Towns, 993 A.2d 367 (Vt. 2010) (denying, in a split decision, a claim that a firefighter's non-Hodgkin's lymphoma was caused by his work).
106. *See* Dobbs et al., *supra* note 2, at § 334.
107. 28 U.S.C. §§ 1346(b), 2402, 2671–2680.
108. 28 U.S.C. §§ 1346(b), 2680(a).
109. Feres v. United States, 340 U.S. 135 (1950).
110. *See* In re 'Agent Orange' Prod. Liab. Litig., 611 F. Supp. 1223 (E.D.N.Y. 1985) (describing the importance of epidemiology to causal issues in the case); In re 'Agent Orange' Prod. Liab. Litig., 818 F.2d 194 (2d Cir. 1987) (affirming dismissal of the veterans' claims against the federal government).
111. 42 U.S.C. §§ 300aa-1–300aa-34.
112. *See* Althen v. Sec'y of HHS, 418 F.3d 1274 (Fed. Cir. 2005).
113. Capizzano v. Sec'y of HHS, 440 F.3d 1317 (Fed. Cir. 2006).
114. *Eg* Hennessey v. Sec'y of HHS, 91 Fed. Cl. 126 (2010) (holding that it was appropriate to consider epidemiologic data in reaching finding of no causation).
115. *See generally* Lloyd Dixon, Geoffrey McGovern & Amy Coombe, Asbestos Bankruptcy Trusts: An Overview of Trust Structure and Activity with Detailed Reports on the Largest Trusts (2010). The asbestos bankruptcy trusts routinely have offered injured claimants only a small percentage of the recovery called for in their original bankruptcy trust fund agreements. *See* S. Todd Brown, *How Long Is Forever This Time? The Broken Promise of Bankruptcy Trusts*, 61 Buff. L. Rev. 537 (2013). The relationship between trust recoveries and the tort system is murky at best. *See* Peggy L. Ableman, *The* Garlock *Decision Should Be Required Reading for All Trial Court Judges in Asbestos Cases*, 37 Am. J. Trial Advoc. 479 (2014) (discussing the findings in *In* re *Garlock Sealing Techs., LLC*, 504 B.R. 71 (W.D.N.C. Bankr. 2014)).
116. *Eg*, Jackson v. Johns-Manville Sales Corp., 781 F.2d 394 (5th Cir. 1986); *see generally* James A. Henderson & Aaron D. Twerski, *Asbestos Litigation Gone Mad: Exposure-based Recovery for Increased Risk, Mental Distress, and Medical Monitoring*, 53 S.C. L. Rev. 815 (2002).
117. *See generally* Robin Kundis Craig et al., Toxic and Environmental Torts: Cases and Materials 685–706 (2011).
118. *Eg*, Hedgepeth v. Whitman Walker Clinic, 22 A.3d 789 (D.C. 2011).
119. Metro-North Commuter Railroad Co. v. Buckley, 521 U.S. 424 (1997).
120. *Eg*, Donovan v. Philip Morris USA, Inc., 914 N.E.2d 891 (Mass. 2009); Potter v. Firestone Tire & Rubber Co., 863 P.2d 795 (Cal. 1993).
121. *Eg*, Caronia v. Philip Morris USA, Inc. 5 N.E.3d 11 (N.Y. 2013).
122. *See generally* Vickie S. Wilson et al., *Utilizing Toxicogenomic Data to Understand Chemical Mechanism of Action in Risk Assessment*, 271 Toxicology & Applied Pharmacology 299 (2013); Luoping Zhang et al., *Systems Biology of Human Benzene Exposure*, 184 Chemico-Biological Interactions 86 (2010).
123. *See* Paolo Boffetta, *Biomarkers in Cancer Epidemiology: An Integrative Approach*, 31 Carcinogenesis 121 (2010); Margaret R. Spitz & Melissa L. Bondy, *The Evolving Discipline of Molecular Epidemiology of Cancer*, 31 Carcinogenesis 127 (2010).
124. Christine B. Ambrosone et al., *Cigarette Smoking, N-Acetyltransferase 2 Genotypes, and Breast Cancer Risk: Pooled Analysis and Meta-analysis*, 17 Cancer Epidemiology Biomarkers & Prevention 15, 21, 25 (2008).
125. *See* Steve C. Gold, *When Certainty Dissolves into Probability: A Legal Vision of Toxic Causation for the Post-Genomic Era*, 70 Wash. & Lee L. Rev. 237 (2013).
126. At least one plaintiff has attempted to introduce expert testimony that the DNA mutation pattern in a brain tumor was consistent with a genotoxic rather than sporadic etiology. Branham v. Rohm and Haas Co., 2013 WL 5763133 (Pa. Super. 2013). In *Branham*, the trial court initially admitted the plaintiff's expert testimony but then dismissed the case and concluded that the testimony was inadmissible. The appellate court ordered a retrial, in part because the trial court committed procedural errors in excluding the challenged testimony.
127. *Eg* Henricksen v. ConocoPhillips Co., 605 F. Supp. 2d 1142, 1148–50 (E.D. Wash. 2009).
128. Tompkin v. Philip Morris USA, Inc., 362 F.3d 882 (6th Cir. 2004). The plaintiff in *Tompkin* had also been occupationally exposed to asbestos.

129. Guzman v. Exxon Mobil Corp., No. 693−606 (La. Dist. Ct., 24th Dist.). The case was tried to a jury and resulted in a defense verdict. Personal communication from counsel to author.
130. *Eg* Snyder v. Sec'y of HHS, 553 Fed. Appx. 994 (Fed. Cir. 2014) (claim denied where court accepted testimony that mutation was sole cause of claimant's condition rather than a cause together with vaccine); Deribeaux v. Sec'y of HHS, 717.F.3d 1363 (Fed. Cir. 2013) (claim denied where genetic cause of symptoms was discovered after claim was filed); Stone v. Sec'y of HHS, 676 F.3d 1373 (Fed. Cir. 2012) (claim denied where mutation that could entirely explain symptoms was discovered after claim was filed); Simanski v. Sec'y of HHS, 115 Fed. Cl. 407 (2014) (claim denied where a possible genetic explanation of claimant's symptoms existed but claimant's parents had declined to obtain genetic testing); Hopkins v. Sec'y of HHS, 84 Fed. Cl. 530 (2008) (claim denied where court concluded claimant's condition was hereditary based on family history although plaintiff had no known disease-causing mutation in the relevant gene); *see also* Blackwell v. Wyeth, 971 A.2d 235 (Md. 2009) (affirming exclusion, under the *Frye* general acceptance test, of testimony that thimerosal in vaccine caused autism, where trial judge found that prevailing view holds autism caused by "a gene or series of interacting genes that have not yet been identified").
131. *See, eg*, Costa v. Wyeth, Inc., 2012 WL 1069189 (M.D. Fla. 2012) (rejecting defense argument that plaintiff's expert on differential diagnosis failed to "rule in" genetics as a cause); In re Diet Drugs Litig., 890 F. Supp. 2d 552, 562 (E.D. Pa. 2012) (identifying genetic mutation as well as diet drugs as risk factors for primary pulmonary hypertension); McMunn v. Babcock & Wilcox Power Generation Group, 2014 WL 814878, *11−16 (W.D. Pa. 2014) (refusing to exclude testimony of plaintiff's differential diagnosis expert who considered genetic predisposition among other risk factors).
132. *See* Susan R. Poulter, *Genetic Testing in Toxic Injury Litigation: the Path to Scientific Certainty or Blind Alley?*, 41 Jurimetrics J. 211 (2001) (arguing against required testing for genetic predisposition); Henricksen v. ConocoPhillips Co., 605 F. Supp. 2d 1142 (E.D. Wash. 2009) (emphasizing plaintiff's failure to determine whether plaintiff's acute myelogenous leukemia exhibited pattern of chromosomal abnormalities associated with chemically induced disease).
133. *See generally* Schultz v. Akzo Nobel Paints, LLC, 721 F.3d 426, 434 (7th Cir. 2013) (reversing summary judgment for defendant that was based on exclusion of testimony by plaintiff's expert who testified that "the mere fact that genetic and/or other environmental risk factors . . . have been identified as probable causes of a particular case of cancer in no way refutes the possibility that chemical exposures being investigated have also played a substantial contributing role at one or more stages of the development of that person's cancer").
134. *See, eg*, Barbara Charbotel et al., Trichloroethylene Exposure and Somatic Mutations of the VHL Gene in Patients with Renal Cell Carcinoma, 2 J. Occupational Med. & Toxicology 13 (2007) (reporting inability to reproduce mutation signature found by other researchers).
135. In re Diet Drugs Litig., 2013 WL 1796989 (E.D. Pa. 2013). The court did not decide the significance of the missing genetic test but ruled that the plaintiff's claim could not proceed because plaintiff, although through no fault of his own, lacked any information regarding family history of PPH.
136. *See supra* note 66.
137. *Compare, eg*, Estate of George v. Vt. League of Cities and Towns, 30 A.3d 1271 (Vt. 2011) (approving use of relative risk greater than two as a "benchmark" tying into preponderance of the evidence standard, but affirming dismissal of plaintiff's workers' compensation claim despite some epidemiologic studies exceeding the threshold) *with* Merck & Co., Inc. v. Garza, 347 S.W.3d 256 (Tex. 2011) (holding that legally sufficient evidence of causation requires at least two studies showing statistically significant relative risk greater than two). A split decision of the Texas Supreme Court in an asbestos-mesothelioma case appears to extend the requirement to proof that *each defendant* (eg, each of several manufacturers of asbestos-containing products) was responsible for an exposure that more than doubled the plaintiff's risk of disease. Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332, (Tex. 2014).
138. The Texas Supreme Court, for example, initially imposed the more than doubling of risk requirement in Merrell Dow Pharmaceuticals, Inc. v. Havner, 953 S.W.2d 706 (Tex. 1997). In Merck & Co., Inc. v. Garza, 247 S.W.3d 256 (Tex. 2011), the court explained that some lower courts had misunderstood language in *Havner* stating that doubling of risk was not a "litmus test," clarifying that this language did not mean that lower relative risks would always be in sufficient proof. In Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332 (Tex. 2014), that court extended the requirement to the "substantial factor" aspect of proof of causation. "Substantial factor" is discussed in the section on "'Substantial Factor' Causation." *Bostic* concerned a claim for mesothelioma by a

plaintiff who had concededly been exposed to asbestos from multiple, relatively small, hard-to-quantify sources. The Texas Supreme Court held that in such a case a plaintiff must prove that the plaintiff's exposure to each defendant's product, alone, was sufficient to more than double the plaintiff's risk of disease. *Id*. at 349–350. For a critique of *Bostic* see Steve C. Gold, *Drywall Mud and Muddy Doctrine: How Not to Decide a Multiple-Exposure Mesothelioma Case*, 49 Ind. L. Rev. 117 (2015).

139. For example, many genomic studies have identified alleles associated with statistically significant risk increases of factors greater than one but less than two. *For example*, Paul D.P. Pharoah et al., *Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer*, 358 New Eng. J. Med. 2796 (2008).
140. Restatement (Third) of Torts: Liability for Physical & Emotional Harm § 28 cmt. c(4) (2010) (noting the possibility that probability of specific causation in an individual case can be "refined" in various ways).
141. *See, eg*, Estate of George v. Vermont League of Cities and Towns, 993 A.2d 367, 393 (Vt. 2010) (Reiber, C.J., dissenting) (dissenting opinion noting that the majority, which rejected firefighter's claim that non-Hodgin's lymphoma was employment-related, did not address testimony that the claimant experienced above-average exposure).
142. Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332, 368 (Tex. 2014) (noting that individual susceptibility to developing mesothelioma after asbestos exposure varies for reasons not yet fully understood).
143. Berg v. Johnson & Johnson Consumer Cos., Inc., 983 F. Supp. 2d 1151, 1153–54 (D.S.D. 2013).
144. For two very different predictions of the likelihood that science will develop highly individualized ex ante risk assessments for use in regulation of toxic substances, *compare* David E. Adelman, *The False Promise of the Genomics Revolution for Environmental Law*, 29 Harv. Envt'l L. Rev. 117 (2005) (expressing skepticism about the potential for toxicogenomics because of the complexity and heterogeneity of human biology and the significance of environmental risk factors rather than genetic risk factors) *with* Jamie A. Grodsky, *Genetics and Environmental Law: Redefining Public Health*, 93 Calif. L. Rev. 171, 269 (2005). ("The new science has the potential both to measure individual genetic susceptibilities to the effects of toxic substances and to provide evidence of toxin-induced injuries long before clinical symptoms emerge.")
145. *Eg* Borg-Warner Corp. v. Flores, 232 S.W.3d 765, 770 (Tex. 2007) (holding that evidence was insufficient to show that automobile mechanic's asbestosis was caused by exposure to asbestos from defendant's brake pads and linings).
146. *See* United States Environmental Protection Agency, Guidelines for Carcinogen Risk Assessment 1–11 to 1–15 (March 2005).
147. For a thorough discussion, see Joseph Sanders, *The "Every Exposure" Cases and the Beginning of the Asbestos Endgame*, 88 Tul. L. Rev. 1153 (2014).
148. *See* Baker v. Chevron USA, Inc., 680 F. Supp. 2d 865, 878 n.9 (S.D. Ohio 2010) (stating, in rejecting a claim that plaintiff's acute myelogenous leukemia was caused by benzene emissions from a nearby factory, that "'one-hit' or 'no-threshold' theory of causation. … has not been accepted as a reliable theory for causation" by courts), *aff'd*, 533 Fed. App'x 509 (6th Cir. 2013); Schultz v. Glidden Co., 2012 WL 968005 at *3 (E.D. Wis. 2012) (stating, in excluding expert testimony in a claim that a plaintiff's acute myelogenous leukemia was caused by occupational exposure to benzene in paint, that the "no-threshold theory of causation in toxic tort cases has been roundly rejected by courts"); Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332, 375 (Tex. 2014) (Lehrmann, J., dissenting) (noting that expert testimony "flatly forecloses" the notion that a single asbestos fiber could cause sufficient genetic errors to cause cancer but that this is not logically inconsistent with the idea that each incremental dose adds incremental risk and even a low dose could be a necessary cause of a particular mesothelioma).
149. Restatement of Torts § 431 (1934).
150. Restatement (Second) of Torts § 431 (1965).
151. *See* Restatement (Third) of Torts: Liability for Physical & Emotional Harm § 26 (2010) ("Conduct is a factual cause of harm when the harm would not have occurred absent the conduct."); *id*. cmt. c (modeling a but-for cause as any necessary element of a causal set that, collectively, are sufficient to bring about a result).
152. An early example was Allen v. United States, 588 F. Supp. 247 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987), in which plaintiffs alleged that radiation from above-ground nuclear weapons tests caused their cancers. *Allen* contrasts notably with more recent cases brought by residents of uranium mining and milling towns, whose cases were dismissed for failure to establish but-for causation. June v. Union Carbide Corp., 577 F.3d 1234 (10th Cir. 2009) (affirming dismissal of claims brought by residents of Uravan, Colorado);

Wilcox v. Homestake Mining Co., 619 F.3d 1165 (10th Cir. 2010) (affirming dismissal of claims brought by residents living near a uranium mill in New Mexico).

153. Rutherford v. Owens-Illinois, Inc., 941 P.2d 1203, 1220 (Cal. 1997).
154. Borg-Warner Corp. v. Flores, 232 S.W.3d 765, 768 (Tex. 2007); *see* Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332, 342–346 (Tex. 2014) (discussing requirement that exposure to defendant's asbestos products be a "substantial factor" in causing mesothelioma).
155. Joseph Sanders et al., *The Insubstantiality of the "Substantial Factor" Test for Causation*, 73 Mo. L. Rev. 399 (2008).
156. *See* Restatement (Third) of Torts: Liability for Physical Harm § 26 cmt. j & rptrs. note.
157. *See, eg*, Bostic v. Georgia-Pacific Corp., 439 S.W.3d 332 (Tex. 2014).
158. Jeff L. Lewin, *The Genesis and Evolution of Legal Uncertainty About "Reasonable Medical Certainty,"* 57 Md. L. Rev. 380 (1998) (tracing the term's spread across jurisdictions from early-twentieth-century use in Chicago).
159. Restatement (Third) of Torts: Liability for Physical Harm § 28 cmt. e & rptrs. note (2010).
160. *Id.*; Mark D. Howard, *Proving Causation with Expert Opinion: How Much Certainty Is Enough?*, 74 Ill. B.J. 580, 584 (1986) ("Most experts, other than professional witnesses, are unfamiliar with the 'reasonable certainty' language used in court.").
161. *See* Restatement (Third) of Torts: Liability for Physical Harm § 28 cmt. e rptrs. note (listing cases) (2010).
162. *Id.* § 28 cmt. e.

Further Reading

Cranor, C.F., 2006. Toxic Torts: Science, Law, and the Possibility of Justice. Cambridge University Press, Cambridge & New York.

Faigman, D.L., Blumenthal, J.A., Cheng, E.K., Mnookin, J.L., Murphy, E., Sanders, J., 2013–2014. Modern Scientific Evidence, 5 vols. West, Eagen, MN.

Goldberg, R. (Ed.), 2011. Perspectives on Causation. Hart Publishing, Oxford.

Goldstein, B.D., Henefin, M.S., 2011. Reference guide on toxicology. In: Federal Judicial Center, National Research Council, Reference Manual on Scientific Evidence, third ed. The National Academies Press, Washington, DC, pp. 633–688.

Green, M.D., Freedman, D.M., Gordis, L., 2011. Reference guide on epidemiology. In: Federal Judicial Center, National Research Council, Reference Manual on Scientific Evidence, third ed. The National Academies Press, Washington, DC, pp. 549–632.

Restatement (Third) of Torts § 26 cmt. c & Rptrs' Note, 2005. American Law Institute, St. Paul, MN.

CHAPTER

3

Methods Used in Forensic Epidemiologic Analysis

M.P. Zeegers, M.J.L. Bours

Maastricht University, Maastricht, The Netherlands

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

OUTLINE

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00003-3

WHAT IS EPIDEMIOLOGY?

Although the epidemiologist is called upon in court more often now than in past decades, there are many who are unfamiliar with the discipline and may confuse it with medicine or statistics. Epidemiology is a core medical and public health science that investigates **probabilities** and **determinants** of health outcomes in human populations. As its name may suggest, epidemiology is indeed also the discipline that investigates epidemics (disease outbreaks), but it is by no means limited to this narrow application. Using knowledge of medicine, research methodology, and statistics as their key tools, epidemiologists are trained to draw causal inferences from empirical data. As such, epidemiology has been a main driver behind the development of evidence-based medicine. In this chapter, we lay out basic principles and methods for the practice of epidemiology. These principles form the basis for forensic epidemiology (FE) practices, in which epidemiologic principles and practices associated with population-based investigations of cause are applied to investigations of individual causation (Box 3.1).

Probabilities of Injuries

The starting point of an epidemiological investigation is always the calculation of the frequency of a health outcome in specifically defined populations. For the benefit of brevity and applicability to the legal field, we will talk about injuries, while the same can be applied to other health outcomes such as treatment (side) effects or disease. The frequency of injury equals the **probability** for an individual to have this injury if the individual belongs to one of the prespecified populations. The epidemiologist compares these population-specific probabilities to start getting insight into cause-and-effect relationships. The **Epidemiological Probability** (Fig. 3.1) is the base function of all factual probability measures described in this chapter.

Determinants of Injuries

Epidemiology recognizes two categories of determinants for injuries: causal and diagnostic. **Causal determinants** influence disease occurrence, recurrence, or prognosis. They are often subject of interest when investigating "disputed legal" matters. Multiple causal determinants often a operate together in a joint fashion and could include, for example, a tortious act under investigation together with other biological and nonbiological causal factors. For example, How likely it is that a plaintiff's whiplash is caused by the accident, taking into account her disease history, age and sex, and the specific circumstances of the accident?

BOX 3.1

EXAMPLES OF INVESTIGATIVE QUESTIONS ADDRESSED BY FORENSIC EPIDEMIOLOGIC METHODS

- What is likelihood that the asbestos exposure that Mr. X experienced during his employment at company Z caused his lung cancer?
- How likely is it that the DNA found on the forensic scene belongs to Mr. X? What is the chance that you are wrong? Could you in your probability calculation take into account the other evidence that points toward the identification of Mr. X?
- What is the probability that the leg amputation of Mrs. Y could have been prevented if the delay in diagnosis would not have occurred?
- How likely is it that the heart failure of Mrs. Y was indeed caused by the side effect of this drug?
- What is the chance that the death that followed the administration of the opiate by 20 min was due to the drug and not to other (unknown) factors?
- What is the chance that Mr. X would have needed neck surgery when he did if he had not been in a minor traffic crash the prior month?
- How likely is it that the bladder cancer of Mrs. Y was caused by passive smoking during her imprisonment given the fact that she was an ex-smoker herself?
- Which liability percentage is reasonable in the given circumstance?
- What would be the life expectancy of Mr. X at the time of his death if the wrongful death not occurred?
- How long is Mr. X expected to survive, given his brain/spinal cord injury, on a more probable than not basis?
- Given the medical and nonmedical evidence at hand regarding the circumstances of this traffic crash, what is the probability that Mrs. Y was the driver?
- Given the medical and nonmedical evidence at hand regarding the circumstances of this car accident, what is the probability that Mr. X was wearing a seat belt?
- What is the probability that Mrs. Y's need for surgery resulted from the crash, versus that it would have occurred at the same time if the crash had not happened?

$$\frac{\text{Number of people (cases) with the injury of interest}}{\text{Total number of the people in the population from which these cases arise}}$$

FIGURE 3.1 **The epidemiological probability.**

Diagnostic determinants on the other hand help to determine injury **status** and, even though their predictive capacity can be stronger, these are generally not causally related to the injury. For example, although X-ray use is highly associated with broken bones, there is no causal relationship between the two. This chapter focuses on the epidemiological methodology related to finding causal determinants and the epidemiological methodology related to diagnostics, when we discuss the application of test accuracy in court. The strength of the **association** between a potential causal determinant and injury can be quantified in several ways. Generally spoken, the stronger the association the higher the probability of causation (PC). This is discussed further later in this chapter.

Epidemiology in the Legal Setting

In legal disputes, epidemiological evidence can be used to evaluate and quantify cause-and-effect relationships. It is concerned with data-driven risk assessment and, therefore, with scientifically quantifiable statements like the following:

- What is most "likely"
- Whether two events happening at the same time are a "coincidence"
- That a particular outcome is "impossible," "certain," "rare," or "common"
- That a particular outcome was caused by an exposure because this is what "usually" is the case
- That a particular outcome was *not* caused by an exposure because the condition is "usually" *caused by something else*
- That the "risk" of injury was increased by the mechanical failure or failure to use a safety device

The result of a forensic epidemiological investigation is an evidence-based probability that a suspected relationship between a cause and an effect is likely to be true on a more probable than not basis (>50%) or at some other level. This calculated probability can support legal decision-making regarding guilt or innocence in criminal actions, and the causation element of proof of liability in civil actions. These evidence-based probabilities can be useful in cases of medical negligence, toxic tort, mass tort, drug side effects, medical device failures, traffic crash-related injury and death, person identification, or life expectancy.

The methods used by a forensic epidemiologist generally consist of four steps:

1. Evaluation of the elements of the legal case, and the opposing theories of causation and liability/guilt and innocence.
2. The application of epidemiologic methods for evaluating causal relationships to the theories proposed by both sides. Are the medical probability statements presented valid and reliable based on solid scientific research and reasoning? Can they be quantified and compared? Do they fit with what is known about the specifics of the case?
3. The conduct of research via systematic literature review and/or database mining and analysis in the event that there are specific circumstances that have not been previously addressed in the literature. Is the investigated relationship plausibly causal? Has the potential effect of confounding and/or bias been accounted for?
4. The quantification of the evidence-based causal probabilities based on an evaluation of all of the relevant and probative evidence.

Currently, many courts still undervalue the role that epidemiological evidence can play in the resolution of disputes about causation in both criminal and tort litigation. A main reason is that the epidemiological report (step 4) takes the form of a probability (of causation) estimate and many courts are inexperienced in the interpretation of statistical data. Indeed, a lack of trust in statistical evidence combined with a lack of knowledge as to how to properly interpret it has resulted in judicial misuse of statistics in some instances. In the United Kingdom, for example, courts have rejected perfectly sound statistical evidence (such as in *McTear v Imperial Tobacco* (UK 2005)) as well as accepted and misinterpreted inaccurate statistical evidence (such as in *Gregg v Scott* (UK 2005) and *Novartis v Cookson* (UK 2007)). The outcome of this misuse is that claims are sometimes wrongly decided and the resulting verdict is unjust. Injustice that is based in a systematic error is a problem for everyone in the legal system; it affects defendants as much as it affects claimants, and it is bound to be repeated.

In 2011, the UK Supreme Court handed down its decision in *Sienkiewicz v Greif*. Controversially, the Supreme Court held that a local authority was negligent and liable for the death of a 55-year-old woman from mesothelioma on the basis that it had exposed her to small amounts of asbestos decades earlier when she had attended primary school. Epidemiological evidence presented by the defendant indicated only a small chance of the primary school exposure having caused the fatal disease. Based on a substantially misinformed description of epidemiology, the Supreme Court rejected the epidemiologic evidence as irrelevant and unhelpful, and made a finding of causation based on the unproven assertion that the local authority had materially contributed to the risk of the victim developing mesothelioma. The case served as an excellent object lesson in why accurate information regarding the utility, and in some cases, necessity, of epidemiologic methods, data, and analysis in addressing questions of causation is needed in the legal system. It is too easy for some courts to reject population-based information as being about "other people" and not the claimant, in the process failing to recognize the fact that causal determinations are typically based, at least in part, on what has happened to "other people." One of the primary purposes of this book is to explain why this is so, and in so doing help provide a more thorough description of the use of epidemiology as a basis for assessing individual causation in a legal or forensic setting.

RESEARCH METHODS TO INVESTIGATE CAUSAL RELATIONSHIPS

What are causal relationships? According to the Oxford Dictionary, a relationship between two things is causal when one thing causes the other to happen. If you turn your ignition key, a sequence of events starts which eventually results in your car starting (assuming you have enough fuel, no engine troubles, etc.). The process of one event causing or producing another event is called causation. A cause is a thing or person that makes some other thing happen; it is the reason for it happening. For example, smoking is a causative factor in several major diseases such as lung cancer and cardiovascular disease. All other things being equal, when nobody would smoke, a large number of cases of disease could be prevented, that is, they would not happen at all. Though litigation is often about accusation rather than causation, valid evidence that quantifies causal relationships can often be of overarching

importance for resolving pivotal legal questions. This is where the methods of FE are most appropriately applied and of greatest utility.

Epidemiologists are intrinsically interested in finding evidence for causal relationships between exposure to a particular agent or harmful event, such as an environmental toxin or a traffic accident, and the occurrence of a certain health outcome of interest, such as a disease or injury. Epidemiologists have been called the Sherlock Holmes of public health, as they use their powers of empirical observation and logical reasoning (ie, evidence-based deduction) in their search for the truth about the existence of a cause-and-effect relation of interest. There are a number of tools that epidemiologists use when investigating causal relationships.

Epidemiologic Methods for Investigating Causal Relationships

To unveil the existence of a causal effect of one factor (the exposure or cause) on another factor (the outcome or effect), epidemiologists study associations between exposures and outcomes in populations. An association is present when the observed occurrence of an outcome differs meaningfully between subgroups of the population that are differently exposed to the putative causative factor. As an example, observing that the occurrence of lethal head injury in motorcyclists is higher among those who did not wear crash helmets compared to those who did, strongly suggests an association between wearing helmets and a fatal motorcycle crash. However, an observed association is by no means always proof of causation. The actual truth may not have been revealed, that is, the observed association might not mirror the underlying true association but only represent the tip of the iceberg. Although in the above example the cause may seem obvious (and probably is), other factors may be responsible for having caused the outcome, such as the type of motorcycle, the motorcyclist's riding experience and ability, the riding speed, or climate conditions, to mention but a few. If besides wearing a crash helmet or not, these other factors are also substantially different between subgroups of motorcyclists, the observed association between wearing crash helmets and lethal head injury may not be as obvious as initially thought. Consequently, erroneous conclusions regarding the observed association lie in wait, resulting in flawed inferences about causal relationships. In a courtroom, this could make the difference between a guilty or not guilty verdict. More on how to calculate the association between potential cause and injury can be found later in this chapter under the section on "Factual Probability."

Epidemiologists face the problems of inference from association to causation on a regular basis. It is their task to tackle these problems and arrive at the best possible evidence for causal relationships. For that purpose, epidemiologists study associations between exposures and outcomes in populations through use of a number of different epidemiologic study designs. These designs enable them to approach the problem at hand from different angles in their attempt to answer the causal question of interest. A study design is a blueprint of the way in which a research question about causal relationships is going to be addressed. It is like a road map used by the epidemiologist to move from cause to effect. Depending on the nature of the research question(s), as well as other important factors such as ethical issues and availability of time and money, different types of study designs can be applied.

Before delving deeper into several major study designs commonly used by epidemiologists—*the randomized experiment, cohort study, and case–control study*—some general terminology related to epidemiologic studies will be introduced first.

The Language of Epidemiologic Study Designs

Epidemiologic research can be performed in a variety of ways. Two major types of epidemiologic studies can be distinguished: *experimental studies* and *observational studies*. In an experimental study, the researcher intervenes in some way in the causal relationship under investigation, for instance by exposing one group of patients to a particular treatment (eg, a medicine) while withholding the treatment from another group of patients. Next, the outcome of interest (eg, the effect of the drug on the course of an illness) is compared between treatment groups to evaluate whether an association exists between the exposure and the outcome. In such studies, exposure to the putative cause, as well as other factors that may influence the outcome (ie, concurrent causes), is primarily under control of the researcher and not the participant. However, this is often not possible or ethical in most real-life circumstances. Recall the previous example about motorcyclists and imagine you are an epidemiologist performing an experiment to assess whether wearing a crash helmet protects against (lethal) head injury. You would then need to provide helmets to only part of the motorcyclists participating in your study, cause all study participants to crash, and compare the extent of the sustained injuries between the participants who wore helmets and those who did not. Even though in theory you could perform this experiment in a highly standardized and controlled way that provides strong evidence for causation, in practice you would soon find yourself sitting in a courtroom not as a forensic epidemiologic specialist but as the accused. In contrast to experimental studies in which the researcher creates an artificial situation, researchers performing observational studies only observe what is happening in reality. The researcher does not intervene and, therefore, has neither control over the exposure nor other factors that influence the outcome and might confound the observed association of interest. As a bystander, the epidemiologist tries to assess whether the exposure has affected the outcome by unraveling the effect of the exposure of interest and effects of other influential factors to which study participants were exposed. This is not always an easy task. More on this subject matter will be discussed in this chapter under the section on "Multiple Concurrent Causes."

Another important distinction that can be made in epidemiologic research concerns the timing of the measurement of the exposure (cause) and the outcome (effect). On the one hand, when both the exposure and outcome are assessed at the same time, epidemiologists speak of a *cross-sectional study*. On the other hand, epidemiologists speak of a *longitudinal study* when the assessment of the exposure and outcome does not occur at the same time but is carried out sequentially. A major disadvantage of cross-sectional research is the fact that the cause-and-effect relationship under investigation may be obscured because the putative "cause" may actually have been affected by the presumed "outcome," instead of the other way around. This is a typical chicken-and-egg problem, which is called "reverse causation" by epidemiologists. Longitudinal research prevents this problem by studying events in the right order of time according to the correct temporal relationship between exposure and outcome, that is, the exposure is determined before the outcome occurs because the cause needs to be present before the effect happens.

Longitudinal studies can be conducted in two ways, namely prospectively or retrospectively (Fig. 3.2). The main difference between the two is that the outcome under investigation has not yet occurred in a *prospective study*, whereas it has already occurred in a *retrospective study*. This means that while prospective studies rely on information about current exposure(s) and future assessment of outcome occurrence, retrospective studies rely on information about past exposure(s) and current assessment of outcome occurrence.

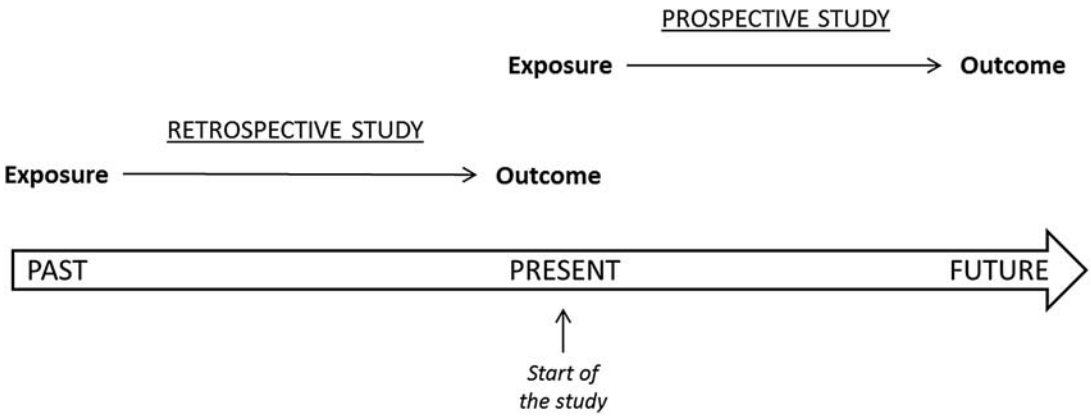


FIGURE 3.2 **Prospective versus retrospective studies.** In a retrospective study, past exposures are determined and the outcome of interest is assessed in the present. In a prospective study, present exposures are determined and the outcome of interest is assessed in the future. In both a retrospective and prospective study, the cohort members are followed up for a specified period of time from the start of the study. The main difference between the two types of studies is that the follow-up period has already occurred at the start of a retrospective study, whereas it has not yet occurred at the start of a prospective study.

The Randomized-Controlled Experiment

A randomized-controlled experiment is an experimental study with a longitudinal (prospective) design. As mentioned before, the researcher has control over the exposure in an experimental study, in which the exposure often is some form of treatment. In a classic randomized-controlled experiment or trial (RCT), this is done by randomly assigning the treatment to the study participants, who could be for instance patients with a certain chronic disease. The random treatment allocation is done in such a way that each patient has a predetermined chance to either receive the experimental treatment (the treatment group) or not (the control group), mostly based on a 50–50 chance. The outcome is assessed in both the treatment and the control group after a certain period of follow-up to evaluate the treatment effect by comparing the outcome, for instance symptom relief, between groups. Observing a meaningful difference provides evidence for the effectiveness of the treatment and, thus, for the causal relationship of interest. The study design of a randomized-controlled experiment is schematically presented in Fig. 3.3.

The key feature and chief strength of a randomized-controlled experiment is the utilization of randomization. In this way, study groups are created that at the start of the experiment have a large contrast in the exposure status of interest (treatment vs. no treatment) but do not differ, or only differ randomly, with regard to other variables that could affect the outcome to be studied and potentially bias the results. These other variables could for instance be important prognostic factors that influence the course of disease. Consider, as an example, about the influence of the age in a group of chronically ill patients. In general, older age is related to worse prognosis. If for some reason the patients allocated to the treatment group are considerably younger than the patients in the control group, a better outcome in favor of the treatment group could just be a result of the difference in age-related prognosis

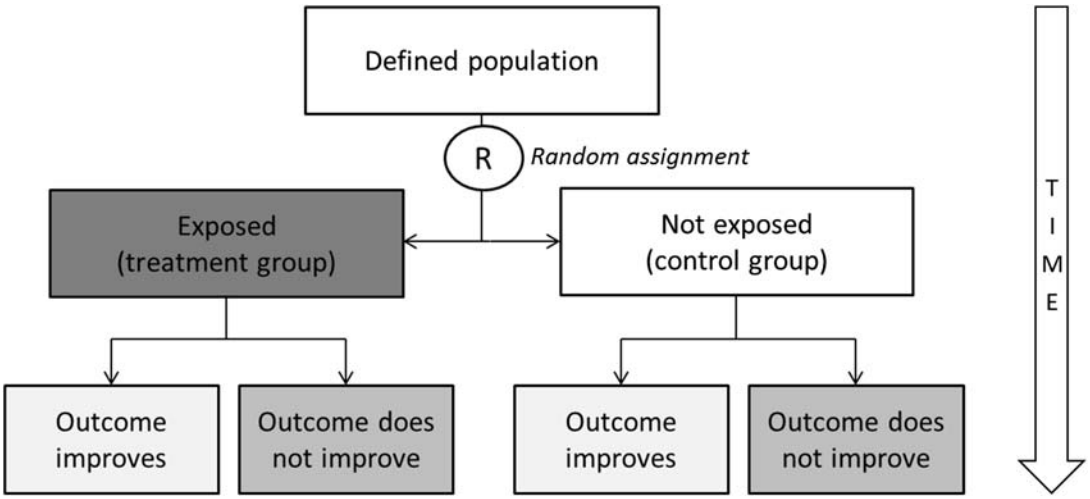
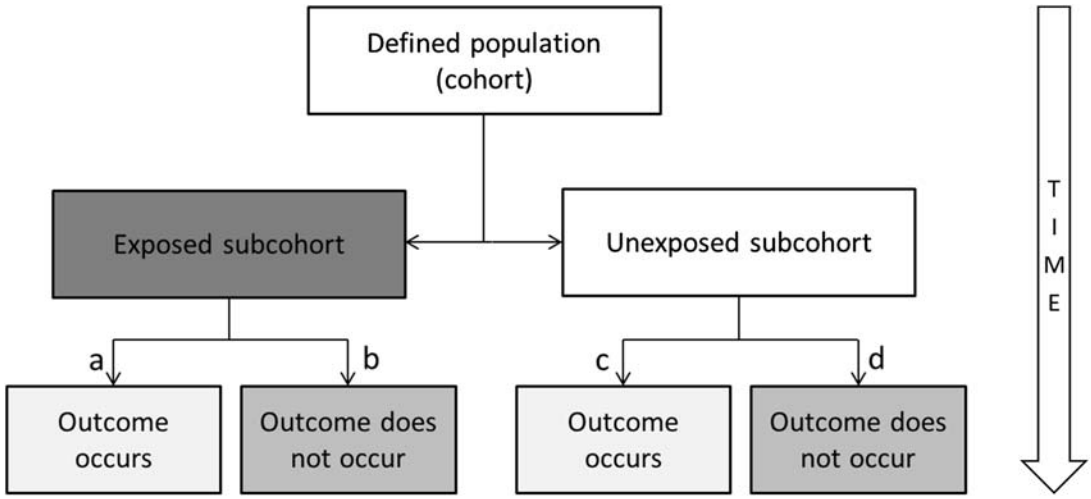


FIGURE 3.3 **Outline of a randomized-controlled study.** At the start of a randomized study, outcome status is determined at baseline in a defined study population. Next, members of the study population are randomly allocated in such a way that one group of participants is exposed to some form of intervention (treatment group) and another group is not (control group). After a prespecified period of follow-up, outcome status in both groups is determined and changes therein relative to baseline are compared between groups to assess the effect of the intervention.

and not of the effect of the experimental treatment. A major advantage of randomization is that it balances the distribution across study groups of not only measurable factors such as age, but also nonmeasurable factors, such as a person's complete genetic makeup. Randomization can be done in several ways, such as by simply flipping a coin or by means of a random number generator.

Another strength of a randomized-controlled study is that under ideal circumstances, depending on the specific characteristics of the experimental treatment (eg, its physical appearance, mode of administration, etc.), the control group could be given a sham treatment. This so-called placebo treatment is identical to the experimental treatment, except for the active ingredient. The study participants can in this way be blinded or masked with regard to the group allocation (treatment or control), so as to prevent them from, knowingly or unknowingly, influencing the study outcome. A person who is aware of being in the control group may, for example, misreport information about the outcome, especially when it is a subjective outcome such as pain or quality of life, or may even dropout of the study. The process of blinding can be applied to not only the participants but also the researchers, data analysts, and any other persons involved who could potentially influence the study outcome. Classically, a study is called single-blind if only the participants receiving the treatment are blinded, and double-blind if the person who gives the treatment is blinded as well.

Because of the above properties of a randomized-controlled experiment, which allow optimal standardization and control over the causal relationship under investigation, this type of study is regarded as the gold standard of study designs. However, as already mentioned, practical and ethical issues often preclude the conduct of an experimental study. In such cases, a cohort study is the next best thing.



	Outcome occurs	Outcome does not occur	Total	Incidence proportion
Exposed	a	b	$N_{exposed}$	$a/N_{exposed}$
Unexposed	c	d	$N_{not\ exposed}$	$c/N_{not\ exposed}$

FIGURE 3.4 **Outline of a cohort study.** A cohort study starts with a defined population (cohort) of individuals at risk for the outcome of interest. Exposure status of the cohort members is determined at baseline. After a prespecified period of follow-up, occurrence of the outcome of interest is determined (eg, the incidence proportion or incidence rate of disease) and compared between exposed and unexposed subcohorts to assess the association between exposure and outcome. A standard crosstable is shown that is used to assess the association between a dichotomous exposure (eg, exposed vs. unexposed) and a dichotomous outcome (eg, diseased vs. nondiseased) in a cohort study. Based on this table, both absolute (risk difference) and relative (relative risk, odds ratio) measures of association can be calculated. Risk difference: $RD = [a/(a+b)] - [c/(c+d)]$; relative risk: $RR = [a/(a+b)]/[c/(c+d)]$; odds ratio: $OR = (a/b)/(c/d) = ad/bc$.

The Cohort Study

A cohort study is an observational study with a longitudinal design (Fig. 3.4), which closely resembles that of an experimental study, except for the fact that the researcher does not (randomly) determine who is exposed and who is not. An important aspect of a cohort study is that it moves from exposure (cause) to outcome (effect), that is, its direction is forward. This means that the point of departure in a cohort study is a defined group (cohort) of persons in whom the health outcome under investigation is not present (yet) at the outset of the study. Each member of the cohort, however, must be at risk of developing the outcome during the course of the study. A cohort study starts with determining the exposure status of the participants, on the basis of which subcohorts are defined (eg, exposed and unexposed subcohorts). Next, all cohort members are followed up for some period of time during which the occurrence of the health outcome of interest (eg, disease or death) is monitored within the subcohorts. Finally, the frequency with which the outcome has occurred (eg, the incidence rate of disease

or survival time) is compared between subcohorts to assess the association between the exposure of interest and the outcome of interest. A meaningful difference in outcome occurrence between exposed and unexposed subcohorts is suggestive of the existence of an association.

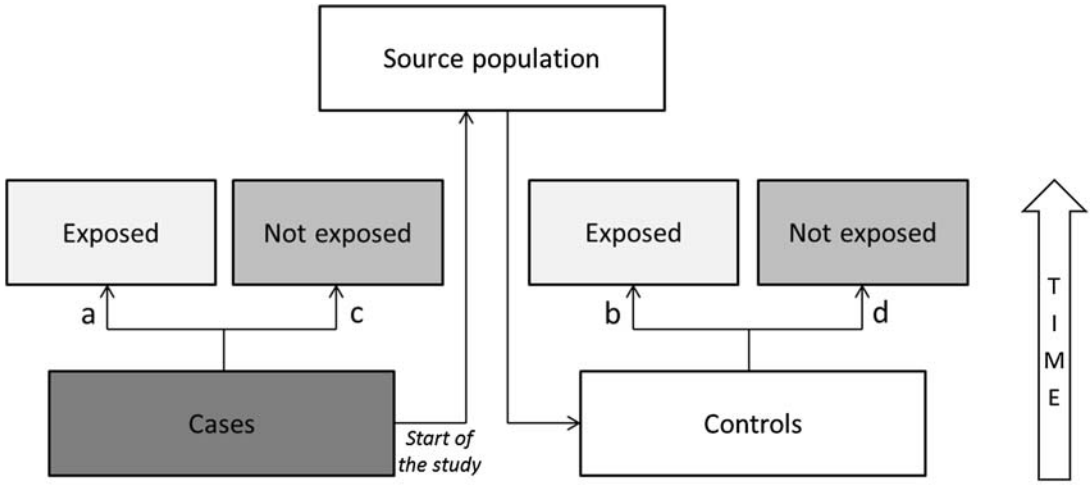
A cohort study can be conducted either prospectively or retrospectively (see Fig. 3.2). Since a retrospective cohort study depends on past information about the exposure history of the cohort members, this type of cohort study is also called a historical cohort study. As a hypothetical example, suppose you want to investigate whether the frequency of cell phone use is associated with the risk of a brain tumor. When studying this prospectively, you would have to begin with selecting a group of persons who are likely to use cell phones and who do not have a brain tumor, but who are at risk of developing such a tumor at some time in the future. The latter simply means that they should have a brain to begin with. You decide to start with a cohort of 40- to 50-year-old working men and women, determine the frequency of their cell phone use (eg, average number of uses in a week), and then follow them up for a period of 15 years, after which you assess who did develop a brain tumor since the start of the study and who did not. Clearly, a prospective cohort study can take a very long time to complete. As an alternative, you could opt to study the association between cell phone use and brain tumors retrospectively. In that case, you have to define the baseline cohort and assess its exposure status somewhere in the past (eg, 15 years ago), and determine the prevalence of brain tumors in the present. Thus, at the start of this hypothetical retrospective study, the 15-year follow-up period would already have taken place. Even though this could be a very time-efficient approach, there are some caveats in conducting such a study, especially with regard to the assessment of exposure status. Cell phone use was not as common 15 years ago as it is today, which means that you probably will have to select a larger group of people to start with. Moreover, accurately assessing the frequency of cell phone use in the past will be no easy task. Under ideal circumstances, there would be some sort of database, perhaps from providers of mobile phone networks that can be used to determine the frequency of cell phone use by the individual cohort members. Another option would be to ask the cohort members themselves about their past cell phone use, but the quality and completeness of information gathered in this way is going to be strongly influenced by their present vital and health status.

Cohort studies have a number of strengths. First, the longitudinal design and direction of the study mirrors the temporal relationship between exposure and outcome. Second, the association of the outcome with several different exposures can be assessed simultaneously, including exposures that are relatively rare. Besides cell phone use in the above example, you could also look at associations between brain tumors and other exposures, such as lifestyle, working environment, and family history of cancer. Third, especially in a prospective cohort study, collecting data on factors other than the exposure of interest, which presumably are related to the occurrence of the outcome, allows adjustment for these factors when analyzing the exposure–outcome association under investigation. In retrospective studies, this depends on the availability and quality of past data on such factors.

Notwithstanding its methodological strengths, setting up and conducting a well-designed cohort study can be severely constricted in practice by limited availability of time and financial resources. Additionally, a cohort study is not the best design for studying infrequent outcomes, such as a rare disease, because the size of the baseline cohort would have to be extremely large to ensure that enough outcome events would happen during the follow-up period. In case of a rare outcome, a case–control study should be the first choice.

The Case–Control Study

A case–control study is an observational study with either a longitudinal (retrospective) or a cross-sectional design (Fig. 3.5). In contrast to a cohort study, the causal relationship of interest is investigated in a backward direction in a case–control study. This means that the study moves from outcome (effect) to exposure (cause). Indeed, a case–control study starts with the selection of a group of cases with the outcome of interest, most often



	Cases	Controls
Exposed	a	b
Unexposed	c	d
Total	N_{cases}	$N_{controls}$
Exposure prevalence	a/N_{cases}	$b/N_{controls}$
Exposure odds	a/c	b/d

FIGURE 3.5 **Outline of a case–control study.** A case–control study starts with the identification and selection of a group of individuals with the outcome of interest (cases). Next, the population from which the cases derive is defined, and from this source population a group of individuals without the outcome of interest is selected to serve as controls. Then, exposure status (past or present) is determined and compared between cases and controls to assess the association between exposure and outcome. A standard crosstable is shown that is used to assess the association between the exposure and outcome in a case–control study. Based on this table, only an odds ratio can be calculated as the measure of association, because the incidence of the outcome of interest cannot be determined in a case–control study (in contrast to a cohort study). However, when the outcome under investigation is rare (ie, it occurs on average in less than 10% of individuals from the source population), the odds ratio can be used as a valid estimate of the relative risk. Odds ratio: OR = $(a/c)/(b/d) = ad/bc$.

individuals who are newly diagnosed with or already have a particular disease or injury. The next step is identifying the source population from which the cases derive. This population can for example be the inhabitants of a certain town, people of a specified neighborhood, or other patients from the hospital in which the cases were diagnosed. From this source population, a group of controls is (randomly) selected. The controls are individuals who do not have the outcome under investigation. The principal idea behind the selection of appropriate controls from a well-defined source population is that these individuals would have been selected into the study as cases if they had the outcome of interest. After the cases and controls have been identified, the prevalence of exposure is assessed in both groups. This can be done cross-sectionally, that is, at the time of the study (current exposure), or retrospectively, that is, at some time in the past (past exposure). The association between the exposure and the outcome of interest is assessed by comparing the exposure prevalence between cases and controls. Observing a substantially and meaningfully different prevalence of exposure among cases relative to controls would suggest the presence of an association between the exposure and the outcome.

An example of a case–control study is the investigation of a relationship between a rare type of skin cancer and deodorant use. For this hypothetical study, cases are enrolled through the dermatology departments of a number of participating hospitals. Controls are enrolled in the same hospitals at the orthopedic and rheumatology departments. By selecting controls in the same hospitals but at departments other than the dermatology department, two assumptions are made. The first is that these individuals come from the same source population as the cases. Hence, they would have been admitted to the same hospitals in case they had been diagnosed with the skin cancer of interest. The second assumption is that the exposure status of the controls is representative of the background exposure experience in the source population. This assumption is reasonable, unless deodorant use is associated with the condition for which the controls were admitted to the other hospital departments. In that case, the exposure status of the controls would misrepresent the background exposure in the group of cases, potentially biasing the observed association between exposure and outcome. To determine exposure status, deodorant use (past or current) is assessed in both cases and controls by means of, for instance, a questionnaire or an interview. Similarly, other factors that may be related to the outcome, such as the frequency of sunlight exposure or sunbed use, can also be assessed to account for these factors in the analysis.

Case–control studies have several strengths. Compared to (prospective) cohort studies, case–control studies generally cost less time and money. Moreover, they are more suitable for studying rare outcomes. Whereas a cohort study would need to include a large number of individuals to get a sufficient number of outcome events during follow-up, a case–control study does not need such a large sample size due to the fact that the selection of study participants is based on the presence of the outcome event. Furthermore, by increasing the number of control subjects selected per case (eg, up to four controls for every case), a case–control study gains statistical efficiency, meaning that the statistical power to find a significant association between the exposure and outcome that is not likely due to chance (ie, beyond reasonable doubt, as it were) increases with a given number of outcome events. Thus, fewer cases would have to be included to detect an association with a sufficient, predefined level of statistical power.

An important issue in the setup and conduct of case–control studies concerns the proper definition of the source population that produced the cases. The selection of an appropriate control group is problematic when this source population is ill-defined. Consequently, this could lead to false conclusions with regard to the observed association, if any, because the comparison between cases and controls with regard to their exposure status is biased. Another limitation of case–control studies is the potential for bias in determining the exposure status. The assessment of exposure can be influenced by the fact that the outcome has already occurred in part of the study participants. For example, patients with a brain tumor might overreport the use of cell phones when they believe that their past cell phone use could be one of the reasons for their tumor. If the controls, however, do not misreport their cell phone use, a biased comparison results that could invalidate the study findings.

Hierarchy of Study Designs

Each of the designs discussed above has its own strengths and weaknesses. Choosing an appropriate study design to address a specific research question about a causal relationship depends on methodological considerations related to the design's specific strengths and weaknesses, as well as on practical issues such as the availability of time and money, as well as data. Methodologically, a hierarchical ranking of the quality of epidemiologic study designs is often proposed (Fig. 3.6). Accordingly, randomized experimental studies are ranked higher than observational studies. Among observational studies, longitudinal studies are better than cross-sectional studies for the majority of (causal) research questions. Prospective studies are better than retrospective studies and, generally speaking, (prospective) cohort studies are ranked higher than case–control studies.

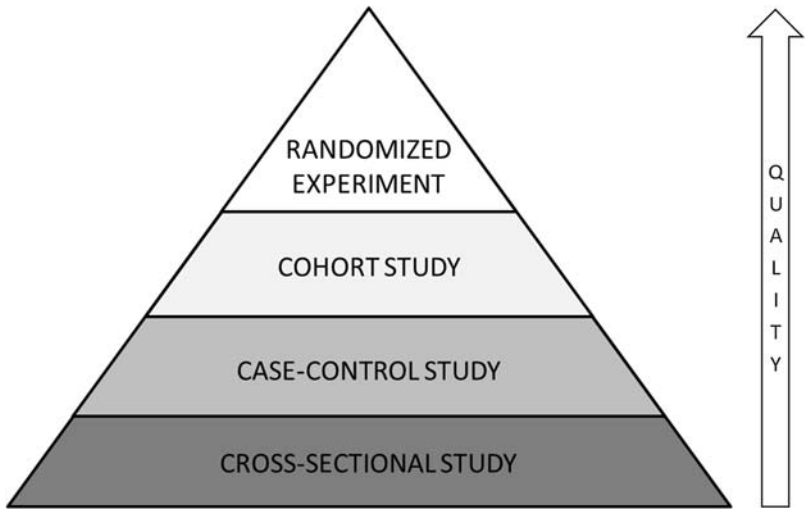


FIGURE 3.6 **Hierarchy of study designs.** The hierarchy is based on the methodological quality of the study design and the strength of causal inferences that can be made on the basis of the findings about an association between exposure and outcome in these studies (given that they are set up and conducted adequately).

The hierarchy of types of studies is based on the methodological quality inherent in the study design. Assuming that the theoretical design is put into practice as it should be, higher quality designs provide stronger evidence for the existence of a causal relationship. This refers to the so-called *internal validity* of the study, which is related to the causal inferences that can be made on the basis of the study findings. The findings of a study with high internal validity can be extrapolated with more confidence to the target population of interest, from which the study population was a representative sample. For instance, an association, or the lack thereof, found between exposure to cell phones and brain tumors in a large prospective cohort study can be generalized more confidently than an association found in a small cross-sectional study.

In general, the internal validity of a study can be compromised by two types of errors that can be made during different phases of the planning and performance of an epidemiological study in practice. The likelihood of such errors occurring depends on the type of study design, at least in part. Different types of designs leave more or less room for error, which can lead to erroneous conclusions about the study findings and flawed inferences about cause-and-effect relations. Later in this chapter the specific types of errors and their consequences will be discussed.

FACTUAL PROBABILITY

Probability of Disease and Injury

In order to make liability claims, absolute probabilities of injury will need to be derived and compared under different scenarios. This is more complicated than it may look at first glance as aspects such as the appropriate time at risk (the longer, the higher the probability), the definition of a population representative for the case under investigation, and the definition of the injury or disease itself (the wider defined, the higher the probability) need to be defined. In it simplest form this probability is given in Fig. 3.1. The epidemiological proportion can be expressed in different ways. Two common expressions are "prevalence" and "incidence" (Figs. 3.7 and 3.8). The prevalence gives an indication of the proportion of cases with the disease or injury at a given time window. This time period can vary from a specific moment in time (ie, point prevalence) to a lifetime (eg, period prevalence) and any time in between. An example of a period prevalence is the number of employees of company X who reported a work-related accident before reaching retirement age. The use of prevalence

$$\frac{\text{Total number of } \textit{existing} \text{ cases in a defined time period}}{\text{Total number of persons in the population to which these cases belong}}$$

FIGURE 3.7 **The epidemiological proportion, expressed as "prevalence."**

$$\frac{\text{Total number of } \textit{new} \text{ cases in a defined time period}}{\text{Total number of persons in the population at the beginning of this period}}$$

FIGURE 3.8 **The epidemiological proportion, expressed as "incidence."**

to prove causality is not ideal and not possible in many cases. Patients may be missed for two completely different reasons, that is, they could have died or got better. For this reason, epidemiologists prefer to work with incidence measures.

Incidence represents the proportion of new cases that will get the disease or injury in a given time window. One could calculate the incidence of a first heart attack, but also the incidence of a recurrence. Incidence can only be calculated in populations of which the members are at risk of getting the disease. One cannot calculate the incidence of prostate cancer in a general population (consisting of men and women), for example.

It is important to understand that the probability calculation of prevalence or incidence can lead to completely different probability quantifications. The relationship between prevalence and incidence can be compared to the workings of a filter coffee machine, in which the incidence is represented by the speed at which the water drips into the filter and the prevalence by the water level in the filter. The microholes in the filter determine how fast the "prevalence pool" is drained and hence the relationship between prevalence and incidence (prevalence = incidence × disease duration).

Probability of Death

Mortality rate is nothing more than the *incidence* of death in a certain time window, often given as an annual rate per 100,000 (Fig. 3.9). In most cases, in a forensic investigation, this time window is much smaller and should be adjusted accordingly. Before mortality rates can be compared, for example, in cases with or without a tortious act, they should nearly always be adjusted for age or presented for different age categories. Age is of course a strong predictor of the probability of death. A perhaps more positive representation of death rates is the calculation of life expectancy. This topic is covered thoroughly in Chapter 10, *Survival Analysis*.

$$\frac{\text{Total number of new deaths in a defined time period}}{\text{Total number of persons in the population at the beginning of this period}}$$

FIGURE 3.9 **The epidemiological proportion, expressed as "mortality."**

$$\frac{\text{Total number of new deaths due to a certain disease or injury}}{\text{Total number of incident patients with this disease or injury}}$$

FIGURE 3.10 **The epidemiological proportion, expressed as a "case fatality rate."**

Case Fatality Rate and Survival Rate

The case fatality rate is the proportion of incident patients dying because of the disease or injury in a certain time window (Fig. 3.10). This probability gives insight into the prognosis of the disease and the efficacy of medical interventions. Both a poor prognosis and inadequate treatment could increase the probability of death. A case fatality rate is also expressed as an *incidence*; however, often the time window to which the rate applies is not given, creating confusion. An example is the case fatality rate of surgical operation that could be applied to the complications of the surgery during the hours or days after the operation, to the days of hospitalization, the year after operation, or a patient's whole life. One has to be careful to first decide on the applicable time window before interpreting a case fatality rate. The flipside of the case fatality rate is the survival rate. This is the proportion of patients with a certain disease or injury that is still alive after a certain time period.

LINKING A POTENTIAL CAUSAL FACTOR TO INJURY

Another source of confusion is the difference between the absolute probability of the disease or injury outcome and the quantification of the association between a potential causal factor and the outcome. In a 2009 Dutch medical negligence court case (NL 105.001.264, 2009), a pregnant plaintiff argued that she suffered a stroke because her doctor failed to give her antihypertensive drugs that could have prevented it. The court concluded that although the Probability of Causation (PC) was established to be 75%, the absolute risk of the stroke in her circumstances only increased from 1 in 13,937 to 1 in 11,315. The high court therefore concluded that on the basis of such a low probability no liability could be inferred. However, as we describe throughout this book, simply ascertaining that an event is rare is unhelpful for an evaluation of cause. We can all agree that the risk of dying in a commercial plane crash is less than 1 in 1,000,000, but this does nothing to tell us the cause of the death of an individual who was last seen boarding a plane that disappeared over the ocean. Causality relies on the comparison of the probabilities of two circumstances; the one under investigation and the one but for the circumstances of the investigation.

There are different ways that probabilities can be compared. It is essential to understand these different approaches as the numeric value can differ substantially from one measure to the other. For example, exactly the same facts could be summarized as a 16% increased risk on an *additive* scale or a 95% increased risk on a *relative* scale. Since these types of calculations are often used as a basis from which to conclude what "more likely than not" (often

interpreted as >50%) or "beyond a reasonable doubt" (sometimes interpreted as >98%), it is important to understand them.

An easy way to understand the various measures of association is to make use of a two-by-two contingency table (also called crosstable), in which the columns represent the positive and negative health outcome and the rows represent the presence or absence of the potential cause under investigation. Depending on the study design (see Figs. 3.4 and 3.5), this table can be used to calculate the *incidence* of disease or injury under different scenarios. The incidence of the negative health outcome when the potential cause is present (I_1) equals $a/(a+b)$. In the example of Table 3.1, I_1 would be calculated as $10/30 = 0.30$. The incidence of the negative outcome when the potential cause is absent (I_0) equals $c/(c+d)$. In the example of Table 3.1, I_0 would be calculated as $7/41 = 0.17$. We could also calculate the incidence of the negative outcome in the total population, which would be $(a+c)/(a+b+c+d)$ or $17/71 = 0.24$.

Risk Difference

One way to investigate the effect of the potential cause is to calculate the difference between these two incidences, $I_1 - I_0$ (see also Fig. 3.4). This is called the *risk difference (RD)* or *attributable risk (AR)*. If those who are exposed to the potential cause are injured more often than those who are not, then this can be interpreted as the extra risk liable to the presence of this potential cause. The RD in the numeric example of Table 3.1 would be $I_1 - I_0 = 0.30 - 0.17 = 0.13$ or 13%.

Relative Risk

A different way to compare I_1 and I_0 is to divide them, which is called a *relative risk* or risk ratio (RR; see also Fig. 3.4). The numeric example would be $I_1/I_0 = 0.30/0.17 = 1.76$. This means that those having been exposed to the potential cause are about 1.8 times as likely to get the disease or injury as those who were not exposed. The same RR is also sometimes expressed as an 80% (not 180%) increased risk ((RR − 1) * 100%). This makes sense, as an RR of 1.0 (or 100%) means no association, that is, the risk of disease or injury is the same for both groups ($I_1 = I_0$). An RR < 1 indicates a protective effect, whereas an RR > 1 indicates a risk-increasing effect. Epidemiologists also use similar ratio measures such as odds ratio (OR, see Fig. 3.5), hazard ratio, and rate ratio that in the strictest sense are different measures, but can be interpreted in a similar way as the RR.

TABLE 3.1 The Two-by-Two Contingency Table

	Negative health outcome	Positive health outcome
Potential cause present (exposed)	10 (*a*)	20 (*b*)
Potential cause absent (not exposed)	7 (*c*)	34 (*d*)

Comparative Risk Ratio

The *comparative risk ratio* (CRR) is a unique metric in FE; it allows for the comparison of probabilities applicable to the investigated circumstances of an individual injury or disease. The CRR is suitable as a basis for a specific (individual) causation opinion to be provided in a legal setting. The metric is denoted as P_1/P_0. Because a CRR is based on the unique circumstances surrounding the injury or disease of an individual, it may or may not be derived from an RR or an OR. An example of an RR that could be used as a CRR applicable to the investigation of injury cause in an individual would be the examination of the frequency of serious injury in 100 randomly selected unrestrained drivers exposed to a 20 mph frontal collision versus the frequency of serious injury in 100 randomly selected restrained drivers exposed to the same collision severity and type. If the frequency of serious injury in the group exposed to the presumptive hazard (failure to use a restraint) was 0.15 and the frequency in the unexposed (restrained) group was 0.05, then the CRR would be $0.15/0.05 = 3$. The populations that P_1 and P_0 are derived from the example are substantially similar in all respects, with the exception of the exposure to the investigated hazard, which is the failure to use a seat belt.

In some instances encountered in a legal setting, P_1 and P_0 are not similar populations, and thus the CRR cannot be derived from either an RR or OR. An example of such a situation occurs when P_0 is a per event risk, and the denominator is a per-time risk (also known as a cumulative risk). An example would be the investigation of a pulmonary embolism (PE) that occurred a week after a patient sustained a lower extremity fracture in a crash. Such complications often result from blood clots forming in the legs and then traveling to the lungs. If the patient had a history of deep vein thrombosis (DVT) in the lower extremities prior to the crash, then a CRR might consist of comparison between a P_0 of a PE following a lower extremity fracture (a per event rate) and a P_1 of the 1-week risk of PE in a patient with DVT (a time-dependent probability).

Another example of a CRR based on dissimilar populations is when there are only a limited number of potential causes to be compared. An example is the investigation of the cause of an adverse reaction in a person who took two different drugs at the same time, both of which could have caused the reaction (and which, for the example, do not interact with each other). In such a situation, the CRR applicable to the unique circumstances experienced by the individual would be estimated by comparing the adverse reaction rate for the two drugs. Numerous examples of how the CRR is estimated for a variety of situations encountered in a forensic setting are provided throughout this text.

Attributable Proportion Under the Exposed

The attributable proportion under the exposed (AP_e) is an indication of the proportion of patients who were exposed to the potential cause and got sick because of this exposure. It can only be used if $RR > 1$ and can be calculated by $(I_1 - I_0)/I_1$ or $(RR - 1)/RR$. When the CRR is based on an RR, these formulae also apply to the CRR. The result of the analysis, given as an RR, CRR, or AP_e, meets the legal standard of what is "more likely true than not," when the RR or CRR is ≥ 2.0 (95% CI > 1.0 lower boundary), or the AP_e is $\geq 50\%$. *NB*: A discussion of the propriety or reliability of the 2.0/50% threshold is not undertaken in this chapter. For more detailed discussions of the legal standards for causation see Chapter 1, *Legal*

Considerations of Forensic Applications of Epidemiology in the United States, Chapter 2, *Epidemiologic Evidence in Toxic Torts*, and Chapter 4, *Causation in Epidemiology and Law*.

In the numeric example given in Table 3.1, the AP_e would equate to 43%. We have to be careful in interpreting the AP_e directly as a PC, since the numeric value of the AP_e is only as reliable as the study design from which it was derived and the associated confidence that the value accurately represents the true magnitude of the relationship. As diseases and injuries are often caused by a multitude of sufficient and necessary causes, one needs to keep in mind that if a single identified cause explains more than 50% of the relationship this does not mean that other potential causes could also explain more than 50% because of the way that they interact. As an example, in a Dutch liability court case that was heard at the high court (CO4/303HR), an employer was held liable for the lung cancer of an employee who was potentially exposed to asbestos, but who was also a lifelong smoker. The AP_e for asbestos exposure was 55%, the court therefore concluded that the employer was 55% liable, ignoring that the AP_e for smoking was 80%. The sum AP_e was therefore not 100%. There are methods to calculate the combined AP_e for more exposures simultaneously, but these go beyond the scope of this chapter.

Attributable Proportion for the Total Population

The attributable proportion for the total population (AP_t) is typically not used in court, but will be discussed here shortly as it is easy to confuse with the AP_e and is a measure that is often cited by epidemiologists. The AP_t can be calculated by $(I_t - I_0)/I_t$ or $(RR - 1)/(RR + 1/p - 1)$, in which p is the proportion of exposed people in the total population on which the AP_t is based. The AP_t can be interpreted as the proportion of disease or injury in a specific population (consisting of exposed *and* unexposed individuals) that can be explained by the potential cause. For example, an AP_t for traffic death and seat belt use of 20% means that 20% of all traffic deaths in a certain population are attributable to the lack of seat belt use or, in other words, that if all occupants used their seat belts then 20% of traffic-related deaths could be prevented. The AP_t is mainly used for public health interventions.

SOURCES OF ERROR IN EPIDEMIOLOGIC RESEARCH

There are two types of error that can occur during the design and conduct of an epidemiologic study: *random error* and *systematic (nonrandom) error*. These errors threaten the precision and validity of the study findings, which is reflected in the associations observed in a study.

Before delving deeper into these sources of error and their consequences, it is important to explain the difference between an observed association and a true association. When epidemiologists undertake research on a hypothesized causal relationship, they go through several consecutive stages. First, they choose the most appropriate study design that best matches the main research question(s). Then, they select and recruit a suitable sample of participants from the target population and, subsequently, conduct the study measurements of the exposure and outcome variables as accurately and precisely as possible. Finally, the collected data is analyzed to assess the association between the exposure and outcome of interest, for instance by determining an RR, OR, or RD. The main goal of such

epidemiologic research is finding an association (ie, the observed association) that is a valid reflection of the underlying true association. In this sense, the true association refers to the real causal relation between exposure and outcome that exists in the target population but which cannot be directly observed. Any random or nonrandom errors made during the stages of this research process, no matter how subtle, can lead to a difference between the observed association and the true association of interest (Fig. 3.11). Consequently, the study findings will be invalid, conclusions will be erroneous, and causal inferences will be wrong. In a courtroom setting, for example a murder trial, this could mean that the actual perpetrator

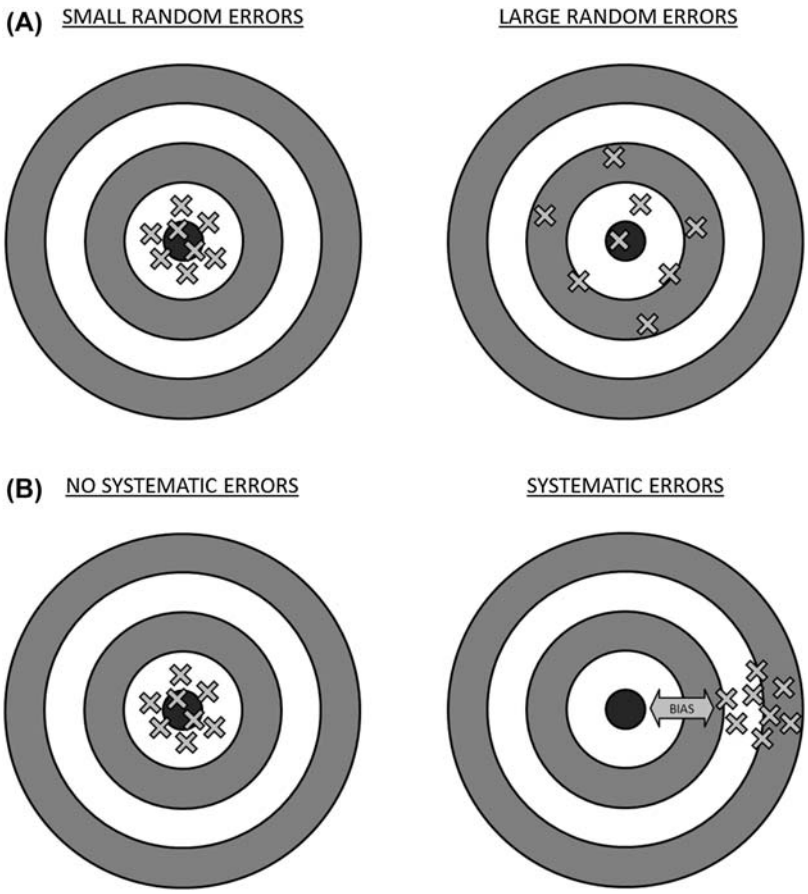


FIGURE 3.11 **The influence of random and systematic (nonrandom) errors on the precision and validity of epidemiologic study findings.** (A) Larger random errors, for instance due to smaller study sample sizes, are reflected by more scatter of the findings of individual studies (ie, the crosses) around the true but unknown association between the exposure and outcome of interest in the target population (ie, the bull's eye). On average, however, the study findings reflect the true association (ie, no bias). (B) Even studies that are not significantly influenced by random errors, as reflected by the small scatter of the crosses indicating a high precision of study findings, can systematically misrepresent the true association between exposure and outcome in the target population. On average, the study findings are invalid due to nonrandom errors (ie, bias) that lead to a systematic deviation of the study findings from the truth.

would be falsely be declared innocent or that the suspect who actually is innocent would falsely be convicted for murder based on flawed evidence.

In general, the consequences of any errors made during an epidemiologic study come to expression as *bias* in the measure of association that is determined on the basis of the study data. According to the Dictionary of Epidemiology, bias is defined as:

> Systematic deviation of results or inferences from the truth. An error in the conception and design of a study – or in the collection, analysis, interpretation, reporting, publication, or review of data – leading to results or conclusions that are systematically (as opposed to randomly) different from the truth.

Bias can influence the study findings in three ways (Fig. 3.12). First, the observed association may appear stronger than the true association, thus overestimating the real effect, if

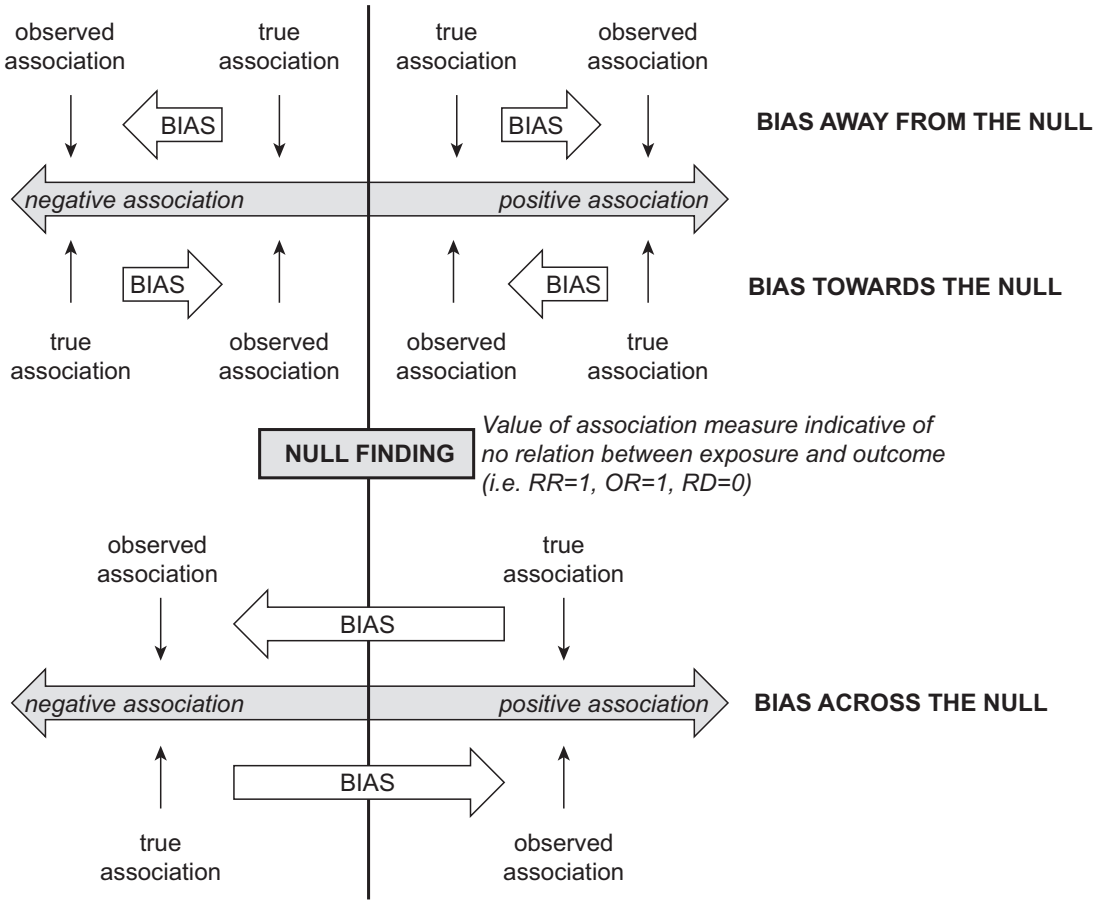


FIGURE 3.12 **Three ways in which systematic errors can bias the findings of an epidemiologic study.** Bias can lead to an overestimation of the true association, including a spurious association when none really exists (bias away from the null), an underestimation of the true association (bias towards the null), or a reversal of the true association (bias across the null).

any, the exposure has on the outcome of interest. This is called *positive bias* or *bias away from the null* finding of no effect (ie, an RR or OR of 1 or a RD of 0). Second, the observed association may appear weaker than the true association, thus underestimating the real effect of the exposure on the outcome of interest. This is called *negative bias* or *bias towards the null*. Third, the observed association may appear opposite to the true association. This can mean that a positive association is observed (eg, RR $>$ 1) while in reality it is negative (eg, RR $<$ 1), or vice versa. This is sometimes called *bias across the null* or switch-over bias. The first two ways in which errors can bias an observed association are purely quantitative, that is, only the quantity of the association appears different (stronger or weaker). In contrast, the third way biases the observed association qualitatively, that is, the quality of the association appears different (eg, a true protective exposure may appear harmful). The major sources of error in epidemiologic research and their potential consequences will be further explained below.

Studying Populations Through Sampling

Epidemiologists aim to study the distribution of disease and its determinants in populations. Because studying complete populations is not possible most of the time, epidemiologic research makes use of *sampling*. Members of the population are selected in such a way, preferably in a random manner, that they constitute a representative sample of the target population. The idea behind sampling is that any association found in the sample, assuming no errors were made, reflects the association in the population and can thus be generalized to all the members of the target population. Because of the sampling process, however, associations observed in samples may differ by chance from the true association in the target population. This random error is called *sampling error*.

To explain sampling error, suppose you have a large opaque vase filled with 10,000 small beads. Half of these beads are red and half are blue, but you do not know this. The true distribution of red and blue is thus 50/50. To find out this distribution for certain, you could take out the beads one by one and count them all. Alternatively, you could also take out part of the beads (eg, 10, 50, 100, or 1000) and determine the red/blue distribution in this sample as an estimate of the distribution in the whole population of beads. You can imagine that every time you would do this, the results may differ from the true distribution just because of chance. For example, the one time you may find that there are more red beads than blue beads in your sample (eg, 52/48, 60/40, or 70/30) and another time you may find the opposite (eg, 48/52, 40/60, or 30/70). These random differences are the result of sampling error. An obvious strategy to reduce the influence of sampling error is to take larger samples. The likelihood that the distribution in the sample differs from the population distribution decreases with increasing sample size. The chance that you will correctly estimate the true 50/50 distribution of red and blue beads is larger with a sample of 1000 beads compared to a sample of 10 beads. Given the true distribution, it is more likely to erroneously observe, for example, an 80/20 distribution by chance with a sample size of 10 (8 red and 2 blue beads) than with a sample size of 1000 (800 red and 200 blue beads).

To apply the above example to epidemiologic studies, the beads are analogous to human beings and the colors reflect the presence or absence of the health outcome of interest, such as

a disease. An epidemiologist tries to determine how the disease is distributed in a sample of the population and explain the observed frequency distribution by also determining whether the distribution of possible disease determinants differs between diseased and nondiseased subsamples. As explained above, differences may occur partially due to chance and the influence of chance can be estimated by statistical parameters. Two widely used parameters for estimating the influence of chance and the precision of the observed association measures are *P-values* and *confidence intervals*. Briefly, *P*-values estimate the probability that the observed association or a stronger one is due to chance, under the assumption that in reality there is no association between the exposure and outcome under investigation (ie, the null hypothesis). Commonly, the observed association is said to be statistically significant when the *P*-value is smaller than 0.05, which means that the probability of a chance finding is less than 5%. Note, however, that there is still a chance, albeit a small one (<5%), of drawing the false conclusion that there is an association while in reality there is none (ie, a Type I error; see Fig. 3.13). Confidence intervals are often preferred by epidemiologists over *P*-values because these intervals say something about the magnitude and precision of the observed association in addition to the statistical significance of that association. Confidence intervals indicate a range of values within which the observed association measure is likely to fall based on a prespecified level of confidence (eg, 95%). For example, if you find an RR of 1.8 in your study sample and the 95% confidence interval ranges from 1.2 to 2.4, this indicates two things. First of all, because the interval does not include the null value of no effect (ie, an RR of 1.0), it indicates that the observed association is statistically significant ($P < 0.05$). Second, the interval indicates that if you were to repeat the study 100 times, each time

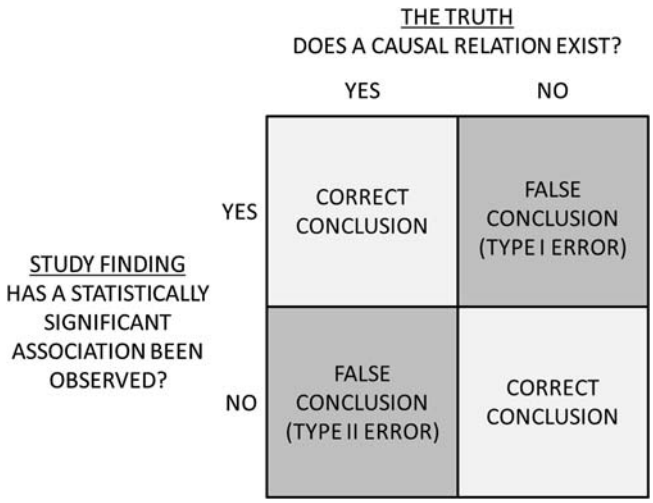


FIGURE 3.13 **The relation between the (unknown) truth about a causal relation between exposure and outcome, and conclusions about the existence of a causal relation based on the association between exposure and outcome observed in an epidemiologic study.** *P*-values estimate the probability that it is correctly concluded on the basis of the study findings that a causal relation exists in reality. The statistical power of a study indicates the probability that a correct conclusion is made based on the study findings, ie, the study has enough sensitivity to find associations that are real.

with a new (random) sample from the target population, you would find an RR within this interval in 95 out of the 100 samples. In other words, under the assumption that there is no association (H_0: RR = 1), it is 95% likely that the true RR in the sample lies within the range of confidence interval.

The most obvious way to reduce the influence of random error is to increase the sample size, thereby increasing the precision of the effect estimate. All other things being equal, when the sample size is larger, *P*-values will be smaller and confidence intervals will be narrower. By increasing the sample size, the so-called *statistical power* of the analysis of an association between exposure and outcome increases. This means that the likelihood decreases of missing an association that truly exists, which in statistical jargon refers to the chance of making a Type II error (Fig. 3.13).

Finally, there are two important things to note about the relation between random sampling error and sample size. The first is that statistical significance says nothing about the meaningfulness or relevance of an observed association. In extremely large samples, even the smallest insignificant differences become statistically significant. An epidemiologist should be able to distinguish what is relevant from what is not, regardless of statistical significance. The second and even more important thing to note is that although random errors can be reduced by increasing the sample size of an epidemiologic study, systematic errors cannot be remedied in this way. Systematic errors threaten the validity of study findings regardless of the size of the study sample. As an example, if you measure body weight with a wrongly calibrated scale, the magnitude of the systematic (measurement) error will remain the same when applying this biased scale in 10, 100, 1000, or 10,000 people, even though the precision of the inaccurate weight estimate will increase with an increasing number of people.

Threats to Validity: Common Forms of Bias

Systematic errors can occur at many stages during the design, conduct, and analysis of epidemiologic studies. The resulting bias that threatens the validity of the study findings can come in different guises. Accordingly, a plethora of possible types of bias can be found in the literature. To prevent confusion about terminology, however, the multitude of bias terms are usually reduced under three common denominators in the epidemiologic literature: selection bias, information bias, and confounding. Selection and information bias will be discussed below, and bias due to confounding is discussed in the section on "Multiple Concurrent Causes".

The first major category of bias is *selection bias*. Selection bias results from systematic errors during the definition and identification of an appropriate target population and the selection and follow-up of a sample of individuals from this population. Ultimately, these errors lead to inclusion of the wrong individuals in the study sample, which thereby misrepresents the true target population and, thus, also the causal relationship of interest. In an epidemiologic study on the association between an exposure and outcome, study findings can be invalidated by selection bias when the probabilities of individuals from the target population being included (and staying included) as subjects in the study population differ according to the exposure and outcome being studied. In Box 3.2, the mechanism by which selection bias can influence the findings of an epidemiologic study is explained in more detail.

Systematic errors leading to selection bias can arise in various ways. One important phenomenon that can lead to selection bias in epidemiologic studies is self-selection of study subjects. Individuals who volunteer to participate in epidemiologic studies (ie, responders) often differ from individuals who do not want to participate (ie, nonresponders). It is known, for example, that responders are on average more health-conscious (eg, do not smoke, are more physically active, utilize health-care services more often, etc.) and have a higher socioeconomic status than nonresponders. This can result in so-called nonresponse bias, which is a form of selection bias that arises if the different characteristics of responders and nonresponders, that is, the motives and factors that affect the decision to volunteer, are related to the exposure and outcome under investigation. Consequently, the responders will not be a valid representation of all individuals from the target population, and the observed association will be biased.

BOX 3.2

SELECTION BIAS IN EPIDEMIOLOGIC STUDIES

Fig. 3.14A–C shows the mechanism by which selection bias can influence the association between an exposure and outcome observed in an epidemiologic study.

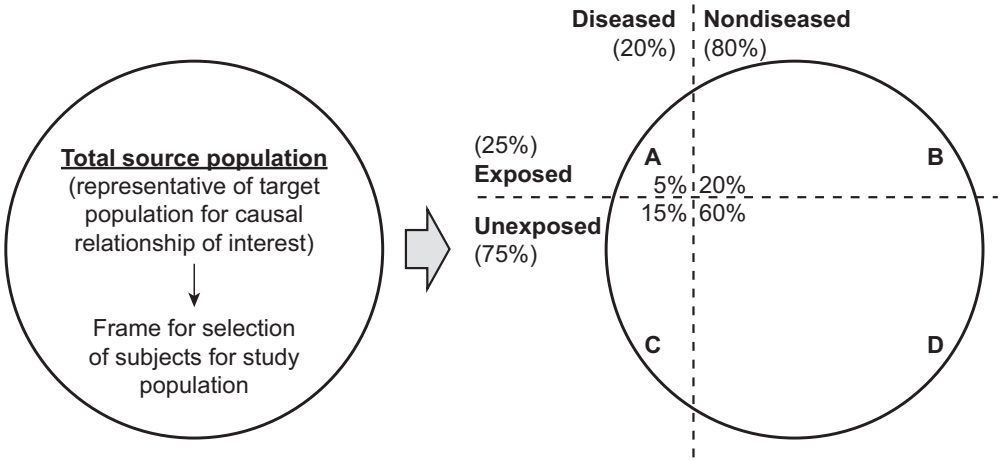


FIGURE 3.14A **The true exposure–outcome relation in the source population.** The total source population represents the individuals who are available for sampling of subjects for inclusion into the study population. The source population is representative of the exposure and outcome distribution in the target population, thereby reflecting the causal relation of interest (ie, the true association between exposure and outcome). For a dichotomous exposure and outcome, a possible measure of the true association is the relative risk (RR = $[A/(A+B)]/[C/(C+D)]$). In this example, the true RR is $(0.05/0.25)/(0.15/0.75) = 1.0$, meaning that in reality there is no causal relation between the exposure and outcome under investigation.

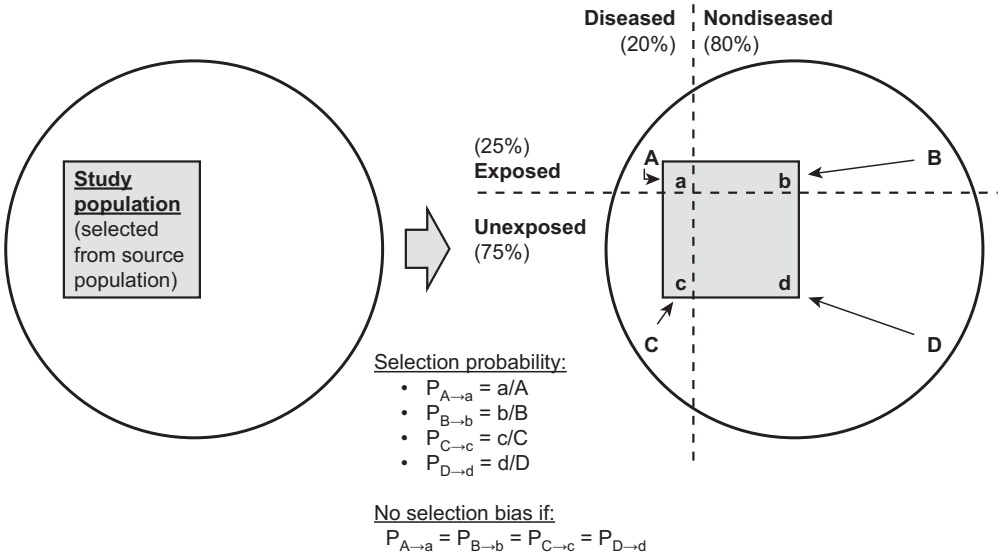


FIGURE 3.14B **The observed association in a sample from the source population.** Individuals selected from the source population constitute the study population, which is free from selection bias when the probabilities of being selected into the study population do not differ according to the exposure and the outcome under study. In that case, the observed association in the study sample will be a valid estimate of the true (but unknown) association in the target population. There is no selection bias because the selection probabilities do not invalidate the study findings as expressed by the observed measure of association. For example, the observed RR $= [a/(a+b)]/[c/(c+d)] = [A(P_{A\to a})/(A+B)(P_{A\to a}+P_{B\to b})]/[C(P_{C\to c})/(C+D)(P_{C\to c}+P_{D\to d})]$, which equals the true RR if the selection probabilities are identical (ie, if $P_{A\to a} = P_{B\to b} = P_{C\to c} = P_{D\to d}$).

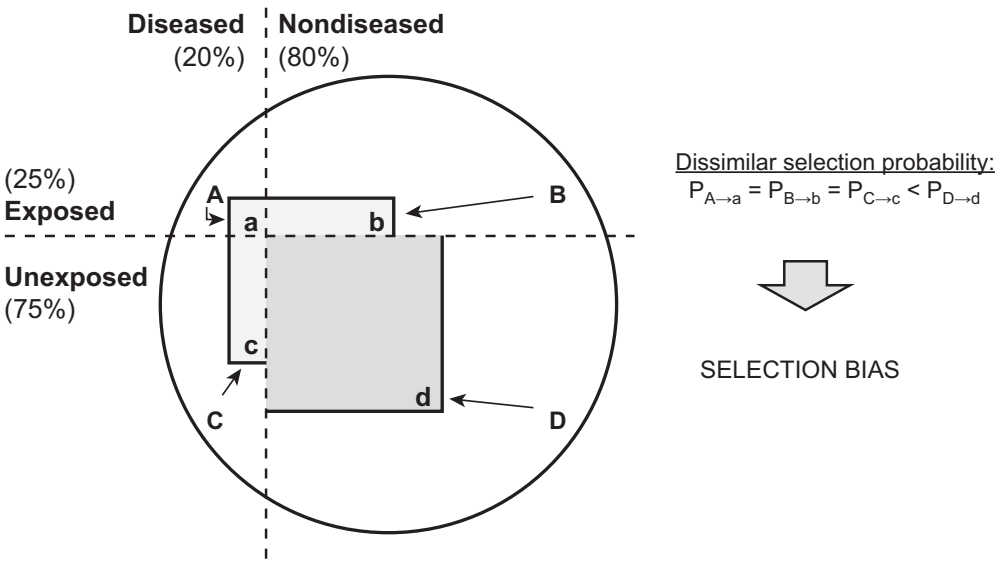


FIGURE 3.14C **Selection bias in the observed association.** Selection bias occurs when the selection probabilities are dissimilar. In the example shown, the probability of being selected from the source population into the study population is higher for individuals who are not exposed and who do not have the outcome (disease) of interest. In this case, as there was no association in reality (true RR = 1.0), the resulting bias causes a spurious association (observed RR > 1.0).

Another important phenomenon that can give rise to selection bias is selective loss to follow-up of study subjects. It is not uncommon in longitudinal studies that individuals who initially agree to participate drop out later on, especially in prospective cohort studies with years of follow-up. The reasons for dropout can be diverse, such as severe illness or death, migration, or withdrawal of consent. This does not lead to bias when the subjects who are lost to follow-up do not differ systematically from those who stay in the study with regard to their exposure and outcome status (and other related characteristics). If the number of dropouts becomes too large, however, statistical power may be compromised by the decreasing sample size. Dropout also gives rise to selection bias when the proportion of people who are lost to follow-up differs according to the exposure and outcome under investigation. Capitalize Analysis of the association between exposure and outcome using only the selective data from the individuals who remain in the study will lead to biased study findings, because the remaining sample of study subjects misrepresents the original target population of interest. This type of selection bias is also called attrition bias.

Case–control studies can, in particular, be susceptible to selection bias because of the difficulties related to the selection of appropriate controls as a valid comparison group for the cases. The importance of selecting an adequate control group in case–control studies has already been discussed in this chapter. Controls should be selected in such a way that they represent the source population from which the cases derive. Ideally, if individuals from the control group had got the outcome (disease) under study, they would have been included as a case. Identification and selection of such an ideal group of controls is often a very difficult task in practice and, therefore, prone to errors that could lead to selection bias. A well-known type of selection bias related to the adequacy of the control group in a case–control study is Berkson's bias, named for the person who first described it. It can occur when the exposure affects the selection of cases but not of controls. If being diagnosed with the disease of interest is (partially) dependent on having the exposure under study, for example, because being exposed causes specific symptoms that lead to (earlier) diagnosis of the disease, then the exposure prevalence among selected cases will be an exaggeration of that in the source population. It follows naturally that the observed association between exposure and disease will be biased in this situation.

The various types of selection bias can lead to incorrect causal inferences and generalizations regarding the findings of an epidemiologic study as a result of systematic errors related to the procedures for selecting individuals into a study (eg, nonresponse bias, Berkson's bias) or related to differential selection of individuals out of a study (eg, attrition bias).

The second major category of bias is *information bias*. Information bias can occur in any type of epidemiologic study, resulting from errors during the collection (measurement) of information about the exposure and outcome of interest. Possible sources of measurement error leading to information bias may be the measurement instrument(s) or method(s) used, the subject(s) being measured, and/or the investigator(s) performing the measurement(s). Information bias is likely to occur when data collection methods are applied which leave room for error; for instance, interviews or self-administered questionnaires to assess subjective measures (eg, pain, fatigue, or quality of life). Additional sources of error that

can result in information bias are not clearly defining and operationalizing the exposure and outcome of interest, not making use of standardized measurement protocols, and inadequate training or blinding the assessors who perform the measurements.

In the special case of a dichotomous exposure (eg, exposed vs. not exposed) and a dichotomous outcome (eg, disease vs. no disease), measurement errors can result in wrongfully classifying the exposure and/or outcome status of a proportion of individuals from the study population. This type of information bias is also called misclassification bias, which arises because misclassified individuals end up in the wrong cells of two-by-two contingency tables. Two forms of misclassification can be distinguished: nondifferential misclassification and differential misclassification. In Box 3.3, an example is given of these forms of misclassification and how the resulting bias influences the findings of a hypothetical case–control study. Nondifferential misclassification results from random measurement errors. This means that any errors in collecting information on exposure and/or outcome are of the same magnitude for every individual participating in the study. The misclassification occurs only in the margins of a two-by-two table. In general, this leads to a dilution of effect or underestimation of the association of interest, that is, bias toward the null. Misclassification of this kind is like shuffling a prearranged deck of cards, with a random crisscross of spades, diamonds, clubs, and hearts as end result. In an epidemiologic study, the individuals in the four cells of a two-by-two table are shuffled instead of playing cards. Moreover, in contrast to a deck of cards that can easily be rearranged in the right order again based on the four suits, the main problem with misclassification in epidemiologic studies is that correcting for misclassification afterward is almost never possible, unless one knows the exact magnitude and direction of the measurement errors. As opposed to nondifferential misclassification, differential misclassification is due to nonrandom (systematic) measurement error. This means that the magnitude of errors in collecting information on exposure and/or outcome is not of the same magnitude and direction for every individual participating in the study. In terms of a two-by-two table, this means that the

BOX 3.3

INFORMATION BIAS IN EPIDEMIOLOGIC STUDIES

Two examples are shown (Fig. 3.15A and B) of misclassification in a hypothetical case–control study on the association between the use of cell phones and getting involved in a car accident. Cases being hospitalized after a car accident are asked whether they had been using their cell phone at the moment of the accident. Controls, who were not involved in a car accident, are selected from the waiting room of the emergency department in the same hospital and asked whether they use their cell phone on occasion while driving. As information on cell phone use is collected via self-report, the exposure measurement is susceptible to misclassification bias.

Continued

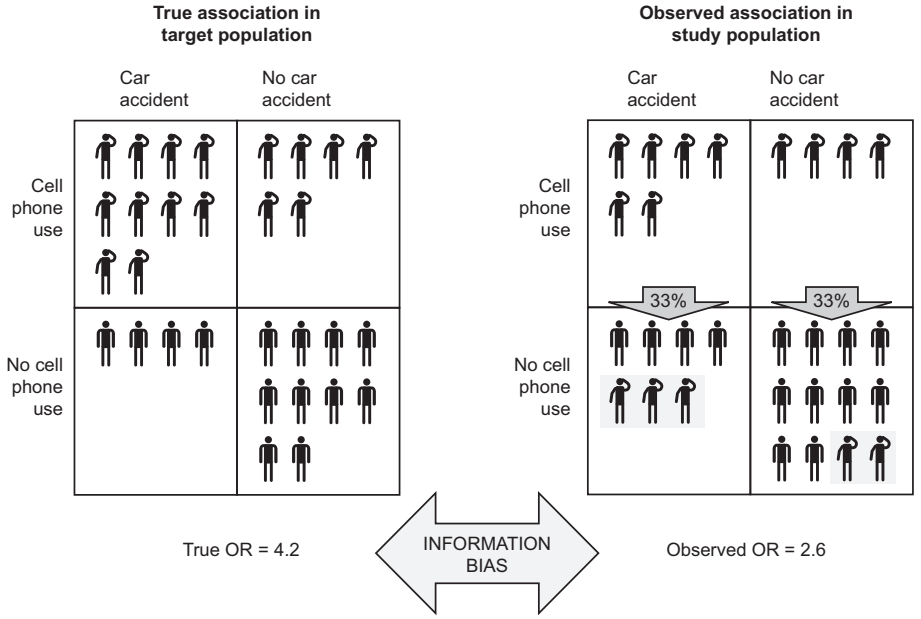


FIGURE 3.15A **Nondifferential misclassification.** Random measurement errors (ie, the same among cases and controls, and/or exposed and unexposed subjects) result in nondifferential misclassification. In this example, 33% of both the cases and the controls misreport their cell phone use. This nondifferential misclassification leads to an underestimation of the true measure of association, ie, the observed odds ratio (OR = ad/bc) is biased toward the null.

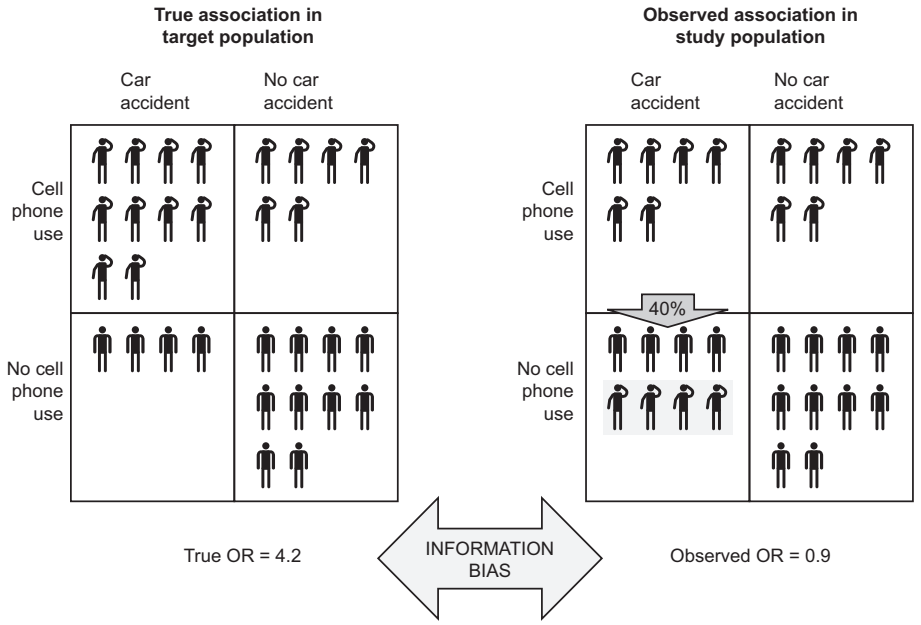


FIGURE 3.15B **Differential misclassification.** Systematic measurement errors (ie, not the same for cases and controls, and/or exposed and unexposed individuals) result in differential misclassification. In this example, cell phone use is misreported by 40% of the cases, but is not misreported by the controls. This differential misclassification gives rise to a strong bias that leads to a reversal of the true association between exposure and outcome, ie, the observed association is biased across the null (cell phone use appears to lower the likelihood of causing a car accident).

misclassification occurs to a different extent within the individual cells of the table. Depending on the nature of the systematic measurement errors, the direction of the bias that results can go either way, that is, an underestimation, overestimation, or reversal of the association under study.

Descriptions of different types of information bias in epidemiologic studies are numerous. A commonly described type of information bias is so-called observer or interviewer bias. This may arise when the outcome status of individuals participating in an epidemiologic study is assessed by a person (eg, the observer or interviewer) who has knowledge of the exposure status of these individuals. If the observer puts more effort, for instance, in determining whether the outcome of interest is present in exposed as compared to unexposed subjects, differential information bias will likely influence the study findings. This could occur in a randomized-controlled study on the effect of a new treatment on symptom relief in chronically ill patients. An investigator who knows the treatment allocation can intentionally or unintentionally assess symptom relief more precisely in subjects allocated to the treatment group than in subjects allocated to the control group. To prevent this type of bias, blinding of outcome assessors for the exposure status of study participants is important. Another common type of information bias is recall bias, which occurs when information about the exposure of interest is provided by subjects whose outcome status is known. This can happen in case–control studies when information about exposure history is gathered from cases and controls via self-report. It is not unusual that cases are more likely than controls to differently recall certain exposures from their past when they believe that these exposures may have caused their disease. Hence, such exposures may be reported with more or less precision by the cases than by the controls, with differential information bias as a result. Examples are cases with lung cancer who underreport their smoking habits compared to healthy controls, or mothers of babies with birth defects (the cases) who overreport infections during pregnancy compared to mothers of babies without such defects (the controls).

Most epidemiologic studies are sensitive to information bias resulting from either flawed data collection procedures or imperfect definition of study variables. Data collection in epidemiologic studies is virtually never completely free of measurement error, unless gold standard measurements can be used for all study variables a rare occurrence. Thus, a principal task of the investigator is to keep the errors as small as possible by using instruments and methods best suited for measuring the causal relation of interest. An even more important task is to make sure that the measurement instruments and methods are applied to all individuals in the study population in exactly the same way. Then, any measurement errors made are most likely nondifferential with a predictable direction of bias in the observed association as result, namely an underestimation of the true association (bias toward the null).

MULTIPLE CONCURRENT CAUSES

Epidemiologic research on causal relationships frequently focuses on how a certain outcome is affected by one potential cause, for example, the influence of being exposed to a particular risk factor on the incidence or prognosis of a disease. In reality, however,

outcomes are usually multicausal, meaning that they do not result from a single cause that operates in isolation, but are determined by the interplay of multiple concurrent causes operating in concert. For example, the great majority of diseases are the consequence of numerous factors, both environmental and genetic, which form complicated webs of causation. Importantly, these webs need to be disentangled before it can be validly judged whether a single exposure variable is causally associated with the outcome being investigated in an epidemiologic study.

To unravel associations between single exposures and outcomes, three different types of effects that can be part of underlying causal networks need to be accounted for in the analysis of the causal relationship of interest. These are mixed effects, intermediate effects, and dependent effects.

Mixed Effects: Confounding

Sometimes the evidence for a particular case seems to suggest that a certain factor is causing an outcome, while in reality the outcome is caused by another factor. This can mean that at first sight the findings of an epidemiologic study point to an association of a certain factor X with the outcome of interest. However, a closer look may reveal that this observed association is not causal but the result, partly or entirely, of another factor (factor Z) that is related to factor X and actually is the true cause of the outcome. This is a major form of bias, called *confounding*, which is due to the mixing of effects of two (or more) different but interrelated factors. Obscuring of observed associations due to mixed effects of several concurrent causal factors is a common phenomenon in epidemiologic research. The challenge that an epidemiologist faces is to separate the causal effect of the factor of interest on the outcome under investigation from the effect of one or more other factors, that is, the confounders. It is thus crucial to identify the factors that could potentially be confounders of the studied association. When is a factor a confounder? Generally, when studying an association between exposure to a factor X and an outcome Y (eg, an injury), a third factor Z can be a confounder if three criteria are met (see Fig. 3.16):

1. Factor Z must have an association with exposure to factor X.
2. Factor Z may be a cause of outcome Y.
3. Factor Z must *not* be part of the effect that factor X has on outcome Y.

Confounding is a very complicated issue in the practice of epidemiologic research. Confounders may mask causal effects, or lead to underestimation, overestimation, or even reversal of observed associations, including inducing spurious associations where none actually exist. Ideally, confounders should be identified before the start of a study, so that these factors can be controlled either at the stage of designing the study (eg, through randomization or restriction) or at the data analysis stage (eg, through stratified or multivariable analysis). Confounding is especially an issue in observational studies, since the investigator has no control over the different exposures. In such studies, accurate measurement of potential confounders is very important in order to enable statistical control for confounding during data analysis. In contrast to observational studies, randomized-controlled experiments are less sensitive to confounding because of the

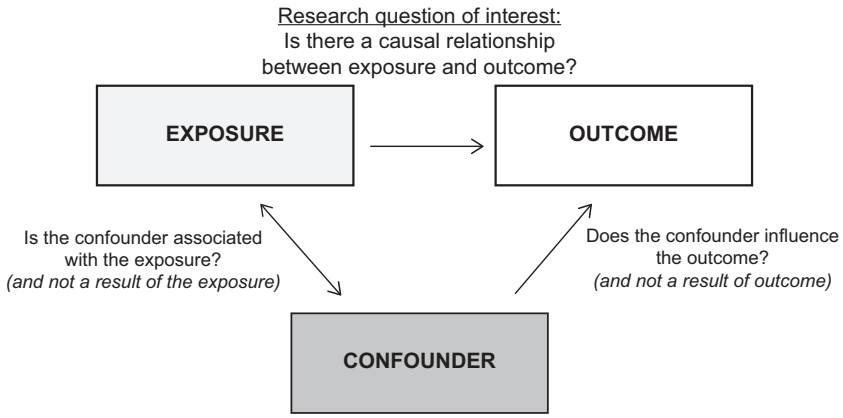


FIGURE 3.16 **Mixed effects: triangular relationship between exposure, outcome, and confounder.** When investigating whether there is a causal relationship between an exposure and outcome of interest, the influence of extraneous variables needs to be taken into account. A confounder is defined as a concurrent cause of the outcome under investigation that is related to, but not a consequence of, the exposure of interest. When an extraneous variable meets the criteria for confounding, the analysis of the causal relationship under study needs to be adjusted for this confounding variable to prevent it from biasing the study findings.

process of randomization. Successful randomization increases the likelihood that potential confounders are equally distributed over treatment groups, thereby eliminating the association of these factors with the treatment under study and thus their potential for biasing the causal relation under investigation because the first criterion for confounding does not hold anymore. A major strength of (successful) randomization is that it can eliminate the biasing influence of both known and unknown confounders, and confounders that are difficult or impossible to measure in practice. This is the reason why randomized-controlled studies are sometimes regarded as the gold standard of epidemiologic study designs.

A few examples of mixed effects:

- Dietary energy intake is a confounder for the relation between exercise and diabetes mellitus risk. Exercise and dietary energy intake affect diabetes risk in opposite directions, and exercise and energy intake are positively related. Consequently, the association between exercise and diabetes risk is confounded by dietary energy intake, meaning that the effect of exercise on diabetes risk will likely be underestimated when not accounting for the effect that energy intake has on diabetes risk.
- Cigarette smoking is a confounder when investigating a drug side effect on heart failure if the side effects of the drug are associated with smoking.

Intermediate Effects

The third criterion for confounding mentioned above refers to a special kind of effect: *intermediate effects* (Fig. 3.17). Intermediate effects are the result of variables that lie in the

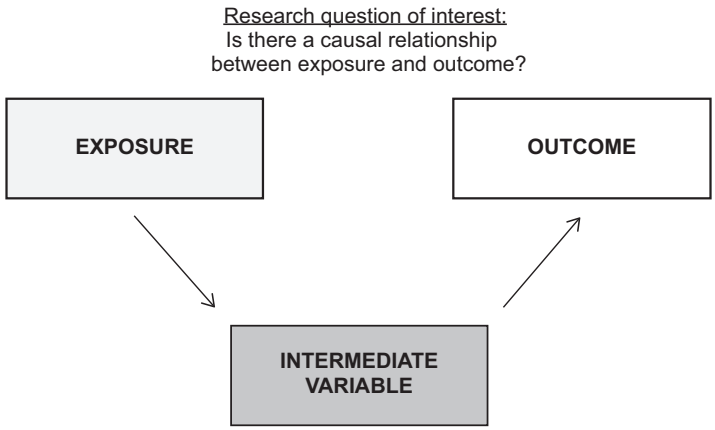


FIGURE 3.17 **Intermediate effects.** Intermediate variables are part of the effect that the exposure has on the outcome. Intermediate variables lie in the causal pathway of interest, meaning that the exposure affects the outcome through affecting the intermediate variable, ie, the exposure has an indirect effect on the outcome.

causal pathway from exposure to outcome, as they are part of the mechanism of action by which the exposure affects the outcome. This means that the exposure affects the outcome of interest, partly or entirely, through an intermediate variable. The exposure thus has an indirect effect on the outcome. Importantly, the question of whether or not an investigator should adjust for intermediate effects depends on the specific purpose of the study. If one is interested in the "pure" direct effect of an exposure on an outcome, then one should eliminate the influence of (known) intermediate variables by statistically adjusting for these variables. On the other hand, if one is interested in the total (direct and indirect) effect of the exposure on the outcome, then intermediate variables should not be adjusted for as they are part of the effect of interest and adjusting for them would bias this effect.

A few examples of intermediate effects:

- Blood pressure is an intermediate variable for the relationship between exercise and cardiovascular disease risk. Exercise affects blood pressure, which in turn affects the risk of cardiovascular disease. Adjusting for blood pressure would hide the effect of exercise on cardiovascular disease risk from view, either completely when exercise affects disease risk only through affecting blood pressure or partly when exercise affects disease risk also through other mechanisms.
- In a medical negligence case, delayed medical treatment is an intermediate variable for the relationship between the putative negligent act, eg, failure to timely diagnose, and the clinical consequence of the delayed treatment.

Dependent Effects: Effect Modification

A final type of special effect that can be revealed by epidemiologic investigations is *effect modification* (Fig. 3.18). Effect modification is also often called interaction in the epidemiologic literature. Although definitions of interaction and effect modification may differ slightly, the underlying principle is the same as they both refer to dependent effects of two or more factors

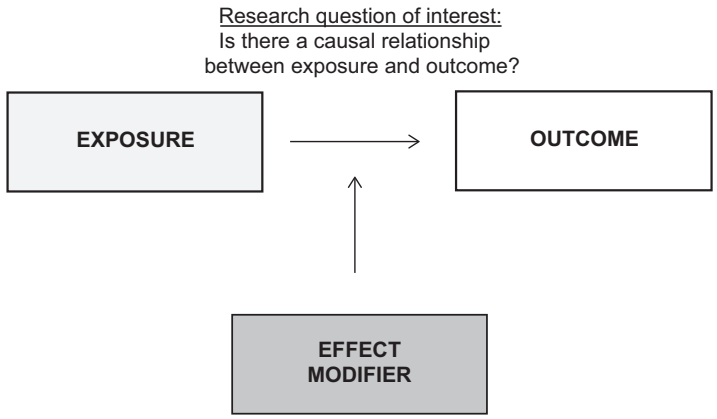


FIGURE 3.18 **Dependent effects: effect modification.** Effect modification is present when the effect of an exposure on an outcome is dependent on a third variable, ie, the effect modifier.

on an outcome. Thus, effect modification is present when the effect that exposure to a certain factor has on the outcome of interest is, partly or entirely, dependent on a third factor, that is, the effect modifier. In this case, the effect of the factor under study cannot be regarded in isolation, that is, without also considering the other factor to which its effect is closely linked.

In contrast to confounders, effect modifiers are not nuisance variables. They should be identified in epidemiologic studies and not controlled. In fact, adjusting for an effect modifier biases the observed association, since the presence of effect modification in the causal relation under investigation is masked. For instance, the effect of a certain drug may be different for men and women, because of genetic or hormonal gender differences. This means that gender is an effect modifier; it modifies the effectiveness of the drug. Actually, there are two different effects of the drug; one for men and one for women. This is why effect modification is sometimes also called heterogeneity of effects. When analyzing the effect of the drug, one should therefore not adjust but *stratify* for gender, as the effect of the drug is not independent of gender.

Examples of effect modifiers are:

- Seat belt use is an effect modifier of the relation between impact speed and severity of injury of the driver after a car crash. The severity of injury after a car crash is higher at greater impact speeds, but injury severity can also be related to whether or not the driver wore a seat belt. The relation between driving speed and injury should thus be studied separately in those who did use seat belts and those who did not.
- For asbestos-related occupational diseases such as asbestosis and mesothelioma, smoking can be an effect modifier. The effect of asbestos exposure is stronger when paired with smoking exposure.

THE HILL VIEWPOINTS

Plausibility of causation can further be assessed via application of the Hill criteria, named for a 1965 publication by Sir Austin Bradford-Hill, in which he described nine "viewpoints"

or "considerations" by which an association described in an epidemiologic study could be assessed for causality. Hill declined to call his viewpoints "criteria" lest they be considered a checklist for assessing causation. The term "Hill criteria" is used widely in the literature, however, and for convenience is used in the present discussion. Of the nine criteria, there are seven that have utility for assessing the plausibility of an investigated specific causal relationship, as follows:

- **Coherence**—A causal conclusion should not contradict present substantive knowledge—it should "make sense" given current knowledge
- **Analogy**—The results of a previously described causal relationship may be translatable to the circumstances of a current investigation
- **Consistency**—The repeated observation of the investigated relationship in different circumstances or across a number of studies lends strength to a causal inference
- **Specificity**—The degree to which the exposure is associated with a particular outcome
- **Biological plausibility**—The extent to which the observed association can be explained by known scientific principles
- **Experiment**—In some cases there may be evidence from randomized experiments (ie, drug trials)
- **Dose response**—The probability, frequency, or severity of the outcome increases with increased amount of exposure

Subsequent authors have added the feature of **Cessation/Dechallenge—Rechallenge** for circumstances when the exposure is repeated over time and there is the ability to observe the associated outcome response, as might occur with an adverse reaction to a medication. Additional considerations when assessing an association are the potential impact of confounding and bias in the data (discussed earlier in this chapter), which can obscure a true relationship, as described earlier in this chapter. Recall that confounding refers to a situation in which an association between an exposure and outcome is all or partly the result of a factor that affects the outcome but is unaffected by the exposure. Bias refers to a form of error that may threaten the validity of a study by producing results that are systematically different from the true results. Two main categories of bias in epidemiologic studies are selection bias, which occurs when study subjects are selected as a result of another unmeasured variable that is associated with both the exposure and outcome of interest; and information bias, which is systematic error in the assessment of a variable.

While useful when assessing a previously unexplored association, there is no combination or minimal number of these criteria that must be fulfilled in order to conclude that a plausible relationship exists between a known exposure and an observed outcome. In many cases there is no need for this first step of the assessment if a general causal relationship is well established. In large part, plausibility of a relationship is entertained once implausibility has been rejected.

The two remaining Hill criteria are **temporality** and **strength of association**. While both criteria have utility in assessing specific causation, temporality is the feature of an association that must be present, at least with regard to sequence (ie, the exposure must precede the outcome), in order to consider a relationship causal. Temporal proximity can also be useful in some specific causation evaluations, as the closer the investigated exposure and the outcome are in time the less opportunity there is for an intervening cause to act.

Another feature of temporality that may have a role in a specific causation evaluation is latency. An outcome may occur too soon or too long after an exposure to be considered causally related. As an example, some foodborne illnesses must incubate for hours or days after ingestion, and thus an illness that begins directly following a meal, and which is later found to be caused by a food borne microorganism that requires >12 h incubation, was not caused by the investigated meal, even if an investigation reveals the microorganism in the ingested food.

Strength of association is the criterion that is used in general causation to assess the impact of the exposure on the population, and is often quantified in terms of RR. In a specific causation evaluation the strength of the association between the exposure and the outcome is quantified by the CRR, as described earlier in this chapter.

TEST ACCURACY

Test accuracy investigation is a standard practice in clinical epidemiology. In this setting, a diagnostic test is scrutinized to determine by various measures how often a test result is correct. In forensic epidemiology, the same principles are used to evaluate the accuracy of proposed tests leading to conclusions that are central to fact finder determinations of guilt or innocence (this is discussed further in Chapter 15, *Criminal Investigation*).

The utility of a test is highly dependent on its accuracy, which is determined by a measure of how often a positive or negative test result truly represents the actual status that is being tested.

For any test or criterion there are typically four possible results: (1) a true positive (TP), in which the test correctly identifies tested subjects with the condition of interest; (2) a true negative (TN), in which the test correctly identifies test subjects who do not have the condition of interest; (3) a false positive (FP), in which the test is positive even though condition is not present, and; (4) a false negative (FN) in which the test is negative even though the condition is present. Fig. 3.19 is a contingency table illustrating the relationships between test results and condition presence, as well as the following test accuracy parameters:

Sensitivity (the rate at which the test is positive when the condition is present) $TP/(TP + FN)$
Specificity (the rate at which the test is negative when the condition is absent) $TN/(TN + FP)$
Positive predictive value (the rate at which the condition is present when the test is positive) $TP/(TP + FP)$
Negative predictive value (the rate at which the condition is absent when the test is negative) $TN/(TN + FN)$

BAYESIAN REASONING

Probability is used to characterize the degree of belief in the truth of an assertion. The basis for such a belief can be a physical system that produces outcomes at a rate that is uniform over time, such as a gaming device like a roulette wheel or a die. With such a system, the

		Condition			
		Yes	**No**		
Test	**Positive**	True Positive (TP)	False Positive (FP)	All Positive Tests (TP+FP)	**Positive Predictive Value TP/(TP+FP)**
	Negative	False Positive (FN)	True Negative (TN)	All Negative Tests (FN+TN)	**Negative Predictive Value TN/(FN+TN)**
		All with condition (TP+FN)	All without condition (FP+TN)		
		Sensitivity TP/(TP+FN)	**Specificity TN/(FP+TN)**		

FIGURE 3.19 **A contingency table, also called a "crosstabulation," of possible test outcomes, and the associated equations for evaluating test accuracy.**

observer does not influence the outcome; a fair six-sided die that is rolled enough times will land on any one of its sides 1/6th of the time. An assertion of a probability based in a physical system is easily tested with sufficient randomized experimentation. Conversely, the basis for a high degree of belief in an asserted claim may be a personally held perspective that cannot be tested. This does not mean that the assertion is any less true than one that can be tested. As an example, one might truthfully assert that "if I eat a banana there is a high probability that it will make me nauseous" based upon experience unknown to anyone but one's self. It is difficult to test such assertions, which are evaluated through collateral evidence of plausibility and analogy.

In medicolegal settings, assertions of belief are often characterized as probabilities, that is, what is most *likely*, for a given set of facts. For circumstances in which a variety of conditions exist that may modify or "condition" the probability of a particular outcome or scenario, a method of quantifying the relationship between the modifying conditions and the probability of the outcome employs Bayesian reasoning, named for Bayes' Theorem or Law upon which the approach is based. Most simply stated, Bayes' Law allows for a more precise quantification of the uncertainty in a given probability. As applied in a forensic setting, Bayes' Law tells us what we want to know given what we do know. Bayes' Law is based on an essay by

Reverend Thomas Bayes (1702–61) on the statistical analysis of probability, presented as a series of 10 propositions. Over the past 250 years, subsequent authors have further defined and refined Bayes' propositions. Although Bayes' Law is known in forensic sciences primarily for its application to DNA evidence, a number of authors have described the use of Bayesian reasoning for other applications in forensic medicine, including identification and age estimation. Bayesian reasoning may be appropriate for assessing the importance or weight of certain types of evidence resulting from an FM investigation. See Chapter 12, *Traffic Injury Investigation: Product Defects*, and Chapter 15, *Criminal Investigation* for examples of applications of Bayesian reasoning to forensic investigation of causality.

Conditional Probabilities and Bayes' Law

The purpose of any forensic test is to "condition" the probability of a particular outcome or result. For example, if the issue of interest is the probability of a particular injury A following a traffic crash, depicted symbolically as $P(A)$, then a conditioned probability would be the probability of injury given the presence of another factor; a positive test B for example. This conditional probability is depicted symbolically as $P(A|B)$; the probability of injury A given the positive test result B. An error that may occur when evaluating a conditional probability is that the assumption is made that the terms are reversible; that $P(A|B) = P(B|A)$. This error is called a conditional probability fallacy. As applied to a diagnostic test, the conditional probability fallacy occurs when it is erroneously concluded that the probability that a test will be positive when a condition is present is the same as the probability that a positive test means the condition is present (symbolically represented as P(test positive|condition) = P(condition|test positive)). As an absurd example, one could devise a test for guilt that was based on body temperature of a suspect; if the body temperature was above +10°C the test would be positive for guilt. The test would, of course, be positive 100% of the time that the individual was guilty and alive (ie, 100% true positive rate), but quite obviously he/she would not be guilty 100% of the time the test was positive (ie, 100% false positive rate). This type of fallacy involves the erroneous conclusion that the true positive rate is the complement of the false positive rate.

A conditional probability fallacy is avoided through the application of Bayes' Law, the principles of which are critical to the evaluation of the potential error rate of a forensic test. Most simply stated, Bayes' Law allows for a more precise quantification of the uncertainty in a given probability. Bayes' Law, as applied to a particular test, can be stated symbolically as:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P\left(A|\overline{B}\right)P\left(\overline{B}\right)}$$

In which $\overline{B}$ (literally, "not B") is the complement of the pretest probability or prevalence of the investigated condition of interest. The term $P(A|B)$ refers to the sensitivity or true positive rate of condition of interest (the probability of a positive test given the presence of condition B) and $P\left(A|\overline{B}\right)$ is the complement of the specificity (1 − specificity) of the test, or the false positive rate. This last term can be narratively described as the probability of a positive test when B is not present.

Posttest Probability and Positive Predictive Value

The posttest probability is a highly useful Bayesian equation that allows for the calculation of the probability that a condition is present when the test is positive, conditioned by the pretest prevalence of the condition of interest. This equation is given as follows:

$$\text{Posttest probability} = \left[\frac{(\text{pretest probability} \times \text{sensitivity})}{(\text{pretest probability} \times \text{sensitivity}) + [(1 - \text{pretest probability}) \times (1 - \text{specificity})]}\right]$$

The equation results in a positive predictive value for a given preevent or pretest prevalence. In a circumstance in which the pretest prevalence is considered "indifferent" (as with the first case study in Chapter 15, *Criminal Investigation*, where the pretest assumption of occupant position (driver vs. passenger) was 0.5) the prevalence and (1 − prevalence) values cancel out, and the calculation is a simplified to a positive predictive value.

CHAPTER

4

Causation in Epidemiology and Law

A. Broadbent

University of Johannesburg, Johannesburg, South Africa

OUTLINE

"Causation" is differently understood in legal and epidemiological contexts. The key difference is that epidemiology deals with the investigation of *general* causation questions, such as whether smoking causes lung cancer, whereas litigation invariably seeks to settle a *particular* historical causal claim, such as whether Jones' smoking caused her lung cancer. Thus in considering how epidemiological evidence may be used in answering individual causation

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00004-5

questions raised in a legal context (ie, the function and purpose of forensic epidemiology), the first goal of this chapter is to set out the exact epistemic significance of epidemiological evidence for singular causal claims. Briefly stated, the significance is as follows: epidemiological evidence is potentially relevant to particular causal claims in that, given certain assumptions, particular causal claims can be assigned a probability given the evidence.

The next goal of this chapter is to examine the relevance of epidemiologic proof of individual cause. If the legal standard of proof is identified with a given probability, epidemiological evidence could, in principle, be relevant for establishing the cause-in-fact element of liability. This relevance is asymmetric, as epidemiological evidence is more suitable for proving (for plaintiff) rather than refuting (for defense) singular claims of causality at any given level of proof. This is due in part to the intrinsic nature of such evidence as a measure of net effects.

The final goal of the chapter is to analyze and consider the various objections that are commonly raised to epidemiologic evidence in assessing specific causation. These objections include resistance to the idea that general evidence by itself has any bearing on particular claims; resistance to the idea that the legal standard of proof can ever be captured numerically; confusion between uses of epidemiological evidence to prove causation and to refute it; confusion between evidence of probability and probability of evidence; and theoretical paradoxes arising from some uses of "naked statistics." While each of these objections should be taken into account when contemplating the use of epidemiological evidence, none of them amount to an obstacle to the use of epidemiological evidence to prove specific causation, provided proper caution is exercised when obtaining the evidence.

BACKGROUND

Epidemiology is a pioneer science: it seeks to discover causal relationships between exposures and health outcomes which were not previously suspected. This is not all it does, of course, but it is one aspect of epidemiology, and historically a crucial aspect for the growth of the discipline. Because of its pioneering nature, it sometimes arises that epidemiological evidence for a causal link between some exposure and some health outcome is the only such evidence, or at least a very important part of what evidence there is. The health outcomes that epidemiology studies are typically harm, and the exposures are sometimes caused or contributed to by human agents. This is one reason why epidemiological evidence might come before a court. A plaintiff has suffered a harm, and there is evidence that the harm was caused by a legal person; and that evidence is epidemiological.

Even where a causal connection is well known to medical science, epidemiology might remain an important source of causal evidence due to limitations on our scientific knowledge of *how* the causal link actually works. Thus even though it is well known that smoking causes lung cancer, we do not know enough about how this happens to be able to tell whether a given smoker's lung cancer was caused by smoking or would have occurred anyway. No medical test can tell us this. Yet the epidemiological evidence is very appealing to a plaintiff seeking compensation for harm caused by smoking, at least in relation to lung cancer. That disease was so rare before smoking became widespread that it did not even have its own medical classification. Nearly all the lung cancer that occurs is due to smoking. Thus it is almost, but not quite, certain that any given plaintiff's lung cancer is caused by smoking, absent any special circumstances of that plaintiff rendering her unusually prone to lung

cancer. This is another reason that courts might be presented with epidemiological evidence: not because it is the only evidence to establish a causal nexus, but because it appears to be the strongest.

It remains the case that there is no stable and fully general judicial position, in any jurisdiction, on the use of epidemiological evidence to prove specific causation. No doubt, there are historical reasons for this, but there are also underlying conceptual challenges for the use of epidemiological evidence to prove causation. Since courts are reacting to cases and not to conceptual questions considered in the abstract, they cannot be expected to consider the conceptual questions fully. A proper understanding of any proposed use of epidemiological evidence requires a basic understanding of the nature of epidemiological evidence, alongside an appreciation of the philosophical issues arising from applying population-level evidence to particular cases. Not all judges understand these things in the same way, if at all. It is therefore unsurprising that judges have not settled on a single position with a singly articulated rationale.

To the extent that there is a unified judicial position on the use of epidemiological evidence to prove specific causation, it is skeptical. The general view appears to be that epidemiological evidence is population-level evidence which may form useful background information but cannot be applied to a particular case, unless normal rules for proving causation are relaxed. This view is uninformed and demonstrably false. In truth, population-level evidence *must* be applied to individual cases in interpreting other evidence, on pain of committing well-documented fallacies (especially the base rate fallacy). Academic commentators have sometimes expressed similar views.

Against this background, it is hard to avoid taking a position on the acceptability of epidemiological evidence of causation in this chapter. There are some probative points of fact and logic which cannot be disputed as matters of opinion. There are other points which are more philosophical in nature. In this chapter, I endeavor to distinguish between these two types of points when discussing them so that the reader understands that some of the arguments presented herein contain opinions rather than generally agreed upon factual or logical assertions. When I am expressing opinions relating to philosophical arguments, it is my goal to present them in a transparent and even-handed manner.

The structure of this chapter is as follows. In the section on "Delimiting the Topic," the topic is delimited. This section will focus on the significance of epidemiological evidence for proving causation *if the evidence is accepted* as proving a general or population-level causal claim. The question addressed in this section is how population-level evidence relates to individual plaintiffs.

In the section on "What Is Causation?" there is a general introduction to the philosophy of causation, by way of essential background. This background established, the section on "What Epidemiological Evidence Says About Particular Causation" which seeks to establish the *epistemic* significance of epidemiological evidence for specific causation, prior to any legal considerations. Since the element of liability to which epidemiological evidence applies is a cause-in-fact, it makes sense to first understand what epidemiological evidence tells us, what information it supplies, about individual causal facts, before considering how this information measures up against legal standards of proof. This section seeks to offer as clear and logical a route through this problem as possible, setting aside objections to be dealt with later on. The section on "How Epidemiological Evidence Relates to Legal Standards of Proof?"

places epidemiological evidence in the legal context, asking how the epistemic significance of epidemiological evidence measures up against legal standards of proof. The section on "Sources of Resistance to Using Epidemiological Evidence" considers various objections to the use of epidemiological evidence in legal contexts, seeking to distinguish between those objections that are simply wrong, and those that amount to legitimate philosophical positions with which this author happens not to agree, but which merit serious consideration. The section on "Conclusion" demonstrates that while there remain unresolved conceptual questions about the use of epidemiological evidence of causation in litigation, there are overwhelming considerations showing that failure to allow or properly employ epidemiological evidence will lead to repeated errors in fact-finding exercise. Further, that epidemiological evidence can and must be employed where available, provided that it is used properly, provided that the proper significance of the evidence and the nature of its limitations are understood.

DELIMITING THE TOPIC

Two limitations on this topic of this chapter should be noted at the outset. First, the sources of law considered in this chapter are in the common law tradition, with cases drawn from England, Scotland, and the United States. The epistemic significance of epidemiological evidence in proof of specific causation will not vary between jurisdictions; the nature of scientific knowledge does not depend on the law of the land. But mapping the epistemic significance of epidemiological evidence onto the kinds of liability and the standards of proof found in other jurisdictions may present jurisdiction-specific challenges.

Second, this chapter concerns only the question of what epidemiological evidence can prove *if it is accepted*, not the question of whether it should be accepted in the first place. In the context of litigation, there are two ways that evidence can be challenged, broadly speaking: one may challenge its *veracity*, and one may challenge its *power* to prove what it is adduced to prove. For instance, in the Scottish case *McTear v. Imperial Tobacco*, Mr. McTear was a lifelong smoker who died of lung cancer, and his widow Mrs. McTear brought a claim against Imperial Tobacco in respect of his death. The judge refused to find, on the epidemiological evidence before him, that smoking causes lung cancer. This is an example of a challenge to the veracity of epidemiological evidence. The judge was not accusing the epidemiologists who testified of lying. Nonetheless, he was saying that what they asserted was false, on the balance of probabilities: that is, that contrary to what the epidemiologists claimed, the evidence failed to show that smoking causes lung cancer. The judge went on to consider the other kind of challenge, concerning power. He said that even if he *had* found that smoking caused lung cancer (a fact accepted by the Scottish Parliament at the time, which was currently legislating on that basis), this would not prove that the pursuer Mr. McTear's lung cancer had been caused by smoking. A number, albeit a miniscule number, of nonsmokers also get lung cancer, and the epidemiological evidence, even if adequate to prove that smoking causes lung cancer in general, would have been insufficient to tip the balance of probabilities in favor of Mrs. McTear's claim that smoking caused Mr. McTear's cancer. At least, so concluded the judge.

In this section, we will confine ourselves to the question of whether epidemiological evidence has the *power* to prove particular causation—whether, in a case like *McTear*, the

judge in that case was correct on the second point. There are good reasons to confine the topic in this way, despite the fact that both kinds of evidentiary challenge may present. Questions about the veracity of epidemiological evidence are not specific to the use of this evidence to prove causation; they are dealt with elsewhere in this text (primarily in the preceding Chapter 3 on *Methods Used in Forensic Epidemiologic Analysis*); and the peculiar difficulties of using epidemiological evidence to prove the causation element of liability are sufficiently complex to call for detailed treatment on their own.

WHAT IS CAUSATION?

Nobody can say exactly what causation is. This is a philosophical problem, but it becomes a practical one in situations where we feel some practical, and not merely philosophical, doubt as to whether a given sequence of events is causal. Because we cannot say what causation is, we have no universally agreed fail-safe test for causation—not even an in-principle, God's eye test. Thus the purpose of this section is not to arrive at a definitive answer to the general question "What is causation?" but rather to set out some of the main practical implications that this philosophical question can have, and also to identify some key ambiguities in the term, which permit it to be used in different ways in legal and epidemiological contexts.

Even though there is no universally accepted, fully general theory of the nature of causation, there are a number of points that are widely agreed among philosophers who have studied the subject. First, the word "cause" does not need to be present for a claim to assert the presence of causation. There are many "causal verbs." For instance, if I say "Drinking gin has made Samantha drunk," I am clearly asserting a causal relation between Samantha's recent intake and her current behavioral state, even though the word "cause" is not present. This is an obvious point but it is useful to emphasis nonetheless, since the word "cause" and cognates are abstract, and thus can promote an air of mystery. This leads to the second widely agreed point, which is that causation is something we encounter very commonly in everyday experience and whose presence or absence we often recognize unproblematically, in everyday experience.

The third obvious point about causation is that it seems to involve some kind of necessity, although whether it truly does, and if so what form that necessity takes, are the topic of widely differing opinions. Thus if I say "Drinking tonic water has made Samantha drunk," I am presumably saying something false, even if it is true that Samantha got drunk after drinking tonic water, along with gin. She got drunk after drinking both, but it was the gin, not the tonic water, that caused drunkenness. There is some sort of *necessity* involved, in the sense that the gin *made* Samantha drunk, given the circumstances. But it seems not to be logical necessity: there is no contradiction in supposing that Samantha downs a bottle of strong liquor and remains sober. The strong feeling that we have that this is impossible arises from our experiences, more or less similar to Samantha's; it does not arise from any sort of logical deductive operation of the kind that drives a mathematical proof, for instance. But what other kind of necessity is there?

The nature of the necessity that seems integral to causation is one of the most troubling mysteries in philosophy. There is no need to go into the many philosophical theories of

the nature of causation, provided it is understood that none of them is universally accepted and that each suffers from objections that appear fatal to all but its advocates.

The practical significance of this mystery is that it makes it unclear *how we find out* about causation. We do not identify it through a process of logical deduction, but rather through reasoning from experience; but then we do not seem to have direct experiences of causation either. Consider the gin case above. The reason we deny that tonic water made Samantha drunk is not because we literally saw, heard, smelled, etc. the gin doing its work. It is rather because we have other experiences of gin, absent tonic, where drunkenness ensues, and other experiences of tonic, absent gin, where all that ensues is sobriety accompanied by a funny taste in the mouth. This leads us to conclude that drunkenness is not a necessary consequence of imbibing tonic, while that it is of gin, in the right circumstances. But at no point do we make direct observations of the causal "glue"; it is not as if we could design a special microscope which can see the causal processes at work.

This point is of crucial importance in considering our topic, namely, the use of epidemiological evidence to prove specific causation, because it means that even the best medical test cannot "see" causation. Knowledge of causation is only ever justified by an inductive inference; even where this inference is extremely strong (or where it is automatic), causation is never directly exhibited.

Moreover, while we can have excellent evidence for causation, that evidence always seems to fall short of deductive proof. Consider the gin case again. Our previous experiences of drunkenness following gin might have been due to some other mixer, or, if the gin was neat, due to some other ingredient in the gin; or, if the gin was chemically identical, due to differences in the constitution of the gin drinkers we have known and Samantha; or, if the previous experiences involved Samantha, in the chemical composition of the air; or, if all those were the same, some curious radiation from a sunspot; and so on. Similarly, consider the previous experiences of tonic; perhaps tonic does cause drunkenness but the previous experiences concerned someone with a stronger constitution than Samantha, or were negated by a sunspot, etc. The further we push the example, the more outlandish the possibilities become; but outlandishness is immaterial in the context of strict deductive proof. The point is that they cannot all be eliminated, which is what a deductively valid proof must achieve. To complicate matters further, there is more than one way that causation may relate two classes of events so that they are always found together in certain circumstances. The trusty barometer falls when a storm is brewing, but neither event causes the other; both are effects of a common cause.

For this dual reason—the lack of direct experience of the causal nexus and the lack of deductive proof of its presence or absence—the epistemology of causation is as troubling in practical contexts as the metaphysics of causation is for philosophers.

The fourth and final theoretical point about causation that is relevant in practical contexts is that causal claims can be either *general* or *singular* (also referred to as *specific* causation elsewhere in this text). Singular causal claims concern particular events that are causally related to each other. Typically, this is the important kind of causal claim in litigation, where the presence or absence of causation in a certain sequence of events is a historical fact that needs to be established for liability to be decided. The Samantha case is a case of singular causation. A general causal claim concerns a whole class of events, for instance, "drinking gin causes drunkenness." This is the sort of causal claim with which epidemiology is primarily concerned; "smoking causes lung cancer," for example.

The analysis of general causal claims is the subject of less philosophical attention than the analysis of singular claims, partly because it is not clear whether the general claim refers to anything over and above a lot of singular causal facts, bundled together. Again, this is a matter of controversy. For present purposes, the central point is to appreciate that, whether or not there are two kinds of causation (one general and one particular), there are clearly two kinds of causal *claim*. The crucial question, as we shall see, that confronts attempts to bring epidemiological evidence to bear in litigation concerns the relation between these two sorts of claims. If general causal claims admitted of no exceptions then there would be no problem. Thus, for instance, "smoking causes cancer" meant that, in every case of smoking, smoking causes cancer, then there would be no problem inferring that smoking caused Jones's cancer from the fact that he smokes. Nor would there be a problem if there was universal implication in the other direction, so that "smoking causes cancer" meant that every case of cancer was caused by smoking. However, such phrases carry neither of these implications in their ordinary use.

Epidemiologists have devised measures that are meant to accurately quantify each of these potential implications. We will cover some of these measures shortly, but they are described in detail in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* of this book. The key point to bear in mind is that it is extremely rare for any exposure of interest to give rise to a truly universal causal claim. Smokers do not always (or even often) get lung cancer, and nonsmokers occasionally do. The crucial issue, then, is how the quantities that epidemiologists use to express the strength of their general causal claims relate to the relevant legal standard of proof.

WHAT EPIDEMIOLOGICAL EVIDENCE SAYS ABOUT PARTICULAR CAUSATION

The first step in deciding whether epidemiological evidence can satisfy any given legal standard of proof in respect of particular causation is to set aside the legal standard of proof, and work out what the epidemiological evidence says about causation in the first place. Once this is clear, we can decide whether what it says, if accepted, might satisfy any given legal standard of proof.

Risk

The most common way that epidemiological evidence will be presented to a court is in the form of a comparative risk ratio, or a *relative risk*. It is of crucial importance that this measure is clearly understood. To understand relative risk clearly, we must first understand risk.

The term "risk" has many meanings, and this can be confusing. When it is used in epidemiological measures such as relative risk, it has a clear technical meaning: the number of new cases of a disease in a given time period, as a proportion of (ie, divided by) the number of persons in that population at the start of the time period. This is a mere statistic and has no connotation of danger. It is not even a measure of probability. A risk is simply a statistical fact about a population. Thus the lifetime risk of lung cancer among male nonsmokers in a given study might be 0.005, or 0.5%. In this usage, "risk" applies strictly to the population.

However, in its nontechnical usage and in the law, "risk" applies to individuals. One might infer that a given member of this population has a 0.5% risk of developing lung cancer in his lifetime. Below, we shall see that there are situations where this inference is justified; but it is important to appreciate that it *is* an inference, and that the epidemiological claim that a given risk is 0.5% is not equivalent to the claim that a lawyer might seek to infer, that a given individual's "level of risk" is 0.5%.

Relative Risk

Epidemiologists use the term "exposure" to refer generically to any potentially causal factor under investigation. They compare groups in order to draw conclusions about how the different exposures of each group affect their health outcomes. *Relative risk* is a simple way of drawing this comparison. Suppose that, having observed that the lifetime risk of lung cancer among male smokers is 0.5%, we find that the lifetime risk of lung cancer among male smokers in our study is 10%. We have found that lung cancer is 20 times as common among smokers in our study as among nonsmokers. This number, 20, is the relative risk, and it is calculated by dividing the risk among the exposed group (in our example, smokers) by the risk among the unexposed group.[1]

$$\mathrm{RR} = \frac{R_{\mathrm{E}}}{R_{\mathrm{U}}}$$

Relative risk is an attractive measure for the epidemiologist for several practical reasons. It is simple; it enables comparison of risks in populations that are of quite different sizes, and where the prevalence of the exposure (eg, number of smokers) is different; and it can be estimated (given certain conditions) from a case–control study, which is a relatively inexpensive kind of study. It is thus very common for epidemiological findings, at least those concerning causality, to be expressed in terms of relative risks.[2]

From Population Risks to Individual Probabilities

As defined, a risk is a property of a population; and a relative risk is a property of two populations (or a relation between them). But it is natural and tempting to seek to infer something about the individual members of those populations. What can legitimately be inferred?

If we select a member of a given population at random, then there is an equal chance that any member of that population could have been selected. If this condition is satisfied, then it is relatively unproblematic to simply translate risks into individual probabilities and relative risks into factors that multiply probabilities. So if a smoker is randomly selected from our study population with a lifetime risk of 10%, then the probability that he develops lung cancer in his lifetime is 10%. This much is not seriously contestable, even if there are philosophical disagreements about what an individual probability is.

The relative risk tells us *how much larger* the exposed risk is than the unexposed. The relative risk thus tells us how much more probable the outcome is in a randomly selected exposed individual than in a randomly selected unexposed individual. In our case, because the relative risk of lung cancer among smokers compared to nonsmokers is 20, we can

legitimately conclude that, given his smoking, McTear was 20 times as likely as a nonsmoker to develop lung cancer in his lifetime.

This kind of inference depends essentially and completely on the randomness of the selection. If an individual is randomly selected, then nothing about that individual affects its chance of selection; it could as easily have been any other member of the population. It is this assumption of random selection, and this alone, that allows us to move so easily from population-level risks to individual probabilities.

In the context of litigation, this point can give rise to both spurious difficulties and pertinent questions. On the spurious side, it is obvious that no litigation begins with the random selection of an individual from an exposed population. There will be reasons that a given individual chooses to litigate, while others could but do not. And even if claimants were somehow randomly selected, any given individual randomly selected from a population of human beings will differ in countless respects from each of those who were not selected. People are not like balls in an urn. But as reasons to disallow an inference from a population-level risk to an individual probability, these are spurious. Provided that there is no reason to suspect that the either reasons for litigating or any other distinguishing features of the individual are potential causes of the disease in question, the individual may be treated as randomly selected from the point of view of assessing the exposure/outcome relation. For example, although information on Mr. McTear's height is not available, he was probably at least slightly taller or shorter than the average height of smokers in the studies from which the epidemiological evidence was drawn in that case. This makes him atypical; but since there is no evidence at all linking height to lung cancer, this atypicality does not undermine the useful fiction that Mr. McTear is randomly selected. Again, he may have had a particularly determined and litigious widow; but since there is no evidence linking the personality of a smoker's wife to his susceptibility or resilience to lung cancer, this respect in which Mr. McTear's claim was not randomly selected does not disturb the assumption that it was, for these purposes.

On the other hand, if the claimant *does* possess features that are either known or reasonably suspected to make her either more or less susceptible to the exposure's capacity to cause the outcome in question, then these features clearly do bear on the transposition of population risks to individual probabilities. The epidemiological term for this phenomenon is *interaction*[3] (or by its statistical name, *heterogeneity of effects*[4] or *statistical interaction*). "Interaction" in this usage does not imply the normal, intuitive idea of one thing literally interacting with another; it implies, rather, that the level of risk of a given outcome is different when two exposures are present together from what we would expect it to be, given the risks observed when the exposures are present singly. For example, exposure to asbestos dust interacts with exposure to tobacco smoke, meaning that very much higher risks for lung cancer are found in populations exposed to both (eg, asbestos workers who smoke) than in populations exposed either to one or the other alone. Essentially, the whole is greater than the sum of the parts.

If the claimant is exposed to potential agents that interact or that might reasonably be suspected to interact with the exposure at issue—for example, if Mr. McTear had been an asbestos worker—this can affect the transposition of population risks to individual probabilities. The effect can operate in either direction, to make the probability either greater or less. Thus if Mr. McTear is an asbestos worker, we can be confident that the probability of his developing lung cancer in his lifetime is considerably greater than the 10% that would be

obtained by simply transposing the lifetime risk among male smokers. Factors that reduce the risk of lung cancer are less well known, but if there were such factors and Mr. McTear exhibited them, the probability of his developing lung cancer would be correspondingly less. From the point of view of litigation, interaction need not present a difficulty if the scale of the interaction is reasonably well estimated (as it is for asbestos and lung cancer); but it could be problematic in cases where interaction is suspected yet reliable quantitative estimates are not available.

To summarize, it is appropriate and valid to assume that an individual probability is equal to the risk in the population from which she is drawn, to the extent that it is reasonable to consider her randomly selected from that population with respect to causative or potentially causative factors of the outcome in question. If this is a reasonable assumption, or approximation, then the use of a population risk to estimate an individual probability is also reasonable. Conversely, if the assumption is unreasonable, so is the use of population risk to estimate individual probability.

Note that, at this stage, we are not considering an "all things considered" probability; we are still only considering the probability given the epidemiological evidence.

From Individual Probabilities to Particular Causation

If it is reasonable to suppose that an individual is randomly selected from a population in the way just described, then it follows that the relative risk tells us how much *more* probable it is that a person from the exposed population will suffer the outcome than an unexposed person. For instance, if the RR for lung cancer among lifetime smokers compared to nonsmokers is 20, this means that there are 20 times as many smokers (proportionally) as nonsmokers. Under the assumptions explained in the section on "From Population Risks to Individual Probabilities", this means that a randomly selected smoker is 20 times as likely to develop lung cancer as a randomly selected nonsmoker.

We can use this fact to estimate how probable it is that the smoker in question would not have got lung cancer otherwise, provided we make two assumptions. Both can be dispensed once we have set the basic inference out. First, let us assume that the RR is causal, since to doubt this is to doubt the scientific evidence, whose veracity we are setting aside. Second, we need to assume that smoking has not *prevented* lung cancer in any case. This might sound like a strange assumption, but it allows us to conclude that the effect the RR measures, which is a net effect, is also the whole effect of smoking on lung cancer. The significance of this point is explained in the next section below.

If we make both these assumptions, then it will follow that the randomly selected smoker is 20 times as likely as a randomly selected nonsmoker to get lung cancer, and that this is because of the smoking; and thus it follows that, for a randomly selected smoker who *does* get lung cancer, the chance that she would have not got lung cancer had she not smoked is 1 in 20. That is because for every 20 smokers developing lung cancer, 1 would have done so anyway. Thus a randomly selected smoker with lung cancer has 1 chance in 20 of being that smoker who would have developed lung cancer anyway.

The first assumption can be dispensed with by interrogating the scientific evidence as to whether the RR in question provides a good measure of the causal strength of the exposure

on the outcome. We have already set this kind of question aside for the present discussion, but in the context of litigation it would obviously be tested.[5] Let us now turn our attention to dispensing with the second assumption.

The Net Effect Problem

Exposures may have different effects on different people. Some may even increase the risk of an outcome in some people while decreasing it in others. If the increase is more than the decrease, then there will be a net difference in the risk of the outcome between exposed and unexposed groups. But the net difference may be less than the number of cases caused by the exposure, since it equals the number caused, less the number prevented. This may sound like an implausible situation, but it is nonetheless a biologically possible one, and one which does occur in some cases. (For example, many drugs have "paradoxical effects" on a small number of patients—effects that are opposite to the desired and/or usual effect.)

If we drop the assumption that all cases of an outcome under exposure would also have exhibited the outcome without the exposure, then we can no longer assume that the probability of causation in a randomly selected case of outcome is *equal* to 1 divided by RR. But if it is correct to interpret RR causally, as we are assuming, then we know that the exposure causes the net difference between the two groups. It might cause more cases than that, if it also prevents some that would otherwise occur absent the exposure. But it cannot cause fewer cases, unless the outcome is partially or totally caused by something else—in which case it is wrong to interpret RR causally, as a measure of the net effect of the exposure.

This means that the epidemiological evidence can give a *lower limit* on how likely it is that a randomly selected person from the exposed group, who has the outcome (eg, a randomly selected smoker with lung cancer) would have experienced the outcome otherwise (eg, would have suffered lung cancer without smoking).[6] If we use PC for the probability of causation (an abbreviation used throughout this text), then the epidemiological evidence says this about such a person:

$$\mathrm{PC} \geq \frac{\mathrm{RR} - 1}{\mathrm{RR}}$$

Thus if $\mathrm{RR} = 20$, and our assumption holds, the chance that a smoker with lung cancer would have contracted lung cancer despite his smoking is 1/20, or 0.05, and thus the probability that lung cancer was causal in this case is $(20 - 1)/20$ or 0.95 (ie, 95%). If we drop our assumption, then the probability of causation may be greater than this, but will not be less, unless we are questioning the validity of RR as a measure of the net effect of smoking on lung cancer in these populations.

Again, none of this concerns the effect that *other* evidence may have on the probability of causation, all things considered. We have however reached a position on the probability of causation, given the epidemiological evidence, on the assumptions that this evidence is accepted as good evidence for general causation, that the measure in question quantifies the population-level causal effect accurately, and that the individual concerned is a randomly selected member of the population in question.

HOW EPIDEMIOLOGICAL EVIDENCE RELATES TO LEGAL STANDARDS OF PROOF?

Epidemiological evidence allows us to estimate a lower bound on the probability of causation in a randomly selected exposed individual with the outcome, given certain assumptions. However, there are two further points to be taken into account when seeking to employ this evidence in a legal context.

First, the epidemiological evidence may itself be subject to some doubt. We have set this aside for the sake of exposition, but clearly, it matters. It is important to be clear about the distinction between two probabilities: the probability of causation given the epidemiological evidence (eg, the probability of causation *given* that RR = 20), and the probability of the evidence itself (eg, the probability *that* RR = 20). Both factors need to be considered in arriving at a conclusion about the probability of causation.

Second, the epidemiological evidence may not be the only relevant evidence, in which case it needs to be weighed against the other evidence. This sort of weighing will potentially affect the second kind of probability just listed—the probability of the evidence itself. It may also effectively undermine the assumption that an individual is randomly selected, for example, by showing that the individual is of higher or lower risk than average.

Assuming that the epidemiological evidence remains credible after considering these two factors, its significance in relation to civil standards of proof is clear. The epidemiological evidence provides a way to satisfy the but-for test on the balance of probabilities.[7] It tells us the lower bound on the probability that an individual would have suffered the outcome in the absence of an exposure. Where RR > 2, this lower bound is over 50%, the balance of probabilities is tilted: it is more likely than not that, had the exposure been absent, the outcome would also have been absent.

A number of judges and legal scholars are highly resistant to using epidemiological evidence. Some of this resistance is simply the result of misunderstanding or poor reasoning; and some of it arises from legitimate doubts that are, in essence, philosophical. The next section explores these doubts.

One interesting consequence of the fact that this is an inequality, rather than an equation, is that epidemiological evidence can be used more easily to prove causation than disprove it. When RR < 2, the lower bound on PC will be below 50%. But this is not enough to show that the exposure is not causal on the balance of probabilities. That will depend on whether the exposure might cause the outcome in some people while preventing it in others, leading to the situation described above where the net effect is less than the total effect. However, the epidemiological evidence could in principle establish causation on the balance of probabilities, if RR > 2.

SOURCES OF RESISTANCE TO USING EPIDEMIOLOGICAL EVIDENCE

Other Evidence Matters

The simplest objection to the use of a formula such as the inequality described in this chapter is that it threatens to obscure the significance of other evidence. Courts need to weigh

all of the evidence. The probability of causation given the epidemiological evidence might differ from the probability of causation given all the evidence.

This is, of course, no real objection to using epidemiological evidence to calculate a probability of causation. It is an objection to ignoring other evidence. The lesson is not that epidemiological evidence should itself be disregarded—that would be an incoherent response to an objection whose basis is that no relevant evidence should be disregarded—but rather, that the probability of causation given the epidemiological evidence must be fed into the fuller evidential picture. If there is no other evidence relevant to proving causation, this is an easy matter; the probability of causation given the epidemiological evidence will equal the probability of causation given all the evidence. It is more complex where other evidence is in play. For example, when a smoker plaintiff also worked in an asbestos factory, or has a family history of lung cancer. The weighing of these different kinds of evidence is ultimately the unenviable task of the judge, and there is very little general guidance that can be offered, except to emphasize, as Haack does, that all evidence must be considered, and that it must be considered together, as a whole interlocking evidential picture.[8]

Probability of Causation Depends on Probability of Evidence

The first assumption set out at the start of the section on "How Epidemiological Evidence Relates to Legal Standards of Proof?" was that the epidemiological evidence is assumed to have been accepted by the court as proving general causation. One objection that has been raised to the use of epidemiological evidence is that the probability calculated for causation confuses two probabilities: the probability of causation given the evidence, and the probability of causation, in and of itself. This issue is addressed in Chapter 2, *Epidemiologic Evidence in Toxic Torts* of this book.[9]

As for the previous objection, this is not so much an objection as a reminder of the need for clarity. There is a prior question concerning the probability of the facts asserted by the epidemiological evidence itself: that smoking causes lung cancer, that the RR is 20, and so forth. If the court finds that these are facts, the next question is what these facts entail for the probability of causation. If the court does not find the epidemiological evidence credible in the first place, then of course no use of that evidence will be compelling. Even if the probability of individual causation *given* the epidemiological evidence is very high, that will not move a court which is not prepared to "give" (accept) the epidemiological evidence in the first place. This is no objection to the use of epidemiological evidence; it is merely a reminder that epidemiological evidence must first be assessed in its own right before it can usefully be applied to the specific causal inquiry.

Confusion Between Proving and Refuting Causation

Another source of resistance to using the PC formula given previously is a concern that it will generate a situation where $RR = 2$ becomes a threshold, leading to all claims where $RR < 2$ equates to no causation.[10] This would be unfortunate because, as already discussed, the epidemiological evidence estimates a lower bound on the probability of causation, but not an upper bound. Thus when $RR < 2$, that does not disprove causation; it only means that the lower bound on the probability of causation given the epidemiological evidence is less than 50%.

Once other evidence is considered alongside the epidemiologic evidence, the court might decide that the lower bound is higher than 50%. To illustrate: if RR = 1.9, but there is strong evidence of a linear dose–response relationship, and the plaintiff was subjected to especially large quantities of the exposure, then the court might well conclude that the plaintiff has a higher probability of causation than a randomly selected member of the exposed population that has been previously studied to generate the RR, and thus the probability of causation is more than 50% even though the probability yielded by the PC formula applied to study results is less than 50%.

One way to mitigate the risk that RR = 2 becomes incorrectly perceived as a threshold for proof of causation is to frame the formula as an inequality, as it has been framed in this chapter. It is indeed incorrect to think that PC $\geq \frac{RR-1}{RR}$ given the epidemiological evidence, even on our assumptions identified in the section on "How Epidemiological Evidence Relates to Legal Standards of Proof?". But this is easily fixed by replacing "=" with "≥". It is, however, an overreaction to foreswear the use of epidemiological evidence merely because of the possibility of misuse.

Skepticism Concerning the Relevance of General Evidence to Proving Particular Claims

It is sometimes said that epidemiological evidence concerns populations, and thus says nothing about individuals. According to Michael Dore, "Epidemiological evidence, like other generalized evidence, deals with categories of occurrences rather than particular individual occurrences… Such evidence may help demonstrate that a particular event occurred, but only when accompanied by more specific evidence."[11] Melissa Moore Thomson wrote that "statistic-based epidemiological study results should not be applied directly to establish the likelihood of causation in an individual plaintiff."[12] According to Andrew See, "Epidemiology studies are relevant only to the issue of general causation and cannot establish whether an exposure or factor caused disease or injury in a specific individual."[13] And in the Scottish case of *McTear*, Lord Nimmo Smith stated: "…epidemiological evidence can be used to make statements about individual causation. … Epidemiology cannot provide information on the likelihood that an exposure produced an individual's condition. The population attributable risk is a population measure only and does not imply a likelihood of disease occurrence within an individual, contingent upon that individual's exposure."[14]

Whatever other difficulties are faced in the use of epidemiological evidence in court, it is important to emphasize that the line of thought represented by the preceding assertions is a non sequitur. Specifically, it does not follow that, because epidemiology studies populations and because epidemiological measures concern populations, epidemiology can deliver no information that affects the probability of causation in an individual case. The easiest way to exhibit this non sequitur is to imagine that *everyone* in a given population has some characteristic, for example, is a smoker. This is a population-level fact. But it enables us to infer, with certainty, that a given individual member of that population also has that characteristic. This is the structure of the logical argument form known as *modus ponens*. Thus, as a general proposition, it is clearly false that information about a population never implies anything about any individual members of that population. It is therefore fallacious to maintain that, because epidemiological evidence concerns populations, it says nothing about

individuals. That conclusion might still be true, for all this refutation shows; but it cannot be established so unthoughtfully.

The situation where a trait is present in 100% of the population is not, of course, the normal one as far as epidemiological evidence is concerned. Normally, traits of interest are present in only some proportion of a population—and often a small proportion; as noted earlier, lung cancer is a rare disease even among smokers. The next question, then, is whether this fact suffices to render epidemiological evidence entirely irrelevant to proving causation in an individual.

The answer is surely negative. The mere fact that a trait is present in only 99% of the population rather than 100% does not immediately render the population-level claim totally irrelevant to individuals. It means, rather, that we are no longer dealing in certainties, but in probabilities. If 99% of people in a population are smokers, this tells us that a randomly selected person is probably a smoker, with a probability of 0.99. And so on. Indeed, if population-level evidence carried no information about individuals, then it would be irrational for an individual to base a decision to stop smoking on an epidemiological study.

It is sometimes said, in this connection, that the but-for test is not satisfied for individuals when epidemiological evidence is appealed to. This appears to be the position of the Korean Supreme Court in tobacco litigation ongoing at time of writing: "Even if an epidemiological correlation between a specific risk factor and the non-specific disease is acknowledged, the correlation simply means that exposure to the risk factor means the existence or increase of the risk of developing the disease and does not necessarily lead to the conclusion that the risk factor is the cause of the disease, as long as there is the possibility that the individual or group exposed to the risk factor is regularly exposed to another risk factor."[15] In effect the Court is stating that epidemiological evidence cannot establish causation in cases where there is some other *possible* cause of the disease; that the but-for test cannot be satisfied.

As we have seen, epidemiological evidence can be applied to particular cases to offer a probability that a disease would not have occurred, but for the exposure (given the assumptions identified in the previous section). Thus it would be wrong to suppose that epidemiological evidence is simply inapplicable to the but-for test. The difficulty must be in meeting the legal standard of proof. There is no doubt that epidemiological evidence can yield a probability of disease but for the exposure. A strong objection, then, must be on one of two grounds: either the probability yielded is insufficient to tip the balance of probabilities; or the nature of the probability yielded by epidemiological evidence is different in kind from the probabilities balanced in legal tests. The former objection is clearly a weak one since it will be refuted whenever the epidemiological evidence accurately and reliably yields a probability over 50%. It is the second objection—the claim that epidemiological probabilities are incommensurate with legal proof—that needs more detailed attention. This is the real conceptual difficulty facing the use of epidemiological evidence in law. Let us turn to it now, in its various forms. Each of the remaining objections in this section amount to the argument that even if epidemiological evidence can generate probabilities about particular cases, these probabilities cannot feed directly into a legal fact-finding exercise.

Particularistic Evidence

Some legal scholars insist that there must be "particularistic evidence" which is specific, in some way, to the individual in question.[16] This is one attempt to express the idea that

epidemiological evidence is somehow unsuitable for legal proof. While that idea is an important one, this particular way of expressing it is difficult to make sense of. Evidence of any kind involves subsuming the individual under a generalization of some sort. To say that some fact is evidence for a particular claim is to say that, in general, facts like this are present when claims like this are true. Suppose that there were a medical test that could determine whether smoking was caused by lung cancer. Surely this would count as particularistic evidence; there is a direct causal nexus between the etiology of the lung cancer in this particular person and the result of the test applied to her on this particular occasion. It is difficult to think of any evidence that could be more particular. Yet the reason that the result of this test amounts to evidence is that results of this kind *generally* accompany cases where smoking caused lung cancer.

What is more, the failure to consider "general" evidence when assessing "particular" evidence can lead to well-known fallacies, such as the base rate fallacy. Suppose that a test has a 5% false positive rate, meaning that 5% of healthy people test positive. Suppose for simplicity that it has zero false negatives. It is tempting to suppose that a person with a positive test result has a 95% chance of having the disease. But in fact nothing can be concluded without information about the prevalence of the disease in the population—the "base rate." Suppose 1 in 1000 people have the disease. Then the probability that you have the disease given that you test positive is not 95%, but just over 2%.[17] This shows that particularistic evidence on its own is often meaningless in the absence of population-level evidence. The idea that population-level evidence is somehow irrelevant to individuals is thus demonstrably false, and even has a name in statistical reasoning (*the base rate fallacy*). See the discussion of posttest probabilities in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* for further information.

Thus this way of objecting to the use of epidemiological evidence is not ultimately helpful, since it does not succeed in pinpointing the difference between epidemiological evidence (or the probabilities it produces) and other evidence. Moreover it may encourage the erroneous overlooking of general evidence in assessing supposed particularistic evidence, which leads to well-known fallacies.

Skepticism About Attaching Numerical Values to Legal Standards of Proof

Some legal scholars have sought to argue that the standard of proof cannot be quantified, or cannot be mapped onto a quantified probability. The claim is that to prove something in law is to show that it probably is the case, and this is not the same as placing a bet that something is the case. The epidemiological evidence may tell us where to place our bets, but not what actually happened.

This line of argument falls to the observation that, in fact, the position of a civil court is very much akin to the position of somebody placing a bet. The court cannot remain agnostic on any point related to the matter in question, since part of the point of having courts is to settle disputes. Thus no matter how terrible the evidence on either side is, the court needs to come to a decision. Even throwing a case out for lack of evidence is still a decision. The mechanism of burden of proof sets some sort of a bar that evidence must clear; but the bar

is relatively low. In effect, a judge or jury needs to select the version of events that is the most probable given the presented evidence. The fact-finder does not have to decide whether either version is very well supported by the evidence, only which version is better supported.

On the face of it, epidemiological evidence may be sufficiently compelling to clear the low bar set by the burden of proof. Epidemiological evidence can give us some reason to think that there was a causal link between exposure and outcome in a particular case. The objection that basing a decision on such evidence would be like placing a bet fails, because in relevant respects, legal fact-finding is indeed as epistemically uncomfortable as placing a bet. The court must take the evidence, decide if it even merits consideration, and if it does, find for the most probable version of events, and treat the finding as fact even if it is not particularly probable. If epidemiological evidence is to be resisted because it is incompatible with legal proof, it must be because it is simply unable to move the burden of proof at all. Likening reliance on epidemiological evidence to betting does not provide us with a reason to think that epidemiological evidence is so incapable. To concede this would be to concede that epidemiological evidence may be ignored; an analogy with betting may make us feel uncomfortable, but we need a stronger, clearer, and less suggestive argument before this can be accepted.

Paradoxical Uses of Statistical Evidence

The most compelling objection to the use of epidemiological evidence in proof of specific causation comes from classic jurisprudential paradoxes about the use of "naked statistics." In many situations, such evidence seems to lead to unjust results if it is allowed to stand at its face value. For instance, suppose 1000 people attend a rodeo but only 499 pay. On the balance of probabilities, a randomly selected person is a gatecrasher. Yet we would not want the rodeo owner to be able to recover automatically from anyone who could not produce a ticket. That problem is known as the *gatecrasher paradox*.

A similar problem is the Blue Bus Problem, with facts similar to those of an actual case.[18] Here, a bus strikes Mrs. Smith's car at night. She does not notice the color of the bus, but a large proportion—say, 80%—of buses in that area are run by the Blue Bus Company. The court in the relevant case held that this was not enough for her to recover from the Blue Bus Company, and most commentators agree that this is correct.

There is no doubt that this kind of use of statistical evidence is disturbing. If the use of epidemiological evidence to assess specific causation is but one typical example of such uses, then indeed that would give serious pause to anyone seeking to rely on it. Thus this class of objections to the use of epidemiological evidence is the most significant.

However, in practical terms, there are considerations that tell in favor of using epidemiological evidence to assess specific causation, even if there are reasons for caution about other uses of statistical evidence.

First, both these paradoxes concern identity, rather than causation. Identity is something for which we can have direct empirical evidence. Causation, on the other hand, is never directly perceived. Every causal claim is established by an inference. The inference is from features of this case that make it like other cases where we already think causation occurs, and unlike others where we think it does not. We never *observe* what would

have been the case but for the defendant's wrongful act, since by definition that situation is counterfactual, and not actual. We infer, however, that this case would have gone like others, on the basis of shared features. Thus the use of statistical evidence to prove causation is different from the use of other kinds of evidence only in the fact that the uncertainty is quantified, and not in respect of the fact that statistical evidence works by subsuming the individual under a larger pattern. That is the case with all evidence for causation, whether statistical or not.

The second point that is worth bearing in mind when considering whether it is just to use epidemiological evidence is much more pragmatic. It is simply that in some situations, this may be one of the only pieces of evidence available to assess causation. In the gatecrasher and Blue Bus cases, part of our repulsion comes from being aware that there is other evidence that could have been considered, had only it been available. A rodeo owner who simply does not bother to collect this evidence is being lazy, and a driver who fails to register the color of a bus that she claims damaged her car when there were no witnesses present is asking us to accept too much on her word. The typical scenario where epidemiological evidence is present, however, is one where limitations in medical science mean we simply could have no other or better evidence about causation in the individual case. There just is no test for determining whether a case of lung cancer in a smoker was caused by his or her smoking. Given the paucity of evidence in such cases, it seems just to allow epidemiological evidence to be considered; and arguably it is the fact that we would normally expect other evidence to be available in the gatecrasher and Blue Bus cases that drives our sense of repugnance at relying solely on statistics in those cases.

The third point against allowing these paradoxes to block the use of epidemiological evidence is simply that statistical evidence is already used in other areas of law. If these paradoxes render it useless then the use of DNA analysis and fingerprints, for example, must also be ruled out; but we can hardly do without those.

CONCLUSION

Epidemiological evidence can be used to estimate a lower bound on the probability of causation as follows.

$$PC \geq 1 - \frac{1}{RR}$$

Epidemiological evidence is capable of satisfying the but-for test for causation on the balance of probabilities, provided that the individual is in effect randomly selected from the population with regard to potentially causative factors. There is potential legal resistance to using epidemiological evidence in this way but that resistance appears to have its basis in misunderstandings or well-known reasoning errors. There remain interesting jurisprudential questions about the use of statistics as evidence, but the situation where epidemiological evidence is typically used can be distinguished from the examples that are usually cited in those jurisprudential discussions.

CASES

American

Smith v. Rapid Transit Inc., [1945] 58 N.E. 2d 754.

English

Barnett v. Kensington & Chelsea Hosp., [1969] 1 Q.B. 428.

Cork v. Kirby Maclean, Ltd., [1952] 2 All E.R. 402.

Korean

Korean Supreme Court Decision 2011Da22092.

Scottish

McTear v. Imperial Tobacco Ltd., [2005] CSOH 69.

ENDNOTES

1. Rothman, Greenland, and Lash, *Modern Epidemiology*.
2. There is a live debate as to whether this tendency is healthy. See, inter alia: Kaufman, "Toward a More Disproportionate Epidemiology"; Poole, "On the Origin of Risk Relativism"; Greenland, "Cornfield, Risk Relativism, and Research Synthesis"; Broadbent, *Philosophy of Epidemiology*, 129–144; Broadbent, "Risk Relativism and Physical Law." Here, no endorsement is intended; the point is simply that lawyers are very likely to encounter epidemiological evidence presented in the form of relative risks.
3. Szklo and Nieto, *Epidemiology: Beyond the Basics*, 183–223.
4. Rothman, Greenland, and Lash, *Modern Epidemiology*, Chapter 5.
5. There is obviously a very large literature on causal inference in epidemiology. Regarding the causal interpretation of measures such as RR, see in particular Broadbent, *Philosophy of Epidemiology*, Chapter 3. For discussions of epidemiological causal inference in connection with toxic tort litigation, see Thomson, "Causal Inference in Epidemiology: Implications for Toxic Tort Litigation"; Haack, "An Epistemologist Among the Epidemiologists"; Haack, "Proving Causation: The Holism of Warrant and the Atomism of Daubert." For a selection of literature on causal inference in epidemiology more generally, see inter alia Hill, "The Environment and Disease: Association or Causation?"; Holland, "Statistics and Causal Inference"; Lipton, *Inference to the Best Explanation*; Bird, "The Epistemological Function of Hill's Criteria"; Broadbent, "Causal Inference in Epidemiology: Mechanisms, Black Boxes, and Contrasts"; Hernán and Robins, "Causal Inference."
6. Greenland and Robins, "Epidemiology, Justice, and the Probability of Causation," 1129.
7. The but-for test states that one can establish a cause by showing that but for the defendant's wrongful act, the harm would not have occurred. See: *Barnett v. Kensington & Chelsea Hosp.*, [1969] 1 Q.B. 428; *Cork v. Kirby Maclean, Ltd.*, [1952] 2 All E.R. 402. The test is not always appropriate but it is the starting point of the causal inquiry.
8. Haack, "Proving Causation: The Holism of Warrant and the Atomism of Daubert."
9. For related discussions see Barnes, "Too Many Probabilities: Statistical Evidence of Tort Causation"; Miller, "Epidemiology in the Courtroom: Mixed Messages from Recent British Experience."
10. Greenland, "Relation of Probability of Causation to Relative Risk and Doubling Dose: A Methodologic Error That Has Become a Social Problem"; Greenland and Robins, "Epidemiology, Justice, and the Probability of Causation"; Parascandola, "What Is Wrong with the Probability of Causation?"
11. Dore, "Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact," 433.
12. Thomson, "Causal Inference in Epidemiology: Implications for Toxic Tort Litigation," 255.
13. See, "Use of Human Epidemiology Studies in Proving Causation," 478.
14. *McTear v. Imperial Tobacco Ltd.*, [2005] CSOH 69.

15. Korean Supreme Court Decision 2011Da22092 at paragraph 4(A).
16. Wright, "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts"; Wright, "Liability for Possible Wrongs: Causation, Statistical Probability, and the Burden of Proof."
17. For every 1000 people tested, 1 will have the disease, on average. The test will detect that person (we are assuming no false negatives). In addition the test will deliver 50 false positives. So of the 51 positive results, on average, 1 will have the disease. The probability of disease given positive result is thus 1/51, or just over 2%.
18. *Smith v. Rapid Transit Inc.*, [1945] 58 N.E. 2d 754.

Further Reading

Barnes, D.W., Too many probabilities: statistical evidence of tort causation. Law and Contemporary Problems 64 (4) (n.d.), 191–212.
Bird, A., 2011. The epistemological function of Hill's criteria. Preventive Medicine 53 (4–5), 242–245.
Broadbent, A., 2011a. Causal inference in epidemiology: mechanisms, black boxes, and contrasts. In: McKay Illari, P., Russo, F., Williamson, J. (Eds.), Causality in the Sciences. Oxford University Press, Oxford, pp. 45–69.
Broadbent, A., 2011b. Epidemiological evidence in proof of specific causation. Legal Theory 17, 237–278.
Broadbent, A., 2013. Philosophy of Epidemiology. New Directions in the Philosophy of Science. Palgrave Macmillan, London and New York.
Broadbent, A., 2015. Risk relativism and physical law. Journal of Epidemiology and Community Health 69 (1), 92–94.
Dore, M., 1983. Commentary on the use of epidemiological evidence in demonstrating cause-in-fact. Harvard Environmental Law Review 7, 429–448.
Gold, S., 1986. Causation in toxic torts. Yale Law Journal 96, 376–402.
Greenland, S., 2012. Cornfield, risk relativism, and research synthesis. Statistics in Medicine 31, 2773–2777.
Greenland, S., 1999. Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem. American Journal of Public Health 89, 1166–1169.
Greenland, S., Robins, J., 2000. Epidemiology, justice, and the probability of causation. Jurimetrics 40, 321.
Haack, S., 2004. An epistemologist among the epidemiologists. Epidemiology 15 (5), 521–522.
Haack, S., 2008. Proving causation: the holism of warrant and the atomism of Daubert. Journal of Health and Biomedical Law IV, 253–289.
Hernán, M.A., Robins, J.M., February 16, 2015. Causal Inference. http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
Hill, A.B., 1965. The environment and disease: association or causation? Proceedings of the Royal Society of Medicine 58, 259–300.
Holland, P.W., 1986. Statistics and causal inference. Journal of the American Statistical Association 81 (396), 945–960.
Kaufman, J.S., January 2010. Toward a more disproportionate epidemiology. Epidemiology 21 (1), 1–2. http://dx.doi.org/10.1097/EDE.0b013e3181c30569.
Lipton, P., 2004. Inference to the Best Explanation, second ed. Routledge, London and New York.
Miller, C., 2012. Epidemiology in the courtroom: mixed messages from recent British experience. Law, Probability and Risk 11 (1), 85–99.
Parascandola, M., 1998. What is wrong with the probability of causation? Jurimetrics 39, 29–44.
Poole, C., January 2010. On the origin of risk relativism. Epidemiology 21 (1), 3–9. http://dx.doi.org/10.1097/EDE.0b013e3181c30eba.
Rothman, K.J., Greenland, S., Lash, T.L., 2008. Modern Epidemiology, third ed. Lippincott Williams & Wilkins, Philadelphia.
See, A., 2000. Use of human epidemiology studies in proving causation. Defence Counsel Journal 67, 478–487.
Szklo, M., Javier Nieto, F., 2007. Epidemiology: Beyond the Basics, second ed. Jones and Bartlett Publishers, Boston, Toronto, London, Singapore.
Thomson, M.M., 1992. Causal inference in epidemiology: implications for toxic tort litigation. North Carolina Law Review 71, 247–292.
Wright, R., 1988. Causation, responsibility, risk, probability, naked statistics, and proof: pruning the bramble bush by clarifying the concepts. Iowa Law Review 73, 1001–1077.
Wright, R., 2008. Liability for possible wrongs: causation, statistical probability, and the burden of proof. Loyola of Los Angeles Law Review 41, 1295–1344.

CHAPTER

5

The Role of the Expert Witness

M. Faure

Erasmus University Rotterdam, Rotterdam, The Netherlands; Maastricht University, Maastricht, The Netherlands

L. Visscher

Erasmus University Rotterdam, Rotterdam, The Netherlands

M.P. Zeegers

Maastricht University, Maastricht, The Netherlands

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

OUTLINE

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00005-7

INTRODUCTION

In this chapter, we discuss general aspects of the role of the expert in court and issues faced by experts. The chapter is written largely from a Dutch perspective, so that references to the Court almost exclusively refer to a judge rather than a jury acting as a fact-finder. Notwithstanding this perspective of the authors, the concepts illustrated herein are widely applicable to issues relating to experts serving in courts outside of the Netherlands.

Although the discussions in this chapter are intended to apply to experts in forensic epidemiology, we draw extensively from writings in Law and Economics, in part because so much has been published in this area on the topic of experts and expert testimony, and in part because these writings translate well to other disciplines.

CAUSAL UNCERTAINTY AND THE EXPERT

In addition to the contributions from the fields of philosophy and epidemiology described in Chapter 4, *Causation in Epidemiology and Law*, law and economics have also provided a foundation for the importance of epidemiologic evidence of causation. Shavell indicates that there is a good economic reason to limit the liability of the injurer to the damage that he has actually *caused*. If the requirement of the causal link did not have this limiting effect, a consequence could be that, especially in cases of strict liability, a potential injurer (eg, an enterprise) would be motivated to abstain from activities that are socially useful and therefore desirable (eg, the production of pharmaceuticals). A potential liability for damage which has not been caused through the influence of the injurer would thus be considered as "crushing."[1]

Many examples exist in tort law of situations in which there is uncertainty concerning the causal relationship between the tort and the damage.[2] The following example concerns the banned drug diethylstilbestrol (DES). The product caused vaginal cancer in offspring of mothers who took DES during pregnancy. The causal link between the use of DES by the mother and the daughter's symptoms was not disputed in the associated litigation. While it was also known which manufacturers had brought DES to the market, there was uncertainty regarding which manufacturer had sold a specific product to a particular mother. Several lawsuits were brought by DES-injured offspring, in which the plaintiffs sued all the producers who had brought DES to the market at the time of their injury, although they could not provide proof of the specific manufacturer from which their mother had purchased the drug. This gave rise to a lengthy debate about whether a proportional liability rule should be used to apportion the burden of liability between the manufacturers.[3] A market share liability would be an example of such a proportionality rule. The Dutch Supreme Court, however, applied a so-called "alternative causation" rule, meaning that the DES daughters were allowed to claim full compensation from any of the manufacturers.[4] A manufacturer could still rebut the presumption by *proving* that he did not sell DES to the particular mother, but this would often be impossible in practice.[5] Hence, the result was similar to a joint and several liability rule.[6]

A second example of shifting the burden of causal uncertainty relates to the employer's liability for occupational diseases. In a well-known Supreme Court case, *Cijsouw v. De Schelde*, a victim of mesothelioma could not prove at what time he had been in contact with the fatal

asbestos fiber that caused his disease. The determination of this moment was crucial for the case since Cijsouw had worked for the defendant firm for several years, but initially the employer could not have known that it was necessary to take measures to protect his employee against mesothelioma and thus could not be liable for the injury. The Supreme Court once more shifted the risk of uncertainty concerning causation to the enterprise by holding that it was presumed that the employee had been in contact with the asbestos fiber that caused his death during the later period of his employment with the defendant.[7] This presumption could have been rebutted if the defendant had been able to prove that it was not during the later period of their employing Cijsouw that the latter was in contact with the fatal fiber.[8] Like with the DES example, the burden that was shifted to the defendant was nearly impossible to meet.

In the third example, *Nefalit v. Erven Karamus*,[9] Mr. Karamus had worked with asbestos for many years in the factory of Nefalit. Almost 20 years after he stopped working, he was diagnosed with lung cancer and subsequently died. His heirs asserted that Nefalit was liable for the cancer because it had taken too few precautions to protect Karamus. Nefalit countered by claiming that the fact that Karamus smoked for 28 years was the cause of the cancer. The district court appointed an expert, who assessed—on the basis of epidemiological data and a calculation model—that the probability that the asbestos had caused the lung cancer was 55%, and therefore the court ordered Nefalit to pay 55% of the losses. The Court of Appeals confirmed this ruling. The Supreme Court, however, argued that despite the expert evidence it was not possible to prove to what extent both factors contributed to the disease, and that the liability of the employer should be reduced by the extent to which the employee contributed to the loss. The result was proportional sharing of the loss between employer and employee.

In the first two cases, there were no expert issues associated with the decision to shift the burden of proof to the enterprise. This eradicated the issues raised in Chapter 4, *Causation in Epidemiology and Law* with regard to the admissibility and probative value of the expert opinion, since the burden was largely insurmountable. In the third case, causal attribution and losses were apportioned between the claimant and the employer.

Causal uncertainty is an important issue routinely encountered by courts outside of the Netherlands, of course. Chapter 1, *Legal Considerations of Forensic Applications of Epidemiology in the United States* and Chapter 2, *Epidemiologic Evidence in Toxic Torts* of this book describe how courts in the United States have historically handled uncertainty and how expert epidemiologic testimony can throw light on disputed issues of causality. Examples from the legal systems of other countries are abundant; causal uncertainty played a major role in the famous British Sellafield case, where an English court had to decide on the causal relationship between childhood leukemia and the nearby presence of a nuclear power plant at Sellafield.[10] Similarly, Belgian courts have been confronted with the question of whether the physical complaints of inhabitants of the community of Mellery in the Walloon Region were caused by emissions from a nearby waste site.[11]

In the preceding cases the critical question concerned how the law should approach with causal uncertainty. One general agreement in all courts relates to the liability rule described earlier, which dictates that no liability is incurred for the background risk that preexisted the harmful exposure, but that the only liability is for the excess risk created by the activity of the injurer.[12] From an economic standpoint liability for the background risk would lead to

"spillage" and thus crushing liability, and would be useless since the investments in the care of the injurer could not reduce the background risk.[13]

The law has to therefore address how the excess risk can be quantified. This can be done by establishing, via epidemiological evidence, the extent to which the risk has been increased by the injurer's activity. Epidemiological evidence can then lead to a probability of causation (PC), which indicates what the probability is that the injurer's activity caused the loss of the victim (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* and Chapter 4, *Causation in Epidemiology and Law*). The concept of the PC is well established in radiobiology to establish the likelihood that an employee in, for example, a nuclear power plant contracted a cancer as a result of his work exposure to quantifiable levels of radiation. In 1992, the American National Council on Radiation Protection and Measurements endorsed a PC method to establish whether a certain disease was caused by radiation.[14] On the basis of this method, the American National Institutes of Health issued a statistical guide indicating the PC for certain exposure doses and circumstances.[15] The guideline from the NIH helps illustrate a streamlined example of how epidemiological measures of association can provide quantitative insights into causal uncertainty. As many situations are not nearly as tidy, the following discussion is an explication of four alternative approaches that can be examined for utility in assessing causal uncertainty in a tort.

We assume, in describing these approaches, that expert opinions exist on the likelihood that a certain activity caused certain damage and that an epidemiologic analysis of the attributable fraction under the exposed (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) resulted in a PC. Thus, the expert opinion could assert that, while it is unknown that a certain exposure *did* cause a certain adverse health outcome, it can be reliable that there is a probability, for example, 30, 50, or 70%, that this is the case. The question of interest is how the legal system might deal with causal uncertainty if expert opinion cannot provide it.

Four alternative methods of dealing with this issue are as follows:

1. One could judge that as soon as it is determined that there is any statistically valid probability that an exposure causes an injury that all victims receive 100% compensation of their damage.
2. The complement of the first option is to deny any claim unless there is 100% certainty that the tort caused the injury.
3. Compensation could be awarded only when the probability that the damage was caused by the tort passed a certain threshold of, say, 50%, as described in the first four chapters of this book. This threshold rule is a form of an "all or nothing" approach: if the probability is lower than the threshold, the victim receives no compensation at all; if the probability is higher than the threshold, the victim receives full compensation. This threshold rule is known in the American system as the "more likely than not" solution, referring to the fact that the plaintiff must convince the judge that it is "more likely than not" that the damage was caused by the tort. In the hypothesis of PC of 30%, there would be no liability.
4. The final alternative is to take into account the probability that the tort caused a certain proportion of the observed injury and to award commensurate compensation based on the PC. In the hypothetical situation in which a PC is 30%, the victim can then receive compensation for 30% of his losses.

The way the law should deal with causal uncertainty has been addressed extensively in the economic literature, for instance by Rosenberg,[16] Kaye,[17] and Shavell.[18] This literature provides some interesting insights concerning the best approach to dealing with causal uncertainty, relative to the list above.

Alternatives 1 and 2

In the first alternative the victim would be awarded total compensation of damage, even if the probability that the loss was caused by the injurer's activity was relatively low (eg, a PC of 10%). If an injurer is held liable for the full amount even if there was only a 10% probability that his activity caused a loss, it can be argued that this would lead to too few incentives to invest in a socially desirable activity, such as the production of pharmaceuticals (hence, crushing liability).

Landes and Posner also provide an example of this first case.[19] Suppose that a nuclear plant wrongfully emits ionizing radiation into the environment, whereby it is established that an exposure of a certain population to the radiation increases the number of cancer cases over a 20-year period from 100 to 111. The problem is obvious that we do not know which of the 111 persons have incurred the cancer as a result of the presence of the nuclear power plant. For every separate victim, the likelihood of incurring cancer as a result of radiation has been increased by approximately 10%. If we now accepted that any of the 111 victims would be allowed to claim complete compensation of their damage from the injurer, this would not only be inefficient, but also unjust. It would indeed mean that, in 100 out of 111 cases, the injurer would have to pay compensation for damage that it never caused, and that only resulted from the background risk. Landes and Posner argue that in that particular case, if one disregards the administrative costs, the efficient solution would be the one in which each of the victims could only recover a proportion of their loss (ie, alternative 4), equal to the excess risk caused by the activity of the injurer, from the injurer.[20] In this case, that would amount to approximately 10% of the damage.

The example demonstrates that the first alternative is inefficient and unjust. By logical extension we can argue the same is true for the second alternative, in which it would be required that the victim proves with 100% certainty that his damage has been caused by the tort. The requirement would mean that in many cases injurers would escape legal consequences of activities that have created additional risk. Thus, the second alternative would result in underdeterrence. It is only the third and fourth alternatives that lead to a result that could be considered balanced.

Alternative 3

The threshold liability leads to a situation whereby the victim's claim is totally accepted if the probability surpasses the threshold of what is true more than 50% of the time, or on a "more likely than not" basis. If the probability surpasses the threshold, compensation is full, but if the probability is lower than the threshold, the victim receives no compensation at all. Despite the prevalence of this approach in many legal systems, the disadvantages of such a threshold solution are rather obvious. One problem, both from the victim compensation as well as from the deterrence perspective, is that the PC could systematically be lower

than the threshold. For example, we can assume that the PC for a certain cancer being caused by a harmful exposure is 40%. At the threshold of 50% this would mean that the enterprise responsible for the exposure leading to 40% of the disease among the exposed would systematically escape liability. Victims would not be compensated and the incentives toward accident reduction would be too low.[21] The result is both inefficient and arguably unjust, since it cannot be denied that the enterprise created certain losses. If we assume that 100 exposed cancer victims from the example file a lawsuit, then 40 out of the 100 cancer cases would have been caused by the emissions emanating from the particular enterprise. Yet, since for every individual the PC would be below the 50% threshold, the enterprise would not be held to compensate the victims in any of these cases.

In practice, the outcome may be more nuanced because the PC may differ between subgroups of victims. Confounding and effect modifying factors such as age, gender, habits (smoking, drinking, diet, exercise), the dose of the exposure (eg, the distance between the source of radiation and the location where the victim lives) may affect the PC per plaintiff. Therefore it is perfectly possible that the PC of some victims exceeds the threshold while the PC of others falls short of it. This approach would hold the enterprise liable in some cases, but not necessarily in all cases where the activity indeed was the cause of the damage.

Alternative 4

The alternative to the threshold approach is the proportional award approach. Applied to the prior example this would mean that if the probability that the victim's damage was caused by the injurer's activity was 40%, the victim would be compensated for 40% of his losses. From an economic perspective, the advantage of this proportional liability is that it exposes the injurer precisely to the excess risk (in this case the additional number of cancer cases) that was caused by the (assumed wrongful) activity of the injurer in each case, provided that all potential injurers and all victims are present in the tort case. The enterprise will then, returning to the previous example, have to compensate 40% of all the damage of every particular victim, which amounts at the aggregate level to the same as compensating 40 out of 100 victims whose illness would have been caused by the enterprise.[22] This approach also accounts for the fact that the PC may differ between subgroups.

The result of the proportional liability approach is that the injurer will receive optimal incentives for prevention, since he is precisely exposed to liability for the risk, which was caused by his activity.[23] A proportional liability rule therefore provides optimal incentives for injury reduction.[24]

There are, of course, several caveats that accompany the proportional liability approach. One caveat is that in the examples we so far assumed that the expert can always make some kind of assessment of the PC. In practice, however, in some cases it can be extremely difficult to make an accurate assessment of probabilities that a certain activity may cause certain damage. Some argue that the PC approach is something like a lottery.[25] It is reasonable to argue that one should not expect too much of statistical evidence, and further that reasonable epidemiologists can disagree about the magnitude of an association that is used for a basis of a PC. Such difficulties will always be present when causation is uncertain, however, regardless of methods. The fact that one cannot rely on the myth of scientific *certainty*

should not be an argument in favor of the unbalanced alternatives described above (alternatives 1 and 2).

The choice between the threshold or the proportional liability approach is one that is disconnected from the expert's process of estimating a PC. The question of interest is how the law should deal with the estimate. We posit that the proportional approach is a better and more balanced alternative to the all or nothing approach associated with the threshold rule.

A second caveat is that a proportional liability rule may incur higher administrative costs than a threshold rule. A threshold rule rules out many liability claims, whereas in case of a proportional liability rule, even the smallest probability could in theory lead to a claim. However, this argument of administrative costs could be an argument in favor of a threshold that is applicable to the proportional liability approach, for example, 10% PC under which no compensation is paid at all and the application of the proportional liability rule only to cases above the threshold. One could equally make the argument that in cases where the PC is, for example, 90%, the claimant receives 100% compensation.

The proportional liability rule has been defended by several American scholars and is also defended in the economic analysis of law.[26] The discussions have noted that an advantage of the approach is that the negative consequences of causal uncertainty are limited, and that a proportional liability rule is less rigorous than the all or nothing threshold approach.[27] The proportional liability rule would indeed mean that all victims (possibly with the exception of those with a very low PC) can claim a proportion of their damage equal to the amount by which the defendant contributed to the loss. Thus the exposure to liability of the enterprise corresponds only with the amount to which it contributed to the risk.[28] This proportional liability rule could, more particularly in cases of product liability, take the form of the market share liability.[29]

THE ROLE OF THE FORENSIC EPIDEMIOLOGIST AS AN EXPERT

So far, we have only lightly touched on the role of the epidemiologist as an expert in addressing causal uncertainty. We now turn to the specific role that the expert could play in that respect.

A first question that one could raise is why an epidemiologist should at all be involved as an expert in assisting the court in solving issues of causal uncertainty. Chapters 1, 2 and 4 of this book have already answered this question rather exhaustively. There are, however, several economic arguments to support such involvement as well. First is the superior knowledge that the qualified epidemiological expert possesses in the evaluation of health risk. The epidemiologist can therefore obtain and process information at lower cost than the parties or Court. An argument can therefore be made that involving an epidemiological expert can lower information costs. Since the economic goal of tort law is to minimize the *total* injury costs (thus including administrative costs) this seems a desirable effect.[30]

Further, by dint of training and experience, the forensic epidemiological expert can provide better quality and more reliable information on causality than the parties or Court. Involvement of an epidemiologist can hence guarantee that issues of causation and more particularly epidemiological questions can be answered in such a way that damages will

be more accurately attributed to the injurer. Hence, involving the epidemiological expert better allows tort law to reach its goal of prevention by better exposing the injurer to the amount of damages that corresponds with his contribution to the accident risk. When asked why the epidemiological expert would be better able than the parties involved or the judge to provide information necessary to let tort law function optimally, the reason is the specialized education and studies followed by the epidemiologist expert, including, for the forensic epidemiologist, unique applications designed to assess questions relating to causal uncertainty applicable to individual circumstances (see the applications of FE described in Part III *Applications of Forensic Epidemiology* of this book for examples). Such an expert is undoubtedly better informed than the court of relevant insights and the latest developments in his area of expertise.

Since a forensic epidemiologist is regularly confronted with similar questions (eg, related to the calculation of a PC), he enjoys the advantage of the experience that benefits the *repeat player*.[31] Being a repeat player he can use prior experience to obtain accurate information more quickly and hence at lower cost than a less experienced expert or a fact-finder. Judges in most legal systems are not specialized, and even if they cannot be expected to have the degree of requisite knowledge and expertise to investigate complex issues relating to epidemiology and causality. From a social perspective one cannot expect a judge to make the educational and time investment that would provide him with the ability to assess the likelihood that a tortious act was a cause of a specified injury. The socially preferable solution is to use the experience and expertize of the forensic epidemiologist who will be able to answer particular questions at relatively low costs, and can hence inform the judge. It is, in other words, the simple economics of labor specialization that explains why it makes sense to use experts in complex cases involving causal uncertainty.

A third advantage can be obtained if an *independent* expert is used. A potential problem with the expert information provided by the parties involved in the tort case is that they have, to a large extent, opposing interests.[32] The victim has an incentive to overstate the likelihood that his injury was caused by the tortious act, and the injurer has an incentive to argue that there is no evidence that the tortious act is related to the claimed damage. An independent court-appointed expert can focus entirely on the question of how the causation issue in the specific case should be assessed as adequately and in the most balanced manner possible. He is not exposed to strategic considerations providing incentives to increase or reduce the probability that the tortious act was causally related to the damage. This option is rarely available in US courts, but is common in European venues.

Some legal scholars have pointed to disadvantages of involving an expert. Meadow and Sunstein have argued that, rather than relying on an expert's impression of the scientific evidence, a judge should rather rely on purely statistical information in cases of, for example, medical malpractice.[33] The suggestion ignores the fact that experts in epidemiology typically compiled the statistics in the first place, and that the determination of whether the statistical information can be validly applied to the unique circumstances of an individual case requires the expert assessment of all the facts and how they fit within what has been previously described. Indeed, this task is precisely the kind of analysis that is described in the latter chapters in this text, in which data are accessed and analyzed *ad hoc* for the unique circumstances of a case as part of a forensic epidemiologic analysis.

IS THE EXPERT ALWAYS AN EXPERT?

There is substantial empirical evidence, as well a common experience, that experts, like all people, are subject to various types of bias. The most common bias from which experts suffer is a too large of trust in their own expertize; ie, overconfidence. In one study of this form of bias it was demonstrated that experts had largely overestimated the precision with which they could predict the likelihood of a meltdown of the nuclear core in a nuclear installation. In another study it was demonstrated that experts showed an irresponsibly large trust in the stability of the Teton Dam based on their personal assessments, notwithstanding a number of problems that had occurred during the construction of the dam. The Teton Dam collapsed in 1976.[34]

Slovic and others point to several factors that may cause experts to underestimate specific risks, including:

- The failure to recognize that human failure can largely influence the effectiveness of technological systems. The authors cite to the nuclear incident at Three Mile Island, where operators repeatedly wrongly assessed problems with the nuclear reactor and failed to intervene in time.
- The underestimation of the integral functioning of technological systems. The authors cite to the failure of engineers to discover that the reason for the failure of the DC-10 after its first flights was a decompression in the cargo part of the plane which led to a destruction of vital control systems.

Translating these technology-related failures to epidemiological terms, in both of the cases described above there were multiple concurrent causes of the problem that were incompletely investigated. The advantage held by experts is that they have more technical knowledge than the parties or the Court. This advantage can be abused to the benefit of the expert. Parker stresses this problem in a detailed study concerning the American rules related to the admissibility of scientific evidence.[35] His reasoning applies to the role of epidemiological experts in the assessment of causal uncertainty. Producing information creates costs for parties and thus profit for those supplying the information. By involving experts in the assessment of causal uncertainty an interest group is created, which can try to serve its own interests, even if these interests do not necessarily match with those of the parties involved or the Court. Parker holds that the goal of procedural law is to exclude these so-called "public choice" problems as much as possible from the civil trial. One way of doing this is by organizing the system in such a way that only the parties involved have an interest in the outcome of the case.[36] However, the expanding scope of liability and increasing amounts of compensation lead to increasing "returns on litigation." This means that parties are incentivized to rely on experts in order to increase their likelihood of a positive outcome in the case.[37]

Another source in expert error is in the systematic over- or underestimation of probabilities associated with a set of facts, or the lack of calibration. As an example, physicians commonly overestimate the survival probability for a patient with cancer. In part the error stems from the fact that experts often base probabilities theoretical models containing subjective assumptions, and in some cases the expert assumptions are no more accurate than assumptions made by non-experts. Appropriately performed epidemiologic investigation, following the

methods described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* are inherently less susceptible to errors of calibration.

Empirical evidence shows that judges are skeptical concerning the results of an investigation by experts when the experts have been hired by the parties involved. On the other hand, empirical evidence also shows that judges often have difficulties in understanding what can be complex and technical testimony provided by experts.[38] Another potential issue is that an expert with an impressive curriculum vitae is more likely to be given credence by the Court than a lesser credentialed but still qualified expert. This can lead to a situation in which the judge believes the testimony provided by the expert, not because of the credibility of the testimony as such, but because it is delivered by the particular expert.[39]

Tomlin and Cooper show that in court cases there is the danger of the following downward spiral: when one party would involve an expert who would, for example, make an objective and correct assessment of causal uncertainty, for example, for the victim, there is a danger that this party loses if the defendant in the same case would use an expert who would subjectively understate the likelihood that the damage of the victim was caused by the defendant. If both parties mutually expect that subjective information concerning the PC will be presented, this leads to a type of prisoner's dilemma, whereby one party cannot afford to "bid" less than the other party with regard to exaggeration of claims. Failure to engage in symmetrical exaggeration on the part of the party that provides objective information could result in losing the case unfairly.[40] The situation can result in a system whereby exaggerations by both parties (and their experts) are not only structurally possible, but neither party has sufficient incentive to provide any objective and accurate information. Like any expert, the forensic epidemiologist is not excluded from this potential conflict of interest. Although he may follow an evidence-based or data-driven approach, bias may still be present in his interpretation and presentation of this information. Conversely, a fair analysis of data should be reproducible and thus verifiable, and carries with it a greater degree of reliability than strictly medical opinions regarding causality.

What cannot be denied is that more wealthy parties (typically the enterprise versus the victim) can hire the most expensive experts, and more of them. While more expensive experts are often (but not always) better qualified and often (but not always) more experienced than less expensive experts, they may also be incentivized to provide an analysis that is favorable to the retaining party.

A problem that has been extensively dealt with in the literature concerns the fact that merely by being paid by one of the parties, the quality of the opinion provided by the expert can change. In the words of Mandel: "money changes everything."[41] This may be especially problematic in cases where no scientific certainty exists yet and hence differences of opinion may exist. This could lead a party to hire an epidemiologist of whom it is known (eg, through his publications) that he has a favorable view on the position defended by that particular party.

In the words of Sales and Shuman:

> There is a great deal of skepticism about expert evaluation. It is well known that expert witnesses are often paid very handsome fees, and common sense suggests that a financial stake can influence an expert's testimony, especially when it is technical and esoteric and hence difficult to refute in terms intelligible to judges and jurors. More policing of expert witnessing is required, not less.[42]

Other literature points to the fact that, assuming an independent expert tries to provide his opinion in an honest and objective way, the attorney acting for the party is only obligated to get the client the desired result. In those cases where differences of opinion may exist and answers are not always clear cut, there may be pressure on the expert to provide testimony that is more favorable to the position of the party who hired him.[43]

The question arises as to whether the fact that experts are hired by adverse parties necessarily jeopardizes epidemiological risk calculations. Tomlin and Cooper hold the somewhat jaundiced view that as long as both sides exaggerate in the same way and the judge subsequently chooses an average position, the outcome may be approximately right.[44] An example, taken from economic experts, illustrates the point. If the actual damage is known to be €100,000, and the expert for the plaintiff estimates the damage at €150,000 and the expert for the defendant at €50,000, there should not necessarily be a problem as long as the judge assumes that the truth is exactly between the two values, and uses €100,000 for the damage. The same principle could be applied to epidemiological measures. However, an important assumption in this example is that the amount to which both sides exaggerate their estimation is equal and that the Court is aware of this. The assumption is not necessarily realistic, however, and may lead to the Middle Ground Fallacy described in Chapter 14, *Medical Negligence Investigation* in which the error rate of two divergent opinions is grossly unequal, and thus using the mid-point between the two opinion is unfairly favors the party with the most exaggerated claim.

REMEDIES

The Judge as "Gatekeeper"

In the United States, an important task is awarded to the judge to verify the reliability of information provided by (party) experts in Federal Rule 702 that reads:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify there too in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

This Rule 702 followed inter alia a 1993 Supreme Court decision in the case *Daubert v. Merrell Dow Pharmaceuticals* (see Chapter 1, *Legal Considerations of Forensic Applications of Epidemiology in the United States* and Chapter 2, *Epidemiologic Evidence in Toxic Torts*).[45] The decision requires the trial judge to verify whether the reasoning followed by the expert as well as the applied methodology are scientifically sound. The judge equally has to verify whether the expert's reasoning and method could also be adequately applied to the facts in the case.[46] On the basis of the Daubert ruling it is hence held that the judge has a role as the evidence "gatekeeper." The process of gatekeeping consists of verifying on the basis of the *Federal Rules of Evidence* and the case law of the Supreme Court whether the reasoning and method followed by the expert was appropriate and a fit for the facts of the particular case. In the American procedural context the judge is mainly considered as a "gatekeeper" since he is the "gate" through which the evidence reaches the jury. However, the criteria

for admissibility of expert testimony developed in the United States in this context are also relevant when the judge has to verify the admissibility of expert testimony in a civil law context in the case.

The core of the American case law on evidence and the legal doctrine that has been developed from it holds that an important task lies with the judiciary to verify whether the information provided by the expert has been obtained according to appropriate scientific standards and methods and whether it is useful to answer relevant questions. This task is considered especially important when party appointed witnesses are used. The result of this verification by the judge could be that the judge would exclude expert testimony that does not meet these criteria. This would in turn provide the parties (and their experts) an incentive to strive for more objectivity when providing expert testimony to the courts. Tomlin and Cooper, however, show that in practice it is often difficult to reach this rather idealistic goal. Research indicates that 48% of the judges in state courts are of the opinion that they are not adequately prepared to assess the quality of what is often very complex and diverse expert testimony that is presented to them.[47]

That judges are, in some cases, quite capable of examining the quality of expert testimony in a critical way is seen in the interesting case *In Re Silica*.[48] In this case the district court in Texas declared expert testimony inadmissible because the expert had not followed applicable scientific standards.[49] The case dealt with a mass tort claim whereby experts would have established that 9000 plaintiffs suffered from silicosis. The Court, however, established that it was not the clinical experts, but rather the lawyers who had determined the criteria by which the diagnosis of silicosis would be established. Moreover, it appeared that most of the plaintiffs had never seen the medical expert who was providing the diagnosis, and that the experts had only examined files. Discovery in the case turned up the fact that the experts reviewed and verified 75 files per day, taking, on average, 4 min per file in order to arrive at the diagnosis of silicosis. The Court held that within the medical profession there are specific scientific criteria that have to be followed to arrive to a diagnosis of silicosis, and the criteria had not been followed in the case.[50] The Court further held that the medical experts never considered the victims as patients, but merely as clients of the lawyer. The district court concluded that it was financial motives that were the reason for the specious methods followed by the experts. The case served as an excellent demonstration that when experts clearly violate a professional and ethical norm, it is not difficult for the judge to reject the testimony.

Posner points to the fact that a second control mechanism for expert methods consists of the fact that even partisan experts should have an interest in following minimal scientific and methodological standards for the simple reason is that if they do not, their expert testimony will be declared inadmissible.[51]

Finally, one can also point at the fact that an expert can be a *repeat player*. As such he can therefore have a reputational motive to provide objective expert testimony. With epidemiologic experts the expert is often also academically active, publishing scientific papers, and maintaining a reputation for scientific integrity in his or her work. It would be highly problematic for an epidemiologist to proffer an opinion as an expert witness that differed with what he had previously asserted in academic publications.[52] Thornton and Ward argue that the market provides incentives to experts to strive at least for a minimum quality in their expert testimony work. When, for example, economists provide expert testimony on lost earnings and it appears that their estimations vary depending upon the retaining party;

this can quickly jeopardize the reliability of that expert in the eyes of the Court.[53] Ideally, market forces should therefore lead experts not merely to say what their clients and clients' lawyers want to hear. In some European courts, court-appointed experts are selected from a predetermined list. To some extent this fact can provide incentives to experts for providing more objective testimony.[54]

Self-Regulation/Certification

The economic theory of regulation as well as the sociology of professions has shown that professional groups will always strive to obtain protection for their professions. When possible they will strive for a protection of their title and for a monopoly on certain services, usually supported by (self-) regulation. The economic argument to support self-regulation has always been the information asymmetry: experts themselves would have excellent information on the quality of the services they provide, but the parties who use those services (either lawyers or the judge) would not be able to sufficiently control the quality of the services.[55] These are the types of arguments that one will often hear from the experts themselves, for example, to defend the mandatory membership of an association in order to be able to call oneself an "expert." Economists warn of the danger of when membership of such an organization becomes a precondition for exercising the profession, the remedy (limitation of competition) is often worse than the disease (information asymmetry). The well-known consequences of concentration on the market may emerge: agreements on prices and price increases can be the result. Minimum standards on quality could be agreed upon, but the experience with the (self-) regulation in other professions shows that they often have few incentives to strive for a high quality.[56] The problem is less of an issue in the practice of forensic epidemiology, mainly because it is a relatively new discipline. Nevertheless a warning here is justified.

It is interesting to notice that also experts active in the domain of assessing personal injury may have the tendency to strive for a recognition, mandatory membership of an organization, and certification. For example, an expert in the Netherlands was noted to have complained about the fact that being an expert in the Netherlands is to a large extent unprotected and hence in principle anyone can call himself a "traffic accident expert."[57] The tendency exists to require membership of a certain professional organization of experts as a quality label. This may create the danger of monopoly to the extent that only those experts would be considered as having the necessary quality. This has occurred in the Netherlands, where there is an institute of "register experts,"[58] as well as in Belgium.[59] From an economic perspective, one should appraise such professional associations with caution. Here, the well-known advice of Adam Smith applies; that when merchants come together the danger always exists that they will use the occasion to make agreements that increase their personal benefit to the detriment of social welfare.[60]

The same concern applies to the practice in many countries to work with the so-called lists of experts. Such lists may be relied upon by courts for appointed experts, or they may be used to bolster the credibility of a party-appointed expert. In some cases, the criteria by which an expert is included on a list is unclear, one is registered on such a list of experts. To the extent that list inclusion is not solely related to professional qualifications and experience, but also being known to the judiciary this may be problematic. Moreover, even if the list only contains

experts of high quality, working only with experts from such a list is already problematic in the sense that it excludes competition. Being on the list then in fact creates a property right that may again increase prices. The list then effectively creates a barrier to entry for new (equally good or perhaps better) experts who enter the market and would not be on the list yet. To the extent that courts use such lists they should thus be sufficiently dynamic to accommodate a changing environment, having free entry, and not be exclusive in the sense that others who equally provide reliable and objective expert testimony should not be allowed and welcomed by the courts as well.

In sum, one has to realize that the danger that a (party-appointed) epidemiological expert will not always serve in all objectivity the public interest should not necessarily be an argument to remedy this danger with the potentially more risky remedy of a restriction of competition (via certification). Other, more proportional, remedies, such as a control by the judge of the testimony provided by party-appointed experts and some of the other remedies discussed below seem more appropriate.

Standardization

Another remedy to consider is standardization of the way in which the expert performs research and analysis, and thus arrives at opinions. The advantage of such standards is that the judge can *ex post* verify whether the examination by the expert has been carried out according to the established professional and scientific standards. This of course supposes that these rules are made explicit in a public standardization. This proposal has received support in the literature. Mandel is in favor of laying down ethical standards for experts.[61] Also Posner indicates that it should be possible for the judge to verify whether the expert testimony corresponds with the methodological standards applicable to the particular profession.[62] This would according to Posner allow excluding "junk science" and could serve as an admissibility test of expert evidence.

Although epidemiology has developed standards on how to conduct scientific research, this text is the first attempt to summarize the methods employed in the reliable application of forensic epidemiology methods to a wide variety of issues involving causal uncertainty.

Court-Appointed Experts

Elliott discusses the problem that the legal system typically considers all experts equal. Hence, it remains difficult for a court to distinguish good from bad information. The fact may provide parties incentives to hire experts of whom it is known that they represent extreme points of view.[63] As a possible solution Elliott suggests that less use be made of party-appointed experts and more of experts appointed by the courts or employed by government.[64] Schwartz is critical of Elliott's suggestion, noting that he has a too rosy view of what experts are able to do.

An argument in favor of court-appointed experts instead of party experts is that party-appointed experts commonly provide diametrically opposing opinions. This contretemps may lead to a costly search for information that may not provide any additional clarity. The situation raises the danger that the court would resolve the contradiction by rejecting the testimony of either party-appointed experts as they are perceived as unreliable and therefore

not useful.[65] Such a judicial reaction ignores the very real possibility that one expert is correct and the other is not.

Posner provides support for a court-appointed expert, but proposes a model that is sometimes used in cases of arbitration.[66] The model he proposes is the one whereby two party-appointed experts would together decide to appoint a third independent expert. The latter would then have the task to increase the reliability of the expert testimony provided, but also to "translate" the expert evidence for the judge. In this respect it is correctly noted that the mere fact that parties have appointed their own expert is as such not necessarily a reason for the judge to appoint a court expert. The court expert should hence not be seen as a "super expert" who should do better than the party-appointed experts, but rather as someone who can intervene to increase the reliability of the expert evidence when the information provided by the party-appointed experts diverges too strongly; a mediator of sorts.

Capacity Building of the Judiciary

From the prior discussion, it is clear that interpreting evidence that is provided by an epidemiological expert can pose difficulties for the Court in some cases. The kind of standardization of methods provided by this text is helpful, however, as it allows the Court to, at a minimum, assess whether the analysis took place according to professional standards and norms. Judges who perform frequent reviews of epidemiologic testimony would benefit from education concerning the methods of the discipline. It is for this reason that the School of Law at Maastricht University, the Netherlands offers courses in basic forensic epidemiology for the judiciary (www.forensicepidemiology.nl).

Peer Review

Peer review is a possible solution to the problem of biased and/or low-quality expert opinions as well.[67] If a panel of peers of the expert could provide expert review of the methods and opinions of the expert and draw the conclusion that the methods are reliable and the opinions sound, this could offer the Court some assurance that the evidence is reliable. The process does not have to be an onerous experience for the expert, as deficiencies that may be identified in the peer review process could be addressed and rereviewed.

CONCLUDING REMARKS

We began this chapter by stressing the fact that causal uncertainty can be reliably addressed via an estimate of the PC following a methodologically sound forensic epidemiological analysis. If the PC is determined to be appropriately assessed, then the Court can use the information by applying a proportional liability or causal threshold approach.

Economic research indicates that from a theoretical perspective the use of party-appointed experts is not necessarily problematic. The mere fact that an expert is handsomely paid by one of the parties does not allow for the inference that the expert will serve the interests of that client and his lawyer in an exaggerated or dishonest way. If such a thing does occur there are market-based correction mechanisms that may work to remedy the problem. The fact that

many epidemiologic experts are not only repeat players as experts (and thus have an interest in representing consistent opinions in varying capacities), but equally have other capacities (eg, in an academic setting) provides incentives to the expert to not solely serve the interests of the client. In addition, the Court may sanction expert evidence that clearly violates professional standards either by excluding the evidence, or the Court may choose to appoint its own expert.

ENDNOTES

1. *See* Shavell, S., 1987. Economic analysis of accident law, 108; a similar point is made by Trebilcock, M.J., 1987. The social insurance − deterrence dilemma of modern North American tort law: a Canadian perspective on the liability insurance crisis. San Diego Law Review (24), 929−1002.
2. *See* on this topic also the many publications of Akkermans and more particularly his dissertation A. Akkermans, 1997. Proportionele aansprakelijkheid bij onzeker causal verband.
3. For an overview of how several European countries, as well as Israel, the United States, and South Africa deal with proportional liability, *see* Gilead, I., Green, M.D., Koch, B.A. (Eds.), 2013. Proportional Liability: Analytical and Comparative Perspectives. De Gruyter, Berlin.
4. For a discussion of alternative causation under German law, *see* Köndgen, J., 1991. Multiple causation and joint tortfeasors pollution cases according to German law. In: van Dunné, J.M. (Ed.), Transboundary Pollution and Liability, the Case of the River Rhine, pp. 99−106, and the interesting article of Bodewig, Th., 1985. Probleme alternativer Kausalität bei Massenschäden. Archiv für die civilistische Praxis (AcP), 505−558.
5. Note, however, that the Dutch Supreme Court only considered the causation question. Formally, it still has to be decided whether bringing DES on the market was in itself wrongful. *See* Hoge Raad October 9, 1992, 1994 Nederlandse Jurisprudentie (NJ), 535 (C.J.H.B.). *See* on this case, Spier, J., Wansink, J.H., 1993. Joint and Several Liability of DES Manufacturers: A Dutch Tort Crisis. International Insurance Law Review (IILR), 176−181.
6. So J. Spier in J. Spier (Ed.), 1996. The Limits of Liability, Keeping the Floodgates Shut, pp. 123−124. For a critical economic analysis of joint and several liability, *see* Tietenberg, T.H., 1989. Indivisible toxic torts: the economics of joint and several liability. Land Economics (65), 305−319 and *see* below 3B.
7. Spier, J. (*supra* note 6), 124−125.
8. Hoge Raad 25 June 1993, 1993 NJ, 686.
9. Hoge Raad 31 March 2006, 2011 NJ, 250.
10. Gardner, M., 1990. Results of a case−control study of leukemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. British Medical Journal, 423−434.
11. For a discussion of that case see Lavrysen, L., 1995. Judicial responses in the nineties to Dutch (and German) shipments of waste to Belgium in the eighties. Maastricht Journal of European and Comparative Law (MJ), 219−243.
12. *See* Bergkamp, L., 2001. Liability and Environment, 287−289.
13. Rosenberg D. (*supra* note 16), 865−866.
14. National Council of Radiation, protection and measurement statements no. 7, September 30, 1992; *see also* Bond, V., 1981. The cancer risk attributable to radiation exposure: some practical problems. Health Physics, 108−111.
15. *See* on these tables also Ketchum, L., 1985. Epidemiologic tables. Law groundwork for future radiogenic cancer claims. The Journal of Nuclear Medicine, 967−972.
16. Rosenberg, D., 1984. The causal connection in mass exposure cases: 'public law' vision of the tort system, Harvard Law Review (HLR), 851−929.
17. Kaye, D., 1982. The limits of the preponderance of the evidence standard: justifiably naked statistical evidence and multiple causation. American Bar Foundation Research Journal, 487−516 and *see* Gold, S., 1986. Causation in toxic torts: burdens of proof, standards of persuasion, and statistical evidence. Yale Law Journal (YLJ), 376−402.
18. Shavell, S., 1985. Uncertainty over causation and the determination of civil liability. Journal of Law and Economic (JLE), 587−609.
19. *See* Landes, W., Posner, R., 1983. Causation in tort law: an economic approach. Journal of Legal Studies (JLS), 109−134 and Landes, W., Posner, R., 1987. The Economic Structure of Tort, 242.

20. Landes, W., Posner, R. (*supra* note 19), 124.
21. Shavell, S. (*supra* note 1), 115.
22. So Shavell, S. (*supra* note 1), 116.
23. So Bergkamp, L. (*supra* note 12), 290–291.
24. *See* this discussion on proportional liability Makdisi, J., 1989. Proportional liability: a comprehensive rule to apportion tort damages based on probability. North Carolina Law Review, 1063; Landes, W., Posner, R., 1984. Tort law as a regulatory regime for catastrophic personal injuries. JLS, 417–434 and Robinson, G., 1985. Probabilistic causation and compensation for tortuous risk. JLS, 797–798. For a discussion of the possible legal foundations of a proportional liability rule see Akkermans, A.J., 1997. Grondslagen voor proportionele aansprakelijkheid bij onzeker causaal verband. In: van Boom, W.H., Jansen, C.E.C., Linssen, J.G.A. (Eds.), Tussen 'Alles' en 'Niets'. Van toedeling naar verdeling van nadeel, 105–115.
25. *See* Estep, E. (*supra* note 9), 259–304. Very reluctant toward accepting statistical evidence in trials is also Tribe (Tribe, L., 1971. Trial by mathematics: precision and ritual in the legal process. HLR, 1329–1393), who argues: "mathematical evidence is more misleading than helpful."
26. Rosenberg, D. (*supra* note 16), 851–929; Shavell, S. (*supra* note 18), 587–609. Also the Dutch Attorney General Hartkamp defended a market share liability in the DES case ([1992] Tijdschrift voor Consumentenrecht (TvC), 241–258). In addition Spier pleaded in favor of a proportional liability for latent diseases in his inauguration address (Spier, J., 1990. Sluipende schade), as did Akkermans, A. (*supra* note 224) in his dissertation.
27. *See* Brüggemeier, G., 1991. Liability for Water Pollution under German Law: Fault or Strict Liability. In: van Dunné, J. (Ed.), Transboundary Pollution and Liability: the Case of the River Rhine, 88–91.
28. Robinson, G., 1985. Probabilistic causation and compensation for tortuous risk. JLS, 798.
29. *See also* Widmer, P., 2000. Causation under Swiss Law. In: Spier, J. (Ed.), Unification of Tort Law: Causation, 112–113.
30. Calabresi, G., 1977. The Costs of Accidents. A Legal and Economic Analysis, fifth printing. Yale University Press, New Haven, 24 ff.
31. Galanter, M., 1974. Why the 'haves' come out ahead: speculations on the limits of legal change. Law and Society Review (9).
32. This is to a large extent also due to the adversarial trial. Thus it is often forgotten that in a way parties also have concurring interests, being that the case is handled quickly and at relatively low costs.
33. Meadow, W. and Sunstein, C.R., 2001. Statistics, not experts. Duke Law Journal (51), 629–646.
34. For a summary of these studies, *see* Slovic, P., Fischoff, B. and Lichtenstein, S., 2000. Rating the risks. In: Slovic, B. (Ed.), The Perception of Risk, Earthscan, London, pp. 109–110.
35. Parker, J.S., 1995. Daubert's debut: the supreme court, the economics of scientific evidence, and the adversarial system. Supreme Court Economic Review (4), 1–56.
36. Parker 1995, 37, *supra* note 35.
37. Parker 1995, 35, *supra* note 35.
38. *See* Tomlin, J.T., Cooper, D., 2006. When should judges appoint experts?: A law and economics perspective. Bepress Legal Series (Working Paper 1699). http://law.bepress.com/expresso/eps/1699, p. 10.
39. *See* Sarvadi, D.G., Blackwood, A.L., May/June 2007. Expert testimony in the silica-cases: the fallacy of scientific objectivity – some observations. Journal of Chemical Health & Safety, 34.
40. Tomlin, J.T., Cooper, D., 2007. Expert testimony, technical advisors, and the determination of damages. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=902434, p. 11.
41. Mandel 1999, 113, *supra* note 1.
42. Sales, B.D., Shuman, D.W., 2005. Experts in Court. Reconciling Law, Science, and Professional Knowledge, American Psychological Association, Washington, pp. 6–7.
43. In the words of Thornton and Ward: "This tends to create strong if sometimes subtle pressure upon the economist to directly or indirectly advocate the position of the side that has hired him." Thornton, R., Ward, J., 1999. The economist in tort litigation. Journal of Economic Perspectives (13), 101–112, at 106.
44. Tomlin and Cooper 2006, 11, *supra* note 38.
45. 113 S Ct 2786 (1993).
46. *See* on this Daubert case law and the subsequent Supreme Court case law also "Reliable Evaluation of Expert Testimony" (note) (116), Harvard Law Review 2002–2003, 2142–2163 and Werden.
47. Tomlin and Cooper 2006, 5, *supra* note 38.

48. *In Re Silica Products Liability Litigation*, no. 2:05-CV-00121 (S.D. Tex. 30 June, 2005).
49. *See* for a discussion of this interesting case Sarvadi and Blackwood 2007, *supra* note 39.
50. These were based inter alia on an X-ray of the chest, a prior history of exposure to silica and the exclusion of other possible sources of the disease.
51. Posner, R.A., 1999. The law and economics of the economic expert witness. Journal of Economic Perspectives (13), 94–95.
52. Posner 1999, 94, *supra* note 51.
53. Thornton and Ward 1999, 108, *supra* note 43.
54. The use of lists does, however, not go undisputed. To the extent that courts would only appoint experts on the list and not others of equally good quality these lists could lead to an effective monopoly. We will discuss this problem in the next subsection. However, removing a bad expert from a list could provide appropriate incentives to the expert for a higher quality testimony, even when being party appointed.
55. *See* generally on self-regulation Ogus, A.I., 2000. Self-regulation. In: Bouckaert, B., De Geest, G. (Eds.), Encyclopedia of Law and Economics, vol. V, Cheltenham, Edward Elgar, 587–602 and Ogus, A.I., 1995. Rethinking Self-regulation. Oxford Journal of Legal Studies 15, 97–108.
56. *See* the contributions in Faure, M., Finsinger, J., Siegers, J., Van den Bergh, R. (Eds.), 1993. Regulation of Professions: A Law and Economics Approach to the Regulation of Attorneys and Physicians in the U.S., Belgium, the Netherlands, Germany and the U.K. Maklu, Antwerp.
57. Bosscha, N.L., 2001. Verkeersongevallenanalyse: een vak apart?, Asser, p. 16.
58. Nederlands Instituut van RegisterExperts (NIVRE).
59. Associatie van Belgische EXperten (ABEX).
60. Smith, A. An Inquiry into the Nature and the Causes of the Wealth of Nations, fifth ed. The Modern Library, New York, reprint 1937, 127–128.
61. Mandel 1999, 119, *supra* note 1.
62. Posner 1999, 94, 95, *supra* note 51.
63. Elliott, E.D., 1989. Toward incentive based procedures: three approaches for regulating scientific evidence. Boston University Law Review (69), 492.
64. Elliott 1989, 501 ff, *supra* note 63.
65. Posner 1999, 93, *supra* note 51. *See also* Lee, T.V., 1988, 484, 488, *supra* note 48.
66. *See* Posner 1999, 96, *supra* note 51.
67. Alemanno, A. Science and Risk Regulation: The Role of Experts in Decision-Making and Judicial Review. In: Vos E. (Ed.), European Risk Governance Its Science, Its Inclusiveness and Its Effectiveness, CONNEX Report Series Nr 06, p. 65 ff.

PART II

AUXILIARY FORENSIC DISCIPLINES

This page intentionally left blank

CHAPTER

6

Forensic Pathology

A. Eriksson

Umeå University, Umeå, Sweden

OUTLINE

INTRODUCTION

Forensic medicine mainly deals with examination and assessment of individuals who have been—or are suspected to have been—injured or killed by external influence such as trauma or intoxication, but also of individuals who are suspected of having injured another person. This means that not only victims and suspects of crime, but also suicidees and accidental fatalities are examined by a specialist in forensic medicine (or forensic pathology). Individuals

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00006-9

with nonfatal injuries after intentionally self-inflicted or accidental injuries or intoxication are, on the other hand, usually handled exclusively within the health-care system. In many countries, forensic medicine represents a medical specialty within the legal system, not within the health-care system.

Forensic pathology is the part of forensic medicine dealing with examination of deceased persons, and this is the focus of the present chapter. In the following, some general principles of the work in forensic pathology are presented. Although the legislation regarding forensic pathology differs between countries, a common principle is that in the investigation of a possible or suspected criminal death, a forensic pathologist is engaged through a formal request from the police or the prosecutor. The task of the forensic pathologist is then to assist in the investigation as a medical expert. This expert role continues throughout the process, including the court proceedings on request of the court and/or one of the parties.

The task is to function as a medical expert for justice, not primarily to support one of the parties in the trial. Hence, the role of the forensic pathologist in the relation to the examined person is obviously completely different from the role of the clinical doctor in his/her relation to the patient, where the physician often becomes an advocate for the patient. The main role of the forensic pathologist is to practise and to mediate a scientific approach to the medical issues raised in a legal context involving death. It is inherent in its very nature that the forensic pathologist, irrespective of principle, strives to assist with impartial assessments, based on "science and tried and tested experience."

The formal organization of forensic medicine and the experts in forensic medicine is somewhat different in different countries. In central Europe, eg, the medicolegal experts are recruited from a university since this has been believed to guarantee a scientific basis, independence, and impartiality. In Sweden and Finland, a national governmental authority is responsible for the administration of services in forensic medicine, whereas in the US, Canada, and several other Anglo-Saxon countries, a variety of systems are applied under the umbrella terms "coroner system" and "medical examiner system," systems that are not always easy to differentiate.

CAUSE AND MANNER OF DEATH

The *cause of death* is the disease or external cause leading to death. The first relevant disease or injury leading to death is called the *underlying* cause of death, which, sometimes through an *intermediary* cause of death, gradually leads to the *terminal* cause of death. These causes are through international agreements arranged in a death certificate in a certain order (1a–c) as illustrated by the following example.

Example of a death certificate includes:
1a/Pulmonary embolism (*terminal* cause of death)
1b/Immobilization (*intermediate* cause of death)
1c/Hip fracture (*underlying* cause of death)
2/Bronchopneumonia (*contributory* cause of death)

In the case above, an external cause (trauma with hip fracture) leads to death from a disease process (deep vein thrombosis with embolism to the lungs). Any *contributory* cause to death is listed under section 2 of the death certificate, eg, 2/bronchopneumonia. In this context, it should be noted that the death certificate is not a list of *all* the decedent's diseases, but a list restricted to what has caused or contributed to death.

Manner of death designates whether the death has been caused by disease (manner = "natural death") or, if caused by an external cause ("unnatural death"; see below), the type of intent (*accident* if no intent, *suicide* if own intent, *homicide* if another's intent). Another manner in this section is *act of war*. If the manner cannot be determined, eg, if it cannot be determined whether a lethal overdose of medicine was intentional (suicide) or accidental (unintentional), the box *undetermined manner of death* is ticked on the death certificate. The same is the case if it cannot be determined whether a fatal stab wound was self-inflicted or inflicted by another person. *NB*: "Homicide" is in the US often defined somewhat differently than in European countries, including some cases which in other countries would be designated *accident*.

The manner of death indicated on the death certificate is the decision of the forensic pathologist, not to be used in a legal process but solely for the purpose of the death statistics. In some countries or jurisdictions, however, the manner of death is decided by a legal process, eg, through a so-called inquest.

Autopsy, Medicolegal

The *medicolegal* or *forensic autopsy* is performed at the request of police, prosecutor, or court by a forensic pathologist—usually in unnatural (violent) deaths, in otherwise sudden unexpected deaths, and in some unwitnessed deaths. The main purposes of a medicolegal autopsy is to reveal the cause of death for the legal system, and in criminal deaths to collect trace evidence and other evidence in order to provide information to reconstruct and to interpret a chain of events, and in some cases to illustrate these findings in a court of law. Other purposes include identification of an unidentified body, and documentation and evaluation of suspected medical malpractice.

In many countries, the main groups subjected to a medicolegal autopsy are not only known and suspected homicides, but all (or most) types of unnatural deaths, such as accidental deaths (eg, traffic deaths, workplace deaths, drownings, falls) and suicides (any method), but also many cases where the decedent is found dead (unwitnessed death) and where there is no known history of (a fatal) disease. Also, unidentified bodies and suspected medical malpractice deaths are usually subjected to a medicolegal autopsy. In many Western countries, the share of such deaths is around ~5% of *all* deaths.

NB: A domestic fall of an elderly person resulting in immobilization and delayed death at an institution will in most countries *not* be subjected to a medicolegal autopsy. This category of deaths forms a substantial part of the unnatural deaths in many countries with an aging population.

The medicolegal autopsy is in most cases supplemented with microscopical examination of inner organs, with toxicological analyses of body fluids, and in special cases ancillary investigations by collection of trace evidence, and samples for DNA analyses, for microbiological

analyses and/or forensic evidence analyses. Further, imaging techniques have been used for more than 100 years as an adjunct to the medicolegal autopsy, mainly in the investigation of unnatural deaths such as the localization of fractures or firearm projectiles, but also for the reconstruction of the event and for presentation in a legal process.

More than occasionally the medicolegal autopsy reveals unexpected findings. In Sweden, for example, among ~5500 annual medicolegal autopsies performed, approximately 5–10 previously unsuspected homicides and ~5 previously unknown suicides are disclosed each year. Furthermore, ~30 suspected homicides (out of a final total of <100) and ~70 suspected suicides (out of a final total of ~1200) are classified as accidents or natural deaths (Hasselqvist and Rammer, 2003; Janko and Druid, 2009).

Autopsy, Clinical

The *clinical autopsy* is usually performed in a hospital setting by a clinical pathologist, not by a forensic pathologist. The main purpose is, usually with the permission of the next of kin, to learn more about the disease(s) for which the decedent was treated, including the cause of death. Hence, the clinical autopsy is in most cases supplemented by microscopical examination of inner organs.

The autopsy has for long been regarded as the "gold standard" for retrospective quality assessment of clinical diagnoses (Graber, 2005), and studies comparing clinical diagnoses and autopsy findings have revealed major discrepancies in one-fourth or more of deceased patients undergoing autopsy (Shojania et al., 2003; Wittschleber et al., 2012; Kuijpers et al., 2014).

Some 50 years ago, the rate of clinical autopsies in Europe and the US was around 60%, but since then, clinical autopsy rates have drastically declined, and is today less than 10% (Charlot et al., 2000; Roulson et al., 2005; Shojania and Burton, 2008). The reasons for this decline are manifold and include the nonreimbursement of autopsies, the clinician's fear of medicolegal problems, reluctance on the part of the family and/or of health-care personnel, increasing number of deaths occurring at long-term care facilities, shortage of clinical pathologists, judicial principles, inadequate and delayed communication of autopsy results to clinicians, the requesting of autopsies being delegated to junior medical staff, organ retention issues, and advances in laboratory diagnostic technology and imaging techniques that result in the erroneous belief among some clinicians that the autopsy has become redundant.

One of the few situations that are working in the opposite direction (toward performing an autopsy) is when the family of a decedent requests an autopsy as a basis for grounds for a malpractice suit.

Low autopsy rates may conceal medical malpractice, thereby preventing an important quality assurance indicator in health care. Further, the decreased reliability of the cause-of-death statistics in turn decreases the usefulness of the statistic for health-care planning and research. The family of the unautopsied decedent is provided with incorrect or insufficient information regarding cause of death, the underlying disease, etc. Hence, the decrease in clinical autopsy rate has negative consequences for the family, for future patients, for health care, and for the society as a whole.

In recent years, the improved quality of postmortem imaging techniques has led to the increasing consideration of such noninvasive techniques as a substitute for the clinical autopsy, or that at a minimum the technology can contribute to a more reliable cause-of-death diagnostic process than the strictly clinical process, with its obvious shortcomings. The diagnostic accuracy of these new techniques is, however, not yet scientifically established, and as yet there is insufficient evidence to consider these techniques as an adequate replacement for conventional autopsy (Eriksson et al., 2015).

Autopsy, General

In both types of autopsy the same basic technique is applied. However, in the medicolegal autopsy more emphasis is laid on the external examination of the dead body than in the clinical autopsy where the background history of the decedent is known and where the death usually (in Western countries often more than 80%) has occurred in a controlled environment such as a hospital or another institution. Further, the medicolegal autopsy is characterized by a more detailed dissection in the search for injuries—eg, fractures—than in the clinical autopsy where injuries often are absent or otherwise previously identified via antemortem radiographic examination.

In both types of autopsy, the body is first subjected to an *external examination*, in which any injury is documented. In the following *internal examination*, the body cavities are opened, the inner organs are eviscerated and dissected—including the brain, the thoracic organs, the abdominal and the pelvic organs. Any injury or disease process is documented and adequate samples from organs, body fluids, and/or other elements are analyzed according to the nature of the case.

Further dissection is performed whenever necessary, eg, to expose skeletal structures when a fracture is suspected or needs to be excluded, to examine the spinal cord, the sinuses, or soft tissues.

The findings are documented verbally in an autopsy report, photographically, radiographically, etc., depending on the specific case. When the results of the ancillary investigations are present, the autopsy report is finalized in its conclusions.

DIFFICULTIES IN DETERMINING THE CAUSE AND MANNER OF DEATH

The difficulty in determining the cause of death without an autopsy is touched on above. But even when an autopsy has been performed it is sometimes difficult or even impossible to establish the cause of death. This is easily understandable if only body parts, such as parts of a skeleton, is available, or if the dead body is in advanced state of decomposition. Even a complete autopsy performed on a "fresh" body is by no means infallible in revealing the (true) cause of death, however. Such "obscure autopsies" are more common in younger decedents, and one well-known such group is "sudden infant death syndrome" (SIDS), defined by (among other things) the absence of pathological findings. Among young adults a negative or inconclusive autopsy is not uncommon as

well. It is suggested that some of these deaths result from genetic aberrations leading to a fatal cardiac arrhythmia. In contrast, in older individuals, pathological findings are almost always found which *can* explain the death, but from a logical point of view some of these deaths may also have "invisible" causes, similar to deaths among younger age groups.

In order to arrive at correct conclusions regarding the true cause of death, information on the medical and social background history of the decedent, medication, substance abuse, etc., is invaluable. Information is also needed regarding the last day(s) alive and on the circumstances of death, including observations on the scene where the decedent was found. As an example, the macroscopical and microscopical autopsy findings in the death of an addict and in a death of a person with a cardiac disease (cardiomyopathy) may be virtually identical. But a detailed background history will initiate further investigations, including toxicological analyses, helping to identify the correct cause of death. In some cases, however, the background history may be overvalued. For example, since it is known that epilepsy may be associated with sudden death in an otherwise healthy person, the diagnosis of epilepsy in the absence of macroscopical findings may be erroneously blamed for the death. In such a case the epilepsy would be more correctly classified as *a possible cause of death*, not *the cause of death*.

A death in a road traffic event is often automatically classified as a "traffic accident," but a considerable share of traffic deaths are in fact caused by disease, almost always due to sudden onset of cardiac disease with a fatal ventricular arrhythmia (Öström and Eriksson, 1987). And some 3–5% of traffic deaths are in fact suicides and even occasionally homicide (using the European definition that includes intent) (Ahlm et al., 2001). In the classification of the manner of death, it is obvious that the autopsy findings can, at best, give an indication of the correct manner. Background and recent history are crucial components in this classification, and shortage of information may result in *undetermined manner of death*.

Classification is also dependent upon legislation and culture. As an example, in a culture or even in a family where suicide is a taboo, the physician may feel forced to assign another cause and/or manner of death than what he/she actually believes is correct. Such inaccuracies aggravate an already difficult comparison of death statistics between countries.

Not only absence of macroscopical or microscopical or toxicological or other important findings may cause problems in determining the cause of death. Sometimes more than one finding may have caused the death, which means that there are *competing* causes of death. For example, the autopsy findings may include both a fatal brain injury and a fatal rupture of the aorta. In this specific case the cause of death can be chosen on formal grounds since in many countries the death of a human being is defined as total and irreversible loss of brain function, and thus the cause of death in the example would be *brain laceration*. But, if the autopsy findings include both a fatal rupture of the aorta and a fatal rupture of the liver, the choice is not quite as easy. Based on the (estimated) swiftness of the exsanguination, one may choose the cause of death as rupture of the aorta. In other cases the choice may be arbitrary or impossible. If a body is totally disintegrated when run over by a train, the cause of death may be assigned "multiple injuries" or *dilaceratio corporis totalis* in Latin.

Second Opinion

In some cases, the next of kin, the prosecutor, or another interested party may want to challenge the findings—or more often the conclusions—in the autopsy report. In most cases, the autopsy findings are not controversial, whereas the conclusions drawn from the findings are more often subjected to individual preferences and variations. In such cases, an independent expert may be engaged to give a "second opinion." In order to be fully respected by the mandator, it is of course for the second expert—likewise as for the primary investigator—of utmost importance to be impartial and base the conclusions only on "science and tried and tested experience." An expert who deviates from this protocol will regress to the kind of "hired gun" partisan expert that the reader is warned about in Chapter 5, *The Role of the Expert Witness*.

The system for obtaining a "second opinion" differs between countries. In some countries there are formal review boards for the assessment of medical expert statements, but in most countries the individual may seek the second expert on the open market, sometimes abroad. The latter is of course not the best prerequisite for finding an impartial expert, particularly since the layman in most cases cannot appreciate the correctness of a medical expert assessment and its scientific basis, and even more so since "expert shopping" may disqualify the expert through bias. Disclosure of conflicts of interest is therefore of importance. An additional problem in this context is that the court in most cases cannot interpret medical expert statements better than any (other) layperson in medicine.

Time of Death

In some cases, particularly in homicides, determination of the time of death can be helpful in further investigation. Description of the different methods of determining time of death is beyond the scope of this chapter, but in short the best method during the first 24-h postmortem is to measure the central core body temperature, combined with some tests of muscular irritability. Lividity (bluish red discoloration of the skin) and rigor (stiffening of the muscles) are findings that are always checked, but which can give only very rough estimates of the time of death. In the days following the death, chemical tests of vitreous fluid from the eye may aid in the determination, and in late stages analysis of insect activities and other external factors may be of importance, depending on the environment in which the body was found. The determination of time of death is however a huge problem if based solely on medical findings, and the accuracy is, at best, not better than $\pm\sim 2$ h from the actual time of death (Madea and Brinkmann, 2003).

In some cases, other medical findings may help. For example, if an injury was caused by a fatal epidural hematoma associated with skull fracture, the alcohol concentration in the sequestered hematoma may, compared to the concentration in blood from the vascular system, give a rough estimate of the time span between trauma and death, and consequently if the time of injury is known, also of the time of death.

Consequently, determinations of time of death based on medical findings must always be considered in the light of the limitations of the methods, and must always be correlated to observations at the scene, including newspapers, digital communication, and other dated material.

NATURAL DEATHS

Annually, ~1% of the population dies in Western countries. Most of these deaths are due to disease, hence called *natural deaths*. Around 4% of these die suddenly and unexpectedly—*sudden natural death*. "Sudden" refers to the time span from the first serious symptoms to death, in most cases defined as $\leq$24 h.

Unnatural deaths are defined as deaths due to something else than disease only; typically an external cause or combination of external cause and disease. An external cause implies in most cases trauma and/or intoxication, but includes also hypothermia, hyperthermia, radiation, etc. See more discussion below.

In Western countries, around 95% of all deaths are natural, and most of these deaths occur in individuals who already are in contact with the health-care system. About 10% of all natural deaths occur suddenly, and some of these deaths may give the impression of being unnatural—which may lead to a medicolegal autopsy.

Among infants (up to the age of 1 year), most deaths occur during the first month and are caused by prematurity and malformations. At 2–4 months of age one important cause of death is sudden infant death syndrome (SIDS), the explanation of which is still unknown in spite of many theories. This syndrome is unique in the sense that its definition depends on the lack of findings in the medical history, in the scene investigation, and in a meticulous autopsy including ancillary investigations. In other words, it is a diagnosis of exclusion. In many countries the rate of SIDS cases was dramatically increasing during the 1970s and 1980s, but after the campaign "back to back," when the prone sleeping position of the child was abandoned, the incidence immediately dropped significantly (see Chapter 13, *Product Defect/Liability Investigation*). The actual mechanism behind this dramatic change is, however, unknown. Known risk factors are smoking in the household, hyperthermia, and bed sharing with an adult.

An alternative cause of death in this age group is asphyxia; however, the findings in some types of asphyxia may be very discrete, even absent, and thus impossible to differentiate from those in SIDS. In addition, asphyxia as a cause of death is also a diagnosis of exclusion—since findings in fatal and nonfatal asphyxia are identical. Other causes of death in this age group, which are also extremely difficult to diagnose based on the medical findings alone, are hyperthermia and hypothermia.

Up to the age of 15 years, natural deaths are rare in the Western world. Deaths due to malignancy, cardiac malformations, and infectious diseases do, however, occur. Among young adults, up to the age of 40, there is a gradual increase in death rate, mainly due to cardiovascular diseases.

Among the middle aged and elderly, there is a dramatic increase in the rate of natural deaths, mostly due to cardiovascular diseases, especially coronary heart disease with its complications like myocardial infarctions and other ischemic manifestations. Second, malignancies cause an increasing number of deaths in both sexes, but also cerebrovascular, infectious, gastrointestinal, pulmonary, and alcohol-related diseases.

DIFFICULTIES IN DIFFERENTIATING BETWEEN NATURAL AND UNNATURAL DEATH

Deaths among alcoholics are often unwitnessed, and the autopsy is often inconclusive, revealing perhaps only liver steatosis, a slightly enlarged heart, and a low or moderate blood

concentration of alcohol. The underlying mechanism is likely a fatal cardiac dysrhythmia caused by chronic effects of ethanol upon the myocardium, the so-called alcoholic cardiomyopathy, and electrolytic alterations caused by abuse may contribute as well. In some alcohol-related deaths, a straightforward cause of death is found, however. These cases include gastrointestinal hemorrhage due to ulceration of the stomach or duodenum, lobar pneumonia, complications of diabetic disease, intoxication, and various forms of trauma.

The death of an alcoholic (or abuser of other substances) may thus be caused by a number of naturally occurring disease processes, but also from various unnatural, external causes—and of course also by combinations of these conditions and circumstances. This is the explanation for why an unwitnessed death of an addict should be reported to the police and why a medicolegal autopsy should be performed in all such cases.

The demarcation between a natural and an unnatural cause of death is not always an easy task. A common situation is that an elderly individual slips or trips, sustains a fracture of the proximal femur, is immobilized (without or after surgery), and finally succumbs to a fatal complication from, for example, pneumonia, deep vein thrombosis with pulmonary embolism, or cardiac decompensation. In addition, there may be other, perhaps preexisting, disease(s) which can further complicate the assessment of the cause of death, maybe also the manner of death.

Another example of the difficulties in the assessment is if a decedent has a history of a brain injury and subsequently develops posttraumatic epilepsy, a condition which may be associated with or even cause death decades later. A physician who is unfamiliar with the traumatic cause of the epilepsy will probably classify such a death as natural, while in fact the underlying cause is unnatural.

Still another example is a death due to allergic reaction (anaphylactic shock), caused by a wasp sting. A forensic pathologist would classify such a death as unnatural, whereas a clinician would likely consider it as a natural death.

UNNATURAL DEATHS

Accidental Death, Suicide, or Homicide

Even after a meticulous death investigation, the manner of death can remain unclear in some cases. This is particularly true when potential witnesses were under the influence of drugs or otherwise impaired at the time of the death and thus cannot accurately relate the sequence of events. An example of such a circumstance would be a drug abuser found dead with a single stab wound located in a part of the body that does not clearly indicate accidental death, suicide, or homicide.

Another example is the difficulties sometimes present in the assessment of various forms of asphyxia. This includes circumstances where it is unclear if the death has been caused by hanging or strangulation, and thus in most cases whether the death represents a suicide or a homicide.

In a situation where a young woman is found hanged but where an honor killing could be suspected, neither medical nor other findings can resolve whether she hanged herself voluntarily or whether she was forced to commit suicide. If there are physical injuries or other findings indicating a struggle, use of bonds, or drugging, a suspicion of homicide is, of course,

strengthened. The medical and technical scene investigation can also contribute to the assessment of the cause of death.

In this context, it should be noted that asphyxial deaths are difficult to assess since petechiae in the eyelids, conjunctivae, etc., and skin injuries on the neck, etc., are identical in fatal and nonfatal asphyxiation. Asphyxiation is thus a cause of death by exclusion, which means that there must be no *competing cause of death* for the diagnosis of asphyxiation to be finally concluded. For obvious reasons this means that may be difficult, if not impossible to decide if the cause of death is hanging if the alternative cause of death is another form of asphyxia, such as strangulation.

The scientific method of *forensiometrics* (Karlsson, 1997) can in some cases contribute to the assessment of the probability of how a certain injury was inflicted. This is a similar investigation technique as that described in Chapter 15, *Criminal Investigation* as injury pattern analysis. In a case of the stab wound, hesitation injuries, linear scars over the nondominant wrist, a suicide note, and the fact that the decedent was found in his/her own home, are findings that favor a self-inflicted injury. In contrast, cut injuries to clothing, defensive injuries, and other or unexplained injuries may favor foul play. These findings, which tend to increase the probability of a homicide versus suicide, may play an important role in a forensic epidemiologic assessment of the manner of death. The court's requirements of evidence that proves guilt "beyond a reasonable doubt" means that such probabilistic evidence has more utility when used by the defense to cast doubt on the prosecution's case rather than as a basis for conviction. Again, see the discussion on the applications of forensic epidemiology in Chapter 15, *Criminal Investigations*.

Another type of problem occurs when the cause of death is undisputed, but where the injury could have been inflicted either intentionally or unintentionally. For example, a fall resulting from an accidental occurrence or from a push can present with identical physical findings. A push normally does not result in a visible injury and can consequently not be identified through a medical examination, but even if does, such an injury may be impossible to distinguish from the injuries caused by the fall. There is an increase in the number of deaths of young women falling from buildings and where a crime of honor is suspected. A medical examination may, in some cases, contribute to an assessment of the probability of a nonaccidental fall; if, for example, evidence of a fist blow is found. Again, uncertainty of the meaning of some findings on autopsy implies the necessity of collateral evidence in the investigation.

The scenario in which a drug-intoxicated person dies suddenly after an altercation with the police is another example of uncertainty regarding the manner of death, but in these cases this relates to uncertainty regarding the *cause* of death. Some of the main alternative causes of death in these cases are *excited delirium* (induced in most cases by stimulants such as cocaine and methamphetamine), *positional asphyxia* (induced by a body position hampering breathing), or *traumatic asphyxia* (induced by hampered breathing due to the weight of another person on the chest). The often well-founded uncertainty regarding the cause of death in such situations expressed by the forensic pathologist is sometimes criticized as an attempt to cover up police brutality. It is, however, important to recognize that the limitations of forensic pathology to divine the cause of death in all cases does not equate with a partisan position favoring a particular theory of how a death occurred.

Accidental Death

The three main groups of manner of unnatural death are *accidental death, suicide*, and *homicide*, and when the manner cannot be determined, the manner is classified as *undetermined*. The manner of death gives information about the type of death, but gives in most cases no information about the cause of death.

One definition of "accident" is *a sudden, unintentional event with an external cause, which results or may result in damage or injury*. "Accidents" thus includes events with or without involvement of a human being.

An example where the classification of manner of death may be disputed is when a known drug addict is found dead due to an "overdose" of an illicit drug. Usually, he/she has not administered an overdose, but rather an "ordinary" dose of a drug with a high risk of fatal complications. Such a death is usually classified as an "accident," since the decedent did not intend to commit suicide. Others will argue that, since the drug was intentionally administered, it is a *suicide*. Still others may argue that is a suicide because of the general self-destructive behavior of the drug addict. If another person administers the drug to the addict who subsequently dies, the manner *homicide* may arise.

So-called "Russian roulette" is another example where it can be argued whether the death should be classified as an "accident" or as a "suicide" because of the extreme risk associated with the undertaking. Some might make the same claim regarding skydiving or hang gliding; however, a death resulting from such activity would never be considered a suicide. A death resulting from risky autoerotic practice by various forms of asphyxia is usually classified as an accident since death was an unintentional side effect—in spite of the intention to generate cerebral hypoxia and the inherent risk involved.

In a fatal single vehicle crash, the driver victim may be found to have moderate or severe coronary artery disease. Without eyewitness information, it may be difficult to determine if the cause of death was natural, or if caused by the crash-related injuries. Or perhaps the injuries are fatal, but the cause of the crash is the coronary heart disease, ie, the terminal cause of death is unnatural but the underlying cause is natural. Obviously, every piece of information may be of importance for a correct classification, and some of the methods described in Chapter 11, *Traffic Injury Investigation* and Chapter 12, *Traffic Injury Investigation: Product Defects* of this test can be helpful in making determinations in difficult cases.

It must be emphasized that the classification made by the forensic pathologist is by no means a "legal decision"; a prosecutor may prosecute a driver for causing another's death even if the death is classified as an accident—and a prosecutor may refrain from an indictment even if a death has been classified by the pathologist as a homicide. Further, in civil actions, insurance companies make their own judgments, typically relying on their own experts, and do not necessarily follow the classification by the pathologist when it is adverse to their interests.

There are many other similar quandaries regarding manner of death—for example, if a patient expires from ventricular fibrillation in connection with the introduction of a stent introduced into a stenosed coronary artery. Is this a mishap because of the surgical procedure, ie, an accident, or a natural death because of the underlying disease?

The results of a medicolegal autopsy can, together with other information, be used to identify fatal injury mechanisms and design technological means of reducing or mitigating similar

events in the future. Prominent examples of safety advancements informed by medicolegal autopsy are safety belts and air bags in passenger vehicles. Other examples are the introduction of window hooks to home safety, which had a dramatic effect on the incidence of accidental falls out of a window; the strengthening of the A-pillars in passenger cars, which in turn decreased the incidence of fatalities in moose–car collisions in Sweden; addition of a protective rollover bar in tractors, which reduced the number of deaths in rollovers; and the identification and banning of particularly dangerous drugs available on the Internet (eg, "Krypton") (Kronstrand et al., 2011), to name just a few.

A factor common to accidents, accidental injuries, and accidental deaths, is the presence of alcohol and other drugs. This applies not only to traffic events, but all potentially injurious events. The importance of alcohol and drugs in causing injury and death varies and relates to sociocultural differences. Although we generally associate alcohol and fatal traffic crashes, it is victims in boat accidents and associated drownings, as well as deaths in snowmobile and all-terrain vehicle–related events who have the highest share of alcohol-intoxicated drivers/riders—60–80%. At the other end of the spectrum, workplace deaths have a 2–4% or less intoxication rate (Sjögren et al., 2000).

ACCIDENTAL TRAFFIC DEATHS

Investigations of fatal traffic events have several objectives, such as establishing cause and manner of death for criminal and insurance purposes, to reconstruct the pattern of injuries and the sequence of events in order to find further preventive measures (disease? drunken/drugged driving?), and to provide information to the next of kins, as well as feedback to the health-care system regarding treatment of previously diagnosed disease. Additional information on the investigation of crash-related death, and the utility of FE in such an investigation, is described in Chapter 11, *Traffic Injury Investigation* of this book.

Some of the questions raised during the autopsy of a traffic death are as follows:

- Did the injuries occur prior to or after death? (Especially important in multivehicle crashes)
- Are the injuries from a primary impact, run-over, or secondary impact with the ground? (Pedestrian crashes)
- What was the position of the decedent in the vehicle? (This is expanded in Chapter 15, *Criminal Investigation*)
- Drunken and/or drugged driving?
- Is there trace evidence connecting the decedent to the culpable vehicle?

The cause of death in traffic crashes in most cases due to high-energy blunt force trauma, but sharp force trauma, burns, chest immobilization (positional or compression asphyxia), aspiration of blood, and drowning are examples of other mechanisms. In blunt trauma deaths, the lethal injuries are most often found in the head or the chest, but sometimes the injuries are massive and the cause of death is assigned "multiple injuries."

DRUNKEN DRIVING

One well-known and important causative or contributing factor in traffic crashes is alcohol, more specifically ethanol which has a hypnotic and anesthetic effect upon the central nervous system. Judgment and motor skills are negatively affected already at even low blood

concentrations of ethanol, and at a blood alcohol concentration (BAC) of 0.15%, a car driver is calculated to run a risk of being injured in traffic which is 300–600 times higher than that of a sober driver. A great share of drunken drivers has a chronic alcohol problem, which must be considered in the preventive work. The legal limits differ between countries, often 0.02–0.05%, with only few countries having a zero tolerance.

DRUGGED DRIVING

Substances other than alcohol can affect driving capability as well. The effect of both licit and illicit drugs upon the driving capability is, however, far more unpredictable than that of ethanol, and as a consequence the laws regarding this vary greatly between different countries.

Suicide

The suicide method used reflects above all the availability of the means by which the suicide is committed. This is sometimes related to the profession of the suicidee, but also the popularity in media, especially from information accessible on the Internet, through which suicide methods can spread widely and rapidly. Males are known to use more violent methods such as shooting, hanging, and intentional traffic crashes, whereas women more often use intoxication and drowning.

HANGING

Hanging is a very common suicide method, especially among men. Contrary to some popular belief the mechanism is, in most cases, not a fracture of the cervical spine, but external pressure upon the carotid arteries, hence blocking the blood supply to the brain. The former mechanism is associated with judicial hangings, in which the body is dropped from a higher level.

The differential diagnosis to consider in some cases is *accidental* hanging, as in autoerotic asphyxia. *Homicidal* hanging is very rare, but can be suspected if the body shows signs of other trauma or if a very high concentration of a drug is detected in the toxicological analysis.

SHOOTING

Shooting is a common suicide method among males, even more so in countries with liberal weapon legislation (eg, the US), and in rural areas where hunting rifles/shotguns are more readily available. A suicidal shot is most often located to the head, more seldom to the chest, and the cause of death is brain laceration and exsanguination, respectively. A *homicidal* shooting can in some cases be difficult to differentiate from a suicide, but depression, a suicide note, and previous suicide attempts are factors which may favor suicide in combination with a meticulous scene and autopsy investigation. An *accidental* shooting injury may be located to any body region and a reconstruction can be an important tool to conclude the manner of death. In some cases the manner (suicide vs homicide) may not be clear from the physical evidence.

INTOXICATION

Intoxication is in many countries the most commonly used method among women. Prescribed psychopharmaceuticals such as neuroleptics and antidepressive agents are commonly used, and may indicate that the suicidee has a history of psychiatric disease, often depression,

or bipolar disorder. The differential diagnosis to consider in these cases is *accidental* intoxication, particularly if the decedent is a drug addict.

DROWNING

Drowning is also a suicide method used by women most commonly; more often so when open water is available. The obvious differential diagnosis to consider in these cases is *accidental* drowning.

SHARP FORCE TRAUMA

Suicide through sharp force trauma is uncommon, although many persons attempt suicide by cutting themself over the radial artery in the nondominant wrist. The method is, however, doomed to fail in most cases since the artery is small and only exceptionally can lead to a fatal exsanguination. To succeed with this method, a larger artery such as the brachial, the femoral or the carotid, must be opened. A stab wound into the heart also has a high risk of lethality, and is sometimes used. A suicide by stabbing or cutting can often be distinguished from a *homicide* through the presence in the former of so-called hesitation injuries, ie, multiple superficial, parallel cuts close to the deeper and fatal injury. The differentiation between suicidal and homicidal cut and stab wounds is not always straightforward, however.

BLUNT FORCE TRAUMA/JUMPING FROM HEIGHT

Suicidal blunt force trauma is uncommon, and almost always requires a high degree of potential energy. Examples are intentionally crashing the car into a heavy or immobile object, jumping or lying before a moving object such as a train or another motor vehicle, or jumping from a high place. Jumping is not uncommon and to some extent reflects that the suicidee lacks other means of readily committing suicide. The popularity of a locale is also a factor to consider; some locations have afforded many the opportunity to commit suicide through the years, such as the Golden Gate Bridge and the Empire State Building, leading to preventive measures such as high barriers that are difficult to scale. A *homicidal* fall from, eg, a balcony, may represent an honor killing, and this fact emphasizes the importance of investigating the decedent's background. A high blood alcohol concentration may indicate an *accidental* fall.

FIRE

Self-inflicted injury through fire, burns, and scalds is uncommon, but is seen more commonly during periods of political protest against an ongoing war or enemy occupation.

Problems in Determining the Manner of Death

As mentioned above, there may be great difficulties in determining if a case of intoxication represents a suicide or an accident. A suicide note, depression, previous psychiatric treatment, and previous suicide attempts are factors which may indicate suicide. Pharmacogenetical findings may, on the other hand, show that the victim of a fatal intoxication was a "slow metabolizer" and that he/she had accumulated the (prescribed) drug in the body over time,

in spite of taking only a dose considered normal for the majority of the population—indicating that the death was accidental rather than suicidal. The ratio between the concentration of the mother substance and its metabolites may also indicate whether the high intake was recent or had accumulated over time. Renal insufficiency is another factor which may lead to an unintended high concentration of a drug, licit or illicit. And as mentioned above, a drug addict may accidentally take an overdose with fatal outcome.

In fatal traffic crashes, it is sometimes difficult to exclude suicide. The literature indicates that the share of suicides in traffic fatalities in the Western world may be around 2–6% of all traffic deaths, but some have expressed the belief that the true figure is closer to 10%. Examples of when the suspicion of suicide is raised are a high-speed frontal collision with a heavy vehicle in the opposite lane of travel on a straight segment of the road and in daylight, and a high-speed collision with a mountain face. The influence of alcohol or drugs, as well as falling asleep, or even an acute disease process, may cause the same situations, however. This means that a meticulous background check, a detailed autopsy, and an extended toxicological analysis is of utmost importance.

In cases where the manner of death is uncertain, the case is classified as *undetermined manner of death*, but of course this group contains a number of suicides. The precautionary principle requires, however, a solid ground for the classification, which in this case means that suicides must not be overdiagnosed, particularly since this classification may have economical and social consequences for the next of kin, but of course also for a sound scientific approach to the cause-of-death statistics.

Homicide

In many countries, homicide is defined as the act of intentionally causing the death of another human being. In some jurisdictions in the US, however, also certain unintentional events resulting from negligent conduct are included in the definition; events that in the other jurisdictions would be classified as "accidental deaths". Different types of homicides are treated differently in human societies and may include murder/manslaughter, euthanasia, judicial execution, and act of war.

The incidence of homicide depends not only upon the definition applied, but more so on sociocultural differences between countries and between regions of the same country. In general, big cities tend to have a higher incidence than sparsely populated rural areas with a higher degree of social control. Globally, the annual number of homicides has been estimated to around 0.5 million and another 0.5 million in acts of warfare. The much lower incidence in Europe as compared with the US has been attributed to differences in weapon legislation. In countries with a low availability of firearms, robbery with murder is rare, but naturally more common in countries with a surplus of firearms. In many Western countries, the incidence of homicide has decreased during the last decades.

METHODS OF HOMICIDE

SHOOTING The incidence of homicide shows some correlation to the availability of firearms, but also other characteristics may differ. In the US, for example, many victims of homicide by gunshot are young men and the firearm is most often a handgun. In the Nordic

countries, on the other hand, shooting deaths are generally uncommon, although more prevalent in gang-related crimes.

To distinguish a firearm homicide from a firearm suicide is sometimes not easy. The location of the entrance wound may help; an entrance wound in the mouth is very rare among homicides, for example. Also the shooting distance may help, as contact wounds are most common among suicides, but rare among homicides. What is true at group level is however not always "proof beyond reasonable doubt" in an individual case, a common limitation of both forensic pathology and forensic epidemiology.

SHARP FORCE TRAUMA Both the incidence and the methods of homicide vary between cultures. In countries where the availability of firearms is low, cutting/stabbing is a common method, sometimes the most common, constituting up to half of all homicides.

ASPHYXIATION AND FATAL PRESSURE ON THE NECK This category includes—among other terms—exclusion of oxygen in the breathing gas, impaired breathing by chest compression or body position, blocking of the outer airway by smothering or gagging, blockage of inner airways (choking) by a piece of food or a foreign body, and aspiration of stomach contents, blood or water (drowning). The primary fatal mechanism in most of these cases is impaired breathing.

Fatal pressure on the neck includes manual and ligature strangulation, arm-lock holds, and hanging. The primary fatal mechanism in most of these cases is occlusion of the blood supply to the brain through compression of the carotid arteries.

The postmortem findings in these cases is often inconspicuous and inconclusive, and may or may not include pinpoint ("petechial") hemorrhages in the facial skin and elsewhere. Blockage of the outer airways (smothering) may leave very few or no signs at all on the dead body. Further, the findings in fatal asphyxiation are identical to the findings in nonfatal asphyxiation, which means that these mechanisms represent diagnoses of exclusion. Another problem is that postmortem findings in ligature strangulation and hanging may be very similar, and consequently hard to differ, particularly if a strangled person is hanged afterward. The pattern and direction of the ligature, the location of the ligature mark, and the presence and amount of petechiae may indicate what really happened.

BLUNT FORCE TRAUMA Fist blows, kicks, blows with some kind of weapon, falls from high level, and vehicle collisions are all examples on this group of homicide methods. Males are often killed by another man, often under the influence of drugs, whereas a female victim often has been killed by a partner or ex-partner.

POISONING This uncommon type of homicide is often hard to detect and even harder to prosecute. The reason for this is that a case of homicidal poisoning may present as a disease process, and that the symptoms may appear days or weeks after the exposition. Exceptionally hard to detect are poisoning homicides in a hospital environment, where the patients already have a disease, perhaps fatal in itself, and where health-care personnel generally are not suspected of such a crime. Another type which is difficult to detect is an "overdose" of a drug injected into the addict by another person.

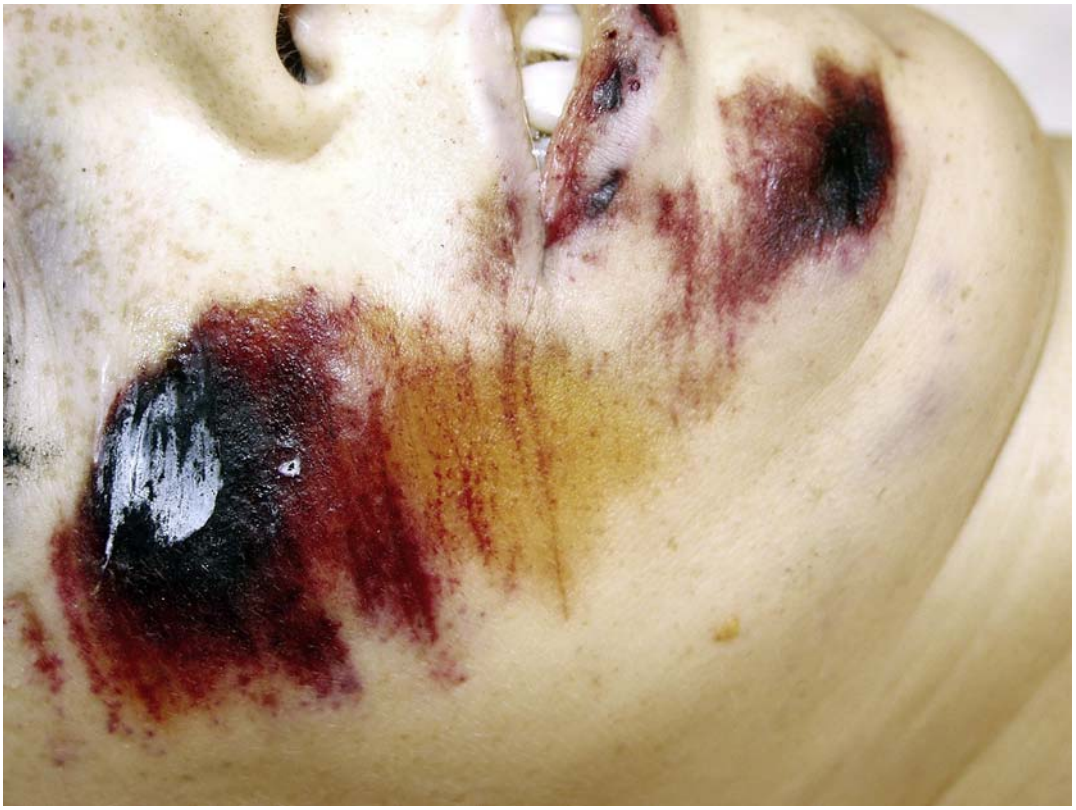


FIGURE 6.1 **Excoriation.** Excoriations of different depth with bleeding in the deeper (upper) parts.

TERMINOLOGY OF COMMON WOUND TYPES

Blunt Trauma

Abrasion/Excoriation

An *abrasion* or *excoriation* (lay terminology: scratch, graze) is a superficial injury of the skin, which usually shows some bleeding (Fig. 6.1). This type of injury is caused by a tangential contact with, eg, a fingernail, a shoe, or gravel. Occasionally, such an abrasion may be patterned, forming a mirror image of the impacting object.

Intradermal and Subcutaneous Hemorrhage

A bruise (subcutaneos hemorrhage) is a collection of blood visible through the skin (Fig. 6.2), emerging from vessels injured by (blunt) trauma—usually veins or small arteries. The medical term *contusion* includes also similar hemorrhages not visible at external examination. Superficial (intradermal) bruises may be patterned and give an image of the

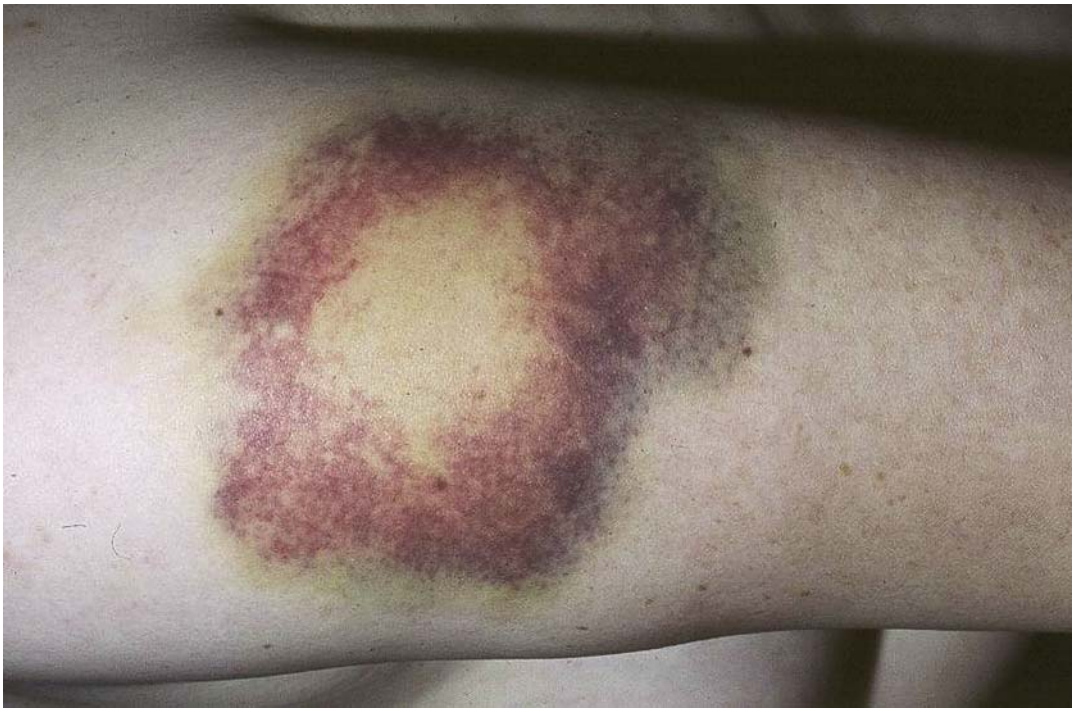


FIGURE 6.2 **Subcutaneous hematoma (bruise).** A few-days-old bruise with yellow margins and cleared center.

impacting object, whereas subcutaneous (deep) bruises usually have a rounded or oval form, independent of the form of the impact object.

A number of factors may affect the ease with which an individual can contract a bruise, eg, age, sex, drug abuse, liver disease, bodily location of the bruise, degree of obesity, blood vessel fragility, and coagulability disturbances. There are also a number of factors affecting the breakdown of the bruise, which means that age determination based on the color of the bruise is imprecise. Most authors agree that only the presence or absence of yellow color is useful in age determination; a yellow color indicates that the bruise cannot be less than 18 h (Dimitrova et al., 2006).

It should also be noted that a bruise can appear at another location than the point of impact. Gravity may shift the location of, eg, a subcutaneous hemorrhage in the forehead to the periorbital area. A "black eye" can also result from a fracture in the skull base after a fall where the back of the head hit the ground.

Laceration

If all layers of the skin are penetrated, the result is a *laceration* (Fig. 6.3). Such an injury results more easily if a bony part is lying under the point of impact, eg, the eyebrow, the shin, or the scalp.

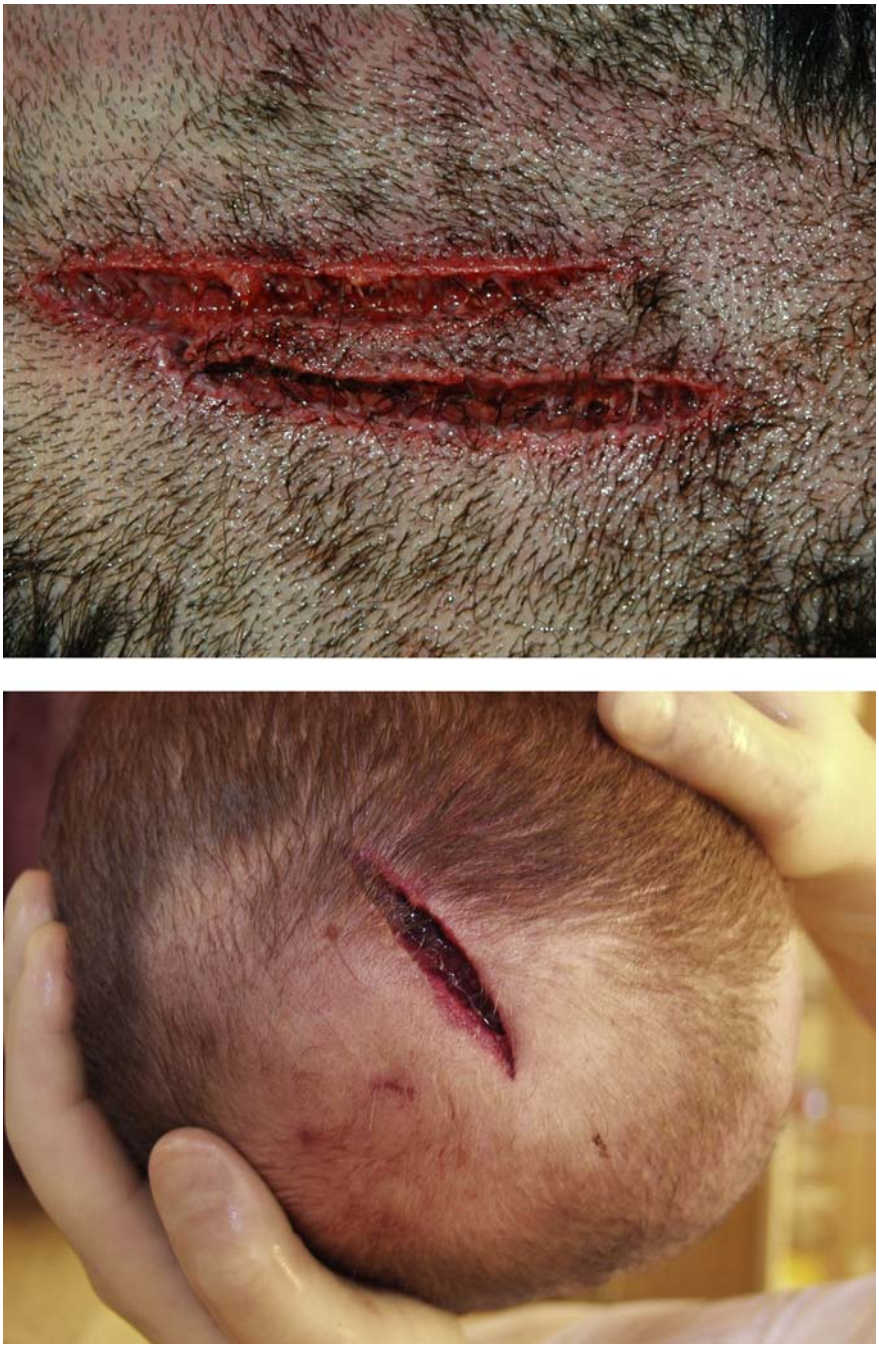


FIGURE 6.3 **Laceration.** Laceration in the scalp after hit with an elongated object. The laceration resembles an incised wound, but has uneven edges and tissue bridges crossing the wound.

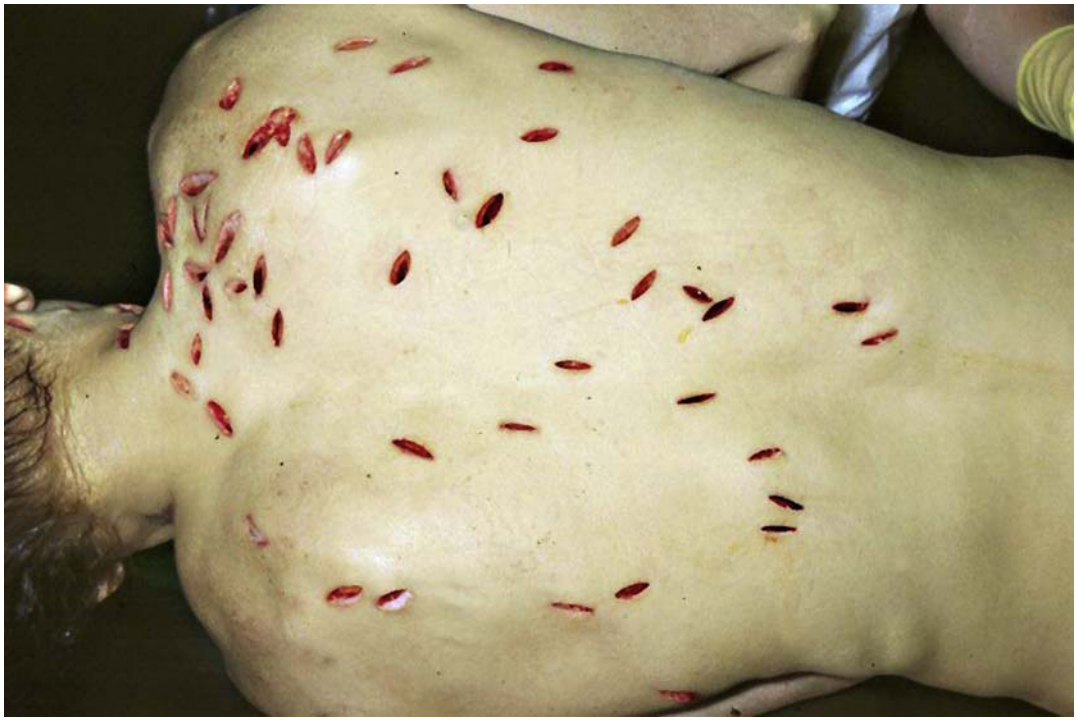


FIGURE 6.4 **Multiple stab wounds.**

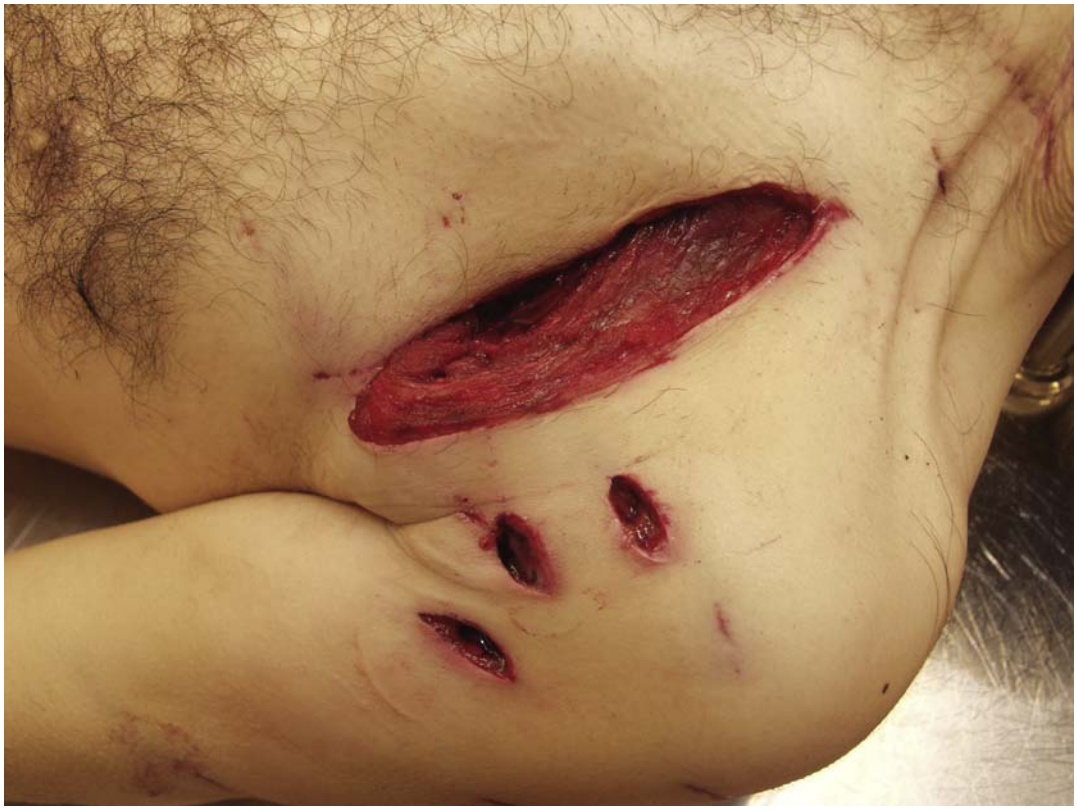


FIGURE 6.5 **Cut (slash) wound with three stab wounds.**

FIGURE 6.6 **Defense wounds.**

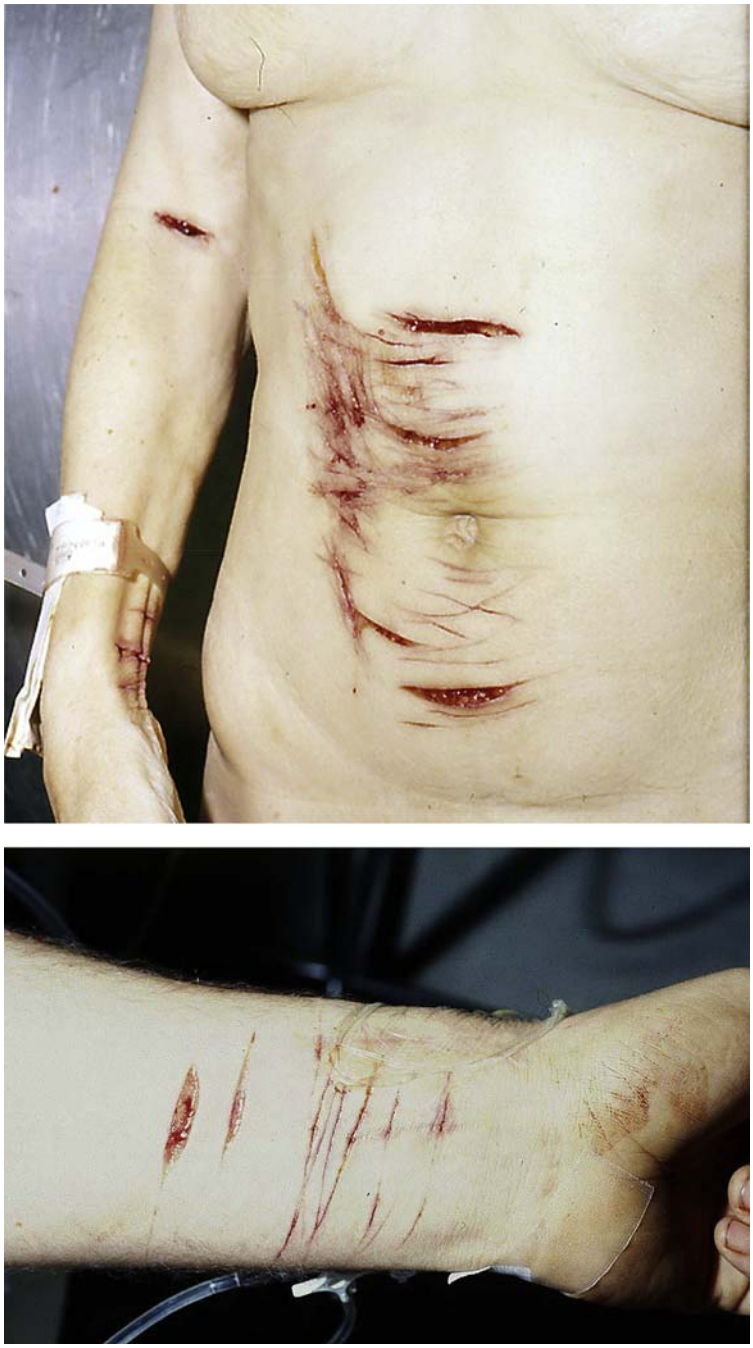


FIGURE 6.7 **Hesitation wounds.**

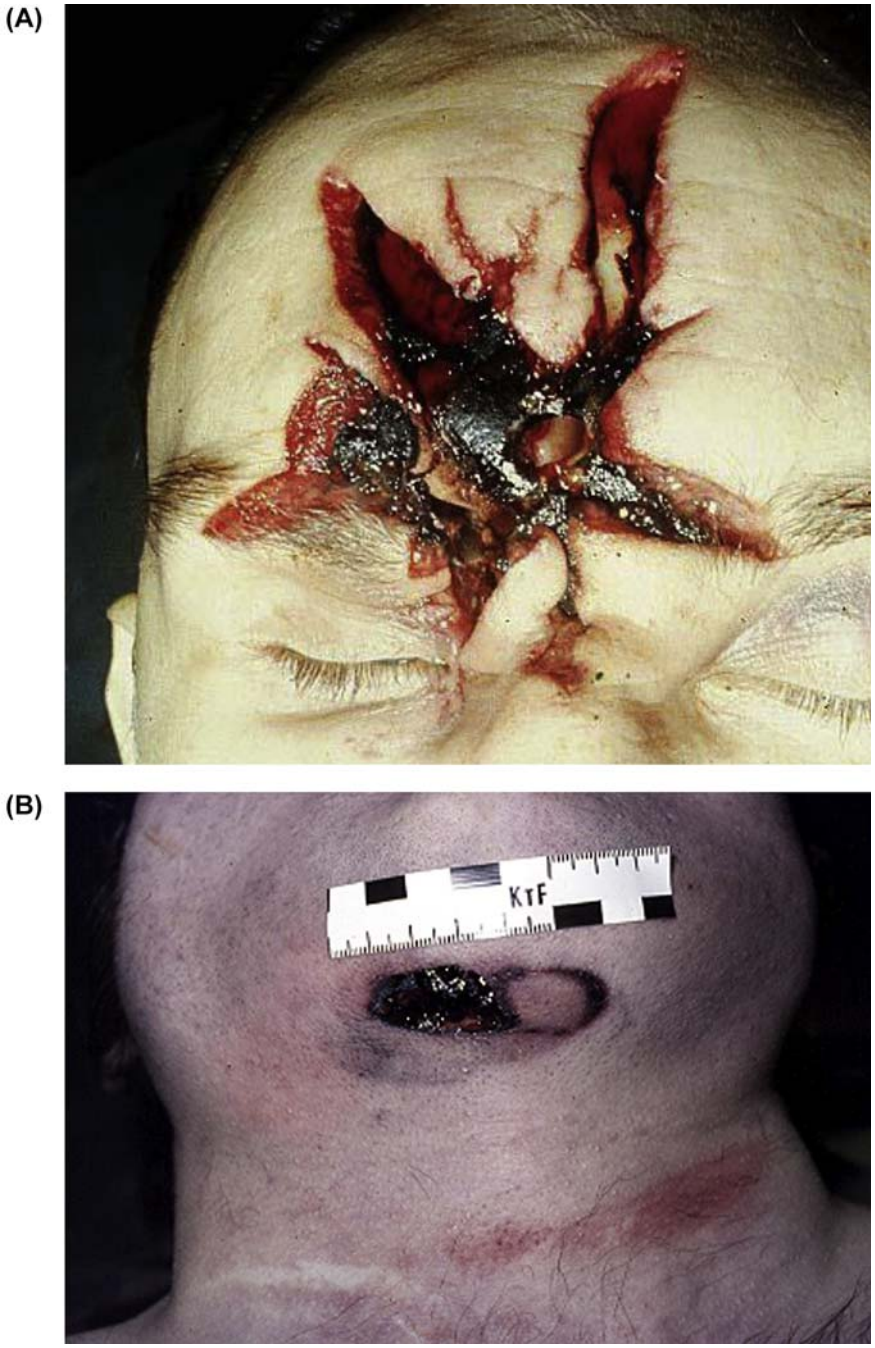


FIGURE 6.8 **Contact entrance wound.** (A) Contact wound over bony skull, where muzzle gases have ruptured the surrounding skin. (B) Contact wound from double-barreled shotgun, with muzzle imprint of the second barrel.

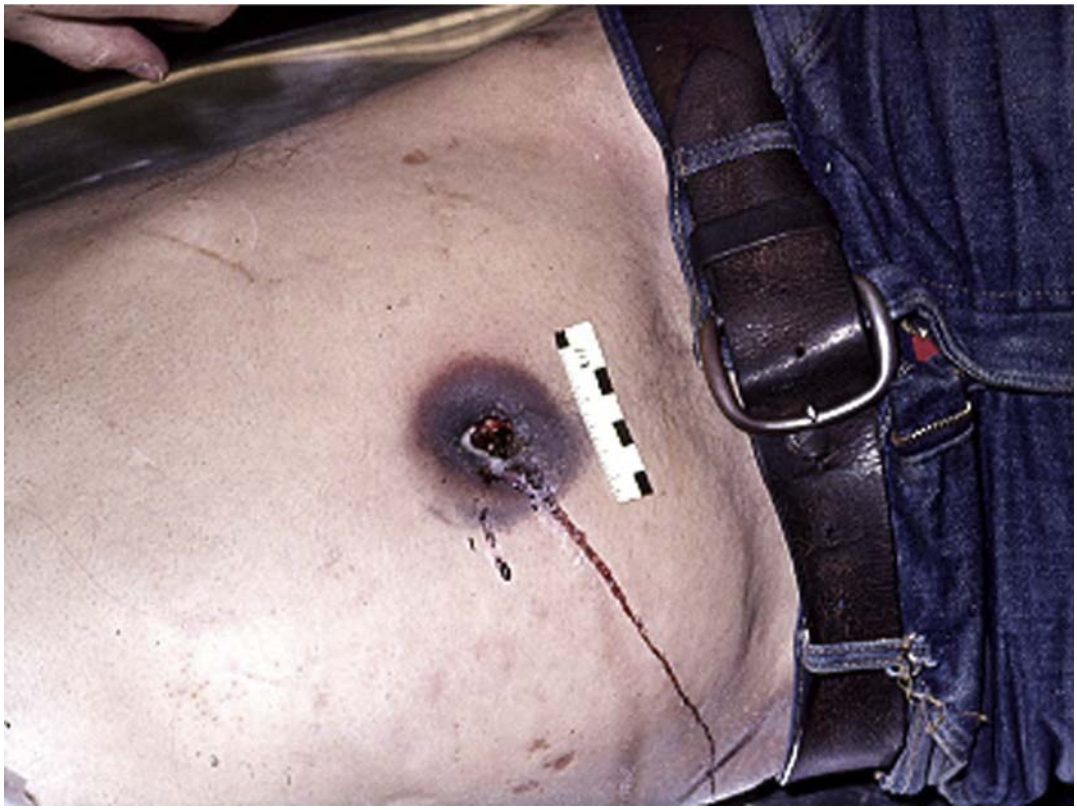


FIGURE 6.9 **Entrance wound, medium range.**

Sharp Force Trauma

These injuries are also called incised wounds, are typically caused by a knife or an other sharp object, and include stab wounds or cut (slash) wounds (Figs. 6.4 and 6.5). The former type is deeper than wide, the latter wider than deep and thus more superficial.

Defense wounds designate the injuries caused on the victim when he/she tries to protect him/herself against a knife attack. Such wounds can be located to the forearms and hands (Fig. 6.6), but a person lying down may have defense wounds on the legs.

Hesitation wounds is the term used when a person attempts to cut him/herself, eg, on the wrist, but starts with a number of superficial, parallel cuts before one or more deeper cuts are made (Fig. 6.7).

Gunshot Wounds

Entrance wound and *exit wound* are self-explanatory terms, but their appearance will vary, depending on shooting distance, type of weapon and projectile, etc. The appearance

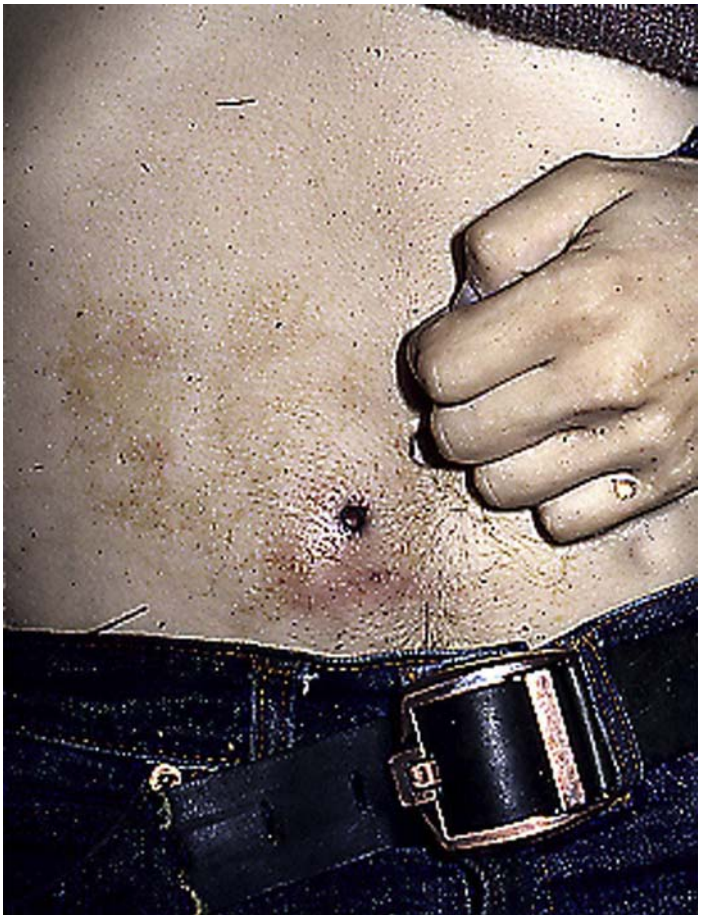


FIGURE 6.10 **Entrance wound, distant range.**

of the entrance wound is highly dependant upon the distance between the muzzle and the skin.

The entrance wound may be a *contact wound* if the muzzle has been in direct contact with the skin, and may be accompanied by a *muzzle imprint*, burning of wound edges, and entry of soot and muzzle gases into the wound (Fig. 6.8). At a slightly greater distance, up to perhaps ~1 m, naked skin can be characterized by burnt/singed hair, soot soiling, and tatooing of unburnt powder flakes (Fig. 6.9). *Distant-range* wounds lack the characteristics of the close-range wounds and are characterized by the properties of the hitting projectile with a dirt ring at the periphery of the rounded or oval entrance wound (Fig. 6.10). The exit wound is often characterized by everted skin flaps (Fig. 6.11).

For details of the characteristics of smooth-bored shotgun wounds and other weapons, we refer to other textbooks on the subject.

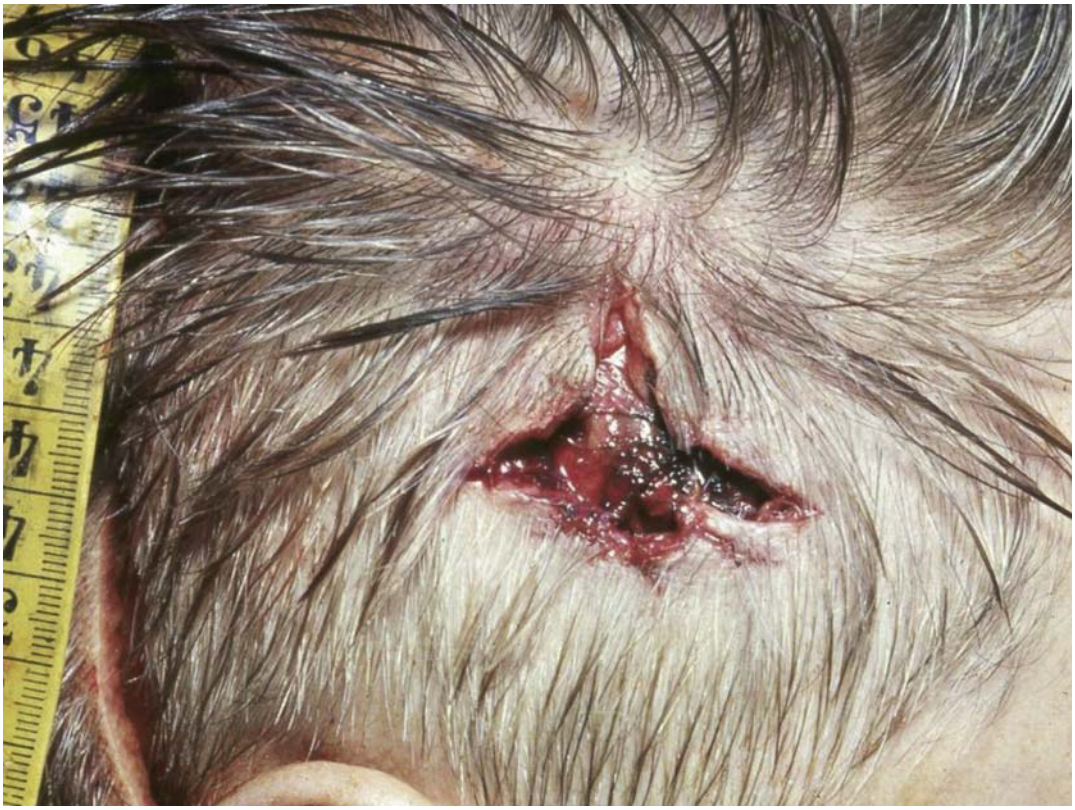


FIGURE 6.11 **Exit wound.**

References

Ahlm, K., Eriksson, A., Lekander, T., Björnstig, U., 2001. Alla dödsfall i trafiken är inte ”dödsolyckor” – en analys av officiell statistik i svensk vägtrafik år 1999 (All traffic related deaths are not “accidents” – an analysis of official Swedish statistics). Läkartidningen 98, 2016–2022.

Charlot, P., Witt, K., Pautot, V., Thomas, G., Zafrani, E.S., Lemaire, F., 2000. Declining autopsy rate in a French hospital: physician’s attitudes to the autopsy and use of autopsy material in research publications. Archives of Pathology and Laboratory Medicine 124, 739–745.

Dimitrova, T., et al., August 30–September 3, 2006. Qualitative visual image analysis of bruise age determination: a survey. In: Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA.

Eriksson, A., Persson, A., Gustafsson, T., Hultcrantz, M., Höistad, M., Jacobson, S., Mejàre, I. Postmortem Imaging. www.sbu/201501.

Graber, M., 2005. Diagnostic errors in medicine: a case of neglect. The Joint Commission Journal on Quality and Patient Safety 31, 106–113.

Hasselqvist, D., Rammer, L., 2003. Criminal death detected at forensic autopsy. Nordisk Rettsmedisin 172, 9–11.

Janko, H., Druid, H., November 27, 2009. Rättsmedicinska obduktioner avslöjar och avfärdar brott (Medico-legal autopsies reveal and rule out crime). In: Annual Meeting of the Swedish Medical Society.

Karlsson, T., 1997. A Multivariate Approach to the Interpretation of Patterns in Homicidal and Suicidal Sharp Force Fatalities (Thesis). Karolinska Institute.

Kronstrand, R., Roman, M., Thelander, G., Eriksson, A., 2011. Unintentional fatal intoxications with mitragynine and O-desmethyltramadol from the herbal blend krypton. Journal of Analytical Toxicology 35, 242–247.

Kuijpers, C.C.H., Fronczek, J., vd Goot, F.R.W., Niessen, H.W.M., van Diest, P.J., Jiwa, M., 2014. The value of autopsies in the era of high-tech medicine: discrepant findings persist. Journal of Clinical Pathology 67, 512–519.

Madea, B., Brinkmann, B., 2003. Handbuch Gerichtliche Medizin. Springer Verlag, Berlin.

Öström, M., Eriksson, A., 1987. Natural death while driving. Journal Forensic Science 32, 988–998.

Roulson, J., Benbow, E.W., Hasleton, P.S., 2005. Discrepancies between clinical and autopsy diagnosis and the value of post mortem histology; a meta-analysis and review. Histopathology 47, 551–559.

Shojania, K.G., Burton, E.C., 2008. The vanishing nonforensic autopsy. The New England Journal of Medicine 358, 873–875.

Shojania, K.G., Burton, E.C., McDonald, K.M., Goldman, L., 2003. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA 289, 2849–2856.

Sjögren, H., Eriksson, A., Ahlm, K., 2000. Role of alcohol in unnatural deaths: a study of all deaths in Sweden. Alcoholism: Clinical and Experimental Research 24, 1050–1056.

Wittschieber, D., Klauschen, F., Kimmritz, A.C., von Winterfeld, M., Kamphues, C., Scholman, H.J., Erbersdobler, A., Pfeiffer, H., Denkert, C., Dietel, M., Weichert, W., Budczies, J., Stenzinger, A., 2012. Who is at risk for diagnostic discrepancies? Comparison of pre- and postmortal diagnoses in 1800 patients of 3 medical decades in East and West Berlin. PLoS One 7, e37460.

This page intentionally left blank

CHAPTER

7

Death Investigation

S.A. Koehler

Forensic Medical Investigations, Pittsburgh, PA, United States

OUTLINE

http://dx.doi.org/10.1016/B978-0-12-404584-2.00007-0

INTRODUCTION

This chapter is designed to be a complement to Chapter 6, *Forensic Pathology* and will provide a brief history of the origin and development of the coroner and medical examiner system and describe how a death investigation is conducted.

HISTORY OF THE DEVELOPMENT OF DEATH INVESTIGATION SYSTEMS

Throughout the world, there are two basic types of death investigation systems. In Europe the coroner system is predominant while in the US there is a split between the coroner system and the more recent developed medical examiner system (referred to collectively as the ME/C office). The primary duties in both systems are to determine the identity of the decedent and establish the cause and manner of deaths that occur in their jurisdiction. See the section on "Reportable Deaths" for a list of the death circumstances that fall under the jurisdiction of the ME/C office. The office has complete legal authority and autonomy over the death investigation and does not require consent from the next of kin to conduct an autopsy or a death investigation.

THE CORONER SYSTEM

The origin of investigation into the cause of death can be traced back to 1194 in England when the office of the coroner was formalized into law by King Richard I. The Articles of Eyre established the office of the *custos placitorum coronae* (the keeper of the pleas of the crown), from which the word "coroner" arose. The coroner was a servant of the king whose main responsibility was the collection of monies owned to the king after the death of a nobleman. The first coroners were royal knights. Throughout the middle ages, the functions of the coroner included conducting coroner inquests, attention to and inspection of the dead, to hear appeals, confessions, and adjure subjects from the realm. Early coroners lacked any training in death investigation. In the early 1600s the American colonists introduced the coroner system from England and it became an integral part of the death investigation system in what was to become the US. Over time, many of the early duties of the coroner were stripped away, and reducing the role of the office to the medicolegal examination of a body and the determination of the cause and manner of death.

Currently, about 52% of the US population is served by a coroner system. The coroner is a public official, elected every four years; the position requires minimal medical experience with training ranging from virtually none to only a few weeks. The majority of coroners have full-time jobs (primarily as funeral directors) and serve only as part-time coroners. Only in the large population areas is the coroner a full-time position. The coroner appoints a number of deputy coroners to assist in the death investigation.

THE MEDICAL EXAMINER SYSTEM

In the mid to late 1800s, there was a shift from the coroner to a medical examiner system. Factors contributed to this change included the public's dissatisfaction with lay coroners, accusations of corruptions, and the growing understanding for the need for highly trained personnel to investigate deaths. The first medical examiner office became operational in New York City in 1918. In the last several decades, the medical examiner system has been slowly replacing the coroner system. A complete state-by-state list of the type of death investigation system is available at www.cdc.gov.epo.dphis/mecips.death_investaigtion.htm.

The structure of a medical examiner's office includes the chief medical examiner (CME) and a number of deputy medical examiners (DME). The CME is a licensed physician and a diplomat of the American Board of Pathology (ABP) in anatomic and forensic pathology with experience in forensic medicine and pathology. They are appointed for a 5-year term. DME are licensed physicians that have completed an ABP-approved fellowship in forensic pathology. The ME offices are staffed by death investigators, autopsy technicians, and forensic specialists (criminalists).

FUNDAMENTALS OF DEATH INVESTIGATION

The Discovery of a Body

All forensic death investigations obviously begin after death, with the discovery of the death occurring in a wide variety of circumstances. The investigation can be initiated by a hiker coming across human remains; a wife discovering her husband unresponsive on the living room floor; an elderly couple whose bludgeoned bodies are discovered in a pool of blood; or a victim of homicide who dies during surgery. The discovery initiates a call to 911 and a response by the first responders, emergency medical technicians, paramedics, the fire department, or police officers. The first duty is to assess the status of the individual. If responders detect any sign of life, they will treat and transport the victim to a hospital. However, if the victim is beyond medical treatment they pronounce the victim dead on arrival (DOA) at the scene. A DOA status indicates no carotid pulse, electrocardiogram (EKG) activity, or the presence of overt signs of death such as rigor mortis, livor mortis, algor mortis (decreased body temperature), decomposition, decapitation, or evisceration.

Once the body is declared dead the jurisdiction of the body and the death scene switches from the emergency medical team to law enforcement at the scene, although only briefly, before being transferred to death investigators from the ME/C office. Prior to leaving the scene, emergency personnel cover the body and provide a copy of their trip sheet to the police office. The EMS-generated trip sheet details the specific of the response (time received, time arrived, assessment, and methods used to determine death), names of the EMS personal, and any observations. The role of the police is to protect the body and the surrounding death scene by encircling it with two layers of crime scene tape. The outer perimeter is the "nonactive" areas, while the inner perimeter is reserved for the "active" investigation. The police

office also generate an Incident Report that contains the names, addresses, phone numbers of the individuals associated with the death as well as noting the position of the victims(s), injuries, and the actions of the EMS. The officer often notes significant physical evidence that maybe destroyed, disappear, or change in some way before other investigators arrive. This includes such things as a cold can of soda on a counter in a warm room, a dry parking space on a rainy day, or tracks in the snow. All deaths should be considered potential homicides until the death has been thoroughly investigated.

The Death Call: The Start of the Forensic Investigation

The ME/C office is notified of a death by a phone call from a patrol officer, paramedic, physician, or nurse. Specially trained death investigators or deputy coroners take these calls. During this initial telephone conversation, the investigators collect the facts and circumstances surrounding the death, including medical history, condition of the body, and other information, which is used to determine whether the case is a reportable death and falls within the ME/C jurisdiction.

Reportable Deaths

Around 20% of all deaths in the US undergo a forensic death investigation. An ME/C office follows a set of guidelines delineating the type of deaths that fall under their jurisdiction called "reportable deaths." An ME/C office takes jurisdiction of a body if the death was sudden, unexpected, unexplained, or suspicious, if the death was the result of violence or trauma; certain types of hospital-related deaths such as stillbirths, criminal abortions, infant deaths; deaths while in custody; or if the body is unidentified or unclaimed.

The Death Scene Investigation: Forensic Death Scene Investigators

Once a death meets the criteria as a reportable death, the ME/C office dispatches forensic death scene investigators to the scene. Death investigations have received extensive training in death investigation methods, including the identification, photo documentation, and collection of forensic evidence. Before entering the actual death scene, they start their investigation by eliciting basic information about the circumstances of the death by interviewing first responders, EMS, police, next of kin and other family members, and witnesses at the scene. During the processing of the scene, numerous photographs are taken of the body and the surrounding scene. These images will later be viewed by the forensic pathologist, homicide detectives, lawyers, and the jury. Next, they turn their attention to the body attempting to make a positive identification of the victim, documenting all visible injuries and trauma, and collecting evidence. They then conduct a detailed search of the immediate and surrounding area for footprints, drug paraphernalia, weapons, shell casing, bullet holes, blood, and other trace evidence. If the scene contains specific types of forensic evidence, specialists called criminalists process the scene. Criminalists are individuals trained in forensic disciplines such as fingerprints, blood spatter, ballistics, trace evidence, or entomology. Once the death scene investigation is completed, the body is wrapped in a white

sheet, and placed inside a disaster bag or "body bag," and transported to the morgue. At the morgue, the body is placed inside a refrigeration storage unit. The forensic death investigation generates three documents: The *death investigation report*, the *final anatomical report* (autopsy/toxicology report), and the *certificate of death*.

Death Investigation Report

The report contains two main sections. The first section contains basic demographic data on the victim such as the age, sex, race, marital status, social security number, residence, occupant, next of kin information, and the phone numbers of key personnel. This section also contains time line information such as time/data last seen alive, time/data/place found unresponsive, time/data/place pronounced dead, and who discovered/pronounced the victim. The second section contains an open-ended narrative called the *circumstance of death*. This narrative, written by the death investigator, offers the investigator's impression of the events occurring before the death, a comprehensive description of the event and circumstances leading up to the death, and the actions of the individual and others after the event. This generates possible scenarios or theories of how the death occurred based on information obtained from interviews; examination of the body and scene; statements from other witness; information from the victim's coworker, family, friend; a review of EMS reports, police reports, hospital and medical records. This section also contains past medical, psychological and social history, past and current medication. If the victim received medical treatment, the following information is collected: arrival time and action of EMS, arrival time at hospital, unit admitted to (ER, OR, ICU), type of medical intervention, and final discharge. The type of death dictates the specific type of information contained within the circumstance of death section of the report. The types of information collected on specific manner of deaths are covered below. In cases of SIDS or industrial deaths, there are standardized data collection forms that are completed by the death investigators.

The death investigation report provides a level of detail about the death not available by examination of the Death Certificate (DC). Note, however, that this report is only available at the ME/C office and is not public information like the DC.

FUNCTIONS OF THE MEDICAL EXAMINER/CORONER OFFICE

ME/C offices have three primary functions: (1) establish the cause of death, (2) determine the manner of death, and (3) to determine or identify the victim. One function of the postmortem evaluation is to determine the *cause of death*, a topic that is discussed at length in Chapter 6, *Forensic Pathology*. The cause of death is the injury, disease, or the combination of these two processes that initiates the train of physiological disturbances that produces the fatal termination of life. This determination is made by reviewing the death investigation report, medical records, and in most cases an examination of the body. The second function of the medicolegal investigation is to determine the *manner of death*, also discussed in Chapter 6. The manner of death refers to the circumstances in which the cause of death occurred. The five manners of death are natural, accident, suicide, homicide, and undetermined. The

determination of the manner of death is a nonbinding subjective opinion based on the preponderance of the available evidence at the time of the death. The manner of death can be changed at any time with the presentation of substantive additional information. With the exception of "undetermined," the manner of death is discussed later in the chapter. The manner of "undetermined" is beyond the scope of this chapter. This designation is only used if there are insufficient physical findings at autopsy to ascertain the manner of death, the toxicology and microscopic examination yield nonspecific or insignificant results, the level of injuries or state of decomposition prevents identification of a specific manner of death, or the circumstances surrounding the death are elusive or impossible to confirm with a degree of medical certainty. The third role of the ME/C office is positive identification of the body. This serves two functions. First, the identity of the decedent may offer clues to the likely cause and manner of death. If the victim was a drug dealer, the investigation would lean toward homicide; whereas, if the victim was a Wall Street trader who lost all his income due to the poor economy, the investigation might lean toward suicide. Second, positive identity ensures that the correct body is returned to the correct next of kin. Methods used to establish identification range from low-tech methods such as visual confirmation, clothing reorganization, documenting matching tattoos and scars to high-tech methods such as fingerprints, dental comparisons, medical and surgical implants, DNA, computer or video overlay identification, and facial reconstruction.

The Forensic Autopsy

The examination of the body after death referred to as an autopsy in the US, morbid anatomy in Britain, and necropsy in Latin America and European countries. The function of a forensic autopsy is to provide information through a postmortem examination of the body and analysis of the fluids to determine the cause of death, manner of death, and mechanism of injury. As mentioned earlier, a medicolegal or forensic autopsy is designed to investigate unexpected, suspicious, or unnatural deaths. These autopsies are required by law and do not require the permission of the family. The completion of the forensic autopsy creates the second key document of a forensic death investigation, which is the final autopsy report.

Types of Postmortem Examinations

Prior to the examination of the body, the forensic pathologist will review all the available information surrounding the death, including the death investigation report, photographs and videos of the scene, medical records, and outside agencies reports (EMS, police, and fire). Based on these information of the circumstances surrounding the death, he will determine the type of postmortem examination to perform. The body will undergo either an external-only examination or a complete forensic examination.

External-Only Examination

An external examination is a postmortem examination that is limited to an external examination of the body, the collection of blood and urine, and a review of the medical records. The body undergoes a head to toe examination documenting the condition of

the body, scars, recent and remote trauma, and medical treatments. The examination in conjunction with the death investigation and information contained within the medical records are used to determine the cause and manner of death without evisceration of the body. External examination are conducted in cases where there is a well-documented past medical history (PMH), deaths by suicide with a through-and-through gunshot wound (GSW) wound to the head or among elderly deaths with significant PMH. However, if there are concerns that the cause cannot be ascertained via external examination the forensic pathologist has the authority to conduct a complete autopsy.

Complete Forensic Examination: The Autopsy Procedure

A complete forensic examination is comprised of an external and internal examination of the body. All forensic autopsies begin by a visual inspection of the body. The height, weight, and condition of the body are noted. The clothing is described and removed layer-by-layer and then transferred to the crime lab to search for trace physical or biological evidence. Photographs are taken throughout the autopsy. The naked body undergoes the external examination described above.

In the US the predominant technique used for dissection of the body involves a Y-shaped incision. Incisions begin at each shoulder that extend downward and meet to the midline of the body in the lower chest, then the incision extends to the top of the pubic bone. The chest plate is removed by cutting the ribs on both sides, exposing the heart and lungs. Samples of blood, bile, urine, and eye fluid are collected. Each organ is first examined in situ, then removed, weighed, photographed, and dissected. Next the heart, lungs, pancreas, spleen, liver, kidneys, prostate, and gastrointestinal tract (small and large intestines) are removed. The brain is removed by first making an incision ear to ear, reflecting the scalp and exposing the skull, then using a reciprocating bone saw to create a circular cut of the skull allowing the removal of the skullcap and the brain. Microscopic slides are made of each organ. The collected body fluids are sent to a forensic toxicologist for analysis. His analysis generates a toxicology report that lists all the compounds by type and concentration detected in the different body fluids.

After competition of the forensic autopsy, the forensic pathologist reviews the results of the autopsy, the microscopic slides, the toxicological analysis, the death investigation report, and medical records to create the final autopsy report. The general format of this report contains the following sections: anatomical diagnosis, external examination, internal examination, and the toxicology report. The anatomical diagnosis section contains three sections: (1) a description of the major findings that contributed to the cause of death, (2) the manner of death, and (3) an opinion statement.

Take for example the death of a 39-year-old white male from advanced arteriosclerotic cardiovascular disease that died in his residence.

The anatomical diagnosis section would read:

1. arteriosclerotic cardiovascular disease:
   - **a.** atherosclerosis, marked, left and right coronary artery
   - **b.** atherosclerosis, mild of the aorta
2. acute pulmonary edema and congestion
3. general congestion of viscera

Manner of death: Natural.

Opinion: A 39-year-old white male died as a result of arteriosclerotic cardiovascular disease.

The external examination section contains the height and weight of the body, condition of the body, and a detailed head to toe description of the body. This section contains a description of trauma, recent and remote medical/surgical treatments, and congenital abnormalities. This section also describes the clothing. The internal examination includes the weights, descriptions, and anatomical findings within the following systems: cardiovascular, respiratory, hepatobiliary, hemolymphatic, gastrointestinal, pancreas, urogenital, musculoskeletal, and the central nervous system. The toxicological report lists all the compounds by type and concentrations detected in the blood, urine, bile, and eye fluid.

Certification of Death

The last official document produced after completion of the forensic investigation is the Certificate of Death or Death Certificate (DC). A DC contains the age, sex, race, marital status, military serves, occupation, level of education, residence, location of the incident, and the place of death. It lists the *immediate cause of death* and the *manner of death*. The Immediate Cause of Death section lists the sequence of events leading to death, proceeding backward from the final disease or condition that resulted in death. The *immediate cause* represents the final disease, injury, or complication directly causing death that can be followed by up to five underlying (due to or as a consequence of) causes of death. For example, the immediate cause of death can read **GSW head** or **sepsis**, *due to*: **pneumonia**, *due to:* **thermal burns**, *due to*: **house fire.** The *manner of death* is the fashion or circumstances in which the cause arose. The five classification are *natural, accidental, homicide, suicide, or undetermined*. The DC is an opinion of the forensic pathologist regarding the cause, manner, and mechanism of death. This is a medical opinion based on the examination of the body, review of the medical records, analysis of evidence, and supplemental information from other forensic expert. The DC is the foundation of national statistical databases such as the CDC and National Centers for Health Statistics (NCHS). The DC is used to monitor the health of the county, communicable diseases, violence crime, and to identify emerging diseases.

Core Epidemiological Data Collected on Medicolegal Investigated Deaths

In all medicolegal investigations conducted by the ME/C office, regardless of the type or manner of death, a core set of information is collected and available about the victim, the circumstance of death, the forensic examination, and the cause and manner of death. Core information about the victim includes the age, date of birth, sex, race, marital status, occupant, and residence. In addition, the victim's PMH, list of past and current medications, and past and current medical treatment are collected. Information regarding the circumstances surrounding the death includes a description of the incident, where the event occurred, the time/date last seen alive, type of activity engaged prior to death, and the time, date, and place pronounced. Information about the level of the forensic examination and the final cause and manner of death is collected.

MANNER OF DEATH

Natural

Natural deaths are caused by a naturally occurring disease processes that interfere or disable vital organ functions, by congenital anomalies, or the degenerative aging processes. The top three causes of natural deaths are cardiovascular disease, neoplasm, and diabetes. A natural death displays no evidence of trauma or indication of foul play and does not require a criminal investigation; however, a thorough examination of the facts leading to the natural death must be undertaken.

The majority of natural deaths are never reported to the ME/C office therefore never undergo any form of medicolegal investigation, and go directly from the residence, nursing home, or hospital to the funeral home for burial or cremation. Reasons for this include the death was expected, the etiology of the disease process was well documented, or the individual's physician felt that the death was natural in nature and willing to complete the DC and list the death as natural. It is important to note that any physician can issue a DC as long as there is a clear documentation in the medical records indicating a natural disease process. However, a physician cannot issue a DC where the manner of death is other than natural. Only the ME/C office is legally permitted to issue a DC with a nonnatural manner of death. A small percentage of natural deaths undergo nonforensic postmortem examination, often at teaching hospitals or veteran's hospitals. These examinations are conducted at the request of the family or physician to confirm preexisting medical conditions, confirm the cause of the death, or for the training of medical personnel. A limited number of natural deaths undergo a forensic investigation by the ME/C office.

The Forensic Investigation of Natural Deaths

While, only a fraction of natural deaths undergoes a forensic investigation they account for the majority of the workload in most ME/C offices. The ME/C office takes jurisdiction of a natural death if the individual lacks a physician to issue a DC or feel comfortable issuing the DC or lack of sufficient PMH to issue a DC.

Typically, the ME/C office is notified of a natural death, by a nurse or physician who also provides detailed information surrounding the death. If the ME/C takes jurisdiction of a natural death, it has great flexibility regarding the level of the investigation and the type of forensic examination. There are three options: first, the ME/C can release the body to the funeral home, but can subpoena the individual's medical records and use that information to determine the cause of death. In this case, the ME/C office will issue the DC, based solely on medical documentation without examination of the body. These cases are called WWI (we will issue) or OWI (office will issue), referring to the DC.

The second option is that the body and medical records are transported to the ME/C office. Reasons to bring in the body include the presence of trauma, lack of a physician, or some questions regarding the circumstances surrounding the death. In this case, a death scene investigation is typically conducted. The forensic pathologist reviews the investigator's report and the medical records and concludes that only an external examination of the body is required. A decision to conduct an external-only examination is predicated on the ability to ascertain the cause of death from information contained within the medical records.

The third opinion is to conduct a complete forensic investigation. Cases that warrant a complete investigation include those where the circumstances of the death are in question, there is a lack of significant medical history, or if the death was sudden and unexpected. In these cases, there is a detailed death scene investigation, the body undergoes a complete forensic autopsy, and toxicological analysis of the body fluids.

The focus of a complete investigation of a natural death is the examination of the internal organs. A complete forensic investigation provides detailed anatomical and pathological data of the internal organs. The internal organ systems examined during the autopsy and the type of data collected are presented in the table below. In addition, toxicological analysis of body fluids provides a list and concentrations of drugs at the time of death.

Data collected on the major internal organs

Organ system	Data collected
Cardiovascular system	• Heart weight • Percentage of occlusion of the coronary arteries • Thickness of left ventricle • Location and size of scarring • Evidence of medical intervention
Respiratory system	• Lung weights: left/right lung • Location/degrees of lung disease • Location, size, and type of tumors
Hepatobiliary system	• Liver weight • Disease state • Location, size, and type of tumors
Central nervous system	• Brain weight • Location/size/type of hemorrhage (epidural, subdural, subarachnoid, intracerebral) • Aneurysm (location, size, type) • Tumors (location, type) • Level of degenerative brain disease
Endocrine system	• Pancreas weight • Kidney weights: left/right kidney • Spleen weight

Accidental

Accidental deaths are those that occur from nonnatural process, which are noncriminal, and not self-inflected. This section will examine six types of accidental deaths investigated by the ME/C office: drug overdoses, motor vehicle accidents, falls, fires, industrials, and medical misadventure.

Drug Overdose

An accidental drug overdose (OD) death is the result of an intentional intake of drugs, medication, or compound at levels that the body cannot detoxify rapidly enough to

prevent death. These deaths are not considered a deliberate attempt to commit suicide (see Chapter 6, *Forensic Pathology* for further discussion). Causes of OD deaths include unintentional overmedicating, binge drinking, or shooting heroin. Common compounds associated with an OD include alcohol, over-the-counter medications (OTM), prescription medications, and illegal drugs.

A possible drug OD death undergoes an investigation like any other suspicious death. The death scene investigators collect core information about the victim, the events preceding the death, actions taken after the indent, a list of current medications, and PMH. The majority of victims of a drug OD death are discovered within a residence either unresponsive or DOA. The death investigation involves a search of the death scene, examination of the body, questioning the next of kin, family and friends, a complete forensic autopsy, and toxicological analysis.

The death investigation starts by searching the scene; however, the search may provide limited cues to the cause of death, as in many cases drug paraphernalia has been removed and the drugs disposed of prior to the arrival of police and forensic investigators.

The forensic autopsy starts with the external examination of the body. The body typically displays no tell tale signs of drug abuse such as needle marks or "track marks," the only visible physical characteristics may be a white foamy froth around the nose and mouth. The internal organs also do not reveal any pathological signs associated with an acute drug OD death. During the autopsy, blood, urine, bile, vitreous fluid (eye fluid), and stomach contents are collected. In some cases, the stomach may contain partially dissolved pills that can aid in the identification of the compound that might have caused the death. Some victims survive long enough to be transported to a hospital where antemortem blood is collected in the emergency department. This blood should be collected in the scope of the death investigation from the hospital because it represents the drug levels before any medical treatment was initiated. Postmortem and antemortem samples are sent for toxicological analysis and interruption. The forensic toxicologist plays a central role in these types of deaths by (1) identifying the type and number of drugs, (2) quantifying the concentrations of each drug, and (3) interpreting how these compounds affect the human body. The goal of the toxicological analysis is to establish first, whether drugs played a role in the cause of death and second, if they did, determine which drug(s) caused the death. The toxicology report contains a qualitative analysis providing the drugs detected within the samples and a quantitative concentration of each compounds.

A forensic toxicologist classifies drug levels into therapeutic, toxic, or lethal levels. Therapeutic levels represent the level prescribed by the pharmaceutical company or physician for a prescription or OTM medication as safe. Toxic levels are levels that exceed the recommended dosage and results in damage to internal organs especially the liver. Lethal levels are levels of a compound that results in death. The range between these three levels varies greatly from drug to drug. One pharmaceutical effect to consider is the synergistic effect. Multiple drugs detected in the blood do not have to each be within the lethal range to cause death. For example, if several CNS depressant drugs were taken each within their respective therapeutic range, their combined or synergistic affects can greatly depress CNS function (ie, respiration and cardiac function), resulting in death. The forensic toxicologist and forensic pathologist review the medical history, the autopsy findings, and the circumstances leading to the death and determine which, if any, of the detected compounds contributed to either the immediate cause of death or contributed to the death.

Motor Vehicle Accident

All fatal motor vehicle accidents should undergo an investigation by the ME/C office. Death scene investigators create an overall description of the accident scene using written narratives, diagrams, photographs, and video. The type and condition of the roadway, weather conditions, and environmental hazards at the accident scene are collected. The make, model, color, traveling direction, estimated speed, point of impact, skid marks, deployment of the airbags, vehicle damage, and final resting position of all vehicles are noted. Occupants are described in terms of position, use of safety systems (belted, unbelted, helmeted, and nonhelmeted), and all visible injuries. The forensic autopsy provides a description of the location, nature, and severity of the injuries and of course the cause of the death. The results of the toxicological analysis determine if alcohol, drugs, or medications played any role in causing the accident. In the US, as in most Western countries, a crash-related fatality is a reportable event, and thus the ME/C office completes and submits a fatality analysis report to the fatality analysis reporting system (FARS) of the National Highway Traffic Safety Administration. In some situations, the local or state police conducts an independent investigation into the accident.

Using information from the EMS report, police and fire reports, witness statements, the medical records, and observation by the death scene investigator offers theories as to possible sequence of events and a likely cause for the accident. Causes of an MVA include impairment of the drive, human error in the form of speeding, reckless driving, inexperience, or falling asleep, environmental conditions such as weather related hazards (ice, snow, rain, or fog), road hazards (poorly maintained roads, potholes), and animals such as deer or dogs on the roadway, preexisting medical disease and/or the effects of prescription and OTM. As the population of elderly driver, those over the age of 65, increases so does the likelihood of sudden deaths behind the wheel caused by cardiac events or cerebrovascular disease increase. In addition, elderly drivers are prescribed numerous medications and the role of these medications either singly or in combination on driving ability has received limited study (See also Chapter 11, *Traffic Injury Investigation*, and Chapter 12, *Traffic Injury Investigation: Product Defects*).

Fall-Related

All deaths caused by a fall must be reported to the ME/C office (especially those that occur in public places, nursing homes, or hospitals). The ME/C has several options regarding the level of investigation conducted on fall-related deaths. One opinion is to release the body directly to the funeral home and issue the DC based on the information contained within the victim's medical records, referred to as OWI DC. The second opinion is to transport the body alone with the medical records to the morgue where the body will undergo either an external-only examination or a complete forensic autopsy. The level of the examination depends on the protocol of the specific ME/C office.

The type of fall-specific information that is collected is dictated by the location, the type of fall, and the circumstances surrounding the fall. The majority of fatal falls occur in the home and among the elderly. Investigators start by determining if the fall was a same level fall or a change in elevation. Cases of same level fall investigators look for the cause for the fall such as loose rugs, wet floors, or other tripping hazards. In addition, they ascertain if the victim

impacted any object during the fall such as a corner of a table, bookcase, or floor. The most common change in elevation fall in a residence is a fall down from a flight of stairs. In these cases information is collected on the distance the individual fell, surface of the landing (concrete, dirt, wood, carpet), the number of and types of steps (wooden, carpeted), level of lighting, and the presence of safety devices (rails). Cases where an individual fells out of a bed, the investigator measures the distance fallen, the type of floor surface, and the presence of bed railing and other safety devices. The most common type of falls encountered in nursing home are falls out of a beds, chairs, or wheel chairs. Falls occurring in a nursing home require a forensic death investigation. Patients in nursing homes are by their nature suffer from a number of health conditions that make them more susceptible to life-threatening injuries from otherwise minor trauma. A less common type of fall is when victims fall out of an open window. The investigation focuses on determining if the victim was pushed (a homicide), deliberately jumped (a suicide), or accidentally fell out the window. The most frequent victims are young children who fall out open windows or through screened windows. Data collected include the distance of the fall, landing surface, type of structure (apartments, warehouse, abandoned building), and safety devices. The investigation forces on the physical and psychological state of the individual to the time of the fall. Information gathered includes the activity the individual was engaged in prior to the fall and alcohol consumption and/or any drugs. The investigation also ascertains the PMH of the victim such as cardiovascular diseases, high blood pressure, diabetes, neurological conditions (dementia, epilepsy), and conditions that affect vision, balance, or walking. The history of past falls is also obtained. The forensic autopsy provides information of the site and severity of trauma caused by the fall. The toxicological analysis provides the type and concentration of drugs at the time of death.

Fire-Related

Deaths by fires are mainly caused from smoke inhalation or burns. A less common cause is blunt force trauma from falling debris. Fire-related deaths undergo a standard death scene investigation, autopsy, and toxicological analysis. In addition, an independent investigation is conducted by the fire marshal and if necessary by the police arson team. Death investigators search the scene for clues as to the cause of the fire such as smoking, cooking, kids playing with matches, space heaters, candles, or electrical overload or malfunctions. The external examination notes the location and pattern of the burns, the types of burns (first, second, third degree), and the percentage of body burns. A critical medicolegal question involved in a fire death is determining whether the victim was alive or dead when the fire started. To make this determination, the airway is examined. If the victim was alive at the start of the fire but failed to escape because they were overcome by smoke or suffered a cardiac event, the victim would have inhaled smoke and other gasses in the atmosphere into the mouth, trachea, and lungs. An examination of the airway would reveal soot deposited in the nostrils, mouth, larynx, trachea, and bronchi. In addition, an analysis of the blood would reveal a positive carbon monoxide level. On the other hand, if the victim was already dead when the fire started (used as a means to conceal a homicide), the airway would not contain soot and the blood would be negative for carbon monoxide. The forensic autopsy is also critical in the determination of the positive identity of the victim especially when the body is badly burned or charred. All fire-related deaths are treated suspiciously until proven otherwise. While the forensic examination can determine the cause of death the

ME/C office relies on the investigation by the fire marshal to ascertain the origin of the fire and determine if the cause was accidental, deliberate, or unknown. Based on their report the manner of death is ruled accidental, suicide, or homicide (arson).

Industrial

An occupational (industrial) death is one that occurs during employment while the individual is engaged in legal activity either at the employment site or when engaged in duties required for their job (off-site). While, most deaths are unintentionally (accidental) caused by individual error, insufficient training, fatigue, or carelessness, a small number are intentionally caused such as suicide at the workplace, or homicide where unhappy employees killing coworkers.

The first role of the death investigation into a possible work-related death is to conform if the death did occurred while the victim was working or "on the clock" at the time of the incident. An industrial death results in two independent investigations: one conducted by the ME/C office, the other by the Occupational Safety and Health Administration (OSHA). In most industry-related death investigations, the circumstances leading to the death are clear. Examples include a roofer fallen off a roof or a construction worker being hit by a piece of equipment. The standard autopsy determines the cause of death typically caused by Blunt Forced Trauma (BFT), crushing injuries, internal injuries, heatstroke, hypothermia, asphyxiation, and acute exposure to poisons compounds. The toxicological analysis determines if alcohol or drugs contributed to the death. If the investigation does not indicate suicide or foul play, the death is an accidental death. If, however, the death was caused by natural conditions there is some debate if the death is a natural death or an industrial death. For example, an account sitting at his desk suffers a cardiac event and dies. Technically, he was on the job but is the manner accidental or natural?

Once the ME has completed their scene investigation, forensic autopsy, and toxicological analysis, they complete an industrial report form that is sent to OSHA. This form contains basis epidemiological data, the name, address and type of industry, a brief summary of the incident, and the cause and manner of death. OSHA investigation on fatal work injuries focuses on collecting information about each workplace fatal injury by cross-referencing the source records, such as DCs, workers' compensation reports, and Federal and State agency administrative reports. The OSHA investigation of fatal work injuries focuses on collecting information about each workplace fatal injury by cross-referencing source records such as DCs, Workers' compensation reports, and Federal and State agency administrative reports.

Medical Misadventure

Medical misadventure deaths result from a medical error by a health-care professional choosing an inappropriate method of care; medication errors; dosage errors; errors occurring before, during, or after a medical or surgical procedure; or failure of a medical device. Unlike the other causes of death, such as homicide, that by law must be reported to the ME/C, deaths by medical misadventures are only brought to the attention of authorities if a physician, nurse, or the next of kin are concerned that something out of the ordinary occurred during medical care/treatment. Otherwise, most investigations are conducted in the civil litigation context (see Chapter 14, *Medical Negligence Investigation*). The investigation is unique in several ways. First, the scene of the death is most often a hospital setting, a physician's office, or a nursing

home. And second, the analysis of whether a medical error was the cause of the death can require highly specialized review of the facts. In some cases the death investigator or forensic pathologist must consult with standard of care experts to determine if the death was a normal complication or caused by inappropriate techniques, medication dosing, or some other standard of care. In these cases the consulting expert is sometimes present during the autopsy.

Suicide

Suicide, also covered to some length in Chapter 6, *Forensic Pathology*, is an intentional act where death occurs as a result of intentional and self-inflicted injuries, and the individual acted alone with a singular motive of killing oneself. Forensic investigators must approach all deaths with a high level of suspicion and especially those appearing as a suicide. The role of the death investigator is to remain neutral and consider all possible explanations for the death including an intentional suicide, a homicide masquerading as a suicide, or an unintentional death (accident). The investigation starts by photographing the location and position of the body, the mechanism of death, and the overall surroundings. The method used for a successful suicide are determined by a number of factors such as ease and availability of the mechanism, level of familiarity with the mechanism, and the history of past attempts. The death scene, personal computer files, and social media such as "facebook" inter alia are searched for an actual or electronic suicide note or posting indicating the intent to commit suicide. While a note demonstrating intent or detailing a plan to take one's life is a strong indication that the death is a suicide; the lack of a note does not indicate against a suicide. It is estimated that between 10% and 37% of suicide victims leave a note. A suicide note may also provide clues to why the individual committed suicide, serves as a final goodbye message to family and friends, or allows the victim the means to vent anger toward a specific individual or society in general. The note can also be a collection of rambling sentences, statements, or figures whose meaning is only known to the author and results from the psychological processes associated with the suicide (ie, psychosis).

The role of the investigation is to differentiate a suicide from a nonsuicide. The statements provided by the next of kin, family, and coworkers about the state of mind of the victim; signs of depression; disengagement from normal activities; communicating suicide ideation; or statements reflecting a willingness to end one's life all can impact the direction of the investigation. The investigation would determine if the victim exercised any major life events that may have triggered a suicide. Key areas of interest include current relationships, financial situation, legal problems, and health status. Relationship issues associated with suicide include a recent separation, divorce proceedings, or the death of a spouse. Financial problems include job loss, mounting medical bills, or a gambling problem. Individuals accused of a serious crime may commit suicide to save the public disclose of the crime. Declining health or the diagnosis of medical conditions (cancer or Alzheimer's) can trigger a suicide.

The forensic examination of the body verifies that the circumstances described at the scene matches the autopsy findings. The internal examination of the organs will indicate the health status and stage of certain types of disease such as cancer, heart and liver disease, and diseases of the brain (stroke, Alzheimer's). The toxicological analysis will determine if the victim was at therapeutic levels for medications prescribed for depression, schizophrenia, or other psychiatric disorders. In addition, a determination can be made if decedent's

judgment was likely impaired by high levels of alcohol or drugs at the time of the suicide. The specific type of information contained within the autopsy report is dictated by the method of suicide. For example, suicides by firearm will contain entrance and exit wounds, path of the bullet, and the organ that caused death.

Homicide

Homicide is the killing of one human being by the act, procurement, or omission of another. Homicide is a neutral term; it merely describes the act and does not pronounce judgment as to its moral or legal quality. The topic is covered at length in Chapter 6, *Forensic Pathology* and Chapter 15, *Criminal Investigation* as well.

Among the many different kinds of death investigated by the ME/C office, homicide is most likely to extend beyond the walls of the ME/C office and enter the judicial system. The death scene investigation, the forensic examination, toxicological analysis, and examination of evidence are handled similarly to other types of deaths with the main differences that the information may be used in a court of law to convict an individual of a crime. It is impossible to cover all methods of homicide and the associated investigation in this chapter, but the proceeding provides a generalized description of a forensic homicide investigation.

A homicide case is initiated typically by an emergency service call regarding an unresponsive victim, or as occurs in the US often, following reports of gunshots. At the scene, EMS personnel assess the victim while law enforcement personnel or patrol officers protect the scene. If the victim is DOA, he is pronounced dead by EMS personnel, and investigators from the ME/C office and homicide detectives are requested to the scene, which is now roped off. The first task of the death scene investigator is to survey the scene to determine the level of investigation and the type of criminalists required to process the scene. The standard death scene investigation is conducted that includes photodocumentation of the scene and body, collecting evidence, and interview witnesses and the next of kin. The photos are critical because they represent the scene close to the time of the event and will later be viewed by the forensic pathologist, homicide detectives, lawyers, and the jury. Criminalists search the scene for fingerprints, impressions, body fluids, and trace evidence. The examination then switches to the body noting its position, clothing, areas of obvious trauma, distinctive odors, and objects around the body. The body is examined for the presence of any trace evidence, which is collected at the scene to prevent loss during transport. The type of evidence collected from a homicide scene varies and is dictated by the mechanism of the homicide. For example, in cases involving a firearm the information collected includes the type of firearm, the location of shell casings, the number of live and spent shells, the path of each projectile, firing distances, gunshot residue, location of entrance and exit wounds, and the level of damage caused by each projectile to the internal organs. After the body is processed, it is transposed to the morgue. Homicide investigators conduct an independent and simultaneous investigation of the homicide. Their focus is on the collection of information by interviewing the victim's family, friends, and coworkers; additionally, they canvas the neighborhood for additional leads. They will compare the characteristics of the crime to others on file searching for any similarities, identify possible suspects, and make arrests.

The body undergoes a standard complete forensic autopsy often with homicide detectives present during the autopsy. Toxicological analysis of the body fluids will determine if the victim was under the influence of alcohol or drugs. Forensic pathologist frequently testify

in court as to how the cause of death was determined, describe injuries to the body, in the case of a shooting the number and path of the bullets, what organs were affected, and in cases of multiple injuries which injury was the fatal.

Further Reading

Adelson, L., 1974. Pathology of Homicide. Charles C Thomas, Springfield, IL.
Altamura, C., VanGastel, A., Pioli, R., 1999. Seasonal and circadian rhythms in suicide in Cagliari, Italy. Journal of Affective Disorders 53, 77–85.
American Academy of Pediatrics, Task Force on Infant Sleep Positioning and SIDS, 1992. Pediatrics 89, 1120–1126.
American Academy of Pediatrics, Task Force on Infant Positioning and SIDS, 1996. Positioning and sudden infant death syndrome (SIDS). Pediatrics 98, 1216–1218.
American Academy of Pediatrics, Task Force on Infant Sleep Position and SIDS, 2000. Changing concepts of sudden infant death syndrome: implications for infant sleeping environment and sleep position. Pediatrics 105, 650–656.
American Medical Association (AMA) Council on ethical and judicial Affairs: confidentiality of HIV status on autopsy reports. In: Presented to the AMA House of Delegates, Chicago, IL, June 1992.
Aufderheide, D.H., 2000. Conducting the psychological autopsy in correctional settings. Journal of Correctional Health Care 7, 5–36.
Bailey, R., 2008. Carbon Monoxide. Available at: http://biology.about.com/library/blco.htm?p=1.
Baker, F.M., 1990. Black youth suicide: literature review with a focus on prevention. Journal of the National Medical Association 82, 495–507.
Baker, S., Spitz, W., 1970. An evaluation of the hazard created by natural death at the wheel. NEJM 283, 405–417.
Barraclough, B.M., Bunch, J., Nelson, B., 1974. A hundred cases of suicide; clinical aspects. British Journal of Psychiatry 125, 355–373.
Beaumont, G., 1989. Suicide and antidepressant overdose in general practice. British Journal of Psychiatry 155 (Suppl. 6), 27–31.
Bergman, A.B., Ray, C.G., Pomeroy, M.A., Wahl, P.W., Beckwith, J.B., 1972. Studies of sudden infant death syndrome in King county, Washington. Pediatrics 49, 860–870.
Bennett, A.T., Collins, K.A., 2001. Elderly suicide: a 10-year retrospective study. American Journal of Forensic Medicine and Pathology 22 (2), 169–172.
Benos, D.J., 2007. The ups and downs of peer review. Advances in Physiology Education 31, 145–152.
Black, H.C., 1990. Black's Law Dictionary, sixth ed. West Publishing Co., St. Paul, Minn.
Blair, P.S., Fleming, P.J., Smith, I.J., Platt, M.W., Young, J., Nadin, P., Berry, P.J., Golding, J., 1999. Babies sleeping with parents: case-control study of factors influencing the risk of the sudden infant death syndrome. British Medical Journal 319 (7223), 1457–1461.
Bootman, J., 2000. To err is human. Archives of Internal Medicine 160 (21), 3189–3192.
Buchholz, U., Mermin, J., Rios, R., 2002. An outbreak of food-borne illness associated with methomyl-contaminated salt. JAMA 288, 604–610.
Buck, U., Albertini, N., Naether, S., Thali, M., 2007. 3D documentation of footwear impressions and tyre tracks in snow with high resolution optical surface scanning. Forensic Science International 172, 157–164.
Buehler, J.W., Smith, L.F., Wallace, E.M., 1985. Unexplained deaths in a children's hospital: a epidemiological assessment. New England Journal of Medicine 313, 211–216.
CDC Medical Examiner/Coroners Map. Available at: www.cdc/gov/mmwr/Preview/mmwrhtml/rr5308a1.htm.
CDC SIDS Data. Available at: www.cdc/gov/SIDS/PDF/SUIDSIforms.pdf.
Criminal and Epidemiological Investigation Handbook, 2011. Department of Justice FBI and Department of Justice.
Diekstra, R.F.W., 1993. The epidemiology of suicide and parasuicide. Acta Psychiatrica Scandinavica supplementum 371, 9–20.
DiMaio, V.J., 2001. Forensic Pathology, second ed. CRC Press, Boca Raton.
DiMaio, V., 1985. Gunshot Wounds. Elsevier, New York.
Drug Abuse Warming Network: Development of a New Design, August 2002. Available at: www.samhsa.gov.
Elfawal, M.A., 1990. Cultural influence on the incidence and choice of method of suicide in Saudi Arabia. American Journal of Forensic Medicine and Pathology 20 (2), 163–168.

FARS, April 2008. Analytic Reference Guide 1975-2007. U.S. Department of Transportation. National Highway Traffic Safety Administration. DOT HS 810 937.
Fatality Analysis Reporting System. Available at: http://www:nrd.nhtsa.dot.gov/departments/nrd-01/summaries/FARS_98.html.
Fatality Analysis Reporting System. Available at: http://www:iihs.org/research/fatality_facts_2006/fars.html
Filiano, J.J., Kinney, H.C., 1994. A perspective on neuropathologic findings in victims of the sudden infant death syndrome: the triple-risk model. Biology of Neonate 65, 194–197.
Flick, L., White, D.K., Vemulapalli, C., Stulac, B.B., Kemp, J.S., 2001. Sleep position and the use of soft bedding during bed sharing among African American infants at increased risk for sudden infant death syndrome. Pediatrics 138 (3), 338–343.
Franks, A., Sacks, J.J., Smith, J.D., 1987. A cluster of unexplained cardiac arrests in a surgical care unit. Critical Care Medicine 15, 1075–1076.
Felthous, A.R., Hempel, A., 1995. Combine homicide-suicide: a review. Journal of Forensic Science 40 (5), 846–857.
Felthous, A.R., Hempel, A.G., Heredia, A., Freeman, E., Goodness, K., Holzer, C., Bennett, T.J., Korndorffer, W.E., 2001. Combined homicide-suicide in Galveston County. Journal of Forensic Science 46 (3), 586–592.
Forensic Epidemiology, 2007. Kentucky Law Enforcement. Fall, pp. 30–34.
Gallerani, M., Avato, F.M., Dal Monte, D., 1996. The time for suicide. Psychological Medicine 26, 867–870.
Glass, T., 2003. Forensic odontology. In: James, S., Nordby, J. (Eds.), Forensic Science. CRC Press, Boca Raton, FL, pp. 61–62.
Gresham, G., Turner, A., 1979. Post-mortem Procedures. (An Illustrated Textbook). Year Book Medical, Chicago.
Goldman, M.B., 1994. Sudden infant death syndrome: back to sleep campaign. Caring 13 (Pt. 2), 52–55.
Goodman, R.A., 2003. Forensic epidemiology: law at the intersections of public health and criminal investigation. Journal of Law, Medication, and Ethics 31, 684–700.
Griffith, E.E., Bell, C.C., 1989. Recent trends in suicide and homicide among blacks. JAMA 16, 2265–2269.
Gunnell, D., Nowers, M., 1997. Suicide by jumping. Acta Psychiatrica Scandinavica 96 (1), 1–6.
Haglund, W.H., Sorg, M.H. (Eds.), 1997. Forensic Taphonomy: The Postmortem Fate of Human Remains. CRC Press, Boca Raton, FL.
Hauck, F.R., Merrick Moore, C., Herman, S., Donovan, M., Kalelkar, M., Kaufer Cristoffel, K., Hoffman, H., Rowley, D., 2002. The contribution of prone sleeping position to the racial disparity in sudden infant death syndrome: the Chicago infant mortality study. Pediatrics 110 (4), 772–780.
Hauck, F.R., Herman, S.M., Donovan, M., Iyasu, S., Merrick Moore, C., Donoghue, E., Kirschner, R.H., Willinger, M., 2003. Sleep environment and the risk of sudden infant death syndrome in an urban population: the Chicago infant mortality study. Pediatrics 111 (5), 1207–1214.
Hayward, R., Hofer, T., 2001. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. Journal of the American Medical Association 286 (4), 415–420.
Hanzlick, R., 2007. Death Investigation: Systems and Procedures. CRC Press, Boca Raton.
Hanzlick, R., Parrish, R.G., 1993. The failure of death certificates to record the performance of autopsies. JAMA 269, 27.
Healey, M., Shackford, S., Osler, T., 2002. Complications in surgical patients. Surgery 137 (5), 611–618.
Ho, T.P., Yip, P.S.F., Chiu, C.W.F., 1998. Suicide notes: what do they tell us? Acta Psychiatrica Scandinavica 98 (6), 467–473.
Homicide Trends in the U.S.: Additional Information about the Data (2008). U.S. Department of Justice, Office of Justice Program, Bureau of Justice Statistics. Available at: http://www.ojp.usdoj.gov/bjs/homicide/addinfo.htm.
Hoyert, D.L., Arias, E., Smith, B.L., 2000. Deaths: final data for 1999. National Vital Statistics Reports 49 (8), 1–114. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr49.
Hunt, L.W., Silverstein, M.D., Reed, C.D., 1993. Accuracy of the death certificate in a population-based study of asthmatic patients. JAMA 269, 1947–1952.
Iserson, K., 1994. Death to Dust. Galen Press, Ltd, Tucson, AZ.
Istre, G.R., Gustafson, T.L., Baron, R.C., 1985. A mystery cluster of deaths and cardio pulmonary arrests in a pediatric intensive care unit. New England Journal of Medicine 313, 205–211.
Isometsa, E.T., 2000. Suicide. Current Opinion in Psychiatry 13 (2), 143–147.
Isometsa, E.T., Heikkinen, M.E., Marttunen, M.J., 1995. The last appointment before suicide: is suicide intent communicated? American Journal of Psychiatry 152 (6), 919–922.

Isometasa, E.T., Heikkinen, M., Henriksson, M., 1997. Differences between urban and rural suicides. Acta Psychiatrica Scandinavica 95 (4), 297–305.
Kemp, J.S., Unger, B., Wilkins, D., Psara, R., Ledbetter, T., Graham, M.A., Case, M., Thach, B.T., 2000. Unsafe sleep practices and an analysis of bedsharing among infants dying suddenly and unexpectedly: results of a four-year, population-based, death-scene investigation study of sudden infant death syndrome and related deaths. Pediatrics 106, 3–9.
Kevan, S.M., 1978. The seasonal behavior of Canadians. Canadian Mental Health 26, 16.
Kevan, S.M., 1980. Perspectives on season of suicide: a review. Social Science and Medicine 14, 369–378.
Kircher, T., Anderson, R.E., 1987. Cause of death: proper completion of the death certificate. JAMA 258, 349–352.
Kircher, T., Nelson, J., Burdo, H., 1985. The autopsy as a measure of accuracy of the death certificate. New England Journal of Medicine 313, 1263–1269.
Klerman, G.L., 1984. Clinical epidemiology of suicide. Journal of Clinical Psychiatry 48 (Suppl. 2), 33–38.
Knight, B., 1996. Forensic Pathology, second ed. Oxford University Press, Inc., New York, NY.
Koehler, S.A., 2008. SIDS deaths: the role of forensic nurses. Journal of Forensic Nursing 3, 87–92.
Koehler, S.A., 2007. The role of suicide notes in death investigation. Journal of Forensic Nursing 87–88, 92.
Koehler, S.A., 2005. Using medical examiner/coroner-generated death certificates in research: advantages and limitations. Journal of Forensic Nursing 1, 133–135.
Koehler, S.A., 2005. The use of Coroner's/Medical Examiner's data by forensic nurses. Journal of Forensic Nursing 1, 37–38.
Koehler, S.A., 2005. Autopsy report. In: Medical-Legal Aspects of Medical Records Analysis. Lawyers and Judges Publishing Co., Tucson.
Koehler, S.A., Brown, P.A., 2009. Foundations of Forensic Epidemiology: Medical Examiner/Coroners Perspective. CRC Press, Boca Raton, FL.
Koehler, S.A., Weiss, H.B., Shakir, A., Shaeffer, S., Ladham, S., Rozin, L., Dominick, J., Wecht, C.H., 2006. Accurately assessing elderly fall deaths using hospital discharge and vital statistics data. Journal of Forensic Medicine and Pathology 27, 30–35.
Kung, H.C., Hoyert, D.L., Xu, J.Q., Murphy, S.L., 2008. Deaths: final data for 2005. National Vital Statistics Reports 56 (10) (Hyattsville MD, National Center for Health Statistics).
Lane, B., 2004. The Encyclopedia of Forensic Science. Magpie Books, London.
Last, J.M., 1988. A Dictionary of Epidemiology, second ed. Oxford University Press, New York.
Leenaars, A.A., Wilde, E.J., Wenckstern, S., 2001. Suicide notes of adolescents: a life span comparison. Canadian Journal of Behavioural Science 33 (1), 47–57.
Loue, S., 2002. Case Studies in Forensic Epidemiology. Kluwer Academic/Plenum Publishers, New York.
Loue, S., 1999. Forensic Epidemiology. Southern Illinois University Press, Carbondale & Edwardsville.
Mino, A., Bousquet, A., Broers, B., 1999. Substance abuse and drug-related death, suicidal ideation, and suicide: a review. Crisis 20 (1), 28–35.
Minino, A.M., Smith, B.L., 2000. Deaths: final data for 2000. National Vital Statistics Reports 49 (12), 1–40. Available:http://www.cdc.gov/nchs/datawh.
Moscicki, E.K., 1989. Epidemiologic surveys as tools for studying suicidal behavior: a review. Suicide and Life Threatening Behavior 19 (1), 131–146.
Moscicki, E.K., 1995. Epidemiology of suicidal behavior. Suicide and Life Threatening Behavior 25 (1), 22–35.
Mortensen, P.B., 1999. Can suicide research lead to suicide prevention? Acta Psychiatrica Scandinavica 99 (6), 397–398.
Moscicky, E.K., 1997. Identification of suicide risk factors using epidemiologic studies. Psychiatric Clinics of North America 20 (3), 499–517.
Moscicki, E.K., 1994. Gender differences in completed and attempted suicides. Annals of Epidemiology 4, 152–158.
Montague, P. Medical Mistakes: Health and beyond. Available online at: http://www.chetday.com/medmistakes.html.
Nakamura, S., Wind, M., Danello, M.A., 1999. Review of hazards associated with children placed in adult beds. Archives of Pediatrics and Adolescent Medicine 153, 1019–1023.
National Institutes of Child Health and Human Development, National Institutes of Health, 2001. From Cells to Selves, Targeting Sudden Infant Death Syndrome (SIDS): A Strategic Plan (Bethesda, MD).
National SIDS/Infant Death Resource Center, 2004a. What Is SIDS? U.S. Department of Health and Human Services, Health Resources and Services Administration, Vienna, VA.

National SIDS/Infant Death Resource Center, 2004b. SIDS Deaths by Race and Ethnicity 1995–2001. U.S. Department of Health and Human Services, Health Resources and Services Administration, Vienna, VA.
Neeleman, J., Mak, V., Wessely, S., 1997. Suicide by age, ethnic group, coroners' verdicts and country of birth. A three-year survey in inner London. British Journal of Psychiatry 171 (11), 463–467.
Neeleman, J., Wessely, S., 1999. Ethnic minority suicide: a small area geographical study in south London. Psychological Medicine 29 (2), 429–436.
Nordenberg, T., September–October 2000. Make No Mistake: Medical Errors Can Deadly Serious. FDA Consumer Magazine. Available online at: http://www.fda.gov/fdac/features/2000/500_err.html.
Omalu, B.I., Macurdy, K.M., Koehler, S.A., Agumadu, U.H., Shakir, A.M., Rozin, L., Wecht, C.H., 2005. Forensic pathology and forensic epidemiology of suicides in Allegheny county, Pennsylvania: a 10-year retrospective review (1990–1999). Forensic Science, Medicine, and Pathology 1, 125–137.
Pear, R., December 20, 1999. Report Details Medical Errors in V.A. Hospitals. The National Gulf War Resource Center, Inc. Available online at: http://www.ngwrc.org/news/content/mondec201200021999.asp.
Pestaner, J.P., Schomburgh, D., 1997. Suicidal gunshot wounds of the head. Laboratory Investigation 76 (1), 6.
Peterson, B., Petty, C., 1962. Sudden natural death among automobile drivers. Journal of Forensic Sciences 7, 274–279.
Pharmacology: Half-life of Drugs, 2008. Available at: www.nottingham.ac.uk/nursing/sonet/rlos/bioproc/halflife/index.html.
Physicians' Handbook on Medical Certification of Death, April 2003. Revision. Department of Health and Human Services. Center for Disease Control and Prevention. National Center for Health Statistics DHHS Publication No. (PHS) 2003-1108. Hyattsville, Maryland.
Ramsay, S., 1996. Suicides among elderly on the up in USA. Lancet 347 (8995), 182.
Rasinski, K., Kuby, A., Bzdusek, S., Silvestri, J., Weese-Mayer, D., 2003. Effect of a sudden infant death syndrome risk reduction education program on risk factor compliance and information sources in primarily black urban communities. Pediatrics 111 (4), e347–e354.
Retterstol, N., 1992. Suicide in Nordic countries. Psychopathology 25, 254–265.
Robbins, C., 1989. Robbins Pathologic Basis of Disease, fourth ed. W.B. Saunders Company, Philadelphia.
Robins, E., Murphy, G.E., Wilkinson, R.H., 1959. Some clinical considerations in the prevention of suicide based on a study of 134 successful suicides. American Journal of Public Health 49, 888–898.
Rosewater, K.M., Burr, B.H., 1998. Epidemiology, risk factors, intervention and prevention of adolescent suicide. Current Opinion in Psychiatry 10, 338–343.
Sacks, J.J., Herdon, J.L., Lieb, S.H., 1988a. A cluster of unexplained deaths in a nursing home in Florida. American Journal of Public Health 78, 806–808.
Sacks, J.J., Stroup, D.F., Will, M.W., 1988b. A nursing-associated epidemic of cardiac arrests in an intensive care unit. JAMA 259, 689–695.
Saferstein, R., 2007. Criminalistics: An Introduction to Forensic Science (New Jersey).
Saferstein, R., 2009. Forensic Science: From the Crime Scene to the Crime Lab. Pearson Prentice Hall, New Jersey.
Scheers, N.J., Rutherford, G.W., Kemp, J.S., 2003. Where should infants sleep? A comparison of risk for suffocation of infants sleeping in cribs, adult beds, and other sleeping locations. Pediatrics 112 (4), 883–889.
Scragg, R.D., Mitchell, E.A., Taylor, B.J., Stewart, A.W., Ford, R.P., Thompson, J.M., Allen, E.M., Becroft, D.M., 1993. Bed sharing, smoking, and alcohol in the sudden infant death syndrome. British Medical Journal 307, 1312–1318.
Sigurdson, E., Staley, D., Matas, M., 1991. A five year review of youth suicide in Manitoba. Canadian Journal of Psychiatry 39, 397–403.
Spicer, R.S., Miller, T.R., 2000. Suicide acts in 8 states: incidence and case fatality rates by demographics and method. American Journal of Public Health 90 (12), 1885–1891.
Spitz, W.U., Fisher, R.S., 1980. Medicolegal Investigation of Death, third ed. Charles C. Thomas Publisher, Springfield Illinois.
Stross, J.K., Shasby, D.M., Harlan, W.R., 1976. An epidemic of mysterious cardiopulmonary arrest. New England Journal of Medicine 295, 1107–1110.
Taylor, B., 1991. A review of epidemiologic studies of SIDS in Southern New Zealand. Journal of Paediatrics and Child Health 27, 344–348.
The Allegheny County Office of the Coroner: Statistical Report, 2005.
The Facts: Vehicular Homicide and the Impaired Driver, 2004 The Facts: Vehicular Homicide and the Impaired Driver, 2004. Available at: http://www.nhtsa.dot.gov/people/outreash/safesobr/13qp/facts/facthom.html.2004.

Torok, T.J., Tauxem, R.V., Wise, P.R., 1997. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. JAMA 278, 389–395.
Unger, B., Kemp, J.S., Wilkins, D., Psara, R., Ledbetter, T., Graham, M., Case, M., Thach, B.T., 2003. Racial disparity and modifiable risk factors among infants dying suddenly and unexpectedly. Pediatrics 11 (2), e127–e131. Available at: http://pediatrics.aappublications.org/cgi/content/full/111/2/e127.
U.S. Census Bureau [State and National Data (C2SS)], 2001. United States Department of Commerce, Washington, DC. Updated September 11, 2001.
U.S. Consumer Product Safety Commission, 1999. Soft Bedding May Be Hazardous to Babies. Consumer Product Safety Alert (Publication number 5049). Retrieved from: http://cpsc.gov/CPSCPUB/PUBS/5049.html. on June 30, 2004.
U.S. Consumer Product Safety Commission. Safe Sleep Campaign 2000. SIDS Awareness Survey. Retrieved from: http//www.cpsc.gov/LIBRARY/SIDSsurvey.pdf, on June 30, 2004.
U.S. Consumer Product Safety Commission Office of Information and Public Affairs, April 8, 1999. Recommendations Revised to Prevent Infant Deaths from Soft Bedding. Retrieved from: http://www.cpsc.gov/CPSCPUB/PREREL/PRHTML99/99091.html.
U.S. Drug Enforcement Administration. Available at: http://www.usdoj.gov/dea/agency/mission.htm.
U.S. Public Health Service, 1999. The Surgeon General's Call to Action to Prevent Suicide. Washington, DC.
Vermulapalli, C., Grady, K., Kemp, J.S., 2004. Use of safe cribs and bedroom size among African American infants with a high rate of bed sharing. Archives of Pediatrics and Adolescent Medicine 150, 286–289.
Wecht, C.H., 2004. Crime Scene Investigation: Crack the Case with Real-Life Experts. Reader's Digest, New York.
Wecht, C.H., 2002. Allegheny County Coroner's Office. Statistical Report for 2002. Coroner's Office, 542 Fourth Avenue, Pittsburgh, Pa. 15219.
Wecht, C.H., Koehler, S.A., 2015. Death in vehicular accidents. In: Wecht, C.H. (Ed.), Forensic Sciences. Matthew Bender & Co., Inc., New York (In Press).
Wecht, C.H., Koehler, S.A., 2007a. History of the Death Investigation System: US. Encyclopedia of Forensic and Legal Medicine. Elsevier Ltd, Oxford.
Wecht, C.H., Koehler, S.A., 2007b. Determination of Fitness to Drive Encyclopedia of Forensic and Legal Medicine. Elsevier Ltd, Oxford.
Wecht, C.H., Koehler, S.A., 2007c. Determination of Fitness to Drive: Driving Offences. Encyclopedia of Forensic and Legal Medicine. Elsevier Ltd, Oxford.
West, I., Nielsen, G., Gilmore, A., 1968. Natural death at the wheel. JAMA 205, 266–272.
Wedgwood, R.J., 1972. Review of USA experience. In: Camps, F.E., Carpenter, R.G. (Eds.), Sudden and Unexpected Death in Infancy (Cot Deaths). Wright, Bristol, England, p. 28.
Welch, S.S., 2001. A review of the literature on the epidemiology of parasuicide in the general population. Psychiatric Services 52 (3), 368–375.
Wigfield, R., Fleming, P., Berry, P., Rudd, P., Golding, J., 1992. Can the fall in Avon's sudden infant death rate be explained by changes in sleep position? British Medical Journal 304, 282–283.
Willinger, M., James, L.S., Catz, C., 1991. Defining the sudden infant death syndrome (SIDS): deliberations of an expert panel convened by the National Institute of Child Health and Human Development. Pediatric Pathology 11, 677–684.
Willinger, M., Hoffman, J.J., Hartford, R.B., 1994. Infant sleep position and risk for sudden infant death syndrome: report of meeting held January 13 and 14, 1994, National Institutes of Health, Bethesda, MD. Pediatrics 93 (5), 814–819.
Willinger, M., Hoffman, H., Wu, K.T., Hou, J.R., Kessler, R.C., Ward, S.L., Keens, T.G., Corwin, M.J., 1998. Factors associated with the transition to nonprone sleep positions of infants in the United States. The National Infant Sleep Position Study. Journal of the American Medical Association 280 (4), 329–335.
Willinger, M., Ko, C.W., Hoffman, H., Kessler, R., Corwin, M., 2000. Factors associated with caregivers' choice of infant sleep position, 1994-1998: the National Infant Sleep Position Study. Journal of the American Medical Association 283 (16), 2135–2142.
Wright, R., 2003. Role of the forensic pathologist. In: James, S., Nordby, J. (Eds.), Forensic Science. CRC Press, Boca Raton, FL, pp. 19–20.
World Health Organization, 1996. Prevention of Suicide: Guidelines for the Formulation and Implementation of National Strategies. World Health Organization, Geneva.

This page intentionally left blank

CHAPTER

8

Injury Biomechanics

D. Kieser
University of Otago, Christchurch, Canterbury, New Zealand
D. Carr
Cranfield University at the Defence Academy of the United Kingdom, Shrivenham, United Kingdom
M. Jermy
University of Canterbury, Christchurch, New Zealand
A. Mabbott
Cranfield University at the Defence Academy of the United Kingdom, Shrivenham, United Kingdom
J. Kieser
University of Otago, Dunedin, Otago, New Zealand

OUTLINE

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00008-2

INTRODUCTION

This chapter is intended to provide an overview of injury biomechanics for use in a forensic epidemiologic analysis of risk or causality. Thus, the purpose of the chapter is to provide basic background information and principles regarding biomechanics, rather than to delve into highly advanced concepts that might be only applicable in an experimental or theoretical setting. We have, therefore, consciously avoided all but the most rudimentary mathematical formulae.

BACKGROUND

Biomechanics is "the study of the structure and function of biological systems by means of the methods of mechanics" (Hatze, 1974).

It is impossible to introduce biomechanics without discussing Sir Isaac Newton's (1642–1727) three laws. These laws are based upon two assumptions. First is the concept of equilibrium, where in an ideal situation all of the forces acting upon an object equate to zero and thus no change in velocity of the object occurs. The second is the conservation of energy, which is simply to say that energy cannot be created or destroyed, it can only be converted from one form to another.

Newton's first law (the law of inertia) asserts that an object either remains at rest or continues to move at a constant velocity, unless acted on by an external force. Note that this inertia pertains not only to linear motion, but also rotation. The mass moment of inertia is proportional to both the total mass of the object and the distribution of this mass around the axis of rotation. Put more simply, there is a greater resistance to a change in rotation if the mass is further from the axis of rotation and hence, why more damage is sustained by a victim hit with the end of a swinging baseball bat than if hit by the handle of the bat (Fig. 8.1).

Newton's second law (the law of acceleration) states that the vector sum of the forces (F) on an object is equal to the mass (m) of the object multiplied by the acceleration vector (a) of the object

$$F = ma$$

This formula allows us to mathematically model injury patterns to predict the severity of injury. For example, a punch sustained by a bare fist will cause more damage than one in a boxing glove, because the time for the impact to occur is longer with the gloved fist, thus reducing its acceleration. In contrast, a victim hit by a standard 2.7 g table tennis ball is likely to sustain less of an injury than if hit by a standard 45.93 g golf ball of the same dimensions and velocity, simply because the weight of the ball is greater.

This also introduces the critical concept of momentum (p), which is defined as a vector of mass (m) times velocity (v).

$$p = mv$$

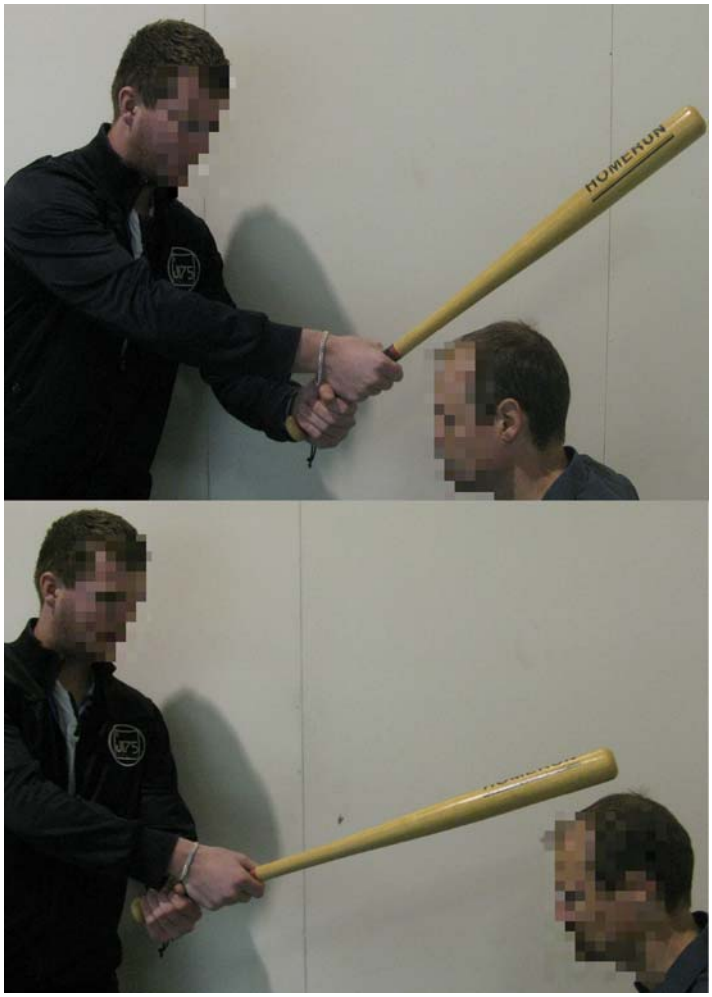


FIGURE 8.1 Illustration of person hit by bat.

Newton's third law (the law of reaction) states that for every action there is an equal and opposite reaction. This law is critical in forensic biomechanics. For example, consider a rock being bashed into a victim's skull; the action and reaction forces are equal in magnitude but opposite in direction, thus the force that the skull experiences is of equal magnitude, but opposite in direction to the force that the rock experiences. Since acceleration is the rate of change of velocity, force equals the rate of change of momentum and thus, for every bit of momentum the rock loses during the collision, the skull gains this momentum so that the total momentum remains the same.

This example also introduces the importance of the material properties of the objects involved in collisions. Obviously the effect on the skull from such an event would be disastrous, whereas the rock may sustain minimal damage. To understand these effects, we need

to understand the often daunting concept of stress and strain. Stress (σ) is a measure of the force (F) acting over a defined area (A)

$$\sigma = F/A \quad \text{measured in pascals } (1 \text{ Pa} = 1 \text{ N/mm}^2)$$

The most common stresses are compressive or tensile, when the force on the object is either crushing or stretching the object, or shear stress, where the forces act parallel to each other, but in opposite directions (Fig. 8.2).

Strain (ε) is a measure of an object's shape change (such as change in length (Δl)) in relation to an external force applied to the object divided by its resting dimensions (initial length (l)).

$$\varepsilon = \Delta l/l \quad \varepsilon \text{ is dimensionless and often reported as a percentage}$$

The relationship between stress (force) and strain (deformation) differs between different materials and is illustrated best on a stress/strain graph (Fig. 8.3).

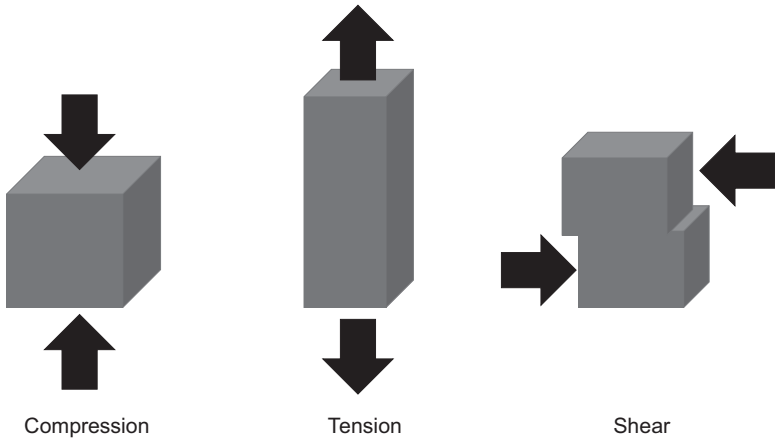


FIGURE 8.2 Illustration of compressive, tensile, and shear stresses.

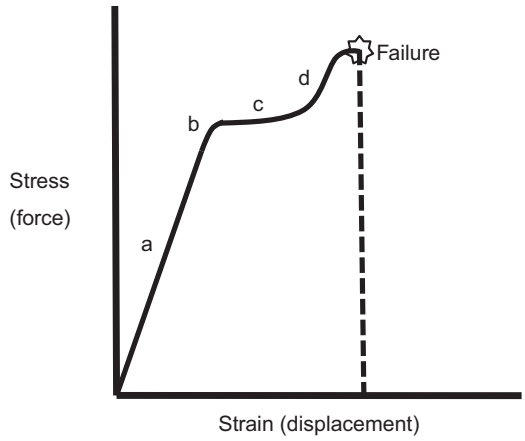


FIGURE 8.3 Stress/strain graph.

By convention, stress is positioned on the *Y*-axis and strain on the *X*-axis. To understand this graph, consider initially, an elastic band. If you gently stretch and release the band, it returns to its original dimensions. This is the initial linear relationship between stress and strain, where the response is termed elastic (a). In other words, if the force (stress) on the object were removed, the strain (deformation) would resolve.

If the elastic band continues to be stressed, there is a small period where the stretching becomes harder without gaining much length. Here the curve is no longer linear but remains elastic (b). At the end of this period, we meet the elastic limit, which is the maximal stress that can be applied before causing permanent deformation to the object.

With a little more force the object deforms permanently and will not return to its original shape; this is termed plastic deformation. If we continue to stretch the object, we hit a phase where we can easily continue to stretch it with minimal force (c). The beginning of this phase is called the yield point, which is the highest point of the stress/strain graph, and the end is termed the ultimate strength, which describes the maximal amount of deformation before the object begins to fail. Most biological tissues progressively fade beyond the yield point and ultimately fail at a variable point beyond the ultimate strength.

Young's modulus or the modulus of elasticity (E) is the ratio of stress (force) to strain (deformation) during the linear elastic phase of the stress/strain graph.

$$E = \text{Stress/Strain} = \sigma/\varepsilon$$

Poisson's ratio describes the ratio of compression on an object to its shape change. To understand this, consider a jelly baby. If you compress the jelly baby, it tends to expand, whereas if you stretch it, it thins. Poisson's ratio (v) is the percentage of expansion divided by the percentage of compression.

Some materials are isotropic, that is to say, that a force in any direction onto an object results in the same stress/strain relationship. For example, glass and metals are isotropic materials.

Other materials are anisotropic, meaning that their stress/strain relationship depends on the direction of force applied. For example, wood, which fractures more easily along its grain than across it, or bone, which in the adult is weaker in tension than compression. Understanding this and simplifying long bones such as the femur, by considering them as tubes, allows us to understand the majority of long-bone fracture patterns seen in forensic medicine. If we bend the bone, we create tension forces on the far side of the bone, but compression forces on the near surface (Fig. 8.4). Knowing that bone is weaker in tension than compression, we can predict that the fracture will initiate on the far surface and proliferate toward the near surface.

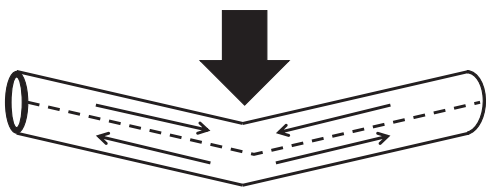


FIGURE 8.4 Graphical depiction of a long bone. Note flexion on the far side of the impact and compression on the near cortex.

It is important to remember that the biomechanical properties of tissues and the injuries they sustain are not only dependent on the mechanical properties discussed previously, but also the shape of the impacted structure, its direction, rate, and frequency of impact and the mode of force applied.

While a material's strength or stiffness is important, the most important factor in prevention of failure is a material's toughness. Toughness is determined by a material's ability to prevent crack formation and propagation when loaded (Gordon, 1991). There are two ways in which materials prevent fracture propagation. Firstly, they directly resist the crack from expanding and thus the crack will only propagate when the stress intensity exceeds the bond strength of the material (termed fracture toughness and relates to the stress intensity factor (SIF)). Secondly, as a crack propagates it loses energy, and thus controlled crack propagation soaks up the energy imparted on the material preventing its catastrophic failure. A good example of this is shatterproof glass as in a windshield, which may form multiple cracks, but does not catastrophically fail and fall onto the driver.

Understanding these few biomechanical principles will aid the forensic investigator in understanding injury patterns. We shall discuss these concepts further throughout the chapter.

TYPES OF TRAUMA

It is important to distinguish between sharp force, blunt force, and ballistic trauma. As the name suggests, sharp force trauma is associated with injuries inflicted by sharp objects such as knives and swords. Similarly, blunt force trauma is caused by blunt objects such as hammers or bats. Ballistic trauma is caused by projectiles such as bullets and arrows.

When assessing injuries, one needs to determine the weapon used. For sharp force trauma, we expect sharp incisions in the skin with minimal surrounding soft tissue injury. Sharp cuts or scoring of the bone may also be seen. In contrast, blunt force trauma is likely to cause significant surrounding soft tissue trauma and skeletal fracture dependent on the extent of force transmitted. Ballistic trauma causes a penetration wound, often with minimal skin injury, but significant deep tissue injury. Associated fractures depend on the path of the projectile and the energy transmitted, but tend to be shattered into multiple pieces (highly comminuted).

BIOMECHANICS OF SKIN AND SOFT TISSUE INJURY

Skin is composed of three main layers. The outer layer is called the epidermis and is composed of layers of cells (keratinocytes) that ultimately die as they get pushed toward the skin's surface and are therefore constantly replaced. This layer is of variable thickness, depending on anatomical location and acts as a protective layer against injury and microbial attack.

Beneath the layer of epidermis is the dermis. The interface between these two layers is undulating to prevent shear forces between the layers and consists of a basement membrane with small projections anchoring the layers together, called rete ridges (Briggaman, 1982). The dermis itself is a well-hydrated (60–70% water composition) vascular region composed of dense fibroelastic tissue, containing both elastin and collagen fibers as well as hair follicles,

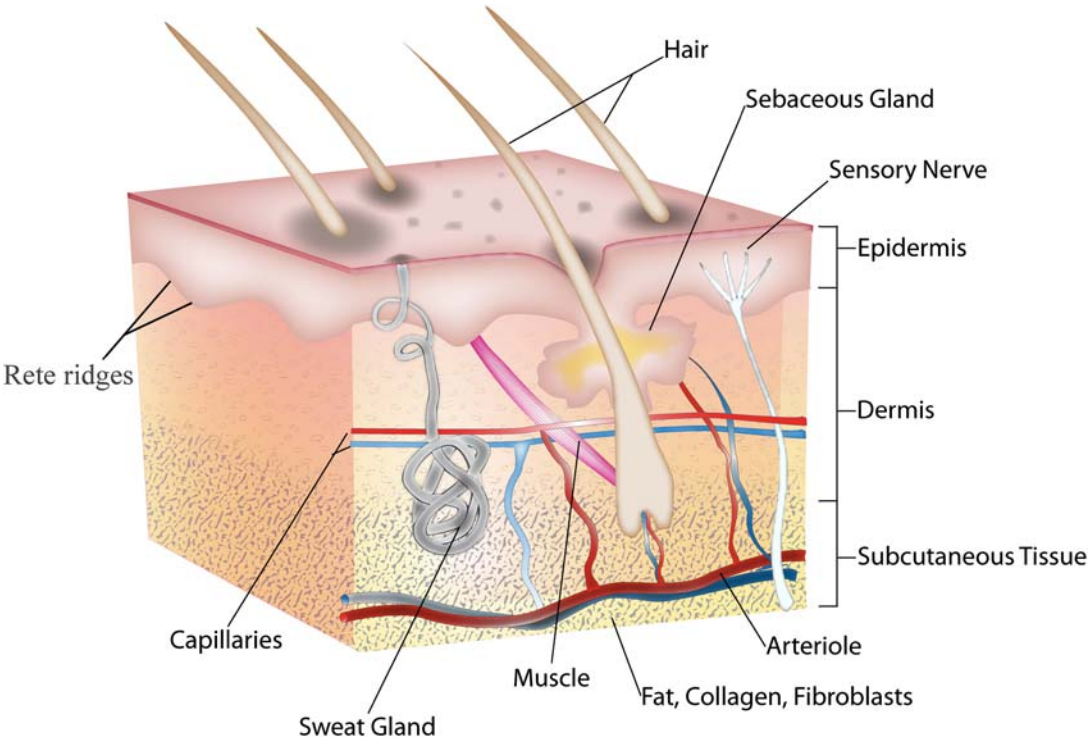


FIGURE 8.5 Graphical depiction of skin. *Modified from Kieser, J.A., Taylor, M., Carr, C., 2013. Forensic Biomechanics. Wiley-Blackwell, Chichester, UK.*

sweat, and sebaceous glands. The collagen fibers provide the skin with its structural integrity, through its tensile strength, particularly during high loads. In contrast, the elastin fibers allow a degree of pliability and recoil at low forces (Sanders et al., 1995).

Beneath the layer of dermis is the hypodermis or subcutaneous fat, composed of 75% adipose tissue held together in a framework of collagen, 20% water, and 5% protein. This layer provides the sponginess of skin as well as an energy store and insulation (Fig. 8.5).

As we descend deeper into the body, we encounter a fascial layer, which is composed of collagen and acts as a covering of the underlying muscles. This fascial layer is generally nonadherent to the underlying muscle, allowing these to move, but is adherent to the overlying fat. Vascular channels sprout through the fascia into the overlying fat, but shear forces in this region are poorly resisted and thus injury can shear through this interval and these blood vessels, causing significant hematomas.

Understanding the biomechanics of skin and soft tissue is highly complex as each layer is composed of multiple other layers and they all vary with the victim's age, sex, ethnicity, hydration and nutrition status, etc. However, we can think of skin as a "three-dimensional matrix of loosely bound coiled collagen and elastin fibers in an amorphous gel of ground substance" (Kieser et al., 2013). Simply, the collagen provides tensile strength, whereas the elastin provides recoil, with the fluid ensuring rapid recovery of shape

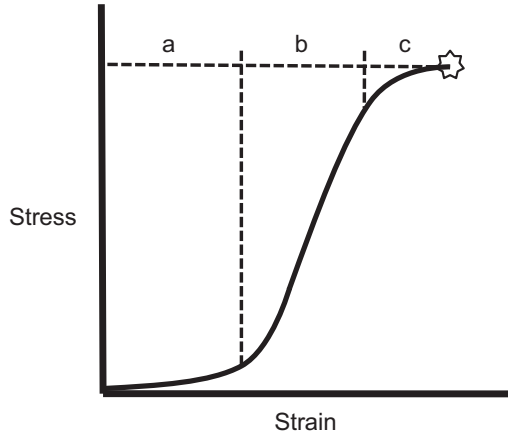


FIGURE 8.6 Classic triphasic stress/strain curve of skin.

(Dunn and Silver, 1983). At low stresses the skin acts as an elastic material, but at higher stresses it becomes increasingly viscoelastic. Thus, with force applied to the skin at a slow rate, the skin provides a low resistant force, whereas at higher rates of force transmission, the skin provides a higher resistant force. The viscous component ensures energy dissipation, whereas the elastic component allows energy storage.

As the skin is stretched, the collagen fibers initially uncoil (Fig. 8.6 region a), providing little resistance and allowing the skin to move (Oxlund et al., 1988). Once straight, the collagen fibers are no longer pliable and thus resist any further stretch (Fig. 8.6 region b). This gives skin its characteristic triphasic stress/strain curve. If, however, the force continues to be increased, collagen fibers start to slip between each other, resulting in the yield region of the stress/strain curve prior to failure (Fig. 8.6 region c) (Payne, 1991; Silver et al., 2001a,b).

This J-shaped stress/strain curve is characteristic of extremely tough materials, because the initial toe region (a) allows large extension with low stress, preventing fracture initiation, the following linear zone has increased stiffness requiring large stresses before a fracture can generate, and finally the area under the curve is far less than a comparable r-shaped curve (such as bone), hence there is less energy available for fracture propagation (Gordon, 2003).

Skin is, however, anisotropic, because the collagen fibers are prestressed in certain directions to allow for normal body function. The lines of this stress have been described in cadavers by Karl Langer in 1861 and validated in living individuals by Kraissl Langer in 1978 and are thus named Langer's lines (Langer, 1861, 1978a,b,c). These lines are important in forensic analysis, because incisions parallel to these lines have wounds that are usually minimally gapped, whereas perpendicular incisions result in significant wound gapping.

Neonatal skin is thin, tight, and delicate, but thickens and matures as the individual ages. In old age the skin loses its mechanical integrity because it becomes less hydrated, thinner, with less elastin and shallower rete ridges as well as reduced collagen quality. This makes skin less able to tolerate tangential forces in the elderly and makes the forensic reconstruction of traumatic events difficult in the elderly (Richey et al., 1988; Kieser et al., 2008a).

Wounding of skin and soft tissues can be classified into groups:

- Abrasions are grazes resulting from the scraping off of the superficial skin layers.
- Contusions are bruises/hematomas resulting from damage to the subcutaneous vessels, most commonly from blunt force trauma.
- Lacerations are tears or rips in the soft tissues usually from blunt force trauma involving the full thickness of skin resulting in the characteristic irregular wound margins, bruising, and skin tagging.
- Incisions are cuts inflicted by sharp objects such as knives.
- Puncture wounds are penetration injuries.
- Bite marks are complex puncture wounds from the application of teeth to the skin.

The factors determining penetration in sharp force trauma are the sharpness of the object, the speed of impact, and the skin resistance (Knight, 1975). Usefully, different weapons tend to inflict different wounds depending on their shape. For example, different screwdrivers (straight, star, Robertson, Pozidriv, and Phillips) leave wounds that mirror their shape, aiding forensic scientists in the determination of the weapon involved (Kieser et al., 2008b).

During blunt force trauma, collagen fibers may fracture, resulting in lacerations to the skin and soft tissue. However, deep soft tissue disruption may also occur with minimal to no skin damage. This makes visual assessment of such injuries subjective and poorly reliable (Altemeier, 2001; Hornor, 2005). The injury sustained from blunt force trauma is dependent on a number of factors including the energy transfer, the shape, mass and size of the object, the angle of impact, the amount of soft tissue, the rate of energy deposition, and the type of force (compression, shear, etc.), not to mention the victim factors such as age, ethnicity, and health.

It has been proposed that during blunt force trauma, the liquid component of the tissues get compressed, but because liquids are incompressible they get forced out into the surrounding tissues, fracturing the cell walls (Whittle et al., 2008). It has therefore been suggested that in contrast to sharp force trauma, which injures tissues from the outside in, blunt force trauma acts in an opposite manner damaging the tissues from inside out.

BIOMECHANICAL PROPERTIES OF BONE AND FRACTURE

Bone is composed of a hard outer shell (cortical bone) and an inner honeycomb scaffold (cancellous bone). This allows bone to be lightweight and tough. It is composed of an organic component (collagen) that is strong in tension, but weak in compression and an inorganic component (hydroxyapatite), which has excellent compressive strength, but is stiff and brittle. The collagen is fibrous and directional. Bone is therefore anisotropic; in other words, its mechanical properties differ depending on the direction of load. The cellular component of bone is made up of three principle cells, namely, osteoblasts, osteoclasts, and osteocytes. These cells allow bone to remodel in response to external forces and repair breakages.

Bone can be classified into two main types. These are tubular bones, such as the femur or tibia, and flat bones, such as the scapula or skull. Tubular bones are essentially long hollow tubes with thick outer cortical shells and varying amounts of inner cancellous bone, whereas

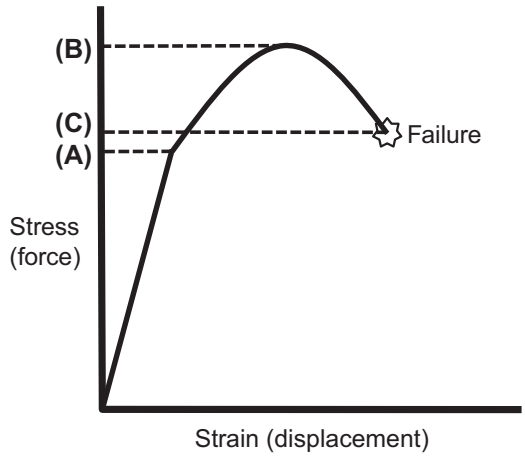


FIGURE 8.7 Stress/strain curve for bone (elastic limit (A), ultimate strength (B), failure (C)).

flat bones tend to be two thin shells of cortical bone sandwiching an inner meshwork of cancellous bone.

The cortical bone can be thought of biomechanically as being a stiff, brittle structure with a high fracture toughness and SIF. Bone has a classical r-shaped stress/strain curve (Fig. 8.7).

However, the bone is more complex, and the cortical bone is made of three main zones. On the outside are rings of bone that circumscribe the bone (outside circumferential system), inside this the bone is made up of columns of haversian systems or osteons which are essentially struts cemented together that run longitudinally along the bone and contain a central open canal. Further inside, the bone is the internal circumferential system, again made of circumferential rings, before it integrates with the cancellous (spongy) trabecular bone (Fig. 8.8).

This architecture optimizes its biomechanical strength and durability. The outer rings act as a stiff, but brittle structure with high fracture toughness. This prevents fracture initiation. However, if the SIF was to be exceeded and a fracture was to initiate, it would lose energy every time it encountered another ring of the outer circumferential system, which would again require the SIF to be overcome. If the fracture breaches this outer wall, it then encounters the tubes of the haversian systems. Here fractures propagate around the rings, through the cement lines between haversian canals, as well as through the central canals themselves. This redirects the fracture around the bone and uses up the energy of the fracture, hopefully preventing it from catastrophically fracturing the bone. Lastly, the inner circumferential system acts much like the outer circumferential system, requiring fractures to exceed their SIF in order to fracture through this layer.

Coating the inner layer of cortical bone is cancellous or spongy bone. Here a seemingly random array of bone trabeculae, appearing like a scaffold, with interdigitating spicules of bone, is laid down to optimally transfer the forces imparted on the bone through daily life. Higher densities of scaffold are aligned along the lines of stress, complying with the so-called Wolff's law. Fracturing through cancellous bone is limited by the fracturing of the individual trabecular struts that act to dissipate the energy and hopefully prevent catastrophic failure of the bone.

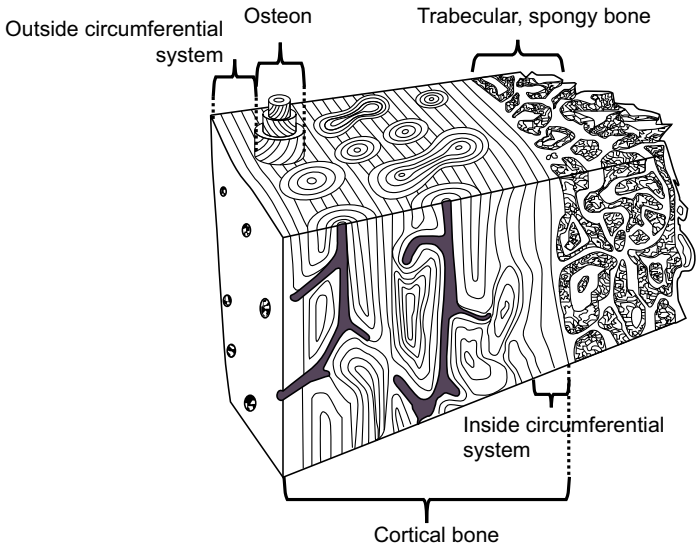


FIGURE 8.8 Graphical depiction of bone. *Modified from Kieser, J.A., Taylor, M., Carr, C., 2013. Forensic Biomechanics. Wiley-Blackwell, Chichester, UK.*

Flat bones have a higher cancellous to cortical bone ratio and, therefore, rely more heavily on the energy dissipation caused by fracturing of the individual trabecular struts, than on the cortical bone itself.

Fractures can be classified according to their material properties. Ductile fractures occur after a period of slow deformation. For example, steel, which initially deforms with gross plastic deformation before snapping. Brittle fractures occur in materials with low toughness, such as glass, and will suddenly fracture when the energy imparted onto the object exceeds its SIF. Viscoelastic fractures are more complex and tend to occur in biological materials. These are rate-dependent fractures that exhibit both a viscous and elastic response to the strain applied.

Remember that bone strength is not only influenced by biomechanical properties, but also anatomical and physiological factors such as size, shape, architecture, bone mineral density, and bone quality. To fracture, a finger takes far less force than the fracturing of a femur. Similarly, fracturing the same bone in a young healthy male is far more difficult than that of an elderly osteoporotic female. Therefore, when analyzing fractures, a degree of logic is required.

FRACTURE PATTERNS

The most basic fracture categories are open (in which the fractured bone is exposed), closed (the skin remains intact over the bone), complete (the fracture extends through the bone), and incomplete (the fracture extends partially through the bone).

How a bone break depends on the nature of the forces applied to it. Bending forces applied to tubular bones tend to cause a transverse fracture on the flexion side of the bone and a

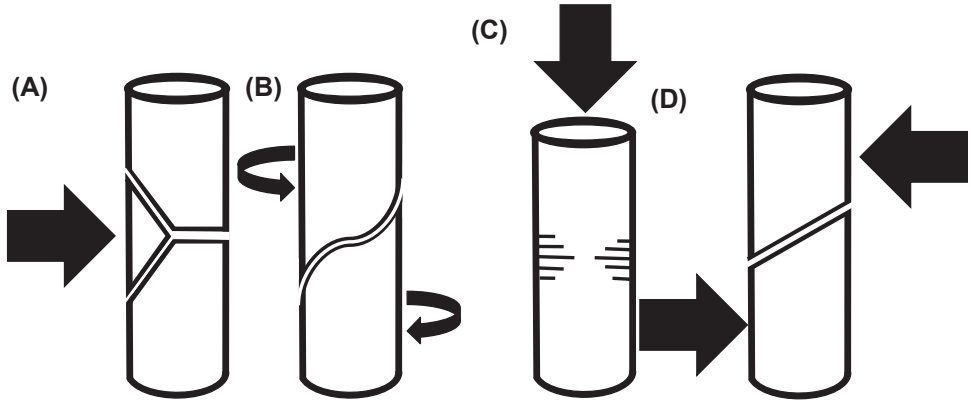


FIGURE 8.9 Simple bone fractures. Fracture sustained from a bending force (A), twisting force (B), compression force (C), and a shear force (D).

compression wedge (also known as a butterfly fragment) that is essentially ejected from the compression side.

Twisting injuries cause fractures that spiral up the bone. Direct compression along the long axis of the bone (axial loading) is well tolerated, but can cause compression of the cancellous bone, similar to a honeycomb candy bar being crushed. Shear forces tend to cause transverse or oblique fractures (Fig. 8.9).

With high-energy impacts, fractures tend to become more comminuted, meaning that they shatter into more than two pieces, as opposed to the low-energy simple fractures described previously. Flat bones tend to fracture with two further patterns. Firstly, stellate fractures which occur when a flat bone is impacted at high energy and shatters around the impact, similar to a windshield hit by a stone. Secondly, depressed fractures which occur when a flat bone is impacted with a blunt object, such as a hammer, creating a sharply punched out fragment, depressed from the surrounding bone (Fig. 8.10).

In children, tubular bones grow from their ends in regions called the growth plate. This is an area of weakness for the bone, and pediatric fractures have a propensity to occur at these sites. Another difference between the pediatric and adult bone is their remarkable pliability. Children's bones tend to bend rather than break. Furthermore, because children's bones are less mineralized than adult bones and contain a higher collagen content, their bones tend to be stronger in tension than in compression. Children therefore have two further fracture types. Buckle fracture, where the bone is bent and buckles on the compressed side as in the crushing of a coke can, and greenstick fractures where a little more force then splits the flexion surface as in breaking a green twig (Fig. 8.11).

As a very broad principle with many exceptions, while most pediatric fractures result from unintentional injury, some fractures are more likely to be associated with intentional injury than others (Kleinman and Schlesinger, 1997; Lonergan et al., 2003). Fractures that at least raise the index of suspicion of abusive injury include long-bone fractures in very young children who are unlikely to fall independently. Children less than 1 year of age are at greatest risk of nonaccidental injury, with an estimated 40–80% of long-bone fractures in this age group attributed to child abuse (Schwend et al., 2000). Another fracture that is more common

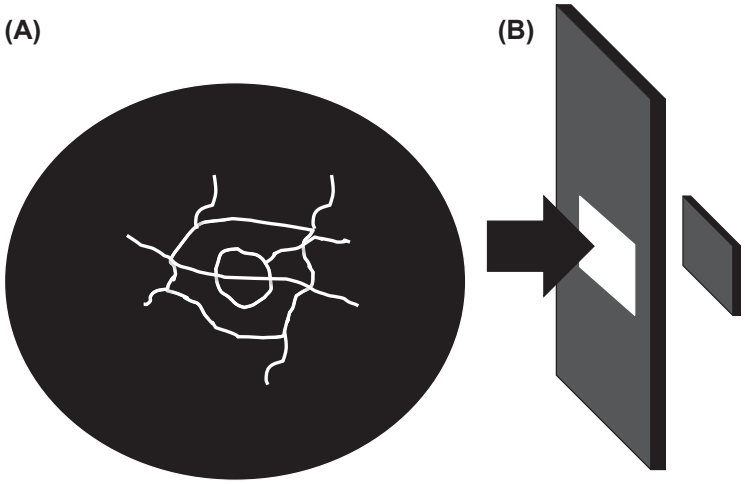


FIGURE 8.10 Additional fracture types seen with flat bones (stellate fracture (A), depressed fracture (B)).

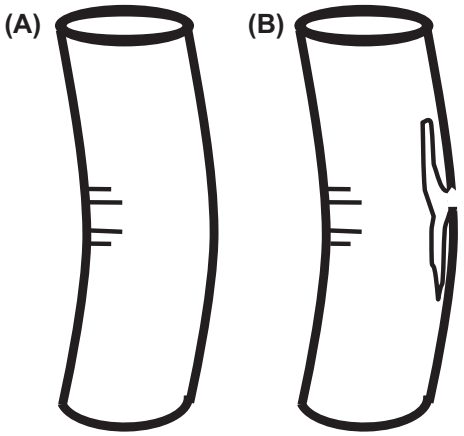


FIGURE 8.11 Additional pediatric fractures. Buckle fracture (A), greenstick fracture (B).

(but not exclusive) to abusive injury is a "classic metaphyseal lesion," which is a fracture of a long bone near but not past the growth plate (Pierce et al., 2004). Spiral fractures have been associated with child abuse, because they require a large torque to be applied to the bone. The magnitude of this torque is rarely encountered in normal play. Similarly, transverse fractures that require high-energy impacts have been associated with child abuse (Scherl et al., 2000). However, these can occur with tripping and falling and are thus not diagnostic of child abuse (Frazier, 2003). Understanding the biomechanics of a pediatric fracture allows us to understand the required force inflicted on the bone and compare this to the given patient history and validate its likelihood (Pierce et al., 2004).

The ability to determine the relationship between the time of fracture and the time of death is an important factor in the evaluation of fatality (Sauer, 1998). Although complex and variable, premortem fractures in "wet" bones tend to occur with less jagged edges than those of "dry" bones (Moraitis and Spiliopoulou, 2006). Unfortunately, this analysis is subjective and not always reliable (Wheatley, 2008).

Remember that this is a very simplistic look at the fracture mechanics of bone, we are only looking at the macroscopic level. Further reference to more dedicated texts should be made if we wish to understand this topic more completely.

FLUID MECHANICS

The term "fluids" includes both liquids and gases, ie, types of matter that can flow. With the exception of air, all forensically relevant body fluids are liquids. This section will focus on blood, though the principles discussed apply equally to other fluids. Liquids are virtually incompressible, which suggests that it takes enormous pressure to change their volume appreciably. (For example, seawater at the deepest point in the ocean, about 11 km in depth, will have a density only 5% greater than that on the surface.)

The behavior of fluids can be explained with only a few forces: weight, pressure, surface tension, and viscous force. Forces applied to an object may act in different directions: compression, tension, or shear (Fig. 8.2).

The simplest force, weight, causes blood to drip downward or simply to flow downhill over a surface (Fig. 8.12).

The strength of the weight force depends on how much mass of fluid is present, so relates to density (mass per unit volume). As liquids are virtually incompressible, their density changes very little pressure or changes in temperature. The density of blood is very similar to water: typically 1050 to 1060 kg/m^3 (Table 8.1).

Static pressure is a pressure experienced by moving or stationary fluids. Atmospheric pressure is one example of static pressure. Static pressure is isotropic ie, it acts equally in all directions. If the pressure is applied on the outside of an object, it compresses the object. If applied on the inside, it tenses the object. The static pressure inside arteries is higher than atmospheric pressure, and it is greatest when the heart is contracting and expelling blood (systole) and less when the heart is expanding (diastole). Hydrostatic pressure is static pressure caused by depth; for example, the weight of the water above a diver pressurizes the water: the deeper, the greater the pressure.

Dynamic pressure is a type of pressure experienced by objects moving through fluids; for example, a paddle being dragged through water or a droplet falling through the air. It is an inertial force caused by the inertia of the fluid. When the paddle is moved, the water must move out of its way: the water has inertia, so some force must be applied to it to move it. This force comes from the paddle. Remembering Newton's third law, the reaction force felt by the paddle is the dynamic pressure. The paddle will experience high dynamic pressure on one face (the forward-moving face). This is responsible for most of the resistance or drag felt by the paddler. Likewise, a droplet falling through the air experiences a high dynamic pressure on the lower face, which resists its motion. This slows its fall: without this

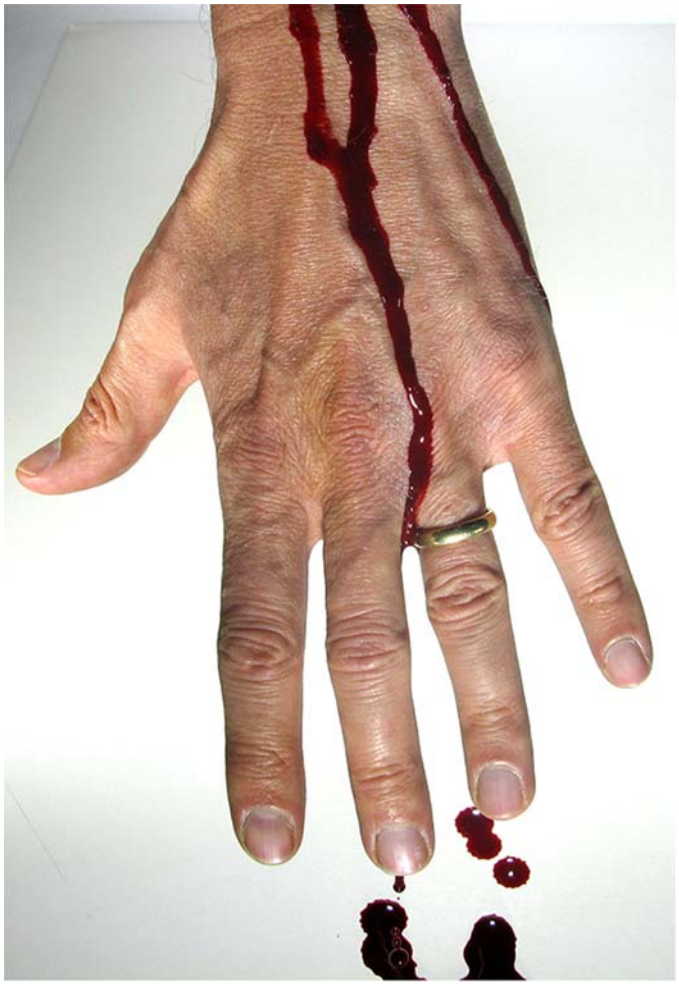


FIGURE 8.12 Blood flowing downward over skin. (Except where indicated, all bloodstains pictured in this chapter were made in the laboratory, using pig blood with anticoagulant to prevent clotting).

TABLE 8.1 Properties of Blood, Water, and Ethanol

	Density (kg/m^3)	Viscosity (Pa s)	Surface tension (N/m)
Water at 20°C	998	0.0010	0.073
Blood at 37°C	1050–1060	Depends on shear rate	0.05
Pure ethanol at 20°C	789	0.0011	0.022

drag, raindrops would travel at a bruising speed of hundreds of kilometers per hour when they reach the ground.

Weight and pressure are experienced by both solids and fluids, even though these two types of matter respond in different ways. The two forces which fluids experience, but which solids do not, are viscous and surface tension forces. These are responsible for the characteristic behavior of fluids. The expression "blood is thicker than water" has a physical truth to it: blood is only slightly denser than water, but it behaves more "thickly" because of its higher viscosity. Viscosity relates to how liquids respond to shear force. As discussed previously, shear forces are forces applied in a sliding or sweeping motion, such as sliding a hot iron over cloth. When a shear force is applied, a solid will deform or break, but a fluid will flow, and when it does so, viscous forces appear. Any shear force applied to a fluid, like blowing a stream of air over the surface of a hot coffee or stirring with a spoon, sets up motion inside the liquid. When a liquid flows over a surface, it does so in layers, each layer slipping over its neighbors (Fig. 8.13). A thin layer of the liquid closest to the surface experiences great resistance from the surface and stays still. The next layer above slips over it, but moves only slowly. The next layer up moves a little quicker, and so on, each adjacent layer exerting a resisting force on the next, a little like the friction of solid objects slipping over each other. This resisting force is the viscous force. Its strength depends on how rapidly the layers slip over each other (ie, the shear rate, Fig. 8.13). The strength of the viscous force also depends on the type of fluid: it is stronger in high-viscosity liquids, such as blood or honey, than it is in lower-viscosity liquids such as water or ethanol.

Many biological fluids have viscosities that change with the type of flow (non-Newtonian fluids). For example, blood is shear thinning: it has a lower viscosity in flows where the velocity varies greatly between layers, ie, as the shear rate increases, the viscosity reduces. Fig. 8.14 shows this for human blood.

When blood drips slowly or oozes from a slight wound, the shear rate is low (to the left of Fig. 8.14) and the blood is highly viscous. On the other hand, a pool of blood struck with a fast-moving weapon will experience strong shear (to the right of Fig. 8.14) and hence lower viscosity: it will flow freely, splashing and breaking up into small droplets easily. This shear-thinning behavior is also seen in ketchup and paint when left to flow under gravity, they are highly viscous, but when shaken or sheared with a paintbrush, they flow freely.

Surface tension is the force in the surface of a body of liquid, which causes that surface to contract into a compact shape. The surface appears to act like the rubber skin of a balloon.

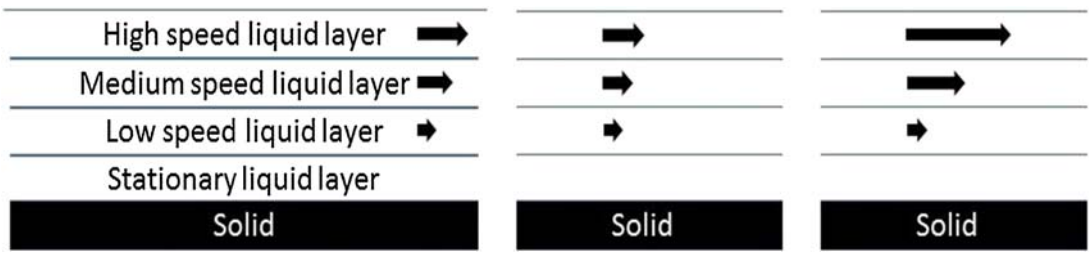


FIGURE 8.13 The liquid flowing over a solid surface moves in layers, which are greater in speed the further they are from the solid surface (left). The liquid speed changes a little as it passes from one layer to the next ie, low shear rate (middle). The liquid speed changes more from one layer to the next ie, high shear rate (right).

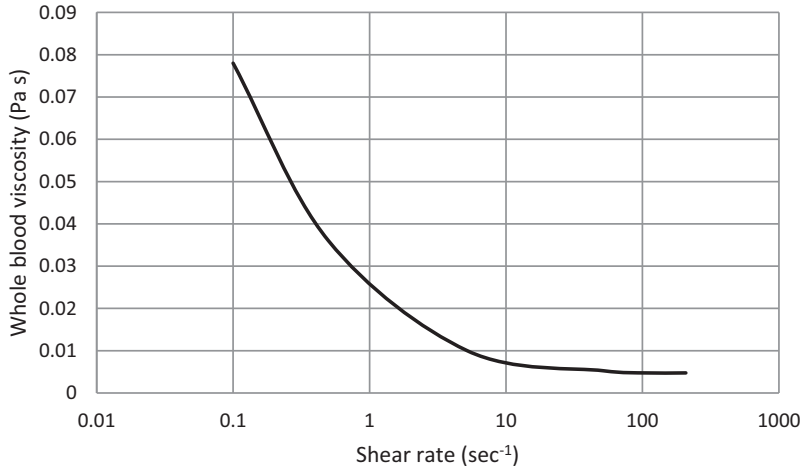


FIGURE 8.14 Viscosity of human blood as a function of shear rate (exact values vary between individuals: representative values taken from de Simone et al., 1990).

The tension in the surface stores energy, just as a stretched rubber band does. The less surface area, the less stored energy, so liquids change shape to minimize the surface area. For a given volume, the smallest surface area occurs when the liquid is in a sphere; hence, droplets and bubbles tend to form spherical shapes.

Where the surface of a drop touches a solid, more effects come into play. Consider a drop of water sitting on a feather. There is surface tension where the water and air meet, and a similar tension where the water touches the feather, but of a different strength. The balance of surface tensions determines the shape of the drop. A feather is hydrophobic (water-fearing), and the tension where the water touches the feather is strong. Hence, the drop takes on an almost spherical shape with as little area as possible touching the feather (Fig. 8.15). For the same reason, the blood in Fig. 8.12 does not spread evenly over the hand, but remains in narrow runnels. Where water sits on a hydrophilic (water-loving) surface, like clean porcelain, the tension is low where the water touches the paper, and the droplet spreads out or wets the surface, forming a lenslike shape (Fig. 8.15).

FIGURE 8.15 Surfaces exist between air and water, and between water and solid: at each of these surfaces, there is tension (left). Drops of colored water on a porcelain surface (hydrophilic) and on a feather (hydrophobic) (right).

Surface tension is a force and also a property of the liquid. Water has a high surface tension, blood less so, and ethanol even less (whisky or vodka will wet a glass differently to water) (Table 8.1).

The tension at a blood–solid interface accounts in part for blood's stickiness or adhesion. The action of proteins and platelets in the blood enhances its adhesion, making it more adhesive than water.

Another important difference between water and blood is that when released from the body blood begins to coagulate into a gel, increasing in viscosity, and finally becomes solid in the form of a clot. Blood is about 90% water, so over time a stain dries. Clotting and drying radically change its properties, but they are too slow to influence the initial bloodstain formation. They do, however, limit the spreading of large pools of blood, such as might result from a serious wound from an unconscious or deceased person. Clotting and drying also affect the final appearance of stains, though less so when the stain soaks into some porous material, like clothing or carpet, where the shape of the stain is fixed before clotting or drying can occur.

This section focuses on blood, but other body fluids include air, blood, cerebrospinal fluid, lymph, urine, semen, feces, and mucus. The last three are not true fluids: especially at low water content, when they behave more like solids, they deform with mild stresses; once the stress is removed they spring back (elastic behavior). Under greater stresses they deform permanently (plastic flow) or break. Soft tissues, such as muscle, fat, and brain, can show plastic flow under the high stresses of blunt weapon impact and gunshot wounding. Under these conditions, they flow as highly viscous fluids. As previously described, the combination of viscous flow and elastic or plastic behavior earns them the name viscoelastic or viscoplastic.

IMPACT MECHANICS

An impact can be defined as "the action of one object coming forcibly into contact with another" (Pearsall, 2001). It is generally assumed that the event described by an impact occurs over a relatively short period of time (eg, a hammer impacting a head, a bullet impacting a torso, etc.). One or both of the objects may deform or fracture depending on relative hardness, stiffness, and strain rate sensitivity of the materials involved. Clearly, from a forensic perspective, impacts can result in injuries; remember these are classified as sharp, blunt, or ballistic trauma. The magnitude of effect of the impact may vary according to the relative velocity of the objects involved (ie, it is influenced by the kinetic energy of the impacting object). Kinetic energy, which is the work needed to accelerate the object from rest to its stated velocity, can be easily calculated:

$$\mathrm{KE} = 1/2mv^2$$

where KE = kinetic energy (J), m = mass (kg), v = velocity (m/s).

If comparing objects of different cross sections, it may be useful to calculate kinetic energy density (KED: $\mathrm{J/m^2}$):

$$\mathrm{KED} = 1/2mv^2/A = \mathrm{KE/A}$$

where A = area ($\mathrm{m^2}$).

Examples of KED for various threats include low-velocity handgun bullets 16 $\mathrm{J/m^2}$, knives 160 $\mathrm{J/m^2}$, and high-velocity rifle bullets 45–75 $\mathrm{J/m^2}$ (Horsfall, 2012).

SPECIAL APPLICATIONS OF BIOMECHANICS IN A FORENSIC SETTING

Body Armor and Wounding Behind Body Armor

Body armor is worn by military personnel, law enforcement officers, first responders, and civilian security guards to provide protection from sharp weapons and ballistic threats (Fig. 8.16). Body armor provides protection to five critical organs of the torso: heart, lungs, kidneys, liver, and spleen (Dixon and Croft, 2007; Breeze et al., 2014). Military body armor provides protection from fragmentation and high-velocity (rifle) ammunition; routine patrol police officers will usually be protected from sharp weapons and low-velocity (hand-gun) ammunition, and a firearm police officer might additionally be protected from high-velocity (rifle) ammunition (Tobin and Iremonger, 2006; Croft and Longhurst, 2007a; Horsfall, 2012).

Body armor typically comprises of "soft" and "hard" elements. Soft armor can be optimized to provide protection from fragmentation, low-velocity (hand-gun) ammunition, and sharp weapons. Hard armor refers to plates that provide protection from high-velocity (rifle) ammunition. Soft armor usually takes the form of a waistcoat or tabard style garment and is typically manufactured using multiple layers of para-aramid fabrics (eg, Kevlar®, Twaron®) or ultrahigh molecular weight polyethylene fibers (eg, Spectra®, Dyneema®).

FIGURE 8.16 Typical military and police body armor. (A) Typical military body armor (B) Typical police body armor.

Testing Body Armor

Body armor is tested using various standard test methods; the most commonly used internationally are STANAG 2920 Ed. 3, the NIJ suite of Standards, and the Home Office suite of Standards (Croft and Longhurst, 2007a,b; National Institute of Justice, (2008); NATO Standardization Agency, 2003).

Fragmentation protection is usually assessed by the use of fragment simulating projectiles (FSPs); the most commonly used of these are the 1.1 g chisel-nosed FSP (CN FSP) and the 1.0 g right circular cylinder FSP (NATO Standardization Agency, 2003) (Fig. 8.17). V_{50} data are calculated for the armor solution; V_{50} is the velocity at which the probability of perforation of the specimen is 50% for a particular specimens and projectiles (NATO Standardization Agency, 2003). Military body armor typically has a mean V_{50} in excess of 500 m/s for 1.1 g CN FSPs. The NIJ and Home Office standards include assessment by sharp weapons, low-velocity (handgun) ammunition, and high-velocity (rifle) ammunition. (Croft and Longhurst, 2007a,b; National Institute of Justice, 2008) (Fig. 8.18).

Behind Armor Blunt Trauma

Body armor protects personnel because it dissipates and absorbs kinetic energy during the impact event (Horsfall, 2012). The projectile is retarded as it perforates the armor; this can

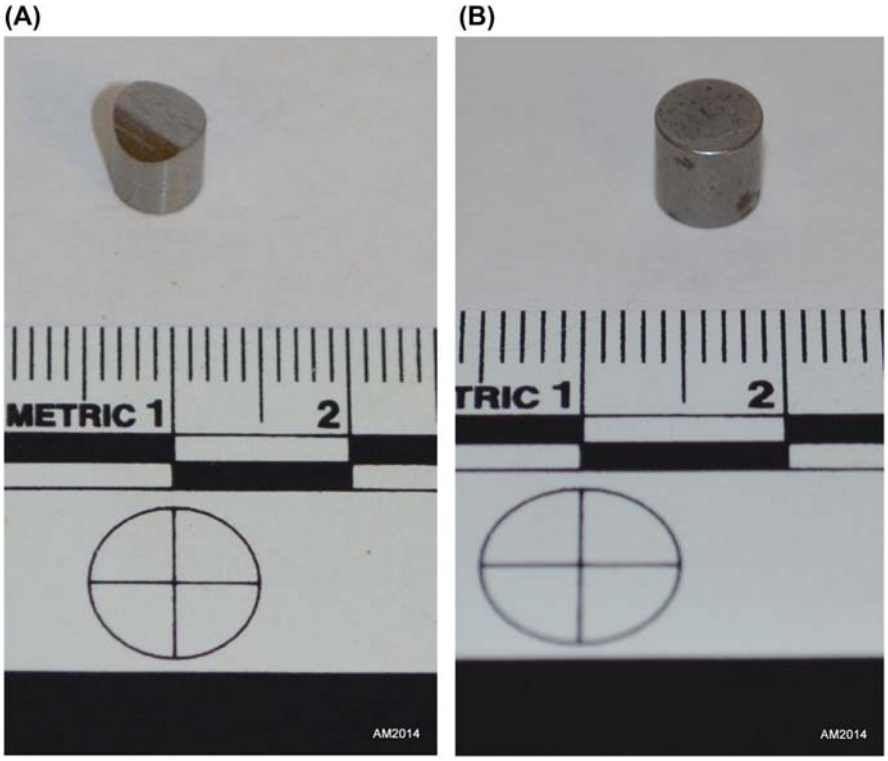


FIGURE 8.17 Fragment simulating projectiles. (A) 1.1 g CN FSP (B) 1.0 g RCC FSP.

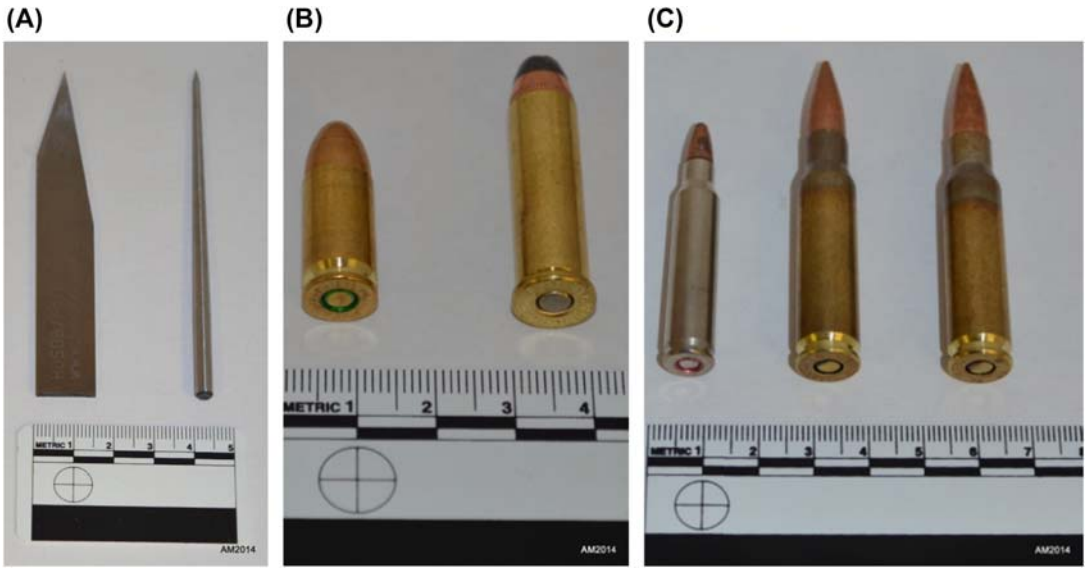


FIGURE 8.18 Examples of weapons that different types of body armor are designed to protect against. (A) Home Office sharp-weapon blade (left) and spike (right) (B) 9 mm FMJ Dynamit Nobel DM11A1B2 (left) and Soft Point Flat Nose Remington R357M3 (right) ammunition (C) Federal Tactical Bonded 5.56 mm (left), 7.62 mm NATO Ball L2 A2 (center), 7.62 mm NATO Ball L40A1 (right) ammunition.

result in both the armor moving into the body and in a behind armor blunt trauma (BABT) injury. BABT injuries are typically skin hematoma, mild lacerations of skin, broken ribs, and occasional hematoma of the lungs (Carr et al., 2013).

Body armor test methods often include assessment of the behind face signature (BFS); typically the armor is mounted on a block of claylike material (usually Roma Plastilina #1) and the depth of the depression formed in the clay due to a nonperforating ballistic impact on the armor measured (Croft and Longhurst, 2007a; National Institute of Justice, 2008). It is recognized that the clay is not a biological accurate representation of the human body, but a standard test medium for assessing BFS. The various standards provide pass/fail data for BFS.

Injuries Due to Perforated Body Armor

The primary role of body armor is to stop a penetrating injury to the wearer, however, it is not viable for body armor to protect the wearer against all forms of attack; the armor will have been designed to provide protection from specific threats only (Tobin and Iremonger, 2006; National Institute of Justice, 2010; Croft and Longhurst, 2007a). When body armor is challenged by a greater threat than it was designed to protect against, it is said to have been "overmatched."

There has been limited open source research into what happens when body armor is overmatched in ballistic events; the authors who have carried out this research are not in agreement. There is some suggestion in the literature that, in situations where a greater threat than expected is present, the specific body armor worn may not aid protection; rather it could

exacerbate the wounding effect (Missliwetz et al., 1995; Breteau et al., 1989). Conversely, it has also been suggested that injury reduction can occur when body armor is overmatched compared to when no body armor is present (Lanthier et al., 2004). Other authors have called for further research in the area (Prather, 1994; Knudsen and Sorensen, 1997).

Examples of military personnel and police officers wearing body armor impacted by ammunition threats greater than expected may not provide information on whether the presence of the body armor exacerbated the damage caused; it is unlikely to ever be a case where identical shot parameters are duplicated, one onto a target with body armor and one without, thus being able to compare the results. However, examples in the open literature do provide some evidence that the overmatching of body armor is an issue (eg, Federal Bureau of Investigation, 2012; Kosashvili et al., 2005); these are described later in further detail.

Reviewing the fatalities that occurred during conflict in the West Bank between the 22nd of March and 30th of April 2002 revealed 22 soldiers (out of 26 cases examined) died while wearing Kevlar® military personnel armor system (MPAS) vests, with no added ceramic protection (Kosashvili et al., 2005). The 26 fatalities suffered a total of 149 entrance wounds; 76 wounds (51%) were from fragments and 73 (49%) were from bullets. Twelve fatalities (46%) were due to injuries from a combination of fragments and bullets, with 14 (54%) due to injuries solely from bullets. Considering only bullet impacts, 16 (49%) occurred in body regions that were covered by an MPAS and 17 (51%) occurred in uncovered regions. Injuries due to fragments in protected regions were very few, highlighting the protective effect of the MPAS. However, very limited protection, if any, was offered against bullets. Statistical analysis of diameters of entry wounds in the covered versus uncovered regions revealed no significant differences (0.79 ± 0.42 cm vs 0.73 ± 0.29 cm; $p = 0.11$), with the authors claiming the presence of armor did not seem to worsen the outcome of bullet injuries. Dimensions of the wounds, other than the entrance diameters, were neither provided nor mentioned. Judging wounding outcomes solely on entrance diameters is an unfair representation of the damage produced by a bullet.

The threats faced by police differ greatly from those encountered by military personnel, although cases of overmatching still occur. In the United States of America from 2003 to 2012, 321 law enforcement officers were feloniously killed due to ballistic attacks (Federal Bureau of Investigation, 2012). Of these incidents, 21 (7%) were due to shots at the torso that used ammunition that was too powerful for the armor that was worn. In all 21 cases, rifle ammunition, ranging in caliber size from 0.223″ to 7.62×39 mm, was used. Whether perforated armor resulted in increased wounding was not discussed.

Bloodstain Pattern Analysis

Once investigators have found bloodstains at a crime scene, the next stage in bloodstain pattern analysis (BPA) is to classify them into types. Each type is caused by a particular mechanism, which can be linked to actions by the assailant(s), victim(s), and any other persons who may have been present. This, together with other evidence, is used to reconstruct the events that took place (crime scene reconstruction). Types of stain and the characteristics used to distinguish each type are described briefly here, but the presentation here is restricted to ideal examples without the complexities typical of crime scenes. For more detail the reader

should refer to the BPA textbooks by James et al. (2014) and Bevel and Gardner (2008), which give an overview of crime scene reconstruction. Jermy and Taylor (2013) discuss the physics behind the formation of each type of bloodstain in greater detail, and Attinger et al. (2013) give the best review of research available at the time of writing. A series of excellent high-speed videos of stain formation are available for free download from the Midwest Forensic Resource Center (https://www.ameslab.gov/mfrc/bpa_videos).

Classification of stains into types should be performed with caution: it is easy for the novice to mistake one type for another. In many jurisdictions, scenes-of-crime examiners must pass a 40-h course in basic bloodstain pattern analysis to qualify to record bloodstain evidence at a crime scene. A further 40-h advanced course is required to qualify to interpret stains and testify with BPA evidence. These courses follow a syllabus set by SWGSTAIN (Scientific Working Group on Bloodstain Pattern Analysis www.swgstain.org), an international group, which sets standards on BPA. The community of bloodstain analysts is linked together by IABPA (the International Association of Bloodstain Pattern Analysts, www.iabpa.org), which organizes training, runs conferences and publishes a journal reporting developments in the field.

Transfer stains are the result of a bloodied object coming into contact with other items. Examples include shoe, finger, and palm prints (Fig. 8.19), which can be valuable in identifying a suspect, but swipes and wipes (smeared transfer stains) can be valuable in reconstructing movements at a scene. Transfer stains occur because of the highly viscous and adhesive nature of blood, which helps preserve fine detail, though this depends on the nature of the surface.

Flow patterns: Blood flowing over a surface is driven by gravity, flowing downhill or flowing from a deep area to a shallow area. It may be channeled into a narrow area, or it may spread out. The flow is resisted by viscosity and surface tension effects, which tend to slow it and limit its spread. Fig. 8.12 shows a flow over a hand: note the blood running in narrow channels over the back of the hand confined by surface tension. In this case the blood runs between the fingers and drips from the fingertips. Flows typically result in lines (which may branch) and pools.

FIGURE 8.19 Transfer stains: a dried handprint on paper, and a wet handprint on a varnished wooden bat.

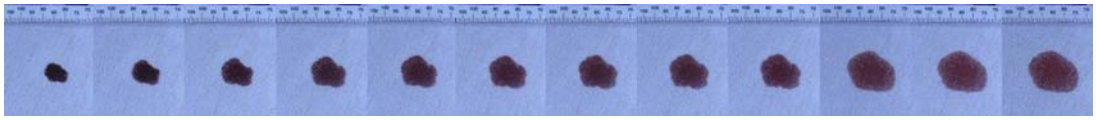


FIGURE 8.20 A sequence of images showing blood on cotton fabric spreading by wicking.

FIGURE 8.21 Drip stains caused by: (left) individual drops from bloodied fingers falling 10 cm onto paper and (right) a series of drops of human blood falling 2.5 m onto concrete, splashing in the pools left by previous drops to create many satellite stains. *Courtesy Matt Noedel, Noedel Scientific.*

Wicking is the spreading of a liquid through a porous medium, for example a textile (Fig. 8.20). The expansion of the liquid through the connected pores in the medium is driven by capillary forces, which are essentially surface tension and the related adhesion forces. If the source of the blood is a deep pool, hydrostatic pressure also assists the blood through the pores. Gravity can also assist the spread if the blood can spread downward.

Dripping (sometimes called passive dripping) and drip stains: blood dripping off some object under gravity, with the resulting droplets falling to strike a lower surface such as a floor (Fig. 8.21). The stains tend to be circular (this depends on the surface) and may have spines (radiating fingers) and may be surrounded by randomly distributed and randomly oriented smaller stains (satellite spatter) caused by accompanying drops or by splashing. In the formation of drips, blood is drawn downward by gravity, against the surface tension forces, which resist dripping and tend to keep the blood suspended together in one coherent body. However, if more blood can flow to the dripping point, a hanging drop will grow until its weight exceeds the surface tension force and it separates and falls. As it separates, it may draw out a string or ligament of liquid, which may ultimately break into yet smaller accompanying drops which may then leave their own stains. If several drops fall onto the same spot, forming a pool, later drops falling into this pool will splash, causing a host of smaller stains surrounding the pool (satellite spatter).

Spatter stains are bloodstains formed by droplets of blood travelling through the air to strike a surface, leaving a stain. Spatter may be further divided according to the action creating the droplets: satellite spatter, cast off, arterial spurt (jetting), impact, gunshot, and expectorate (or expirated) spatter.

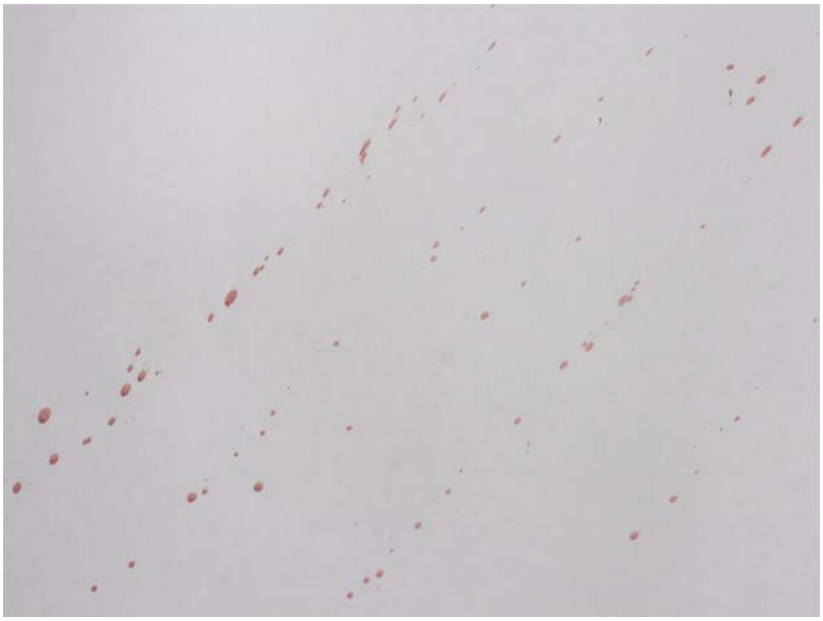


FIGURE 8.22 Cast off from the tip of a bloodied baton, swung back and forth at 45 degree to the vertical.

Swing cast-off stains are formed when a bloodied object, like a weapon, is swung. As the object moves in an arc, blood pools at the tip of the object. If the inertia of the pooled blood is greater than the adhesive and surface tension forces binding it to the object, droplets are released, initially travelling at a tangent to the path of the swing, until gravity and drag alter their trajectory. Castoff leaves a series of stains on nearby surfaces (eg, walls or ceiling) usually in a line (Fig. 8.22). Where the drops strike the surface at right angles, they form circular stains: elsewhere, the stains are elliptical. The pattern thus has information on the direction and number of swings, relating to the number of blows being struck. Cessation castoff is formed where the object stops or reverses rapidly at the end of the swing. The rapid acceleration casts off blood from the weapon, again at a tangent to the swing.

Jetting is the rapid acceleration of a fluid under pressure (Fig. 8.23). In BPA an important example is arterial spurt, in which blood is forced out of a wounded artery by blood pressure. Another example is the early, fast ejection of a thin-liquid film from under a blunt impact weapon as it strikes a bloodied surface. In this case, pressure caused by the impact causes the blood to squeeze out through the narrow gap between the weapon and struck surface. Blood moves rapidly out from this point and may thin out to form stringlike ligaments of blood. These ligaments may break up into droplets. This early jetting may leave streaks of blood radiating out from an impact point (eg, the radial lines near the black cross in Fig. 8.24). These radial lines are not the true impact spatter though: that is a coarser stain pattern spread over a larger area (Fig. 8.24) and is dealt with next.

Impact spatter occurs when a blunt object strikes something with a deep enough covering of blood: it sends sheets of blood flying out from the impact point. These sheets break up into droplets, which radiate out from the impact point, leaving a characteristic pattern (Fig. 8.24).

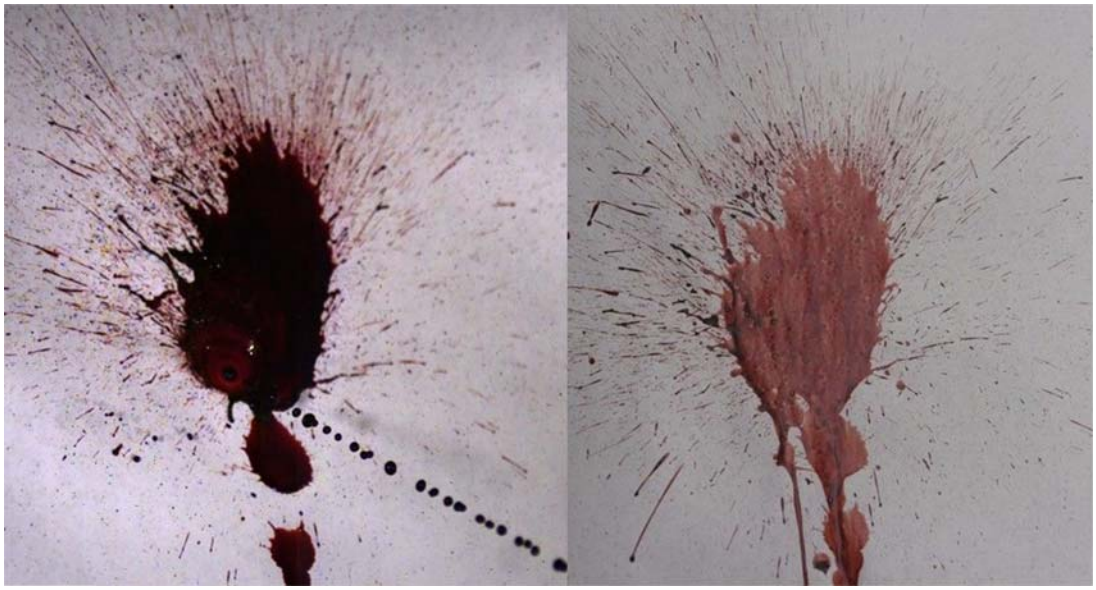


FIGURE 8.23 Jetting or spurt: a stream of blood ejected from a syringe enters from the bottom right and splashes when it hits the paper target (left). The resulting stain when dry: note the central large stain and the surrounding satellite spatter caused by splashing (right).

One of the most useful features of such a pattern is that the droplets, when they strike a surface, leave elliptical stains. The shape of the stains is related to the angle at which they strike the surface, according to the sine law first described by Balthazard et al. (1939) and Rizer (1955). The ratio of the stain's width to its length is related to the sine of the angle of impact (Fig. 8.24). The angle of impact can thus be calculated to an accuracy within a few degrees.

If the angles of impact of several stains are measured, and if the droplets, which caused those stains, can be assumed to have travelled in straight lines, the path of the droplets took can be tracked back to their source. Where the paths cross is the probable location of the wound at the time the blow was struck (Fig. 8.24). This is valuable information in crime scene reconstruction: it offers clues as to the position of the victim at the time of the assault.

This procedure is known as "backtracking" or "stringing." The analysis is less laborious if the stains are photographed and analyzed with software such as HemoSpat (www.hemospat.com) or BackTrack (http://people.physics.carleton.ca/~carter/).

The procedure depends on the assumption that the droplets have followed a straight trajectory. In fact, all trajectories are curved, though the initial part of their flight deviates only slightly from a straight line. BPA analysts are trained to select stains caused by upward-moving drops close to their point of origin, which will have followed the straightest paths. Due to inevitable small errors in measurement, and the straight-line assumption not quite holding perfectly, the extrapolated trajectories rarely cross at a point, instead converge in a volume termed the "area of origin" of a few centimeters in diameter.

The spreading and splashing of a drop striking a surface is complex and is not described here in full, but it is reasonably well understood from decades of research in sprays used for

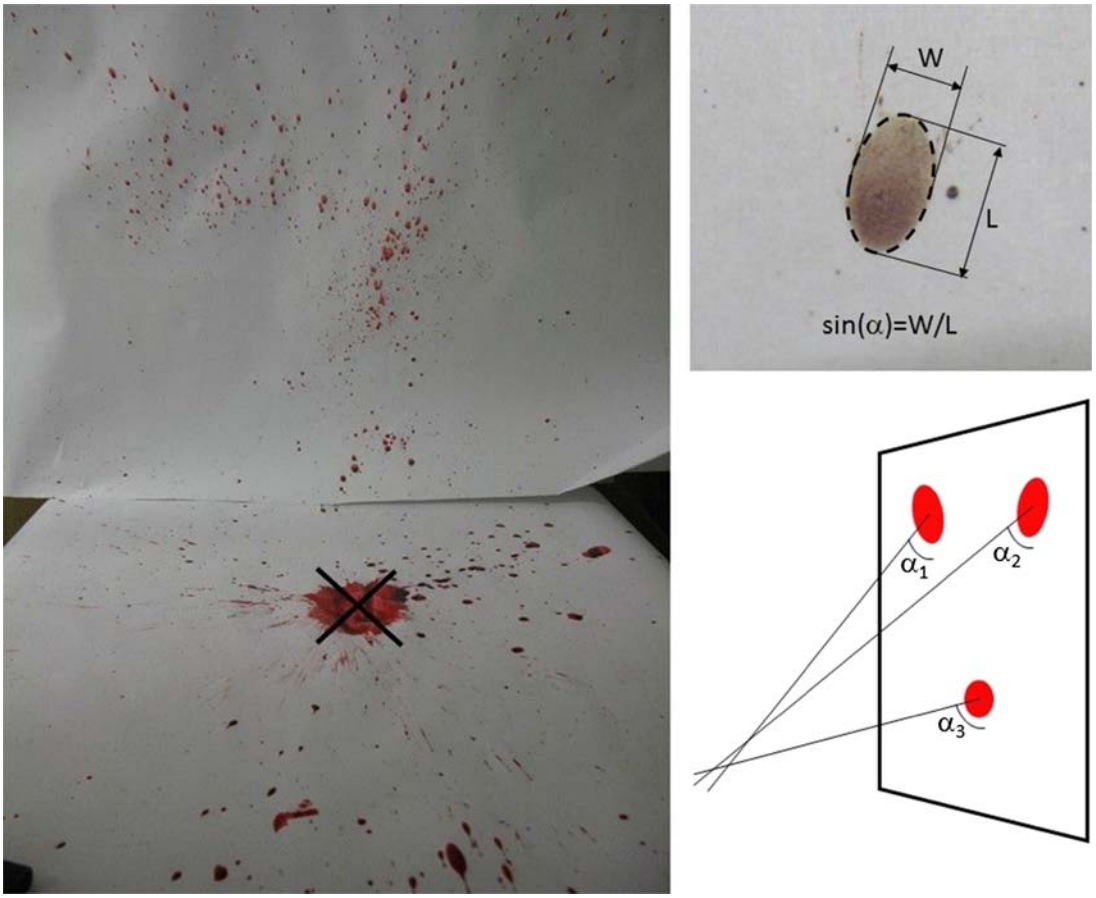


FIGURE 8.24 A hammer impacting a pool of blood sends droplets radiating out from the point of impact (black cross) to create a pattern of elliptical stains on the wall behind (left). The angle of impact α can be calculated from the width to length ratio of an elliptical stain (right, top). Straight-line trajectories can be followed back from several stains (right, bottom). The impact point will be located in the region where these trajectories cross. In practice, many stains are used, each selected to be likely to have followed a straight path little affected by gravity.

cooling, coating, and fuel injection. Some of the physics is described in Jermy and Taylor (2013) and Attinger et al. (2013).

Winds or drafts may blow a droplet off course (side wind), speed it up (tail wind) or slow it down (head wind), and correspondingly affect the distance it travels and the location of the resulting stain. These effects may diminish close to a wall, due to slowing of air near a solid surface.

Gunshot spatter is caused when droplets radiate out from entrance and/or exit of gunshot wounds. The physical processes involved in gunshot spatter are very similar to those in impact spatter, but gunshot spatter typically generates smaller drops (Fig. 8.25), and hence patterns typically have dense arrays of small stains, sometimes referred to as a "mist." These smaller droplets (relative to typical impact patterns) form due to the greater energy density in

FIGURE 8.25 Early stages of the formation of a gunshot spatter pattern: a snapshot taken as a 0.22 caliber lead pellet moving at 290 m/s passes through a sponge soaked in blood. Note that spatter is projected backward toward the gun as well as forward.

a gunshot wound (greater kinetic energy and/or delivered over a smaller area of contact between the projectile and the wound). The directionality may be hard to ascertain due to the small size of stains, but the pattern will radiate from the bullet trajectory, and the bullet is often recovered, giving at least one fixed point on the trajectory. Kneubuehl (2011) treats gunshot wounding in some detail.

Expirated spatter (or expectorate spatter) occurs if a person is injured such that blood pools in their mouth or nose, when they exhale, blood can be entrained in the outgoing breath and carried with it to form expirated blood patterns. Although the processes involved in the breaking up of blood into droplets are complex and depend on the quantity of blood, how it pools and adheres to the mouth and airway walls, as well as the speed of the air passing over it; in general, smaller droplets (a fine spray) are formed by more violent exhalations. Expirated patterns may sometimes be positively identified by the presence of mucus, air bubbles, or oral bacteria.

Acknowledgments

We would like to thank Dr. Michael Taylor of ESR, New Zealand, for his inspiration and guidance and Glynny Kieser for her editorial support. We are grateful to Matt Noedel (Noedel Scientific), Albert-Menno Laffra, and Nathan Hoogendorp for providing photographs.

References

Altemeier, W.A., 2001. Interpreting bruises in children. Pediatric Annals 30, 517–520.

Attinger, D., Moore, C., Donaldson, A., Jafari, A., Stone, H., 2013. Fluid dynamics topics in bloodstain pattern analysis: comparative review and research opportunities. Forensic Science International 231, 375–396.

Balthazard, V., Piedelievre, R., Desoille, H., Derobert, L., 1939. Etude des gouttes de sang projete. In: XXIIe congres de medicine legale de langue francaise. Paris.

Bevel, T., Gardner, R.M., 2008. Bloodstain Pattern Analysis with an Introduction to Crime Scene Reconstruction. CRC Press Inc.

Breeze, J., Allanson-Bailey, L.J., Hepper, A., Midwinter, M.J., 2014. Demonstrating the effectiveness of body armour: a pilot prospective computerised surface wound mapping trial performed at the Role 3 hospital in Afghanistan. Journal of the Royal Army Medical Corps 161 (1), 36–41. http://dx.doi.org/10.1136/jramc-2014–000249.

Breteau, J., Fackler, M., Sendowski, I., Martin, P., 1989. The personal protective equipment provided for combatants: the part played by wearing a protection vest in the behaviour of projectiles wounding outcomes. In: 11th International Symposium on Ballistics. International Ballistics Society, Brussels, Belgium.

Briggaman, R.A., 1982. Biochemical composition of the epidermal-dermal junction and other basement membrane. Journal of Investigative Dermatology 78, 1–6.

Carr, D., Horsfall, I., Malbon, C., 2013. Is behind armour blunt trauma a real threat to users of body armour? A systematic review. Journal of the Royal Army Medical Corp. http://dx.doi.org/10.1136/jramc-2013–000161.

Croft, J., Longhurst, D., 2007a. HOSDB Body Armour Standards for UK Police. Part 2: Ballistic Resistance. Publication No. 39/07/B. Home Office Scientific Development Branch, Sandridge, UK.

Croft, J., Longhurst, D., 2007b. HOSDB Body Armour Standards for UK Police. Part 3: Knife and Spike Resistance. Home Office Scientific Development Branch, Sandridge, UK.

de Simone, G., Devereux, R.B., Chien, S., Alderman, M.H., Atlas, S.A., Laragh, J.H., 1990. Relation of blood viscosity to demographic and physiologic variables and to cardiovascular risk factors in apparently normal adults. Circulation 81, 107–117.

Dixon, C., Croft, J., 2007. Body Armour Good Practice and Quality Framework. Publication No. 44/07. Sandridge, UK. http://www.bsst.de/content/PDF/44-07-Body_Armour_Good_Pra1.pdf.

Dunn, M.G., Silver, F.H., 1983. Viscoelastic behaviours of human connective tissues: relative contribution of viscous and elastic components. Connective Tissue Research 12, 59–70.

Federal Bureau of Investigation, 2012. Law Enforcement Officers Killed and Assualted. http://www.fbi.gov/about-us/cjis/ucr/leoka/2012/officers-feloniously-killed/felonious_topic_page_-2012.

Frazier, L.D., 2003. Child abuse or mimic? Consultant Paediatricians 2, 212–215.

Gordon, J.E., 1991. The New Science of Strong Materials. Penguin Books, London.

Gordon, J.E., 2003. Structures, or Why Things Don't Fall Down, second ed. Da Capo Press, Cambridge, MA.

Hatze, H., 1974. The meaning of the term 'biomechanics'. Journal of Biomechanics 7, 189–190.

Hornor, G., 2005. Physical abuse: recognition and reporting. Journal of Paediatric Health Care 19, 4–11.

Horsfall, I., 2012. Key issues in body armour: threats, materials and design. In: Sparks, E. (Ed.), Advances in Military Textiles and Personal Equipment. Woodhead Publishing and The Textiles Institute, Oxford.

James, S.H., Kish, P.E., Sutton, T.P., 2014. Principles of Bloodstain Pattern Analysis: Theory and Practice. Taylor & Francis Group, CRC Press Inc., Boca Raton, Florida.

Jermy, M.C., Taylor, M.C., 2013. Forensic biomechanics (developments in forensic science). In: Kieser, J., Taylor, M.C., Carr, D. (Eds.), The Mechanics of Bloodstain Pattern Formation. Wiley, ISBN 9781119990116, pp. 99–136.

Kieser, J.A., Whittle, K., Wong, B., Waddell, J.N., Ichim, I., Swain, M., Taylor, M., Nicholson, H., 2008a. Understanding craniofacial blunt force injury: a biomechanical perspective. Forensic Pathology Reviews 5, 37–51.

Kieser, J.A., Bernal, V., Gonzalez, P., Birch, W., Turmaine, M., Ichim, I., 2008b. Analysis of experimental cranial wounding from screwdriver trauma. International Journal of Legal Medicine 122, 179–187.

Kieser, J.A., Taylor, M., Carr, C., 2013. Forensic Biomechanics. Wiley-Blackwell, Chichester UK.

Kleinman, P.K., Schlesinger, A.E., 1997. Mechanical factors associated with posterior rib fractures: laboratory and case studies. Pediatric Radiology 27, 87–91.

Kneubuehl, B., 2011. Wound Ballistics: Basics and Applications. Springer.

Knight, B., 1975. The dynamics of stab wounds. Forensic Science 6, 249–255.

Knudsen, P.J.T., Sorensen, O.H., 1997. The destabilising effect of body armour on military rifle bullets. International Journal of Legal Medicine 110, 82–87.

Kosashvili, Y., Hiss, J., Davidovic, N., Lin, G., Kalmovic, B., Melamed, E., Levy, Y., Blumendeld, A., 2005. Influence of personal armor on distribution of entry wounds: lessons learned from urban-setting warfare fatalities. Journal of Trauma 58, 1236–1240.

Langer, A.K., 1861. Zur Anatomie und Physiologie der Haut. I. Uber die Spaltbarkeit der Cutis. Sitzungberichte der Akademie der Wissenschaften 44, 19–46.

Langer, K., 1978a. On the anatomy and physiology of skin I. British Journal of Plastic Surgery 31, 3–8.

Langer, K., 1978b. On the anatomy and physiology of skin II. British Journal of Plastic Surgery 31, 93–106.

Langer, K., 1978c. On the anatomy and physiology of skin III. British Journal of Plastic Surgery 31, 185–199.

Lanthier, J.M., Iremonger, M.J., Lewis, E.A., Horsfall, I., Gotts, P.L., 2004. Is the wounding potential of high velocity military bullets increased after perforation of textile body armour. In: Personal Armour Systems Symposium. The International Personal Armour Committee., The Hague, Netherlands.

Lonergan, G.L., Baker, A.M., Morey, M.K., Boos, S.C., 2003. Child abuse: radiologic-pathologic correlation. Radiographics 23, 811–845.

Missliwetz, J., Denk, W., Wieser, I., 1995. Study on the wound ballistics of fragmentation protective vests following penetration by handgun and assault rifle bullets. Journal of Forensic Sciences 40, 582–584.

Moraitis, K., Spiliopoulou, C., 2006. Identification and differential diagnosis of perimortem blunt force trauma in tubular long bones. Forensic Science, Medicine and Pathology 2, 221–229.

National Institute of Justice, 2008. Ballistic Resistance of Body Armour NIJ Standard-0101.06. U.S. Department of Justice, Washington DC.

National Institute of Justice, 2010. Selection and Application Guide to Ballistic-resistant Body Armor for Law Enforcement, Corrections and Public Safety. NIJ guide-0101.06. http://198.77.71.164/Documents/BA-Selection-Application-Guide-2011-17-2010.pdf.

NATO Standardization Agency, 2003. Standardization Agreement (STANAG) 2920. Ballistic Test Method for Personal Armour Materials and Combat Clothing, second ed.

Oxlund, H., Manscot, J., Vidiik, A., 1988. The role of elastin in the mechanical properties of skin. Journal of Biomechanics 21, 213–218.

Payne, P.A., 1991. Measurement of properties and function of skin. Clinical Physics and Physiological Measurement 12, 105–129.

Pearsall, J., 2001. The Concise Oxford Dictionary. Oxford University Press Inc., New York.

Pierce, M.C., Bertocci, G.E., Vogeley, E., Moreland, M.S., 2004. Evaluating long bone fractures in children: a biomechanical approach with illustrative cases. Child Abuse and Neglect 28, 505–524.

Prather, R.N., 1994. Small arms vs. soft armour - a pilot study. In: Personal Armour Systems Symposium. Stores and Clothing Research and Development Establishment, Colchester, UK.

Richey, M., Richey, H.K., Fenske, N.A., 1988. Aging related skin changes: development and clinical meaning. Geriatrics 43, 49–64.

Rizer, C., 1955. Police Mathematics. Charles C. Thomas, Springfield, IL, pp. 72–73.

Sanders, J.E., Goldstein, B.S., Leotta, D.F., 1995. Skin response to mechanical stress: adaptation rather than breakdown – a review of the literature. Journal of Rehabilitation, Research and Development 32, 214–226.

Sauer, N.J., 1998. The timing of injuries and manner of death: distinguishing between antemortem, perimortem and postmortem trauma. Springfield, IL. In: Reichs, K.J. (Ed.), Forensic Osteology, pp. 321–332.

Scherl, S.A., Miller, L., Lively, N., Russinoff, S., Sullivan, C.M., Tornetta, P., 2000. Accidental and non-accidental femur fractures in children. Clinical Orthopaedics and Related Research 376, 96–105.

Schwend, R.M., Worth, C., Johnston, A., 2000. Femur shaft fractures in toddlers and young children: rarely from child abuse. Journal of Paediatric Orthopaedics 20, 475–481.

Silver, F.H., Freeman, J.W., De Vore, D., 2001a. Viscoelastic properties of human skin and processed dermis. Skin Research and Technology 7, 16–23.

Silver, F.H., Christiansen, D.L., Snowhill, P.B., Chen, Y., 2001b. Transition from viscous to elastic-dependency of mechanical properties of self-assembled collagen fibres. Journal of Applied Polymer Science 79, 134–142.

Tobin, L., Iremonger, M., 2006. Modern Body Armour and Helmets: An Introduction. Argros Press, Canberra, Australia.

Wheatley, B.P., 2008. Perimortem and postmortem bone fractures? An experimental study of fracture patterns in deer femora. Journal of Forensic Sciences 53, 69–72.

Whittle, K., Kieser, J.A., Ichim, I., Swain, M.V., Waddell, J.N., Livingstone, V., Taylor, M., 2008. The biomechanical modeling of non-ballistic skin wounding: blunt force injury. Forensic Science, Medicine and Pathology 4, 33–39.

CHAPTER

9

Biomechanical, Epidemiologic, and Forensic Considerations of Pediatric Head Injuries

W.E. Lee III

University of South Florida, Tampa, FL, United States

J.D. Lloyd

Brains, Inc., San Antonio, FL, United States

OUTLINE

INTRODUCTION

Pediatric head injury is a prominent public health and legal issue in which test validity plays a critical role in causation assessments. As an example, if a child suffers skull fracture in an accidental event described only by a parent, the determination of whether the injury

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00009-4

could have occurred as the parent described plays a critical role in assessing the probability the parent is telling the truth about the circumstances of the injury. A forensic epidemiology approach to injury causation can provide helpful insight into the strength of the association between an observed injury and competing explanations for how the injury occurred. Such an analysis cannot be accomplished without an understanding of the biomechanical aspects of the competing explanations as they relate to the observed injury. In this chapter an overview of the epidemiologic and biomechanical aspects of pediatric head injuries, with special attention on the types of injuries that are encountered in a forensic setting, is presented.

In adults the hardened skull serves as a protective helmet to the brain, whereas the skull of a baby is comprised of several soft and flexible plates, and not nearly as protective against blunt trauma. Sutures between the plates (see Fig. 9.1) are unfused in the newborn, (facilitating vaginal delivery), and do not fully fuse until the second or third year of life. Further, the size and mass of an infant's head and brain is disproportional to its tiny body and equivalent to one-third of a full-grown adult's body. Brain tissue is highly incompressible but relatively deformable in both adults and infants. Without the protection of the hardened skull of an adult, however, infants and young children are especially susceptible to injuries of the head and brain due to deformation, even from seemingly innocuous forces.

Head trauma is the most common cause of traumatic death in children (Quayle et al., 1997; Shane and Fuchs, 1997; CDC, 2010). Approximately, 600,000 children experience blunt head trauma per year in the United States and 5700 children die as a result (Cheung and Kharasch, 1999; Kraus et al., 1990; CDC, 1990). The Centers for Disease Control (CDC) include, under the definition of head trauma, penetrating injuries, traumatic brain injury (TBI), skull fracture, contusions, concussions, and various types of hematomas. In general, head trauma is accidental or intentional in nature.

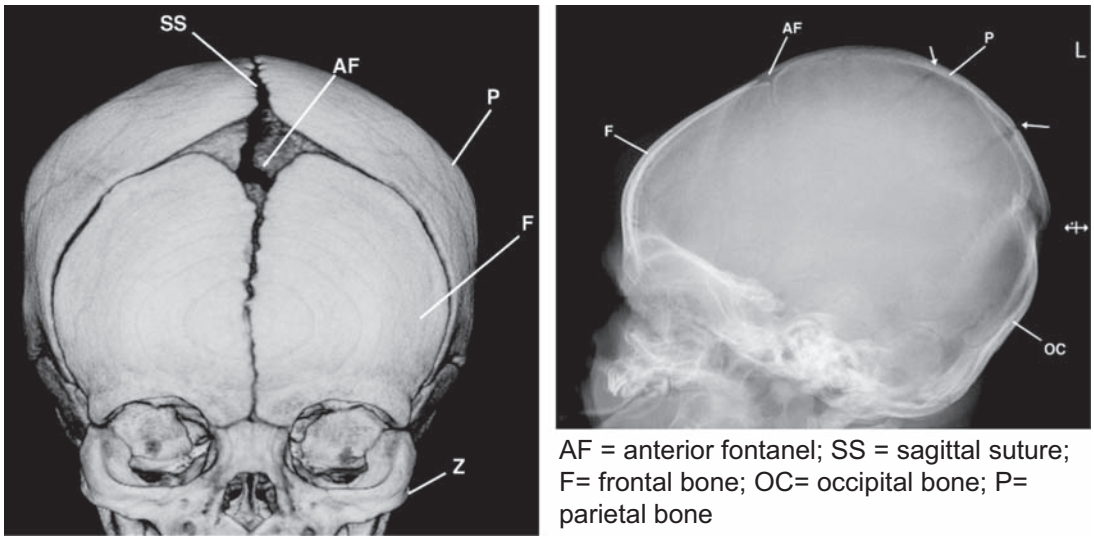


FIGURE 9.1 The infant skull and its associated suture structures (Collins, 2013). (X-ray image illustrates how thin the infant skull is, particularly in the parietal region.)

Falling is inevitable in childhood, and particularly as infants transition to toddlers, and injury can result even with the most attentive adult caregiver. This is not to say that fall-related injuries are unpreventable, and parent education in effective prevention strategies plays an important role in reducing morbidity and mortality from falls.

Health-care visits for pediatric head trauma is common; the CDC reported that there were 251,545 reported emergency department (ED) visits for TBI in children aged 0–4 in the United States during 2002–06, and that falls were the most common cause of the injury. The CDC also notes that falls are the leading cause of TBI in children of this age group. Soreide et al. (2009) note that 11% of pediatric fatal trauma is due to falls. Reece and Sege (2000) report that 81% of pediatric head injuries are accidental, with 58% of these injuries due to falls (more than half of which are considered short falls, ie, less than 4 ft). For the remaining 19% of alleged abusive injuries, 17% were claimed to be due to falls (89% of these falls less than 4 ft). In their analysis of US children hospitalized for TBI, Shi et al. (2009) found that among <1-year-olds, falls on/from stairs or steps accounted for 8.9% of all falls, fall from a chair 5.7%, falls from a bed 22.8%, falls from other furniture 8.7%, and falls (same level) due to slips, trips, or stumbles to be 2.7%. Falls from height were found to account for about 35% of pediatric skull fractures (Kraus et al., 1986). Ersahin et al. (1996) found that depressed skull fractures accounted for 15–25% of children with skull fractures, with falls from height being the most common cause. In a similar study, Holsti et al. (2005) found that of 827 child patients seen in the ED for closed head injuries, of which 39% had skull fractures, 60% were the result of falls. Leventhal et al. (2008) determined that among children less than 3 years of age who required hospitalization during 2003 for fractures, about 12.1% were the result of abuse.

As noted previous, falls are a leading cause of pediatric injuries, with falls being responsible for the majority of TBI cases. Very young children are more restricted in the possible causes of head trauma, relative to adults. In the adult population, falls account for only 28% of TBI cases. Other causes include motor vehicle collisions (20%), and being struck by or against other persons or objects (19%), and assaults (11%) (Langlois et al., 2006). In contrast with the generally accepted premise that falls at ground level can and do produce serious and sometimes fatal head injuries, whether or not the same mechanism can produce similar injuries among infants and toddlers is more controversial.

In the legal realm, child abuse cases involving alleged abusive head trauma are often claimed by the defendant to be the result of an accidental fall from height, typically less than 4 feet—a so-called "short fall." The fall may be the result of the child falling off a piece of furniture or an accidental drop from the caretaker's arms. Alternately, the child may have experienced a trip-and-fall or have been knocked to the ground by another child or family pet. Falls downstairs are more complex and potentially involve a series of head impacts. The injuries resulting from such unintentional events can include the same injuries seen in abusive head trauma cases, and include skull fractures, TBI, retinal hemorrhages (RHs), and injuries to other parts of the child's body.

Many alleged child abuse events involving a fall claim are unwitnessed, aside from the testimony of the care provider. In a judicial setting, prosecution medical experts may completely disregard the claimed fall event, instead relying on an assertion that the observed injuries could not have resulted from the history given by the caretaker. It has been previously noted that when infants present with head trauma associated with a claimed short

fall, medical personnel tasked with assessing the cause of the injury for the purposes of a criminal investigation will generally presume that the incident was nonaccidental (Rorke-Adams et al., 2008). The concept of true and false positive rates associated with the assessment of the cause of a pediatric head injury is of paramount importance in assessing the reliability of an opinion given in a forensic setting. Despite this fact, opinions regarding the cause of a pediatric head injury are often given with no means of weighing the likelihood that the opinion is accurate.

Over the past three decades, opinion leaders in child-abuse medicine have taken the position that a short-distance fall is unlikely to result in serious head injury in an infant or a toddler. In 1991, for example, Chadwick et al. (1991) suggested a minimum height threshold of 1.5 m for a fatal pediatric fall. After examining the records of all children admitted with a reported fall over a 43-month span at the Children's Hospital-San Diego trauma center, the authors determined that the deaths from alleged low falls were actually due to abuse. Noting high survival rates in one-, two-, and even three-story falls, the authors wrote, "Long falls outside of buildings are more likely to provide accurate data points for studies of children's injurability, and research on children's injurability should utilize these longer falls rather than short indoor falls witnessed by just one person." In the same era, Williams (1991) concluded from a series of 398 consecutive pediatric falls treated at Children's Hospital-Oakland emergency room (ER) that "falls of less than 10 feet are unlikely to produce serious or life-threatening injury." After examining a series of 363 stairway falls treated at Children's Hospital of Philadelphia ER, Joffe and Ludwig (1988) wrote, "Children frequently injure themselves in falls down stairs but usually not seriously." Mainstream child-abuse texts still feature advice from the late pathologist and pediatrician Dr. Robert Kirschner who wrote "Falls within the home, down stairs, and so forth, are relatively trivial events and produce trivial (non-life-threatening) injuries. Blows, shaking, and forced impact are violent events and produce potentially lethal injuries."

An often-referenced epidemiological study of pediatric calls that is frequently referenced in medicolegal cases concluded that the probability of an injury from a short fall was very, very low (Chadwick et al., 2008). The authors noted that their study objective was to develop an estimate of the risk of death in children aged 5 and less, resulting from short falls of <1.5 m in height. The authors found that a large injury database in California indicated six possible short fall-related fatalities in a population of 2.5 million children over a 5-year period. The authors arrived at an annual mortality rate of <0.48 deaths per 1 million children aged 5 and under. The conclusion is easily (and often) misinterpreted as indicating that the risk of death from a short fall is less than 1 in a million. In actuality an estimate of the risk of death from a short fall would have to use all short falls as a denominator, not the total population.

Researchers from a range of disciplines have continued to explore the possible effects of short and long falls through epidemiological studies, biomechanical analysis (including recreations), and computer simulation techniques such as finite element modeling. Some analyses have concluded that short falls can be dangerous. In 1989, Hall et al. (1989) published a review of autopsy records over 4 years in Cook County, Illinois, in which 18 pediatric fatalities attributed to falls of less than 3 feet were noted, with two of them occurring in children who were under medical observation at the time. In 2001, Plunkett reviewed a decade of head injuries reported from the use of playground equipment to the US Consumer Product Safety Commission database, identifying 18 fatal falls in children aged 1–13 years

from heights of between 2 and 10 feet. One of the falls was videotaped, allowing for a detailed biomechanical analysis of the forces of the fall using dummies as surrogates.

There are a number of variables that can influence injury risk in short falls. Fall orientation is a significant determinant of outcome, since this affects the relative contributions of linear and angular kinematics. While a horizontal fall might result in a moderate impact force due to linear accelerations, the contributions due to angular accelerations would be negligible. Whereas, a roll fall from the same height would result in significantly greater forces due to angular accelerations (Lloyd, 2013).

Presented in the following sections of this chapter are a detailed discussion of the biomechanics of head and brain injury, including original research from the coauthors. In the concluding section of the chapter the authors' perspective on the role of biomechanical analysis in abusive head trauma cases is presented.

BIOMECHANICS OF HEAD AND BRAIN INJURY

There are two primary mechanisms associated with TBI—impact loading and impulse or inertial loading. The latter may occur in conjunction with impact loading or may be the result of head acceleration only. Impact loading (blunt trauma) involves a direct blow transmitted primarily through the center of mass of the head, resulting in extracranial focal injuries, such as contusions, lacerations, and external hematomas, as well as skull fractures. Shock waves from blunt force trauma may also cause underlying focal brain injuries, such as cerebral contusions, subarachnoid hematomas, and intracerebral hemorrhages. In contrast, concussion results from impulse or inertial loading, which is caused by sudden movement of the brain relative to the skull. Inertial loading resulting in differential movement at the surface of the brain can cause subdural hemorrhage due to bridging vein rupture, whereas diffuse axonal injury results from a similar mechanism occurring to deeper structures throughout the brain. Holbourn (1943) was the first to cite angular/rotational acceleration as an important mechanism in brain injury. Gennarelli et al. (1971, 1981, 1982), Gennarelli and Thibault (1982), and Thibault and Gennarelli (1985), in a series of studies using live primates and physical models, investigated the role of rotational acceleration in brain injury. They concluded that angular acceleration contributes more than linear acceleration to the generation of concussive injuries, diffuse axonal injuries, and subdural hematomas.

Skull Fracture

Hobbs (1984) studied 89 pediatric skull fractures in children less than 2 years of age, of which there were 20 fatalities. The author concluded that accidental events often were associated with a single, narrow, linear fracture with no associated intracranial injury, and that multiple or complex fracture patterns were more characteristic of abusive situations.

Weber (1984) conducted a systematic biomechanical evaluation using 15 human cadavers (age 3.4 $\pm$ 2.18 months; weight 5.13 $\pm$ 1.6 kg). The infant cadavers were dropped from a height of 0.82 m onto three different floor surfaces: (1) stone tile floor, (2) carpet floor with foam backing on tile floor (carpet thickness 0.3 cm, foam thickness 0.3 cm), (3) linoleum floor with foam backing on tile floor (linoleum thickness 0.2 cm, foam backing 1.0 cm).

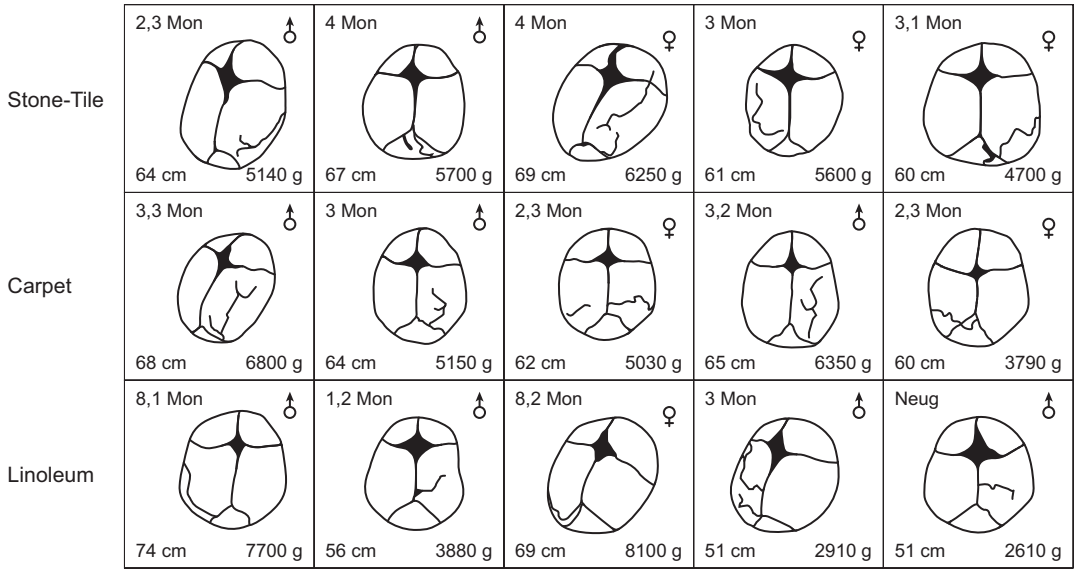


FIGURE 9.2 Infant skull fractures as observed in Weber's (1984) investigation.

Cadavers were equally divided into the three impacted surfaces (5 each). All cadavers were dropped such that the body and skull impacted the floor simultaneously. The skull impact point was the occipital-parietal region. All skulls showed various degrees of skull fractures (see Fig. 9.2). In three cases the skull fractures crossed the skull sutures. Two of these were dropped on the tile floor and the third on the carpet. In the rest of the cases, fractures were confined within the suture boundaries.

In a follow-up study, Weber (1985) conducted 35 additional falling tests where the contact surface was softly cushioned ground, including 10 falls onto a 2-cm-thick foam rubber mat and 25 falls onto a double-folded 8-cm-thick camel hair blanket. As in the previous 1984 study, the fall height was 82 cm. One parietal skull fracture was observed in the falls onto the 2-cm-thick foam rubber mat and four in the double-folded 8-cm-thick camel hair blanket. Reflecting on the combined study results (50 total falls), Weber concludes: "These results indicate that it is no longer possible to assume that the skull of infants is not damaged after falls from table height."

Prange et al. (2004) investigated the mechanical properties and anthropometry of the pediatric head. They hypothesized that the pediatric skull would exhibit viscoelastic properties, unlike that of the adult skull. Their protocol included quasi-static and dynamic compression tests on neonatal cadaver skull specimens and also drops of pediatric cadaver skulls from heights of 15 and 30 cm onto a flat smooth anvil. Compression testing included several strain rates, with 0.05 mm/s considered quasi-static, and two different orientations of the skull were employed: anterior–posterior and right–left. Stiffness (N/mm) values were obtained from the compression tests. The fall test protocol was designed to minimize any head rotation during the fall. Five different impact locations were studied for the falls: the vertex, occiput, forehead, right parietal bone, and left parietal bone. Fig. 9.3 presents the

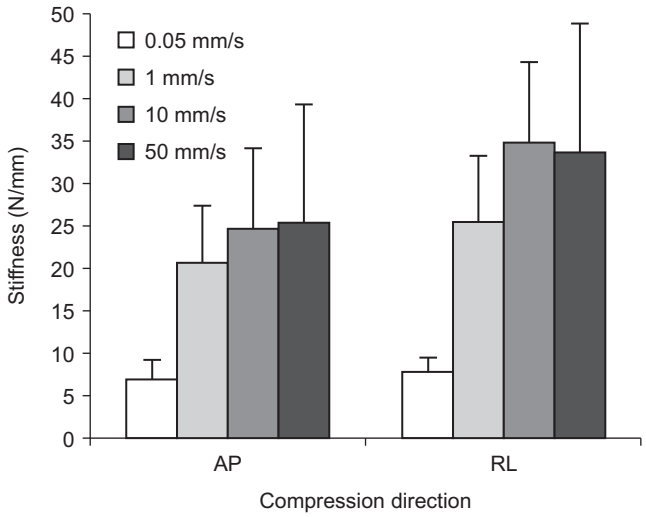


FIGURE 9.3 Summary of the compression testing of pediatric cadaver skulls from Prange et al. (2004). Three specimens were studied. Testing was done for the anterior–posterior (AP) direction and the right–left (RL) direction. Standard deviation error bars are included.

results of their compression testing. The authors found that the infant specimens displayed an increase in stiffness with increasing rate of compression, but there was no directional dependence. Other studies have disagreed on this point, finding a directional dependence (Hodgenson et al., 1967; Thomas et al., 1968). The authors also note that the stiffness values for the pediatric specimens are significantly lower than that for adult specimens. A graphical summary of their results for the drop studies is presented in Fig. 9.4. The average head accelerations were 38.9 and 55.3 g for the 15 and 30 cm drops, respectively. The average pulse durations were not significantly different for the two drop heights, with an average pulse duration of 18.3 ms.

Building on the findings from Weber (1984, 1985) and Prange et al. (2004), Van Ee et al. (2009) reproduced many of the earlier studies and added some additional conditions. Based on the collective results, Van Ee developed a probability of skull fracture versus linear acceleration curve presented in Fig. 9.5. Translating the accelerations into the associated fall heights, the threshold is somewhere between 12″ and 32″.

Concussion

Ommaya and Gennarelli published a landmark study on cerebral concussion and traumatic unconsciousness in 1974. The researchers devised an inertial loading apparatus (Fig. 9.6) that induced pure translational (linear) or rotational (angular) loading on the heads of primates without producing contact phenomena (impact).

Their results show that both linear and angular forces contribute to injurious effects on the brain. Pure translation produced focal injuries, such as contusions, while diffuse effects, including concussion and subdural hematoma were only produced when the rotational loading was present.

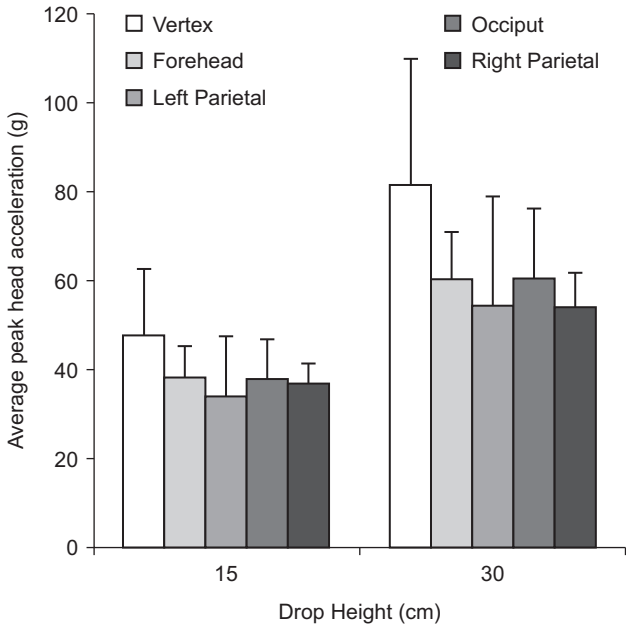


FIGURE 9.4 Summary of the fall experimental data from the study of Prange et al. (2004). Three specimens were evaluated. Standard deviation error bars are included.

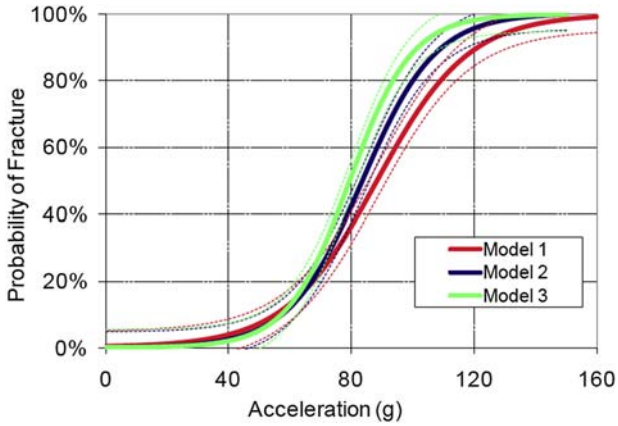


FIGURE 9.5 Injury risk models for the prediction of skull fracture based on peak linear head acceleration (Van Ee et al., 2009).

Cerebral Edema

Malignant cerebral edema (aka posttraumatic brain swelling) is a rare, but often fatal complication of TBI characterized by an excess accumulation of fluid in the intracellular or extracellular spaces of the brain. Trauma or injury can cause cerebral blood vessels to rupture,

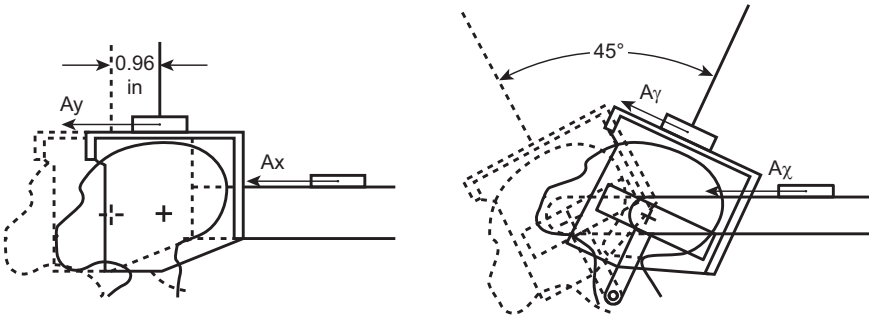


FIGURE 9.6 The test device used by Ommaya and Gennarelli (1974). (The left device measures translational motion and the right device rotational loading.)

which changes the water, sodium, and potassium levels inside the brain, which continue to rise in the first several days after injury. The condition is diagnosed by a rapid and uncontrollable increase in intracranial pressure (ICP) within hours after an injury.

Posttraumatic edema can be divided into two *temporal* phases, along with two *pathophysiologic* mechanisms. There is immediate (early) swelling, within minutes of injury, and delayed (late) swelling, which *peaks at 3–5 days postinjury*. Furthermore, there is vasogenic edema (ie, accumulation of extravasated plasma in the interstitium due to derangement of the vascular epithelium) and cytotoxic edema (due to accumulation of intracellular water as energy failure takes place).

Vasogenic edema occurs due to a breakdown of the tight endothelial junctions which make up the blood–brain barrier (BBB). This allows intravascular proteins and fluid to penetrate into the parenchymal extracellular space. Once plasma constituents cross the BBB, the edema spreads; this may be quite rapid and extensive. As water enters white matter, it moves extracellularly along fiber tracts and can also affect the gray matter. This type of edema may result from trauma, tumors, focal inflammation, late stages of cerebral ischemia, and hypertensive encephalopathy.

Mechanisms contributing to BBB dysfunction include physical disruption by arterial hypertension or trauma, and tumor-facilitated release of vasoactive and endothelial destructive compounds (eg, arachidonic acid, excitatory neurotransmitters, eicosanoids, bradykinin, histamine, and free radicals). Subtypes of vasogenic edema include hydrostatic cerebral edema. This form of cerebral edema is thought to result from direct transmission of pressure to cerebral capillaries with transudation of fluid from the capillaries into the extravascular compartment.

In *cytotoxic edema*, the BBB remains intact. It occurs due to a disruption in cellular metabolism that impairs functioning of the sodium and potassium pump in the glial cell membrane, leading to cellular retention of sodium and water. Swollen astrocytes occur in gray and white matter. Cytotoxic edema is seen with various toxins, including dinitrophenol, triethyltin, hexachlorophene, and isoniazid. It can occur in Reye's syndrome, severe hypothermia, early ischemia, encephalopathy, early stroke or hypoxia, cardiac arrest, and pseudotumor cerebri.

During an ischemic stroke, a lack of oxygen and glucose leads to a breakdown of the sodium–calcium pumps on brain cell membranes, which in turn results in a massive build

up of sodium and calcium intracellularly. This condition causes a rapid uptake of water and subsequent swelling of the cells (Rosenberg, 1999). It is this swelling of the individual cells of the brain that is seen as the main distinguishing characteristic of cytotoxic edema, as opposed to vasogenic wherein the influx of fluid is typically seen in the interstitial space rather than within the cells themselves (Klatzo, 1987). While not all patients who have experienced a stroke will develop a severe edema, those who do have a very poor prognosis (Hacke et al., 1996).

In most instances, cytotoxic and vasogenic edema occur together. It is generally accepted that cytotoxic edema is dominant immediately following an injury or infarct, but gives way to a vasogenic edema that can persist for several days or longer (Rosenberg, 1999). The use of specific MRI techniques has allowed for some differentiation between the two mechanisms and suggests that in the case of trauma, the cytotoxic response dominates (Barzó et al., 1997).

Two additional components of cerebral edema are an increase in the cerebral blood volume (CBV) due to acid–base disturbance (hypercarbia), which results in vasodilation, and interstitial edema due to posttraumatic hydrocephalus. Astrocytes, which have larger number of mitochondria in children, compared to adults, are thought to be the cellular origin of malignant edema, as:

1. they are responsible for maintenance of the interstitial ionic milieu;
2. they are capable of swelling to 20 times their normal size;
3. they have the largest number of ATP-dependent ionic exchange pumps;
4. mitochondrial dysfunction plays a primary role in cytotoxic edema, as demonstrated in animal and pharmacologic knockout models.

Two distinct processes: (1) increased CBV and (2) possible loss of cerebral vascular autoregulation may contribute to the pathophysiology of malignant cerebral edema.

Malignant cerebral edema can occur within hours of the head injury. It is more common in children and young adults and has a high mortality rate (Malignant cerebral edema, n.d.).

Treatment: Aggressive management is needed to preclude devastating neurological deficit and mortality. Goals of treatment are to keep ICP $<$20 mm Hg and CPP $>$60 mm Hg. It is unclear whether patients *may* benefit from early surgical decompression. Since TBI-related edema represents catastrophic failure of mitochondrial energy exchange, prognosis for meaningful outcome is often guarded and uncertain.

Subdural Hematoma

According to Gennarelli and Thibault (1982) the most common type of acute subdural hematoma results from tearing of veins that bridge the subdural space as they travel from the brain's surface to the various dural sinuses. The severity of injury associated with bridging vein rupture has led to several studies of mechanical failure properties (Lowenhielm, 1974, 1975, 1978; Lee and Haut, 1989; Meaney, 1991; and Depreitere et al., 2006). The results from these studies are summarized in Table 9.1.

Lowenhielm (1974) performed mechanical tests on 22 human parasagittal bridging vein samples from 11 decedents between the age of 13 and 87 years with no previous brain injury.

TABLE 9.1 A Summary of Failure Properties of Human Bridging Veins

Study	Samples	Age range	Maximal stress (N/mm^2)
Lowenhielm (1974)	22	13–87 years	1.5–1.55
Lee and Haut (1989)	139	62–85 years	1.513
Meaney (1991)	59	9–62 years	1.55

He hypothesized that bridging vein rupture is a consequence of rapid brain deformation in response to head angular acceleration. When the head is exposed to blunt trauma, the brain is displaced with respect to the dura, causing bridging veins and surrounding connective tissue to be stretched. Lowenhielm states that maximal shear stresses in the brain matter does not necessarily occur during impact, but about 7 ms after impact, where bridging vein disruption coincides with the occurrence of these large shear stresses. He concluded that gliding contusions are not likely to arise if the maximal angular acceleration does not exceed 4500 rad/s^2 or the change in angular velocity does not exceed 70 rad/s.

Lee and Haut (1989) performed mechanical tests to investigate the strain rate dependence of human bridging vein properties. They tested a total of 139 parasagittal bridging veins from eight unembalmed human cadavers aged between 62 and 85 years without noticeable head trauma or cerebrovascular disease. Bridging veins were stretched beyond failure by a servo-controlled hydraulic testing machine. These results indicate that bridging vein ultimate strain is independent of loading rate. Meaney (1991) also performed dynamic tests of human bridging veins. Results, from the bridging veins of 59 unembalmed cadavers ranging in age from 9 to 62 years with no sign of head injury, indicate an average stretch ratio of 1.55 across all strain rates tested.

Results from Lowenhielm (1974), Lee and Haut (1989), and Meaney (1991), described in Table 9.1, indicate failure properties of human bridging veins are consistent across children and adults.

Depreitere took a different approach to examining the consequence of subdural hematoma due to bridging vein rupture. His 2006 paper described a methodology in which 10 unembalmed cadavers were subjected to 18 occipital impacts producing head rotation in the sagittal plane with varying rotational acceleration magnitudes and pulse durations. Bridging vein ruptures were detected by injecting contrast dye into the superior sagittal sinus under fluoroscopy and by autopsy procedures. Bridging vein ruptures were produced in six impact tests. The data suggest a tolerance level of approximately 10,000 rad/s^2 for pulse durations shorter than 10 ms, which appeared to decrease for longer pulse durations. The major finding of in the paper was the determination that bridging vein rupture is a function of peak angular acceleration and pulse and impact duration, whereby the threshold in terms of angular acceleration decreases with increased impact duration. Fig. 9.7 presents data from the research by Depreitere and Lowenhielm, including a line of best fit and error bars based on the standard error of the mean.

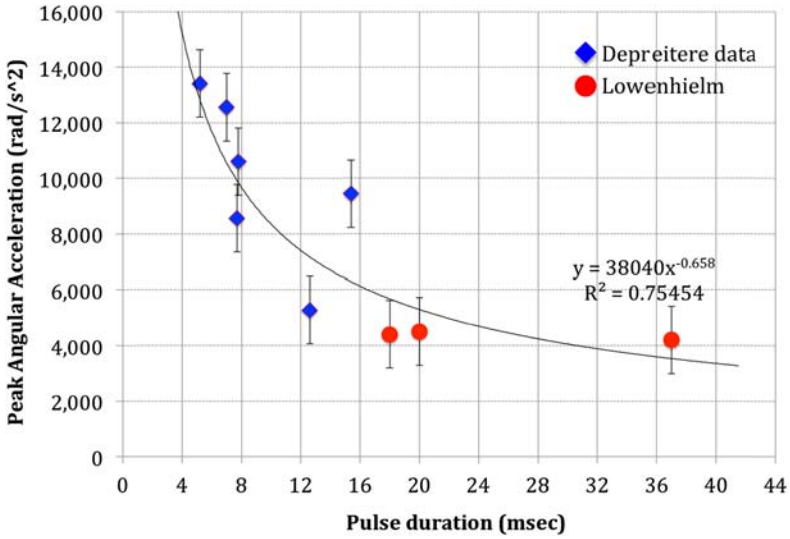


FIGURE 9.7 Bridging vein failure as a function of peak angular acceleration and pulse duration. *Data from Depreitere, B., Van Lierde, C., Sloten, J.V., Van Audekercke, R., Van der Perre, G., Plets, C., Goffin, J., 2006. Mechanics of acute subdural hematoma resulting from bridging vein rupture. Journal of Neurosurgery 104 (6), 950–956 and Lowenhielm, P., 1978. Tolerance level for bridging vein disruption calculated with a mathematical model. Journal of Bioengineering 2 (6), 501–507.*

Optic Nerve Sheath and Retinal Hemorrhages

RH is a disorder in which bleeding occurs in various layers of retinal tissue at the posterior wall of the eye. Fig. 9.8 presents the basic anatomy and vasculature of the eye. A RH can be caused by hypertension or central retinal vein occlusion, a common condition. The condition is also commonly found in pediatric victims of abusive head trauma. RH is one of the original criteria of the triad comprising the pathologic signs of shaken baby syndrome (along with intracranial bleeding and swelling), and it is thought to result repetitive rotational acceleration/deceleration associated with abusive shaking. It has been proposed that the injury mechanism results from inertial forces resulting in traction between the retina and the vitreous humor, the fluid-filled chamber that comprises the majority of the eye and is adjacent to the retina.

Vitreoretinal traction describes pathology of the eye in which the adherent vitreous fibrils in the vitreous humor pull against the internal limiting membrane of the retina. It has been suggested that acceleration/deceleration associated with shaking produces vitreoretinal traction, thereby causing RHs in infants in whom shaking is suspected as a cause. However, biomechanical study of the effects of shaking has only demonstrated relatively small forces at the eye, with the greatest forces seen in the lower cervical spine.

It is now generally understood and accepted that RH can be caused by any mechanism that increases ICP of the brain. The eye is the only externally visible element of the central nervous system, where the optic nerve and retinal vasculature are sheathed within the dura (Fig. 9.8). Since arterial pressure exceeds venous pressure, returning blood through the central retinal vein will be impeded with increases of arterial pressure, thereby increasing

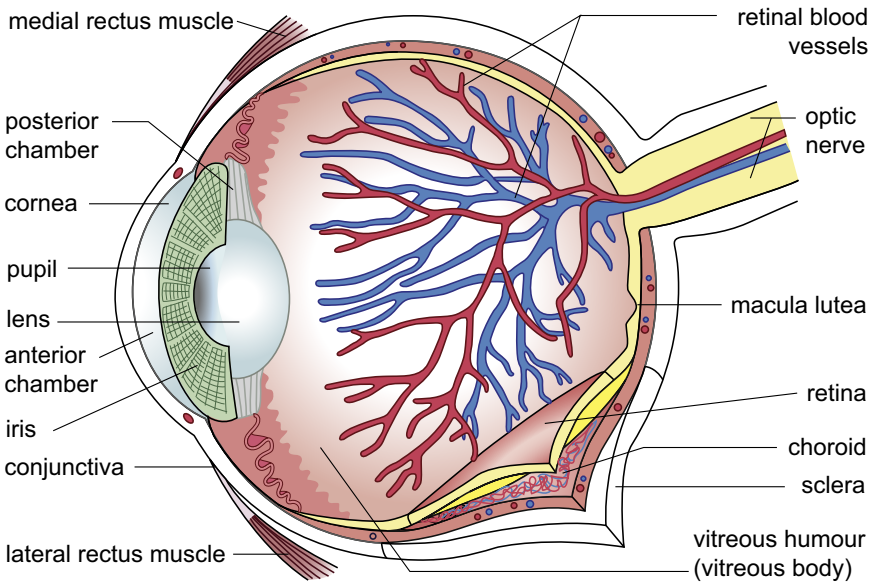


FIGURE 9.8 Basic anatomy and vasculature of the human eye.

back pressure within capillaries in the retina, which consequently burst producing RHs. Since the underlying etiology is increased ICP, bilateral observations would be typical, where the degree of hemorrhage would be proportional to the ICP.

It has been suggested that RHs associated with abusive head trauma in infants are distinct from those associated with other forms of trauma in that they tend to be more extensive, evident in various retinal and subretinal layers, and extending throughout the retina to the ora serrata (anterior border of the retina).

Another determinant of the severity of RH is the duration over which increased ICP is experienced. In cases of alleged abusive head trauma in infants, ophthalmologic examinations are routinely performed, but since RHs are not life threatening, the examination may be delayed until all emergent evaluations and procedures have been completed. During this time, the number and degree of RHs may continue to increase until ICP is relieved with the passage of time, medical intervention, or death.

While RH is associated with abusive head trauma, the association is not exclusive to this cause in pediatric deaths. In a one retrospective review of 137 pediatric deaths in children <3 years of age at a metropolitan medical examiner's office, the authors noted that the incidence of RH and optic nerve sheath hemorrhage (ONSH) was greater in cases following advanced life support and cerebral edema than with any other etiology (Matshes, 2010 AAFS proceedings). A recent publication described a finding of bilateral RH and ONSH in a documented short fall of a 7-month-old infant (Lantz and Couture, 2011). RH is observed in non-pediatric populations as well, particularly those who are subjected to large variations in externally and internally generated pressures, including deep sea divers and weight lifters.

TABLE 9.2 Summary of Biomechanical Pediatric Head and Brain Injury Thresholds

Injury	Source	Description	Linear acceleration (g)	Angular acceleration (rad/s^2)
Skull fracture	Van Ee et al. (2009)	25% Probability of fracture	70	
		50% Probability of fracture	82	
		75% Probability of fracture	94	
Subdural hematoma	Lowenhielm (1974)	Bridging vein failure threshold		4500
	Ommaya (1984)	AIS 5 critical injuries		≤4500
	Depreitere et al. (2006)			4000–6000

While RH and ONSH are sensitive indications of abusive head trauma (ie, the findings are present in a majority of cases), they are not particularly specific for the condition (ie, they are also present in cases with no intentional injury mechanism).

Summary of Biomechanical Pediatric Head and Brain Injury Thresholds

Table 9.2 is a summary of biomechanical pediatric head and brain injury thresholds.

PEDIATRIC HEAD INJURIES AND FALLS

Because the alternative explanation in investigated cases of abuse is often a fall described only by a caregiver, it is helpful to understand what has been previously published on the types of injuries observed in falls occurring at varying heights.

A study by Helfer et al. (1977) described the investigation of short falls in 246 children, with 219 occurring at home and 95 in a hospital. Most of the home falls were 90 cm or less, but the impact surface was not specified. Two of the children sustained skull fractures, judged to not be serious in nature. For the hospital falls, the contact surface was presumed to be a noncarpeted floor. One child in this group sustained a skull fracture. A more recent study by Lallier et al. (1999) described 1410 pediatric patients who went to an ER after a fall, with the fall distance for the children in the study 10 feet or more. For the 507 subjects in the 0–4 year old group, there were five skull fractures, two concussions, and three intracranial injuries. However, the mechanics of the falls and the contacted surfaces were not reported. In a study reported by Warrington and Wright (2001), 11,466 parent-reported descriptions of fall events during the first 6 months of age were evaluated. It was found that 33.0% of the infants fell from a bed, 12% fell from arms while being carried, and other specific fall types were identified such as falls from baby chairs, tables, changing units, baby walkers, and from a caregiver's lap. The authors' analysis indicated 14% experienced some visible injury, of which 97% involved the head, and 21 fall events (<1%) lead to a concussion or fracture.

Thompson et al. (2011) described 79 falls from furniture occurring children 0–4 years of age and involving a visit to the ED. Injuries were reported as follows: 15 had none, 45 had minor injuries (AIS 1), 17 had moderate injuries (AIS 2) including 6 skull fractures, and 2 had serious injuries (AIS 3) of which both were subdural hematomas, one accompanied by skull fracture.

Wheeler and Shope (1997) reported a case where a 7-month-old infant fell 24 inches out of a bed and sustained a depressed skull fracture. A toy car found on the floor at the location of the fall corresponded to the dimensions and contour of the depression observed in the infant's skull. Reiber (1993) reported a fatality involving a 17-month-old child who fell backward from a rocking chair approximately 2–3 feet, sustaining a left subdural hemorrhage, a subarachnoid hemorrhage, and a left parietal cerebral contusion without skull fractures.

There are numerous reports in the literature of witnessed pediatric falls with significant injury. Arnholz et al. (1998) report an event witnessed by a noncaretaker involving a 6-week-old baby who fell 2–3 feet out of a stroller and impacted a concrete step. Biparietal linear skull fractures were found after an ER visit along with two areas of subgaleal (between the scalp and skull) hemorrhages. Results from a skeletal survey (whole body X-ray) were negative for other fracture. In a study reported by Williams (1991), 106 fall incidents involving infants and small children that were witnessed by a person other than the caretaker were described. Documented injuries were as follows: 15 patients had no injuries (7 of these children fell more than 10 feet), simple fractures occurred in 77 patients (43 of these fell more than 10 feet), and serious injuries, such as intracranial hemorrhages, cerebral edema, and skull fractures, occurred in 14 patients falling between 5 and 40 feet. No life-threatening injuries were observed in the 3 patients who fell less than 10 ft. Plunkett (2001) performed a retrospective study on playground-related fatal head injuries as documented by the United States Consumer Product Safety Commission over an approximately 10.5-year period. The author described 18 fatal head injuries resulting from falls, with the age ranging from 12 months to 13 years and the fall heights from 2 to 10 ft. Twelve of the 18 fatal events were witnessed by someone other than the care provider. Four of the falls had a finding of RH (although not all of the cases provided information regarding the finding), 13 patients had a subdural hemorrhage, and 5 patients had skull fractures. In addition to documenting that finding that fatalities due to head injury can occur in short falls, the author noted that such fall-related injuries may be associated with a lucid interval after the injury, as well as bilateral RHs. The former finding was important, as the timing of injury to loss of consciousness can be critical evidence in determining when a caregiver had access to a child.

Kelly and Hayes (2004) reported the results of a retrospective study of infants younger than 2 years who experienced a subdural hematoma with RH. The authors identified 64 cases, of which 41 were nonaccidental trauma and 23 were accidental. One of the accidental cases was a corroborated fall from an adult's arms. Another witnessed incident involved a fall of an adult onto a child.

Hall et al. (1989) reported on 18 children who died from accidental falls of <0.9 m, of which 8 falls were witnessed by 2 or more people in public places, with 2 of the falls occurring under medical observation. Reported injuries included 15 subdural hematomas (5 of these also had linear skull fractures), 1 epidural hematoma, and 1 case of cerebral edema. Child abuse was ruled out in all 18 cases.

Lantz and Couture (2011) described a case where an 8-month-old male infant died after falling downstairs, sustaining a subdural hematoma, a subarachnoid hemorrhage, and hemorrhagic retinopathy. The death was deemed accidental, as two witnesses heard a thud and found the infant lying on the landing of the basement steps.

Reichelderfer et al. (1979) reported on injuries documented at playgrounds, noting that a significant percentage of the injuries were due to falls, and that many of the falls were at a low height. The authors calculated that for a fall height of 1 foot, the resulting acceleration at impact would be in the range 475–525 g for a fall onto concrete and 140–160 g for a fall onto asphalt. The authors advocated the use of materials such as sand, wood chips, and pea gravel, which would significantly reduce the calculated impact force of short falls.

Powell et al. (2002a) conducted a retrospective study on children 3 years and younger regarding stroller-related injuries occurring during 1994–98; this was essentially a short fall study. Data were obtained from the National Electronic Injury Surveillance System (NEISS) of the United States Consumer Product Safety Commission. The authors reported that 76% of the estimated 64,373 injuries involved falls. Injuries included closed head injuries (22%) and skull fractures (<1%). Using similar methods, the Powell and colleagues studied high chair-related injuries (Powell et al., 2002b), again resulting in a short fall study. In the 40,650 estimated high chair-related injuries, 44% involved the head, with 21% of the injuries described as a closed head injury, 3% were intracranial hematoma, and <1% skull fracture.

Kurinsky et al. (2013) investigated injuries to children 3 years and less associated with falls from high chairs and regular chairs, using the same NEISS database accessed for the Powell et al. studies, but for a different time period (2003–10). In their methodology, "high chairs" included high chairs, booster seats, and attachable or reclining feeding seats. The authors noted that high chairs tend to be at a height above a typical chair. Among an estimated 402,479 cases, falling was the most common event with 92.8% of the high chair cases described as falls and 87.3% of regular chair cases associated with falls. The authors noted that in most cases, the fall from a high chair was the result of the child standing up, twisting, or climbing out of the seat. Among the high chair-related injuries, 37.3% of the cases were diagnosed with a closed head injury and for chairs there were 20.7% with a closed head injury. The authors observed that patients <1 year old were 1.44 times more likely to experience a closed head injury than children in the 1–3 year age group. Schalamon et al. (2006) reported similar results regarding high chairs and the results of short falls.

Other authors have reported on injuries associated with nursery-related furniture generally. Ozanne-Smith and Heffernan (1990) divided nursery-related furniture into six categories: strollers and prams, baby walkers, high chairs, changing tables, cots (excluding portables), and baby exercisers. The authors noted that almost all of the injuries occurred during the first 3 years of life (985 total cases) and that 588 of these injuries occurred during the first year of life. For falls from baby walkers (168 cases), most of the injuries were to the head, including 10 skull fractures. For changing tables, almost all of the injuries were related to falls. Watson et al. (1998) described essentially the results using similar methods.

Prange et al. (2003) conducted a biomechanical study of short fall mechanisms using an anthropomorphic test device that modeled a 1.5-year-old infant. The authors examined the forces associated with falls of 1, 3, and 5 feet, and compared these with along with intentional injury mechanisms, including vigorous shaking and blunt head impact following shaking. The fall impact surfaces included concrete, a carpet pad, and a foam mattress. The authors

found that vigorous shaking produced rotational velocities and accelerations comparable to those produced in minor falls. The blunt impacts produced rotational responses that were significantly higher than even the 5-foot fall onto concrete, however. The authors concluded that intentionally inflicted blunt impacts against hard surfaces were more likely to lead to inertial brain injuries in comparison to short falls or as a result of vigorous shaking.

Using a similar approach, Coats and Margulies (2008) employed an instrumented anthropomorphic infant surrogate to study short distance (0.3–0.9 m) falls onto a mattress, carpet pad, or concrete, focusing on linear and angular parameters in addition to impact force. Their anthropomorphic device featured a deformable skull/suture skull case and a 3-D mobile neck. The authors concluded that skull fractures may occur in head-first falls over heights of 0.9 m onto carpet and 0.6–0.9 m onto a concrete surface.

Biomechanical Analysis of Fall Events

Fall events are subject to the laws of physics. Central to the analysis is the determination of the force associated with the claimed fall event. When such forces are known (or can at least be estimated), the analysis can compare the impact forces with the injury biomechanics literature regarding the type of injuries observed. An analysis of the fall event can be conducted when the fall vertical distance and the nature of the surface contacted are known. These two parameters allow for an estimate of the impact velocity and the impact force as a function of the extent to which the impacted surface is deformed (ie, padded).

The physics equations needed to conduct such an analysis are straightforward. If available, it is also helpful to know the fall motions, ie, which body part or region contacted the surface first, and a rotational component was present, but such information is rarely known with much certainty in the absence of video evidence. For a simple fall from height (with no rotation), the impact velocity is given by:

$$v_{\text{impact}} = \sqrt{2gh_{\text{fall}}}$$

where v_{impact} is the impact velocity (m/s or feet/s), g is gravitational acceleration (9.81 m/s^2 or 32.17 feet/s^2), and h_{fall} is the vertical fall distance (m or feet). The equation assumes zero velocity at the beginning of the fall.

If rotation is involved (eg, a forward fall such as a trip), the angular velocity can be calculated with $\sqrt{3gh_{\text{fall}}}$ given certain assumptions, however, such an analysis can be complicated when examine the movement of a falling body.

The deceleration at impact is given by:

$$a_{\text{decel}} = \frac{v_{\text{impact}}^2}{2d_{\text{stopping}}}$$

where d_{stopping} is the stopping distance (m or feet) and a_{decel} is the deceleration in feet/s^2 or m/s^2. The calculation assumes zero velocity at the end of the deceleration. In most cases, the stopping distance is an estimation and is usually a function of the material that is contacted by the falling object with an additional contribution from the deformation of the surface that is doing the contacting (eg, the head). It should be noted that the deceleration calculated from the equation is the average deceleration. The peak deceleration can be 2–2.5 times higher

in falls from height situations (Schulz et al., 2008). Most investigators divide a_{decel} by the gravitational acceleration to provide units in terms of earth's acceleration, or g's.

EXPERIMENTAL STUDIES

In this section, we (the authors) describe the results of our original research, along with others (Margulies, Prange, Ibrahim, etc.) and compare our findings to the injury threshold data represented in Fig. 9.7 (the Depreitere graph) to assess the relationship between falls, shaking, and injury risk. We note that others might come to different conclusions than we have; however, we present these findings so that readers can assess the evidence for themselves.

STUDY 1: BIOMECHANICAL EVALUATION OF HEAD KINEMATICS DURING INFANT SHAKING VERSUS PEDIATRIC ACTIVITIES OF DAILY LIVING

A biomechanical study was performed to quantify kinematic variables associated with various infant shaking mechanisms/techniques absent head impact, and compare these values to a series of pediatric activities of daily living, in the context of previously described biomechanical injury thresholds. The purpose of the study was to examine the potential disparity between the forces of shaking and the types of forces infants ordinarily sustain.

Using Intersense™ sensors attached to the heads and torsos of two infant surrogates, the investigators collected linear and angular motion data during resuscitative, aggressive, and gravity-assisted shaking as well as during various nonabusive activities normally experienced by infants, such as burping, rough play, etc. Tasks were performed by nine adult subjects, ranging in age from 20 through 77 years and included two females and seven males. The researchers also collected data from a 7-month-old infant child spontaneously at play in a commercial jumping toy (see Fig. 9.9). Raw data including orientation, angular velocity, and linear acceleration were acquired wirelessly. Using MatLab (The MathWorks. Natick, MA), data were filtered by the fourth order Butterworth low-pass filter. Angular accelerations were subsequently derived and head injury criterion (HIC) values were computed. The HIC is a unitless measure of serious head injury risk related to linear acceleration. The experimental findings were compared with previously published biomechanical studies of shaking, as well as with experimentally derived biomechanical injury thresholds.

The average peak rotational acceleration generated in the biofidelic mannequin (a child restraint airbag interaction (CRABI)-12 test dummy) by the nine adult subjects was 1068 rad/s^2. This value was consistent with the prior published reports from biomechanical studies, and statistically undifferentiated from the angular accelerations observed in a normal 7-month-old infant at play in a commercially available jumping toy (954 rad/s^2).

Additionally, the measures of angular acceleration generated during the experiment

STUDY 1 *(cont'd)*

FIGURE 9.9 Male subject demonstrating aggressive shaking with a doll designed by the National Center for Shaken Baby Syndrome (left) and a 7-month-old infant at play in his Fisher Price Jumperoo (right).

were substantially below previously published injury thresholds. Thus, in the absence of head impact, we were unable to experimentally demonstrate the level of forces that would explain the triad of diagnostic findings that have been attributed to shaken baby syndrome.

STUDY 2: BIOMECHANICS OF SHORT FALLS IN CHILDREN

This study involved the systematic assessment of accelerations associated with falls from heights ranging from 2 to 6 feet onto varying flooring surfaces including concrete, linoleum, apartment grade carpeting with underlay, berber carpet with underlay, commercial carpeting without pad, and wood laminate.

A CRABI-12 biofidelic mannequin (29.5 inches/22 lbs) and a Hybrid III 3-year-old (37.2inches/35.65 lbs) biofidelic mannequin were used during this systematic evaluation of short falls. A triaxial piezoelectric accelerometer was installed at the center of mass of the CRABI head, in accordance with convention, along with an InvenSense triaxial digital gyroscope. Still photography and high-speed video were used to record the fall sequences.

A height adjustable platform was used to represent the fall surface (see Fig. 9.10). The platform has trapdoors, which are held in place by electromagnets. Interruption of power to the electromagnets causes the sprung trapdoors to open instantaneously, thereby initiating the fall sequence. One-hundred-and-seventy-five trials were completed to

Continued

STUDY 2 *(cont'd)*

investigate biomechanical mechanisms of injury associated with short falls in children.

Accelerometer data were acquired at a rate of 10,000 samples per second using LabView software. Data from the gyroscope were acquired at 3800 Hz per channel. Raw data were displayed on screen for visual verification. These data were analyzed using MatLab, including fast Fourier transform analysis to visualize the frequency spectrum of the data, followed by phase-less filtering using the fourth order low-pass Butterworth filter with a cutoff frequency of 1650 Hz. The peak magnitude value of head linear and angular acceleration components were derived and HIC computed.

Tests conducted on both of the test dummies exceeded injury threshold values from a fall height of only 2 feet (61 cm), based on peak linear acceleration and HIC results. Our findings provide a biomechanical explanation for many of the head injuries described in the literature by prior authors and reviewed earlier in this chapter. Based upon these results, we conclude that household short falls present a real risk of head and brain injury among infants and toddlers.

FIGURE 9.10 Height adjustable test platform and data collection instrumentation.

STUDY 3: BIOMECHANICAL EVALUATION OF SHAKEN IMPACT SYNDROME

A biomechanical evaluation of shaken impact syndrome was performed to evaluate the risk of injury to an infant when shaking was combined with head impact. Injury risk was measured as a function of linear and angular head kinematics of a biofidelic infant surrogate during a biomechanical recreation.

Two adult males performed the shaken impact and impact activities. A CRABI-12 biofidelic mannequin, fitted with the same instrumentation as described in the previous studies, was utilized as the infant surrogate.

A number of conditions were explored using a height adjustable test apparatus (see Fig. 9.11). These included noncontact shaking, shaken impact (which implies a brief shaking episode, followed immediately by impact), and impact only. For the shaken impact and impact only scenarios, participants were instructed to impart gentle, moderate, and vigorous impacts on the infant surrogate. In addition, the act of dropping the mannequin onto the surfaces was explored for the impact only technique. Surfaces impacted included a

FIGURE 9.11 CRABI-12 mannequin impacted onto changing table pad.

Continued

STUDY 3 *(cont'd)*

standard infant crib mattress, a standard changing table pad with cover, and a hard wooden tabletop. Mattress height was set at a standard bed height of 23 inches, whereas the changing mattress and tabletop were both studied at 35 inches, as measured from the floor. Both participants performed five repeated trials for each condition, for a total of 230 trials.

Data from the instrumented dummy were acquired in the same manner as described in prior experiments. Our results indicated that angular accelerations associated with intentional impact and shaken impact are typically below 8000–10,000 rad/s^2. The exception was a vigorous impact against the changing pad or wood tabletop. It was also noted that dropping the infant surrogate onto the test surface produced accelerations and thus injury risk similar to the moderate impact condition.

Across all events where sufficient rotational brain motion was recorded to produce significant brain injury, sufficient linear acceleration to cause skull fracture in an infant was also documented. These findings suggest a high degree of association between underlying brain injury and skull fracture for the tested scenarios.

STUDY 4: BIOMECHANICAL EVALUATION OF INFLICTED HEAD TRAUMA

In cases of abusive head trauma in infants and young children, it is often alleged that the perpetrator struck the child victim with their hand or with or against a hard surface. In the present study the investigators performed a biomechanical evaluation of head and brain injury risk associated with such intentional injury mechanisms.

Two adult male investigators served as participants; neither had any physical disabilities that might affect their performance. An instrumented CRABI-12 dummy was again used as an infant surrogate. Eight conditions were investigated; these included striking the head with both an open hand and a baseball bat, as well as impacting the mannequin head against interior wall structures both on and between supporting studs (see Fig. 9.12). The walls were constructed for the purpose of the study using 2 × 4 inch vertical wood studs set at 16 and 24 inches separation, and covered with 3/8 and 1/2 inch gypsum drywall was fastened in accordance with local building codes. Participants were instructed to impart gentle, moderate, and vigorous impacts to the infant surrogate, repeated five times for each of the eight conditions, resulting in a total 240 trials.

Data from both the dummy instruments were acquired as previously described. The results of these experimental intentional impacts are described in the following section.

STUDY 4 *(cont'd)*

FIGURE 9.12 CRABI-12 mannequin vigorous impact into drywall.

Summary of Experimental Studies

In Fig. 9.13 below is a graphical presentation of the risk of subdural hematoma associated with the results of the experimentally recreated accidental events and intentional acts in the four studies described earlier. Peak angular acceleration observed in the studies is plotted against pulse duration according to the following equation by Depreitere et al. (2006):

$$t = \dot{\omega}_P / 0.625 \times \ddot{\omega}_P$$

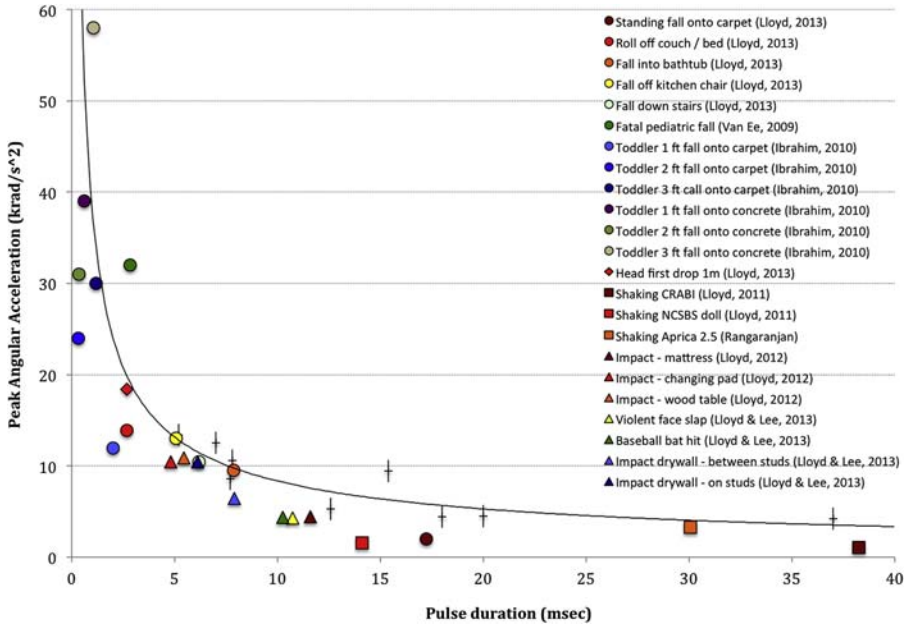


FIGURE 9.13 Risk of pediatric subdural hematoma due to accidental versus intentional events.

where $\dot{\omega}_p$ is the peak change in angular velocity, $\ddot{\omega}_p$ is the peak angular acceleration, and t is pulse duration.

Data points for bridging vein failure threshold, based on Lowenhielm (1974) and Depreitere et al. (2006), are indicated by cross markers. A line of best fit was calculated through these points, which is represented by the darker line. This limited data set has a considerable standard error, as illustrated by error bars. A light gray line outlines the lower range of values that might produce subdural hematoma.

The accidental events are denoted with circle markers, the intentional events are indicated by triangle markers, and the shaking events by square markers. A fall onto head from 1.0 m may be either accidental or intentional, therefore is represented by a diamond marker.

These results indicate that both accidental and intentional events have the potential to cause subdural hematoma in a pediatric population. Specifically, based on research by Lloyd (2013), rolling off a couch or bed from a height of 18 inches or falling from a kitchen chair onto a hard surface are both potentially injurious, as is a standing fall into an empty bathtub, or falling down a short series of stairs. Results from Ibrahim and Marguiles (2010) suggested that a 1–3 foot fall onto concrete or a 3-foot fall onto carpet may result in a subdural hematoma, consistent with a documented fall of a 23-month-old toddler from a home play set onto carpet-covered concrete floor (Van Ee et al., 2009).

Conversely, our findings indicate that shaking an infant, no matter how vigorously, is unlikely to produce a subdural hematoma (Lloyd et al., 2011). Furthermore, while violently slamming an infant down onto a hard table or changing pad can potentially cause bridging vein failure, the same action with a crib or bed mattress is unlikely to do so, as the thickness

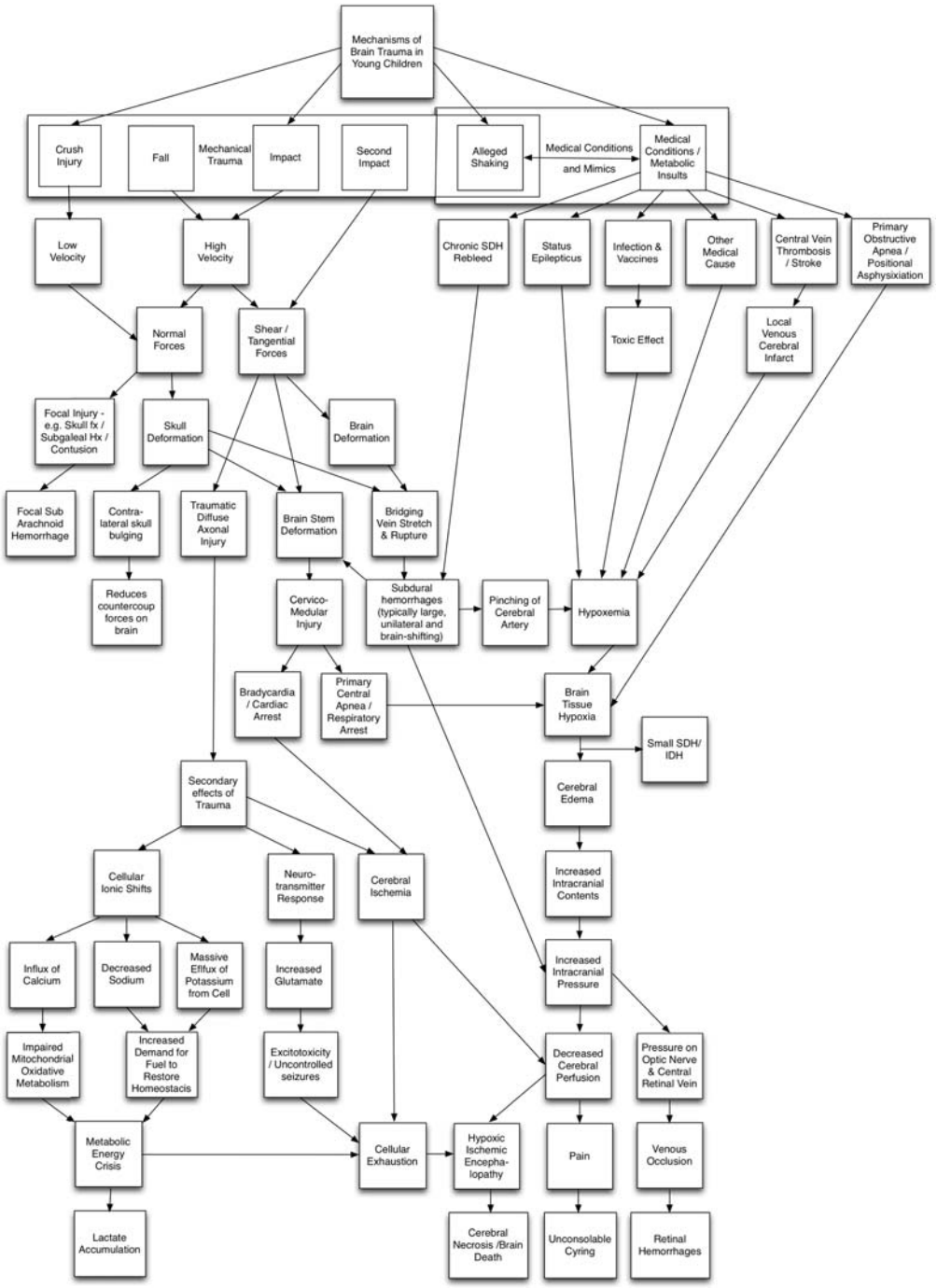


FIGURE 9.14 Flow chart summary of various mechanisms of brain trauma in young children.

of the padding would mitigate rotational accelerations of the brain to below injurious levels (Lloyd, 2012). Unsurprisingly, slamming an infant into a drywall-covered wall has the potential to cause significant injury, especially if the impact is to an area directly supported by underlying wooden framework. We observed that impacts severe enough to generate forces with high-serious injury risk also resulted in visible damage to the wall structure (Lloyd and Lee, 2013). Finally, we observed that the forces generated when the dummy was from a height of 1.0 m or more were consistent with those required to cause serious head injury (Lloyd, 2013).

DISCUSSION

The current legal landscape regarding abusive head trauma cases is mired in controversy. Prosecutors and medical personnel tasked with the investigation of child abuse and protection of the most vulnerable members of society often reject a history of injury resulting from a short fall given by a caretaker as incredible, based on the somewhat counterintuitive belief that a short fall is incapable of producing the same injury as shaking, and that the injury risk of the latter is only comparable to a fall from more than 10 or 20 feet. While the protection of children by the legal system is of paramount importance, this must be balanced with the protection of innocent parties from the personal devastation that a wrongful prosecution can bring. The reliance by prosecutors on speculative and frankly unscientific speculation regarding the biomechanical aspects of pediatric head trauma by some experts has, in the experience of the authors, resulted in the overzealous prosecution of innocent parents and caretakers (Cowley et al., 2013). As demonstrated in Fig. 9.14, pediatric head injuries can have a variety of presentations, causes, and outcomes. The present chapter has focused only on the most difficult of these cases, in which the only evidence that an injury resulted from intentional trauma is the expert assertion that the injury history provided by the caretaker was not a plausible cause of the observed injuries, based on the assertion that the injury is "impossible" absent intentional abuse. (An example of such a case is presented in Chapter 15, *Criminal Investigation* in this book.) An adequate defense to a prosecution based on such assertions requires a multidisciplinary team of experts knowledgeable in the biomechanics and epidemiology of not only abusive head trauma but also short falls and other unintentional head injury mechanism, as well as the medical aspects and differentiating characteristics of both abusive and unintentional head trauma.

References

Arnholz, D., Hymel, K.P., Hay, T.C., Jenny, C., 1998. Bilateral pediatric skull fractures: accident or abuse? Journal of Trauma Injury Infection and Critical Care 45 (1), 172–173.

Barzó, P., Marmarou, A., Fatouros, P., Hayasaki, K., Corwin, F., 1997. Contribution of vasogenic and cellular edema to traumatic brain swelling measured by diffusion-weighted imaging. Journal of Neurosurgery 87 (6), 900–907.

Centers for Disease Control, Division of Injury Control, 1990. Childhood injuries in the United States. American Journal of Diseases of Children 144 (6), 627.

Center for Disease Control, 2010. Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths, 2002–2006.

Chadwick, D.L., Chin, S., Salerno, C., Landsverk, J., Kitchen, L., 1991. Deaths from falls in children: how far is fatal? Journal of Trauma 31 (10), 1353–1355.

Chadwick, D.L., Bertocci, G., Castillo, E., Frasier, L., Guenther, E., Hansen, K., Herman, B., Krous, H.F., 2008. Annual risk of death resulting from short falls among young children: less than 1 in 1 million. Pediatrics 121 (6), 1213–1224.

Cheung, D.S., Kharasch, M., 1999. Evaluation of the patient with closed head trauma: an evidence based approach. Emergency Medicine Clinics of North America 17 (1), 9.

Coats, B., Margulies, S.S., 2008. Potential for head injuries in infants from low-height falls. Journal of Neurosurgery: Pediatrics 2 (11), 321–330.

Collins, J.D., October 01, 2013. A Month-Old Infant Misdiagnosed with Child Abuse. http://www.hcplive.com/publications/family-practice-recertification/2013/October2013/Misdiagnosis-of-Rib-and-Skull-Fracture-in-a-Month-Old-Infant.

Cowley, L., Tempest, V., Maguire, S., Mann, M., Naughton, A., Wain, L., Kemp, A., 2013. Implementing scientific evidence to improve the quality of child protection. BMJ Quality Improvement Reports u201101.w703.

Depreitere, B., Van Lierde, C., Sloten, J.V., Van Audekercke, R., Van der Perre, G., Plets, C., Goffin, J., 2006. Mechanics of acute subdural hematoma resulting from bridging vein rupture. Journal of Neurosurgery 104 (6), 950–956.

Ersahin, Y., Mutluer, S., Mirzai, H., Palali, I., 1996. Pediatric depressed skull fractures: analysis of 530 cases. Child's Nervous System 12 (6), 323–331.

Gennarelli, T.A., Ommaya, A.K., Thibault, L.E., 1971. Comparison of translational and rotational head motions in experimental cerebral concussion. In: Stapp Car Crash Conference. Fifteenth Proceedings, SAE Paper No. P-39, pp. 797–803.

Gennarelli, T.A., Adams, J.H., Graham, D.I., 1981. Acceleration induced head injury in the monkey I: the model, its mechanistic and physiological correlates. Acta Neuropathologica Supplementum 7, 23–25.

Gennarelli, T.A., Thibault, L.E., 1982. Biomechanics of acute subdural hematoma. Journal of Trauma 22 (8), 680–686.

Gennarelli, T.A., Thibault, L.E., Adams, J.H., Graham, D.I., Thompson, C.J., Marcincin, R.P., 1982. Diffuse axonal injury and traumatic coma in the primate. Annals of Neurology 12 (6), 564–574.

Hacke, W., Schwab, S., Horn, M., Spranger, M., De Georgia, M., von Kummer, R., 1996. 'Malignant' middle cerebral artery territory infarction: clinical course and prognostic signs. Archives of Neurology 53 (4), 309–315.

Hall, J.R., Rayes, H.M., Horvat, M., Meller, J.L., Stein, R., 1989. The mortality of childhood falls. Journal of Trauma 29, 1273–1275.

Helfer, R.E., Slovis, T.L., Black, M.B., 1977. Injuries resulting when small children fall out of bed. Pediatrics 60, 533–535.

Hobbs, C.J., 1984. Skull fracture and the diagnosis of abuse. Archives of Disease in Childhood 59, 246–252.

Hodgson, V., Gurdgian, E., Thomas, L., 1967. Development of a model for the study of head injury. In: Proc. 11th Stapp Car Crash Conf., pp. 432–443.

Holbourn, A.H.S., 1943. Mechanics of head injuries. Lancet 242 (6267), 438–441.

Holsti, M., Kadish, H.A., Sill, B.L., Firth, S.D., Nelson, D.S., 2005. Pediatric closed head injuries treated in an observation unit. Pediatric Emergency Care 21 (10), 639–644.

Ibrahim, N.G., Marguiles, S.S., 2010. Biomechanics of the toddler head during low-height falls: an anthropomorphic dummy analysis. Journal of Neurosurgery: Pediatrics 6, 57–68.

Joffe, M., Ludwig, S., 1988. Stairway injuries in children. Pediatrics 82 (3 Pt 2), 457–460.

Kelly, P., Hayes, I., 2004. Infantile subdural haematoma in Auckland, New Zealand: 1988–1998. New Zealand Medical Journal 117 (1201).

Klatzo, I., 1987. Pathophysiological aspects of brain edema. Acta Neuropathologica 72 (3), 236–239.

Kraus, J.F., Fife, D., Cox, P., Ramstein, K., Conroy, C., 1986. Incidence, severity, and external causes of pediatric brain injury. American Journal of Diseases of Children 140 (7), 687.

Kraus, J.F., Rock, A., Hemyari, P., 1990. Brain injuries among infants, children, adolescents, and young adults. American Journal of Diseases of Children 144 (6), 684.

Kurinsky, R.M., Rochette, L.M., Smith, G.A., 2013. Pediatric injuries associated with high chairs and chairs in the United States, 2003–2010. Clinical Pediatrics 20 (10), 1–8.

Lallier, M., Bouchard, S., St-Vil, D., Dupont, J., Tucci, M., 1999. Falls from heights among children: a retrospective review. Journal of Pediatric Surgery 34 (7), 1060–1063.

Langlois, J.A., Rutland-Brown, W., Wald, M.M., 2006. The epidemiology and impact of traumatic brain injury – a brief overview. Journal of Head Trauma Rehabilitation 21 (5), 375–378.

Lantz, P.E., Couture, D.E., 2011. Fatal acute intracranial injury, subdural hematoma, and retinal hemorrhages caused by a stairway fall. Journal of Forensic Science 56 (65), 1648–1653.
Lee, M.C., Haut, R.C., 1989. Insensitivity of tensile failure properties of human bridging veins to strain rate: implications in biomechanics of subdural hematoma. Journal of Biomechanics 22 (6–7), 537–542.
Leventhal, J.M., Martin, K.D., Asnes, A.G., 2008. Incidence of fractures attributable to abuse in young hospitalized children: results from analysis of a United States database. Pediatrics 122 (3), 599–604.
Lloyd, J.D., Willey, E.N., Galaznik, J.G., Lee III, W.E., Luttner, S.E., 2011. Biomechanical evaluation of head kinematics during infant shaking versus pediatric activities of daily living. Journal of Forensic Biomechanics 2, 1–9.
Lloyd, J.D., August 3, 2012. Biomechanics of pediatric brain injury. In: Presented at Evidence Based Medicine and Social Investigation Conference, Vancouver, Canada.
Lloyd, J., September 18–21, 2013. Biomechanics of brain injuries associated with short falls in children. In: Presented at 11th Annual Conference on Brain Injury, New Orleans, LA.
Lloyd, J.D., Lee, W.E., June 27–28, 2013. Biomechanics of inflicted head trauma. In: Presented at 4th Penn State Hershey International Conference on Pediatric Abusive Head Trauma, Burlington, VT.
Lowenhielm, P., 1974. Dynamic properties of the parasagittal bridging veins. Zeitschrift fur Rechtsmedizin 74 (1), 55–62.
Lowenheilm, P., 1975. Mathematical simulation of gliding contusions. Journal of Biomechanics 8 (6), 351–356.
Lowenhielm, P., 1978. Tolerance level for bridging vein disruption calculated with a mathematical model. Journal of Bioengineering 2 (6), 501–507.
Matshes, E., 2010. Presented at the American Academy of Forensic Sciences Conference in Chicago, IL on February 24th 2010 that Retinal and Optic Nerve Sheath Hemorrhages (ONSH) Are Not Pathognomonic of Abuse Head Injury.
Malignant cerebral edema. (n.d.). In Wiki.CNS. Retrieved August 4, 2014, from: http://wiki.cns.org/wiki/index.php/Malignant_Cerebral_Edema.
Meaney, D.F., 1991. Biomechanics of Acute Subdural Hematoma in the Subhuman Primate and Man (Ph.D. dissertation #AAI9125717). University of Pennsylvania.
Ommaya, A.K., Gennarelli, T.A., 1974. Cerebral concussion and traumatic unconsciousness. Brain 97 (1), 633–654.
Ommaya, A.K., 1984. Biomechanics of head injury: experimental aspects. In: Nahum, A.M., Melvin, J. (Eds.), The Biomechanics of Trauma. Prentice-Hall, pp. 245–270.
Ozanne-Smith, J., Heffernan, C.J., 1990. Child injuries associated with nursery furniture. Monash University Accident Research Center, Report No. 12.
Plunkett, H., 2001. Fatal pediatric head injuries caused by short-distance falls. American Journal of Forensic Medicine & Pathology 22 (1), 1–12.
Powell, E.C., Jovtis, E., Tanz, R.R., 2002a. Incidence and description of stroller-related injuries to children. Pediatrics 110 (5), e62.
Powell, E.C., Jovtis, E., Tanz, R.R., 2002b. Incidence and description of high chair-related injuries to children. Ambulatory Pediatrics 2 (4), 276–278.
Prange, M.T., Coats, B., Duhaime, A.C., Margulies, S.S., 2003. Anthropomorphic simulations of falls, shakes, and inflicted impacts in infants. Journal of Neurosurgery 99 (1), 143–150.
Prange, M.T., Luck, J.F., Dibb, A., Van Ee, C.A., Nightingale, R.W., Myers, B.S., 2004. Mechanical properties and anthropometry of the human infant head. Stapp Car Crash Journal 48, 279–299.
Quayle, K.S., Jaffe, D.M., Kupperman, N., Kaufman, B.A., Lee, B.C., Parks, T.S., McAlister, W.H., 1997. Diagnostic testing for acute head injury in children: when are head computed tomography and skull radiographs indicated? Pediatrics 99 (5), E11.
Reece, R.M., Sege, R., 2000. Childhood head injuries: accidental or inflicted? Archives of Pediatrics and Adolescent Medicine 154 (1), 11.
Reichelderfer, T.E., Overbach, A., Greensher, J., 1979. Unsafe playgrounds. Pediatrics 64 (6), 962–964.
Rieber, G.D., 1993. Fatal falls in childhood. How far must children fall to sustain fatal head injury? Report of cases and review of the literature. American Journal of Forensic Medicine and Pathology 14 (3), 201–207.
Rorke-Adams, L., Duhaime, C.A., Jenny, C., Smith, W.L., 2008. Head trauma. In: Reece, R.M., Christian, C.W. (Eds.), Child Abuse: Medical Diagnosis & Management, third ed. American Academy of Pediatrics.
Rosenberg, G., 1999. Ischemic brain edema. Progress in Cardiovascular Diseases 42 (3), 209–216.

Schalamon, J., Ainoedhofer, H., Saxena, A.K., Petnehazy, T., Singer, G., Hollwarth, M.W., 2006. Falls from high chairs. European Journal of Pediatrics 165, 732–733.

Schulz, B.W., Lee III, W.E., Lloyd, J.D., 2008. Estimation, simulation, and experimentation of a fall from bed. Journal of Rehabilitation Research and Development 45 (8), 1227–1236.

Shane, S.A., Fuchs, S.M., 1997. Skull fractures in infants and predictors of associated intracranial injury. Pediatric Emergency Care 13 (3), 198.

Shi, J.X., et al., 2009. Costs, mortality likelihood and outcomes of hospitalized US children with traumatic brain injuries. Brain Injury 23 (7–8), 602–611.

Soreide, I., Krüger, A.J., Ellingsen, C.L., Tjosevik, K.E., January 2009. Pediatric trauma deaths are predominated by severe head injuries during spring and summer. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 17, 3.

Thibault, L.E., Gennarelli, T.A., 1985. Biomechanics of diffuse brain injuries. In: Stapp Car Crash Conference. Twenty-nine Proceedings, SAE Paper No. 856022, New York.

Thomas, L., Sezgin, Y., Hodgson, V., Cheng, L., Gurdjian, E., 1968. Static deformation and volume changes in the human skull. In: Proc. 12th Stapp Car Crash Conf., pp. 260–270.

Thompson, A.K., Bertocci, G., Rice, W., Pierce, M.C., 2011. Pediatric short-distance household falls: biomechanics and associated injury severity. Accident Analysis and Prevention 43, 143–150.

Van Ee, C.A., Moroski-Browne, B., Raymond, D., Thibault, K., Hardy, W., Plunkett, J., 2009. Evaluation and refinement of the CRABI-6 anthropomorphic test device injury criteria for skull fracture. In: Proceedings ASME International Mechanical Engineering Congress & Exposition, IMECE2009–12973.

Warrington, S.A., Wright, C.M., 2001. Accidents and resulting injuries in premobile infants: data from the ALSPAC study. Archives of Disease in Childhood 85, 104–107.

Watson, W., Routley, V., Ozanne-Smith, J., 1998. Nursery furniture injuries. Hazard. Edition no. 37.

Weber, W., 1984. Experimental study of skull fractures in infants. Zeitschrift fur Rechtsmedizin 92, 87–94.

Weber, W., 1985. Zur biomechanischen Fragilitat des Sauglingsschadels. Zeitschrift fur Rechtsmedizin 94, 93–101.

Wheeler, D.S., Shope, T.R., 1997. Depressed skull fracture in a 7-month-old who fell from bed. Pediatrics 100, 1033–1034.

Williams, R.A., 1991. Injuries in infants and small children resulting from witnessed and corroborated free falls. Journal of Trauma 31 (10), 1350–1352.

This page intentionally left blank

CHAPTER

10

Survival Analysis

H.D. Tolley, J.M. Barnes

Brigham Young University, Provo, UT, United States

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

OUTLINE

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00010-0

INTRODUCTION

Survival analysis is one of the primary statistical methods for analyzing data on time to an event such as death, heart attack, device failure, etc. Such data analysis is essential for many facets of legal proceedings including apportioning cost of future medical care, estimating years of life lost, evaluating product reliability, assessing drug safety, measuring viability of medical therapies and devices, assessing actuarial loss, etc. This branch of empirical science entails gathering and analyzing data on time until failure or death. Survival analysis includes a variety of specific type of data analysis including "life table analysis," "time to failure" methods, and "time to death" analysis. Reliability methods and life contingencies are based on the same fundamental principles of survival analysis.

Use of survival analysis in legal proceedings, either in court or for settlement negotiations, often entails determination of loss or damage. Common questions in such proceedings are "how long will the patient live?" or "how much has the victim's life been shortened?" An exact answer is not possible to either question, and only an answer as a statistical probability can be obtained. A statistical answer entails an average or "expected value" with an associated level of uncertainty. This level of uncertainty is often illustrated by the use of a confidence interval.

This chapter is intended to serve two purposes. First is a description and illustration of the assumptions and basic methods of survival analysis. Second is to present a statistical model of survival analysis, which includes the inherent uncertainty of the estimate, for use in legal proceedings. Survival analysis is a mature scientific discipline with a variety of statistical methods and associated computer programs available to the analyst. The type of data available, the manner the data were obtained, the mathematical models used to analyze the data, and the integrity of the conclusions can be very confusing for someone not steeped in the latest developments. The first purpose of this chapter is to present the basic assumptions and definitions of survival analysis and illustrate how these assumptions may alter the credibility of data use in forensics. The details of statistical analysis and the use of computer programs are covered in a variety of papers and text books.[11] Here we restrict our attention to the fundamentals that underly all of these procedures, fundamentals that are essential for the forensic epidemiologist to understand in order to explain the results to lay fact-finders.

Legal proceedings revolve around the premise that the process will be to discover the facts and align them with legal statute and precedent. To include scientific evidence into the legal process is not a trivial process. This difficulty is exacerbated when the scientific data are not absolute facts but are *probable* results with a level of uncertainty. We show how the legal concept of "more probable than not" can be used as a tool to reliably describe survival analysis results in a legal setting.

We will illustrate the fundamentals by focusing on one statistic commonly used, that is the median life expectancy (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*). We selected the median for illustration for two reasons. First, many other statistical measures can be obtained from survival analyses including 5-year survival probability, excess mortality, increased risk, mean residual lifetime, present value of a life annuity, present value of an insurance at death, and so forth. Regardless of which of these or other statistical measures one uses, the issues regarding the assumptions are the same as one

encounters with use of the median lifetime. The median is easily understood as the point in time where half of the population is still alive, and thus commonly used. Second, the median lifetime and its confidence interval play a key role in evaluating what is "more probable than not" for the future survival of an individual.

DEFINITIONS

The following word definitions are of concepts commonly defined in the survival literature using probability formulas and/or statistical jargon. These are the basics for survival analysis. The more technical statistical definitions are in the Appendix of this chapter. Knowledge of these definitions is important to be able to evaluate or challenge a survival analysis result.

1. *Time to failure.* This is the duration of time from an initial event to the event of interest. Usually this is the time from birth or from an accident until death of an individual. Other types of events can be used to measure survival time such as failure of a medical device, cardiovascular event, failure of an electronic component, and so forth. The "beginning" is, itself, an event date, usually the time of an accident or the point in time of a surgery or beginning of a medical treatment. It is not simply the time that an individual enters the study. For example, if a person is in an automobile accident but is not formally followed until a year after the accident, then the "beginning" is the date of the accident and not the date that the individual enters a follow-up study.
2. *Censoring.* The process by which complete survival times are not observed. Censoring means that the observation interval is not complete. This can happen because the individual was not in the study from the beginning, but say joined after a year or so (left censored) or the individual was not followed until death because the individual was either lost to follow-up, left the study, or the study was terminated (right censoring). There are two common types of right censoring according to the manner in which the study is terminated:
   a. *Type I censoring.* In this case the cohort of individuals is followed until a fixed number of individuals are observed to die. The length of time that the cohort is followed until the fixed number of deaths has occurred is random.
   b. *Type II censoring.* In this case the cohort of individuals is followed for a fixed amount of time. The actual number of individuals that are observed to die is random.

   In most survival studies, the Type II censoring is the most common. The reason is that a cohort of individuals, such as the cohort of spinal injury patients, is continuously followed over a time. The censoring arises because an analysis of the time to death must be made at a fixed point in time. Patients whose date of death exceeds that time are censored. For example, in a court case, experts need to determine the median lifetime, with the data to date, and cannot wait until all individuals in the cohort have died. Usually, in registration type data sets, such as the spinal injury patients, new patients are continually being added to the cohort.

3. *Survival function.* The survival function is a formula, a graph, or a table from which one can determine the probability of surviving to any age. Although much of the technical literature uses mathematical formulas for the survival function, actuaries have traditionally used tables that give the probability of surviving a year given the individual's age and gender. Graphs are often used which plot the percentage of the population alive after a fixed time period. Fig. 10.1 is a plot of three estimated survival curves for the time (years) from injury date to death. To determine the probability of survival for any number of years, locate that value on the x-axis. The y-axis value of the survival curve at that x-axis value is the probability of surviving the specified number of years.
4. *Hazard function.* The hazard function (sometimes called the force of mortality) is the instantaneous rate of failure or death at any particular time or age, given that the individual or device is alive (has not failed) at that time. This is a more technical component of survival analysis but is important in assessing the impact of various risk factors (see below).
5. *Risk factors.* These are characteristics, including blood pressure, smoking status, lifestyle, genetics, environment, etc., that may increase the likelihood or probability of death. For example, smoking is known to reduce lifetime or increase the likelihood of death. Smoking is thus a risk factor for survival. Other risk factors include diabetes, sedentary lifestyle, exposure to polluted water and air, etc. It is important to not ignore factors that increase survival probabilities, however. These include lifestyle, social situation (ie, married), and others. Further, depending on the age of the individual, it may be reasonable to draw inferences about their mortality risk based on their personal characteristics relative to those who die in their age and gender cohort.

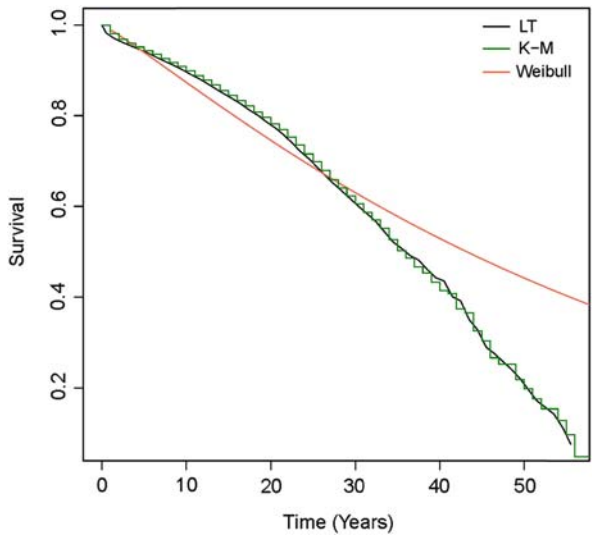


FIGURE 10.1 Estimate of survival using the three methods. *LT* is the life table method, *K-M* is the Kaplan–Meier method.

For example, if approximately half of the deaths at age 75 result from cancer, and an individual is known to be cancer free, there is a large proportion of the cancer deaths in the population that are due to long-term disease that can be eliminated from the risk applicable to the individual.

6. *Covariates.* These are characteristics of individuals that are measured in conjunction with lifetime that may influence time to death. For example, age, gender, level of injury, level of smoking, etc., would be covariates that one records on each individual as well as time to failure or death. Note that presence or level of risk factors can be covariates. The purpose of measuring and including covariates in a survival analysis is to "adjust for" any effects of the covariate variables on survival or death.
7. *Median lifetime.* This is the time point at which 50% of the cohort of individuals (or devices) will have failed. Median survival time is often used in survival analyses rather than the mean or average survival time because survival data are usually skewed to the right and use of the mean is considered to be less representative of the center of the survival distribution than the median. The median lifetime can be estimated in many cases even though less than half of the patients have died.
8. *Confidence interval* is an interval estimate of the statistic of interest, for example the median, based on the observed survival data. The interval is estimated in such a manner that over repeated samples of survival data, with a confidence interval calculated for each sample, the probability that the interval includes the true population parameter (eg, the median lifetime) is at least a fixed given value, such as 95%.[1]
9. *Bias.* The amount that the average value of the estimate, say the estimated median, over repeated samples of survival data differs from the true value, the population median lifetime.
10. *Parametric model.* A parametric model is a mathematical model of known form that can be used to calculate the probability of death in any time interval. The form of the survival function or the distribution of survival times is determined by this known model. Only the parameters of the survival function are unknown and estimated from the data. The most common forms assumed or the distributions of survival times are the Exponential, Weibull, Gompertz, and Gamma.
11. *Nonparametric model.* A nonparametric model does not specify the probability of death in any interval to be of a known mathematical form. Nonparametric models usually specify weak assumptions in order to estimate survival curves. The advantage of nonparametric models is that they require less from the analyst. On the other hand, since they require fewer or weaker assumptions, they are not as efficient in their use of the data as parametric models are, *if the parametric assumptions are correct.*

USING SURVIVAL ANALYSIS IN A FORENSIC SETTING

The "More Probable Than Not" Criterion

In a forensic or judicial setting where the harmful effect of an exposure, such as a faulty pharmaceutical agent or device is being assessed, inter alia, it is important to present evidence that an individual's harm was caused by the exposure to the agent or use of the device, on a

"more probable than not" basis (>50% probable). Of course, the majority of this book is devoted to how individual causation is assessed using population-based information.

This same goal of ascertaining what is "more probable than not" can be used to assess future damage or loss resulting from an accident or injury. For example, determination of damage and loss often entails predicting the number of years of medical care to be incurred by the injured party or the number of years of life lost by the injury. Both of these examples involve the estimation of future lifetime. When evidence is in the form of a life table or survival data, the median lifetime plays an important role as an estimate of future lifetime, but not to the extent that it can be concluded on a "more probable than not" that the individual will survive to the time indicated. What the median value indicates is that more than 50% of the time the individual will survive *at least* to the value, but this is not the same thing as opining on what happens in more than 50% of cases like the individual.

To illustrate this concept, we can choose any duration of time that is shorter than the median. We can conclude that it is more probable than not that the individual will have a survival time or time to death that exceeds the specified duration. The same thing is true for a duration that is longer than the median, even by a matter of months or even days. But, herein lies a difficulty. It is not "more probable than not" that the person will die exactly at the median survival. In fact, in almost all cases, they will live longer or not as long as the median. In assessing the probable survival of an individual, the median should be seen for what it is, which is a measure of the middle of the lifetime length of a population. Using a median or average makes perfect sense when one is assessing the future needs of a large group of people when there is a limited pool of funds, as the use of a measure of central tendency of a population means that for every person who lives more than the average and requires more resources, there is another person who lives less than the average and thus requires fewer resources than the median person. Thus, there is sufficient symmetry that the future needs of the entire group, and thus all of the individuals in the group, can be planned for, and if the assessment of life expectancy is accurate, there is minimal risk that there will be insufficient resources to take care of the needs of every individual in the population.

This approach to life expectancy for a population was never designed to be an accurate method for prediction of the survival of an individual. Using the estimated *median* or *average* value for survival as a basis for a maximum estimate of how long an individual would have lived is insufficiently reliable as a methodology to meet the minimum standard for the provision of expert medicolegal testimony, as such an estimate is uncertain, varying from one sample to another. In fact, using the median as a maximum survival projection will be wrong 50% of the time, as, by definition, 50% of the time the individual will outlive the estimate. Thus using a population-based methodology that *only* provides a median or average value does not provide a life expectancy prediction that is "more likely than not" correct for an individual, as it only covers the range of 0–50% of the survival probabilities possibly applicable to the individual (assuming a reliable and valid basis for the projection). Although the median value (50th percentile) is often used for survival projections, it is misleading to present a precise population-based value like a median or an average when assessing the exact age of death for any individual. It is not any more likely that he or she will survive to the same age as the 50th percentile of the comparison population than it is that he or she would survive to, say, the 53rd or 47th percentile.

In order to be applicable to an individual, a survival opinion meant to apply to an individual in a forensic setting must cover >50% of the probabilities of what will occur in the future in order to be "more probable than not". In actuality, the range of survival values that can be considered "more likely than not" is the middle >50% of the bell curve, ranging from slightly less than 25% to slightly more than 75%. In order to meet the standard of what is "more probable than not" without favoring one side or the other, the >50% probability must be equally distributed both above and below the 50% threshold. Thus, slightly more than the 75th percentile is the maximum age the individual will survive on a more probable than not basis, and slightly less than the 25th percentile is the soonest the individual will die on a more probable than not basis (we will just refer to the 25th and 75th percentile from here on, with the caveat that slightly more and slightly less than these values, respectively, is necessary to meet the *more* probable than not standards). It is critical that the expert providing testimony on a survival projection inform the fact-finder that the median survival, while in the range of what is more probable than not, will underestimate the individual's survival in one out of two cases in which it is used. In comparison, the 75th percentile projection, serves as the upper boundary of the range of most probable survival, and will only underestimate individual survival in one out of four cases. Thus, although the remainder of this chapter largely discusses methods for arriving at a median survival value and the associated error rate, it is a sound and appropriate practice to present and explain both the median and 75th percentile survival projection in a forensic setting. The more fully informed fact-finder can make the determination regarding the acceptable amount of error in the estimate that is used in making a decision on damages. Note that we have not discussed the lower bound of the range of what is more probable than not, the 25th percentile. This is because the most typical application of the methods described in this chapter is to project how long an individual might live or might have lived, *not how soon the individual would be expected to die*. For a survival project, the 25th percentile project will underestimate an individual's life span in three out of four cases, and thus has no utility for presentation in a forensic setting.

Developing a Threshold for a Survival Projection Using the Median

What is the median lifetime? No exact answer is possible from life table or survival data. Only a statistical estimate of the median lifetime with a specified level of uncertainty can be obtained. Put in other words, we cannot make the statement that "it is more probable than not" the patient will live to a specific value unless we quantify the inherent uncertainty in the estimate.

If we use an estimation of median survival probability as our sole survival estimate (thus ignoring the 50th to 75th percentile range practice described above), we need to include the level of confidence in arriving at the estimate in order to provide the fact-finder with the upper bound of the estimate given the uncertainty inherent at arriving at the value. We propose here to use the "value at risk" paradigm used in managing financial risk.[2] "Value at risk" is a measure of a random loss at some level of confidence. To apply this paradigm to determine a "more probable than not" threshold, we adjust our value using the confidence interval bounds for the estimated median lifetime. Here, again, the type of loss dictates which bounds we use. If we use the median lifetime as the "more probable than not" threshold at a

specified risk of underestimation (per the above discussion), as we only have an *estimate* of the median lifetime, we must hedge this threshold value. We do this by replacing the estimated median lifetime by the upper confidence bound of the median lifetime.

For example, suppose an injured patient has an estimated median survival of 20 years with a 95% confidence interval of between 16.5 and 24.5 years. Assume also that the issue is to determine the loss to be incurred by the patient for medical care and excess living expenses. In this case we use 24.5 years rather than 20 years as the "more probable than not" threshold associated with the median. We are then 95% confident that the actual losses calculated assuming a future lifetime of 24.5 years will, more probably than not, be sufficient for the patient (ignoring the inherent error rate of projected survival of the median per the prior discussion). If, on the other hand, we use the estimated median lifetime of 20 years, we are only about 50% confident that "more probably than not" the funds will be sufficient. That is, there is a 50% "chance"[3] that the funds will be insufficient to cover the patient's years of medical care.

In the alternative, if the issue is to determine the amount of life lost,[4] we would use the value of 16.5 years as the upper bound in calculating years to live. If we assume that the median lifetime of an uninjured person is 30 years, we subtract 16.5 years from 30 years to get 13.5 years of life lost. In this case, we are 95% confident that using 13.5 years as the number of years of life lost is "more probable than not" a sufficient estimate of years of life lost.

SURVIVAL FOLLOWING A SPINAL CORD INJURY: AN EXAMPLE

Population and Data Cohort

To illustrate the definitions above, and to underscore some of the issues in survival analysis methods and their effects on the use of the more probable than not standard, we will provide the following example of projected survival for non-Hispanic males who sustained a spinal cord injury between the ages of 20 and 35 for whom we have follow-up data. The data also contain other covariate information such as race, etiology of spinal cord injury, neurological variables, Frankel score (spinal cord injury severity), and spinal cord level of injury, inter alia, all of which we ignore for the present example, for simplification. The analysis is based on a data set of >45,000 spinal cord injury patients in the US who have been followed over time.

The duration of interest in the analysis is the time between spinal cord injury and death. Thus "age" in this example is replaced by the time since injury. In effect, we are monitoring the "age" of the injury for a cohort of patients between 20 and 35 years old, not the age of the patient at injury or the current age of the patient. The data set, when isolated to the characteristics of interest, consists of 14,353 individuals. Of these, 22.2% have been reported as dead, and the rest were alive at the time the data set was constructed. Individuals that are still alive at the time the data are collected are censored, in that they did not die but are "lost to follow-up" relative to the data file. This censoring is right censoring since the patients were followed-up to a certain point and then lost.

SURVIVAL MODELS

There are many different methods used to model survival data. Here will illustrate the life table method, the Kaplan–Meier nonparametric model and the Weibull parametric model.

Three Models

The Life Table Method

The first method is the life table method, traditionally used by actuaries in determining life insurance rates, annuity premiums, requisite reserves, and so forth. Life tables date back many centuries and are a simple method of representing the mortality experience of a cohort of individuals. Although the life table (sometimes referred to as a "mortality table") has a long history in actuarial methodology, it is still often used today to summarize mortality.[5]

The very simple life table consists of a set of rows, each row representing a discrete period of time, say a year, and columns for different mortality status for the time period. Table 10.1 is a typical example of a simple life table constructed from the spinal injury data. Here the first column represents the age of the injury (number of years since the injury) at the beginning of the time period. The second column represents the number of individuals at the particular age who are alive at the beginning of the time period. Here the third column represents the number who died during the interval. The next column represents the number who left the cohort during the year for reasons other than death. In this case, it is the number of males with a spinal cord injury of the specified age who are in the data set at the time the data set was constructed.[6] Another column *could be* the number of individuals who enter the cohort at that age during the time period. We do not have such a column here as everyone

TABLE 10.1 Life Table of Raw Counts From Spinal Cord Injury Data File

Age	Number living	Number dead	Number censored	Probability of death
0	14,353	2	0	0.0001
1	14,351	248	1855	0.0173
2	12,248	155	758	0.0127
3	11,335	102	217	0.0090
4	11,016	89	128	0.0081
5	10,799	90	223	0.0083
6	10,486	87	396	0.0083
7	10,003	99	265	0.0099
8	9639	92	245	0.0092

"Age" is the number of years since injury.

is followed from the time of their injury. The last column is the probability of dying during the 1-year period starting with the age of the injury.[7]

In this table, we can calculate the probabilities of surviving during any one year by taking the probability of death during the year and subtracting that from 1. That is either the individual dies during the year or survives. The probability of surviving 2 years is the product of the probability of surviving in the first year times the probability of surviving the second year. To illustrate, suppose we wish to calculate the probability of surviving from "age" 2 to "age" 4. This implicitly assumes that the patient survived the first 2 years of the spinal injury and we wish to determine the probability that the patient will survive the next 2 years. This probability is given as: $(1 - 0.0127) * (1 - 0.0090) = 0.9784$.

Similarly, we can calculate the time until half of the population has died using this same technique. In this case, we multiply one minus the probability of death for each time interval until the product drops below 50%. The time at which this occurs is approximately the median lifetime.[8]

Kaplan–Meier Method

The Kaplan–Meier method resembles a variant on the life table method. Recall that in the life table method the time axis is divided to many discrete time intervals, usually years. The number at the beginning of the year, the number dying in the year, and the number censored or lost to follow-up in the year are all tabulated. The Kaplan–Meier method also divides the time axis into many discrete intervals. However, in this case the intervals are not defined by a fixed length but by the occurrence of an event. The two events of interest here are death of a patient and when the patient is censored or lost to follow-up. The data file could be constructed where all of the patients were entered in order according to how long they lived after those who died or were lost to follow-up. Looking at Table 10.1, the first year would be divided into two intervals, according to the time of the two deaths. The second year would be divided into 2103 intervals corresponding to the times of death of the 248 patients who died and according to the censoring times that the 1855 patients who were censored.

The probability of death during any of these intervals is simply the number who died divided by the number of individuals who started the interval. The probability of surviving the interval is one minus the probability of dying in the interval. The median time to death is calculated as in the life table method by taking the product of the probabilities of surviving each of the intervals of time from the first until the product reaches 50%.

Parametric Methods

Both of the previous methods are nonparametric in that they assume no particular mathematical form for predicting the probability of death in a time interval or the number expected to be alive after a fixed time interval. Parametric models assume a fixed mathematical form or equation for calculating these two values. The mathematical equation has one or more parameters that give the corresponding survival curve a form that is similar to that of the data.

Although there are many possible parametric models, we will limit our illustrations here to the Weibull model.[9] A typical survival curve using the Weibull model is given in Fig. 10.2.

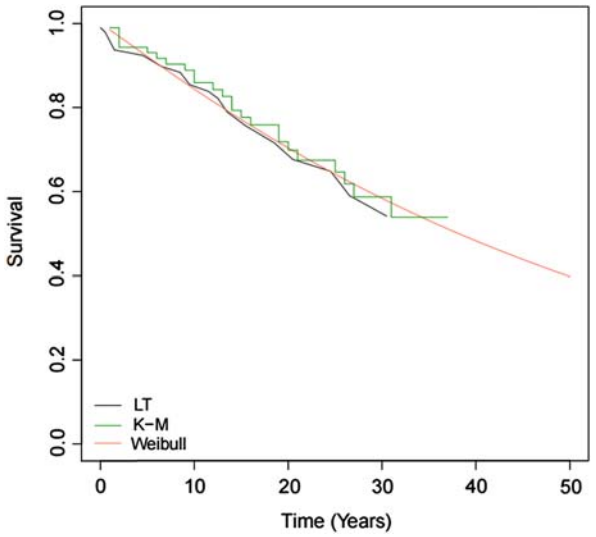


FIGURE 10.2 Estimated survival curve using a random sample of 100 patients. *LT* is the life table method, *K-M* is the Kaplan–Meier method.

Note that in this figure we have plotted two different survival curves generated by using two different set of parameters. Note that the curve can take different shapes.

Application to the Example

We now use the three methods to estimate a survival curve for spine injury patients using the data described above. Computer codes for fitting these data are in the programming language called "R" and are described in the Appendix. The three survival curves are plotted in Fig. 10.1. The three graphs give the proportion of the cohort of spinal injuries alive of the 14,353 patients who are followed.

Note that the life table method and the Kaplan–Meier method produce very similar survival curves. After about 30 years the Weibull survival curve, however, is very different. The reason for this is that the Kaplan–Meier method and the life table method both use the data available for each year to estimate the probability of death during the year. If there are only 50 people who have injuries that are 30 years old, for example, then the probability of death for injuries between anniversary date 30 and anniversary date 31 is based solely on the experience of these 50 patients. The reason for this is that the mortality during the interval is not "parameterized" as a function of some overall set of variables that can be estimated from all of the data.

The Weibull model, on the other hand, uses the data from *all* of the injury age intervals to estimate the two parameters that describe the mortality rates for all of the intervals. Since there are relatively fewer patients (about 2000) with injuries over 30 years compared to about 10,000 with injuries less than 30 years, the Weibull model tries to determine two parameters

that describe the mortality experience of the 10,000 observations *and* the mortality experience of the 2000, simultaneously. In this case, the mortality experience of those with injuries less than 30 years old dominates the model fitting process.

Which model is best? Although the survival curve estimated by the life table method and the Kaplan–Meier method represent tabulated deaths, the number of individuals is small. Consequently, these tabulations may be unstable. In other words, if we had a data file with the same underlying mortality process, the actual observed counts of deaths may be different than observed in the current case. Additionally, if the estimated Weibull mortality pattern seen for injuries under 30 years persists, then the Weibull-estimated survival curve for injuries that are 30 years old and older will be more stable. The reason for this is that for older injuries the estimates of the mortality curve are not based only on the data for injuries over 30 years but also on the deaths in patients with injuries under 30 years.

This example illustrates one of the major reasons for using a parametric model to estimate a survival curve. Parametric models estimate the survival curve by "borrowing" information about mortality in one time interval and applying that to mortality in other intervals in order to increase the stability of the model. On the other hand, both the life table method and the Kaplan–Meier method make no assumptions about previous patterns and basically just tabulate the sparse number of deaths observed. This is one of the fundamental differences in parametric methods as compared with nonparametric methods. The parametric methods "borrow" information where there is a lot of data to "fill in" mortality patterns where the data are sparse. Nonparametric methods do not "borrow" such information. When data are sparse in the region of interest and the mortality patterns can be assumed to persist, parametric methods are better. However, if the patterns of mortality do not persist, borrowing information using parametric models can result in biased estimates of the median lifetimes, as may be true in Table 10.2.

To illustrate this effect of borrowing information, we took a random sample of 100 individuals from the spinal injury data set and determined the survival curves in the same manner as in Fig. 10.1. However, in this case the number of patients with injuries that are older is greatly reduced. In fact, there are no recorded deaths in this sample for patients with injuries over 30 years. Fig. 10.2 gives these survival curves. As we see from this figure, neither the Kaplan–Meier nor the life table methods have sufficient data to estimate a survival curve beyond 30 years. This is to be expected since these methods use the number of deaths in the current interval to estimate mortality and survival.

TABLE 10.2 Estimated Median Lifetimes Using Each of the Three Methods for Modeling the Spinal Cord Injury Data

Method	Median lifetime	95% Lower CI bound	95% Upper CI bound
Life table	35.98	34.81	37.76
Kaplan–Meier	35.20	34.40	36.40
Weibull	43.23	41.59	44.84

The bounds for the 95% confidence intervals (CI) are also included.

The Weibull model, on the other hand, uses all the data in previous years to estimate an annual rate of death as a function of the age of the injury and then applies this rate to each year. Since there is no mortality observed in the sample for injuries over 30 years old, the survival curve in Fig. 10.2 is an extrapolation of what should be observed if there had been sufficient numbers of patients. However, as we noted from Fig. 10.1, the survival curve estimated by the Weibull model for durations after 30 years using the entire data set seems to be high. If this is true, the median lifetime will be overestimated. Unfortunately, there is no way to determine from the sample of 100 patients if the Weibull model estimate of survival and median lifetime is accurate, an overestimate or an underestimate without examining all the data.

MEDIAN SURVIVAL AND THE CONFIDENCE INTERVAL

Parametric and Nonparametric Estimates

The estimated median lifetimes using each of the three methods is given in Table 10.2. Note that the estimated median lifetime using the Weibull model is considerably longer than estimates using the other two models. This result is not surprising given the survival curves estimated by each of the three methods as displayed in Fig. 10.1.

In Table 10.3, we give the estimates of the median lifetime using the random sample of 100 observations taken from the spinal injury data file. As in Table 10.2, the median lifetime for the sample using the Weibull method is considerably longer than the median lifetime estimated using the life table model. Note that the Kaplan–Meier method could not even estimate a median lifetime for the data because of the heavy censoring. As noted above, there are no deaths for injuries older than 30 years in this sample.

SIMULATION STUDY

Simulation Setup

In the preceding subsection, we noted that the parametric model gave different results than either of the two nonparametric methods. We conclude from this that the parametric assumption can be a critical assumption, possibly producing results very different than a

TABLE 10.3 Estimated Median Lifetimes Using a Sample of 100 Patients

Method	Median lifetime	95% Lower CI bound	95% Upper CI bound
Life table	33.86	25.22	56.64
Kaplan–Meier	NA	25.70	NA
Weibull	38.14	24.67	56.90

Each of the three methods for modeling the spinal cord injury data was used. The bounds for the 95% confidence intervals (CI) are also included.

nonparametric assumption. Unfortunately, in the example above we do not know which of the estimates of median lifetimes given in Table 10.2 is closest to the true value. Here we examine how each of the methods of estimating median lifetime performs under known simulated conditions.

To simulate a cohort of individuals to evaluate the three methods of estimation, we will construct a hypothetical survival curve using a known density function. A density function gives the relative frequency of lifetimes in the population. Here we depict a density function as a graph. The height of the graph gives the relative likelihood of a lifetime equal to the value on the x-axis. Fig. 10.3 gives a plot of the Weibull density function. From this figure we see that the highest point of the graph is at the x-axis value, labeled time, here, of about 3. That means that most of the lifetimes of individuals will be between about 2 and 5 units of time. We simulate using this density function by drawing lifetimes, as if drawing from a hat, with the relative number of lifetimes of various lengths following the same curve as Fig. 10.3. When we do, most of the lifetime we randomly draw will be between 2 and 5 units long. We will draw very few over 10.

We randomly generate survival time values by drawing a sample of values from the specified density function. At the same time we also randomly select a time that the patient enters our study. If he/she enters late and the patient's lifetime exceeds the "end of the study" time, then the death of the patient is not observed. In this case the patient's information on lifetime is censored. These randomly selected values are considered to be real data, times to death, or time to censoring, for hypothetical patients. These data are then fit to a survival distribution using the Kaplan–Meier, life table, and Weibull models illustrated above. Since we know the "real" survival curve from which we generated the data, we can compare the estimated survival curve to determine how well the procedures work.

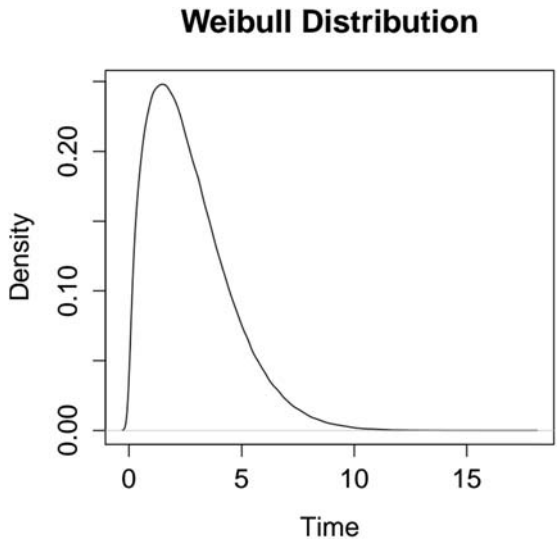


FIGURE 10.3 Weibull density.

In this case we use three different density functions to generate the survival times of the hypothetical patients. These are the Weibull, the gamma, and the bimodal distributions. The Weibull density is the one that the Weibull model assumes is true. The gamma density is close to the Weibull but different. It is not plotted here. The bimodal density functions is plotted in Fig. 10.4. The bimodal is clearly different than the either of the other two. It arises when the cohort of patients is made up of samples from two different populations, say, a frail population and a robust population.

For each simulation we get a hypothetical sample of patients with their lifetimes and whether or not they are censored. Each simulated sample contained lifetimes for either 100 patients or 1000 patients. We repeated the simulation 1000 times, each time recording the "real" lifetime, and the estimated lifetime, and the confidence interval estimated using the sample data.

Censoring

Here we will consider only Type II right censoring. That is, individuals will be lost to follow-up after being in the study some (possibly random) length of time, only. Type II censoring means that the censoring mechanism was based on time and not on obtaining a sufficient number of observed deaths. There are two types of time-based mechanisms. The first happens for a particular patient because of external random influences. For example, the person may move or be accidentally lost to the study. The second is the most common and is the type for the data considered here. This type of censoring occurs when the study is terminated. For example, in a follow-up study wherein each patient's mortality status is regularly monitored, and new patients are continually entering the study, if a researcher or analyst needs to obtain an estimate of the median lifetime using the data to date, then all

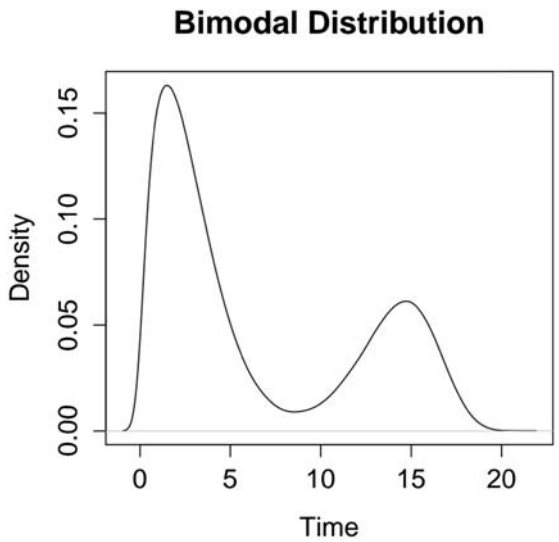


FIGURE 10.4 Bimodal Weibull density.

individuals in the follow-up cohort who have not died would be censored by "termination of the study." In the simulations below we consider only this second type of censoring.

Simulation Results

Since we know the theoretical median lifetime from the density function we use to simulate the hypothetical data, we know the true median lifetime generating the data. For each simulated data set we use all three of the estimation procedures to estimate the median lifetime and the 95% confidence interval.[10] We tabulate three statistical measures for each method of estimation of median lifetime for each of the 1000 simulations:

1. Coverage of the confidence interval. This answers the question regarding whether the confidence interval is really a 95% confidence interval or not. If the coverage is too far below 95%, it is an indication that the method does not estimate the median lifetime very well. It also indicates that there may be difficulty in determining the "more probable than not" threshold.
2. The width of the confidence interval. Wider confidence intervals are not using the data as efficiently as methods with narrow confidence intervals, assuming that both methods have the same coverage. Widths of confidence intervals here should be considered relatively. Since the amount of censoring and the number of observations directly impact the width of the confidence interval, the values here should be interpreted relatively. That is, are the widths for the different ways of estimating the median lifetime about the same or not?
3. Bias of the estimate. This is tabulated as a proportion of the median. If an estimate is biased, it overestimates or underestimates the median lifetime. The actual amount over or under can be calculated by adding the bias to 1 and multiplying the sum by the median. For example, if the median is 20 years and the bias is 0.05, then on average the estimate of the median is $20 * (1 + 0.05) = 21$. If there is a large bias one way or the other, it is an indication that there is a problem with the model.

Because of the close similarity of the life table method to the Kaplan–Meier method we expect similar results in the simulations. In simulations we noted that the coverage, width, and bias in the tabulations are nearly equal for these two methods. Consequently, we tabulate only the Kaplan–Meier method results in the following tables.

Table 10.4 summarizes the results for the case where there is no censoring. All patients were followed to death. In Table 10.4, the density we used was the Weibull density. The Weibull model of estimating median lifetime assumes, in fact, that the underlying density is Weibull. Note that the Kaplan–Meier and the Weibull model are similar in their output. We did not include a table for the gamma density as it was very similar to Table 10.4. Looking at the coverage of the intervals we see that the Kaplan–Meier is about the same as the Weibull. Put in other words, even though the assumptions of the Weibull model are fully met, the bias and coverage is about the same for this parametric and nonparametric method. Note, however, that the width of the confidence interval for the Weibull model is significantly narrower than the Kaplan–Meier estimate. The efficiency of the Weibull model obtained by a parametric model when all assumptions are met is represented by this narrower confidence interval.

TABLE 10.4 Confidence Intervals and Bias of Estimated Median Lifetimes From Simulated Data With No Censoring, Using the Weibull Method

Method	Count	Coverage	Width	Bias
Kaplan–Meier	100	0.954	0.475	0.004
Weibull	100	0.953	0.390	0.004
Kaplan–Meier	1000	0.946	0.472	0.001
Weibull	1000	0.956	0.388	−0.00003

The Kaplan–Meier and the Weibull methods for modeling the spinal cord injury data were used. Coverage is based on an assumed 95% confidence intervals.

When we use the bimodal density for lifetimes, the bias in the Weibull model increases dramatically, as much as 60% biased. Kaplan–Meier method has much less bias in these cases. In all cases the coverage of the confidence intervals, after adjusting for bias exceed 95%. Additionally, the width of the confidence interval is only about 12% wider for the Kaplan–Meier than the Weibull method.

The conclusion of this simulation is that the Kaplan–Meier method works well in these cases with no censoring. We now look at the cases with censoring.

Table 10.5 gives the same results as Table 10.4 when there is 30% right censoring. In this case nearly 30% of the deaths are not observed because the study is terminated. The underlying density is the Weibull density. Here we see about the same coverage and bias between the two methods. In Table 10.6, however, with 30% right censoring and an underlying bimodal density we see a considerable bias in the Weibull method and nearly no bias in the Kaplan–Meier method. Additionally, the Kaplan–Meier confidence intervals are about the same as those of the Weibull.

Tables 10.7 and 10.8 give the results for 50% censoring. Here we see the same pattern as Tables 10.5 and 10.6. That is, the Weibull model works well if the underlying density is the Weibull, or similar to the Weibull and is severely biased if the bimodal model is correct.

TABLE 10.5 Confidence Intervals and Bias of Estimated Median Lifetimes From Simulated Data With 30% Right Censoring, Using Weibull Method

Method	Count	Coverage	Width	Bias
Kaplan–Meier	100	0.935	0.474	−0.001
Weibull	100	0.941	0.392	−0.001
Kaplan–Meier	1000	0.947	0.152	0.001
Weibull	1000	0.939	0.124	0.001

The Kaplan–Meier and the Weibull methods for modeling the spinal cord injury data were used. Coverage is based on an assumed 95% confidence intervals.

TABLE 10.6 Confidence Intervals and Bias of Estimated Median Lifetimes From Simulated Data Using 30% Right Censoring and Bimodal Method

Method	Count	Coverage	Width	Bias
Kaplan–Meier	100	0.989	0.409	−0.006
Weibull	100	1	0.414	0.300
Kaplan–Meier	1000	0.990	0.409	0.001
Weibull	1000	1	0.416	0.304

The Kaplan–Meier and the Weibull methods for modeling the spinal cord injury data were used. Coverage is based on an assumed 95% confidence intervals.

TABLE 10.7 Confidence Intervals and Bias of Estimated Median Lifetimes From Simulated Data Using 50% Right Censoring and the Weibull Method

Method	Count	Coverage	Width	Bias
Kaplan–Meier	100	0.934	0.494	0.003
Weibull	100	0.945	0.411	0.005
Kaplan–Meier	1000	0.941	0.154	0.001
Weibull	1000	0.935	0.129	0.001

The Kaplan–Meier and the Weibull methods for modeling the spinal cord injury data were used. Coverage is based on an assumed 95% confidence intervals.

TABLE 10.8 Confidence Intervals and Bias of Estimated Median Lifetimes From Simulated Data Using 50% Right Censoring and the Bimodal Method

Method	Count	Coverage	Width	Bias
Kaplan–Meier	100	0.981	0.406	−0.008
Weibull	100	1	0.498	0.370
Kaplan–Meier	1000	0.994	0.406	−0.002
Weibull	1000	1	0.497	0.376

The Kaplan–Meier and the Weibull methods for modeling the spinal cord injury data were used. Coverage is based on an assumed 95% confidence intervals.

TABLE 10.9 Confidence Intervals and Bias of Estimated Median Lifetimes From Simulated Data Using 70% Right Censoring and the Weibull Method

Method	Count	Coverage	Width	Bias
Kaplan–Meier	100	0.989	0.497	–
Weibull	100	0.958	0.483	0.002
Kaplan–Meier	1000	0.946	0.174	0.001
Weibull	1000	0.957	0.150	0.001

The Kaplan–Meier and the Weibull methods for modeling the spinal cord injury data were used. Coverage is based on an assumed 95% confidence intervals.

On the other hand, the Kaplan–Meier model is nearly unbiased in both cases. Coverage of both methods is about 95% for the Weibull model but too high for the bimodal model. This means that the method of estimating the confidence interval for either of these methods when the densities are similar to the bimodal is too conservative (too wide).

Table 10.9 is the tabulation for the Weibull density case when the censoring is 70%. This case is similar to the spinal injury data file. Note that for samples of size 100, with 70% censoring, the Kaplan–Meier bias is blank. This indicates that the method could not come up with an estimate of the median lifetime. This is not a surprise since there is no time to death on 70% of the patients. In the case of 1000 patients, there are enough observations that the procedure can often estimate a median lifetime, but not always. The Weibull method, on the other hand, does generate estimates that are nearly unbiased. However, at 70% censoring, however, it is difficult to determine if this is the right model. The table for the bimodal density is not given here as the stability of both estimates was very poor giving high bias in both cases. The simulations in the case seemed to indicate that neither method was reliable.

Conclusions and Applications

The conclusion from the simulation is that for most cases where the density of the survival time is similar to the Weibull model, both the Weibull and the Kaplan–Meier model work well. Bias is low and the confidence intervals are either close or slightly conservative. Note that the confidence intervals of the Kaplan–Meier estimates are wider than those estimates based on the Weibull distribution. On the other hand in the case of the bimodal density function the Weibull method is biased but the Kaplan–Meier is not, provided that the censoring is 50% or less. In the case of 70% censoring, any estimate of the median lifetime depends heavily on the assumption that the density function of the lifetimes is approximately Weibull.

We now apply these results to the spinal cord injury example using the "more probable than not" argument. In Fig. 10.1, we plotted the survival curve for a subset of the spinal injury data. Note that in that curve the survival distribution estimated by the Kaplan–Meier estimate tapers off much faster than the curve estimated by the Weibull method.

Additionally, the estimated median lifetimes of the two procedures, 35.20 years versus 43.23 years, show that the Weibull method estimate is about 21% higher than the Kaplan–Meier estimate.

As shown in the previous section, if the density of lifetimes follows a bimodal distribution, then the Weibull method will be biased high from 20% to 40%. On the other hand, the previous section also demonstrated that when there was a 70% right censoring, as in this data set, the Kaplan–Meier estimate of the median could be unstable.

For purposes of illustration, we will assume that Kaplan–Meier estimate is usable. We wish to determine a length of time that a patient in this group will live. The purpose here is to provide for calculations based on this lifetime, such as the expected medical care costs, years of life to be lived as disabled, and so forth, so that "more probably than not" the calculations will be sufficient. In this case, as noted above we use the 95% upper confidence bound. Put in other words, for a patient in this group we use as an estimate of lifetime the value of 36.40 years, the 95% upper bound given in Table 10.2, rather than estimated median of 35.20 years, given in Table 10.2. If we forecast a lifetime of 36.40 years, we are 95% confident that the calculations of resources based on 36.40 years will more probably than not be sufficient.

Suppose, on the other hand, the issue is the years of life lost as a productive member of society. For this example, suppose that the life table of a person with the same demographic, risk factors, and socioeconomic status has a median lifetime of 45 years. In this case the years of life lost is calculated as 45.00 years minus 34.81 years = 10.19 years. The 34.81 is the lower bound of the 95% confidence interval for the median lifetime given in Table 10.2. In this case, we are 95% confident that a settlement based on 10.19 years of life lost will "more probably than not" be sufficient.

Note that in either case choosing the median rather than the upper or lower confidence limits will result in conclusions for which we are only 50% confident of that any amount will "more probably than not" be sufficient. Note also that if the entire data set had been that represented in Fig. 10.2 with only 100 individuals followed-up, with a 70% censoring, we would be unable to choose an upper bound. In this case, methods described below would have to be used to adjust an existing life table to this circumstance.

INCLUDING RISK FACTORS AND SEVERITY MEASURES

In the previous sections we have illustrated three different estimates of the median lifetime and their confidence interval. These can be used to assess loss using the "more probable than not" levels of loss, either in cost, loss of life, and so forth. However, in the examples above, recall that we grouped all non-Hispanic patients between ages 20 and 35 into a single cohort and modeled the time to death, where time is the number of years since injury. One would expect that a 35-year-old patient would have a shorter lifetime than a 20-year-old patient. In addition, patients with more severe injuries or patients with other risk factors, whether associated with the injury or not, will have shorter lifetimes. None of these life shortening conditions have been considered so far.

There are two related ways to extend the methods we have discussed so far. The first is to assume that the various risk factors and the levels of severity of injury collectively define

different strata or risk groups in the cohort. This is a method traditionally used by actuaries for both life and health policies. Collectively these factors define either a normal risk group or one of several levels of "rated" risk groups. Each of the risk groups are then defined for every year of age. The second way, a modern approach used by epidemiologists and biostatisticians, is to model the risk factors and levels of severity as regressor variables in a generalized linear model. This method, in essence, defines a multitude of risk factor groups for all values of all of the risk factor variables considered simultaneously.[11]

One can incorporate risk strata using any of the methods here simply by fitting a survival function to each of the risk strata groups. One of the advantages in the risk strata groups is that the division of individuals is very coarse, usually resulting in relatively large numbers of individuals in each risk strata. A disadvantage is that with coarse groupings the effects of key risk factors may not be evident.

To implement a generalized regression model is more involved. The life table method uses the general contingency table setup to estimate the survival function and its variance.[12]

The Cox proportional hazards model resembles an extension to the Kaplan–Meier method. One may view this model similar to a risk strata model. Each level of the regressor variables defines a risk stratum that one can fit the Kaplan–Meier survival curve to. However, to get the ability to estimate the various survival curves for each of the risk strata groups, an underlying hazard function is assumed to be common to all risk groups. The different risk factors are assumed to proportionally change (increase) the baseline hazard function to model the mortality of the risk strata group.[11]

The Weibull model, or most other parametric models incorporate regressor variables using the "accelerated life testing" approach, is commonly used in reliability analysis.[11] In this case a risk factor "accelerates" the aging process similar to the manner in which one might proportionately increase the hazard function in the Cox model. In general, accelerated testing methods have been only marginally successful when dealing with human populations.

ADJUSTING EXISTING LIFE TABLES

Often, extensive follow-up data on individuals with a certain injury or event does not exist. In these cases one cannot construct either a life table or even a parametric or nonparametric model of lifetime. In such cases one may be inclined to use readily available life tables for uninjured individuals and adjust them according to some pattern seen among the injured. For example, if a cohort of 100 individuals who have been injured is followed for 2 or 3 years, there would clearly not be enough information to determine a median life expectancy or a survival curve. However, one could determine the excess mortality of this cohort relative to an uninjured cohort of the same age and demographic profile. This excess mortality can then be used to increase the mortality rate per year in the life table for the uninjured individuals. The life table created by increasing the mortality of the standard life table at each year using the observed excess mortality in the sample cohort might then be used for estimating the median lifetime of an injured patient. In this case the original life table would be assumed to have little or no error. Thus most of the variation in the estimate would come from the variation in the estimate of excess mortality. This variation would be used to determine the "more probable than not" threshold.

We note here that excess mortality is only one method of adjusting an existing table to model an injured subset of the population. There are a variety of methods, the strengths and weaknesses which have been discussed in the literature.[13,14,15]

TECHNICAL APPENDIX

Most computer programs that provide estimates of survival curves also provide estimates of the variance of these survival curve estimates. It is also common for these computer programs to provide an estimate of the median and the variance of the median.

For the examples considered here we used two functions in R:

1. flexsurvreg() with dist = "weibull"
   This can be obtained without cost at http://CRAN.R-project.org/package=flexsurv
2. survfit() with type = "kaplan-meier"
   This can be obtained without cost at http://CRAN.R-project.org/package=survival

However, in some cases this variance may not be readily available. For example, if one estimates the survival curve by using a standard curve and then multiplying the mortality rates by an estimated measure of excess mortality, the variance of the estimated median lifetime is not directly available. In this case, the baseline survival curve is estimated with very low variance. However, the excess mortality factor may be estimated with significantly higher variance. To implement the "more probable than not" argument, we present here the method of obtaining the variance of the median lifetime given that one knows the variance of the survival curve.

DEFINITIONS

We will use the following notation:

$S(x)$ = the survival function. This is the function that indicates the probability that an individual currently alive at time 0 will survive to time t. The survival function can be represented as a graphical curve, a table, or a formula. Clearly the survival function may be dependent on several parameters of the group of individuals it represents. For example, age, race, sex, and general health of the individuals may alter the likelihood an individual survives to a specified length of time.

$\widehat{S}(x)$ = estimate of the survival function, $S(x)$. This estimate is based on data. The kind of data available, eg, right or left censored, discrete, aggregated, continuous, etc., dictate to some extent the methods that can be used to estimate the survival function.

$f\left(\widehat{S}(x)\right)$ = denotes a function of the survival function of interest. For example, the average lifetime, the probability of living over 50 years, the cost of a series of annuity payments from now until death are all functions of an estimated survival curve. Some of the functions are simple straightforward calculations using the survival function. Others, such as the median and other quantiles are more complex. Some actuarial calculations are very involved. At

the base of each of this, however, is the estimated survival function. Thus, the accuracy of these functions depends directly on the accuracy of the estimate of the survival function.

$Var\left(f\left(\widehat{S}(x)\right)\right)$ = the variance of some function of the estimated survival function.

m_p = the pth percentile of the survival function. That is, m_p is the time at which to which 100*p% of the population will not survive. Put in other words, 100*p% of the population will die before time (age) m_p.

$$S(m_p) = 1 - p \qquad [10.1]$$

The estimate of the pth percentile, denoted $\widehat{m}_p$, satisfies Eq. [10.1] with $S(m_p)$ replaced by $\widehat{S}(m - m_p)$.

VARIANCE OF A QUANTILE

Confidence intervals of the quantiles are based on the normal approximation using the square root of the variance of the estimated quantile. Given the variance of the estimated survival function, $Var\left(f\left(\widehat{S}(x)\right)\right)$, the variance of a quantile is determined using Greenwood's formula as described here.

Recall, using Taylor's expansion, that the variance of some function of a random variable, X, say $g(X)$, is given as

$$Var(g(X)) \approx \left\{\frac{dg(x)}{dx}\right\}^2_{x=\mu} Var(X) \qquad [10.2]$$

Applying this result to the estimate of the median we know that

$$Var\left(\widehat{S}(m_p)\right) \approx \left\{\frac{d\widehat{S}(x)}{dx}\right\}^2_{x=m_p} Var\left(\widehat{m}_p\right) \qquad [10.3]$$

Now, recall $\frac{dS(x)}{dx} = -f(x)$, where $f(x)$ is density function of the time to death model. Substituting this into Eq. [10.3] and rearranging terms we get,

$$Var(m_p) = \frac{1}{f(m_p)^2} \cdot Var(S(m_p)) \qquad [10.4]$$

The result in Eq. [10.4] gives the variance of m_p as a function of the variance of $S(m_p)$ divided by the factor $f(m_p)^2$. If we set $p = 0.5$, then Eq. [10.4] gives the variance of the estimate of the median. One problem with this is that one must either know or have an estimate of $f(m_p)$. If the survival function is a parametric model such as the Weibull or Gompertz model, then forming an estimate, $\widehat{S}(x)$, entails estimating one or more parameters. In this case forming an estimate of $f(x)$ simply entails using these estimates in the derivative of $S(x)$. This is relatively straight forward. However, for nonparametric models approximate methods must be used.

ENDNOTES

1. The calculation of the confidence interval here is based on the assumption that the estimated life expectancy quantile is normally distributed and that the variance of the quantile can be reasonably approximated by the Greenwood formula. As such the procedure presented here is a large sample procedure, meaning that it is an approximation that becomes more accurate as the accuracy of the estimated survival function is increased by increasing the number of observations in the data set used to estimate the survival function.
2. Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk, third ed. McGraw-Hill, New York.
3. Note here that a 50% chance is calculated as follows: If we use the median lifetime estimate on each of many times and on each time we use the "more probable than not" argument, then the argument is correct all of the time. However, the threshold that we use to determine the number of years for which medical care is needed, this number of years is sufficient is only correct 50% of the time.
4. Alternatively, the number of years of earning potential that is lost.
5. The life table illustrations here are for survival time since back injury. Actuaries traditionally use life tables that summarize the mortality experience based on the analysis of one or more "experience" studies. Consequently, the entries in an actuarial life table are often not actual counts of individuals but are standardized counts that are expected based on the experience used to model the data. Typically, the number of individuals in the first age group, denoted the "radix," is arbitrarily set to 100,000.
6. Although all of the patients will eventually die, we only observe the patient until the "end of study" or the time at which the data set is assembled.
7. For example, the row beginning with Age = 0 lists 2 deaths between Age = 0 and Age = 1, that is during the first year after a spinal injury, with a probability of death during the year of $\frac{2}{14,353} = 0.0001$.
8. The median lifetime is actually usually slightly less than this time, but more than the previous time duration. For example, if the first time that the product drops below 50% is time 20, then the median is between time 19 and time 20. The exact value of the median is formed by interpolating between these two times. Interpolation can be linear or can be based on assuming a constant rate of death during the 1-year period.
9. We use the common two-parameter Weibull model. A third parameter can be used to change the time index. However, since we are considering time since an event, where the event time is set to zero, the two-parameter Weibull provides all the flexibility available for the Weibull.
10. Recall that the upper bound of the 95% confidence interval is used as a threshold for the "more probable than not" estimate.
11. Collett, D., 2003. Modelling Survival Data in Medical Research, second ed. Chapman & Hall/CRC, Boca Raton.
12. Koch, G.G., Johnson, W., Tolley, H.D., 1972. A linear models approach to the analysis of/survival and extent of disease in multidimensional contingency tables. JASA 67, 783–796.
13. Strauss, D., Shavell, R., 1998. Life expectancy of persons with chronic disabilities. Journal of Insurance Medicine 30, 96–108.
14. Strauss, D.J., Vachon, P.J., Shavelle, R.M., 2005. Estimation of future mortality rates and life expectancy in chronic conditions. Journal of Insurance Medicine 37, 20–34.
15. Shavelle, R.M., Strauss, D.J., Paculdo, D.R., 2006. Computing exact excess death rates from a published mortality study. Journal of Insurance Medicine 38, 105–110.

PART III

APPLICATIONS OF FORENSIC EPIDEMIOLOGY

This page intentionally left blank

CHAPTER

11

Traffic Injury Investigation

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

OUTLINE

INTRODUCTION

Road traffic crashes are the most common cause of serious injury and death in the US and Europe (Jacobs et al., 2000). As of 2015, it was estimated by the WHO that 1.2 million people die and between 20 and 50 million people sustain some degree of injury in a traffic crash each year worldwide[1]. Risk factors for traffic crashes often involve the negligent action of a driver (inattention, speeding, impairment due to alcohol or drugs, etc.).

Because those injured in traffic crashes are often not the negligent party, litigation associated with claims for damages is common[2]. Epidemiologic issues are commonly raised in crash-related litigation, most often relating to causation of injury. Traffic crash-specific issues related to multiple concurrent causes (see in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) include the effect of an unused or inoperative safety device such as a seat belt or air bag, or the contribution of an alleged defect in the construction of the vehicle, among others. The circumstances of such events may make the need for a forensic

http://dx.doi.org/10.1016/B978-0-12-404584-2.00011-2

epidemiology (FE) investigation of cause moot. An example would be the finding of a fractured wrist following an unintended air bag deployment in a parked vehicle. It could not be said that the injury could have occurred if the air bag had not deployed and thus the cause of the injury is indisputable. If, on the other hand, the air bag deployment had followed a moderate speed crash, the injury may have occurred regardless of the air bag deployment due to the forces of the crash. Only by comparing the risk of the injury for the same crash type and severity, both with and without an air bag, can the contribution of the unintended air bag deployment to the cause of the injury be quantified. This is an ideal application of the FE methods described in this chapter.

Assessment of the cause of injury associated with a traffic crash is a multidisciplinary undertaking. Traffic reconstruction methods are needed to assess the nature and severity of vehicle forces (typically quantified in terms of crash-related speed change, also known as "delta V") and thus accurately characterize the frequency of injury associated with similar collisions. Medical knowledge is needed to understand the nature of the collision-related injuries and their sequelae. Biomechanical methods are used to link the medical documentation of injury with the reconstructed crash forces, in order to assess how the collision may serve as a plausible explanation for the observed injuries.

As described elsewhere in this text, it is the lack of widely adopted standards regarding what constitutes scientifically valid evidence of injury causation, as well as a means of quantifying and weighing evidence of causation that is accessible and clear to judge and jury fact-finders, that has created a vacuum of information in crash injury litigation that has been filled, in part, with misplaced and/or unreliable opinions from the fields of crash reconstruction, biomechanics, and medicine (Freeman and Kohles, 2011).

Causal assessment of crash-related injuries, outside of a legal arena, is always performed in a clinical setting and largely based on a patient history-based assessment of temporal order and proximity between a crash and an injury; ie, if a diagnosed injury follows a crash relatively closely in time then the injury is attributed to the crash. Causation of serious injury following a traffic crash is typically assumed, based on the high degree of correlation between the injury and a high-energy event. As an example, for a femur fracture observed in a front seat occupant following a frontal crash, the injury is highly correlated with the type of sudden load to the thigh that might occur in a frontal collision, even one occurring at a relatively low speed. (Tencer et al., 2002). Further, traffic crashes are the most common cause of nonosteoporotic femur fractures, and thus the injury can be said to be "specific" to the exposure, and the exposure is specific to the injury (Böstman et al., 1989). Readers will note that the use of specificity here means simply "unique to," in contrast to the specialized way the term is used in the Hill viewpoints (Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*).

Another aspect feature of the causal analysis of serious injuries following traffic crashes comes from a counterfactual perspective, which approaches the cause of the injury by examining the likelihood that the injury would have occurred at the same time, but-for the exposure to the hazard. Femur fractures are neither insidious nor spontaneous in healthy bones, and therefore if the only extraphysiologic load occurring at the time that the injury became apparent is a traffic crash, then the crash is not only the most likely cause, but also the sole plausible cause. This approach holds true even if a femur fracture were exceedingly unlikely or virtually unheard of in a very low-speed collision, as the suggestion that the injury was present but undetected prior to the collision would not pass a common sense threshold. Equally implausible is the suggestion that the injury resulted from some other, unknown

traumatic source of an extraphysiologic load that must have exceeded that of the collision, but which occurred at some time between the collision and the first recognition of the injury. As a result of these features, the cause of injuries that are highly specific for high-energy trauma are rarely, if ever, disputed as resulting from a particular collision.

A far more common source of dispute over the cause of serious injury concerns the effect that the presence or absence of a safety device, such as an appropriately used seat belt or an air bag deployment, might have had on the cause of an injury outcome. These disputes are often fueled by the improper application of generally true principles that are only moderately correlated with injury outcomes. As an example, while it is accurate to assert that seat belt use decreases the risk of serious injury in traffic crashes (Orsay et al., 1990), the role of impact severity, crash type (rollover vs planar crash), impact vector, and supplemental restraint (air bag) presence and deployment, inter alia, may play an equal or greater role in serious injury risk (Crandall et al., 2001; McGwin et al., 2003).

Caution should be exercised with causal evaluations that are based on the application of generally true principles that are widely accepted (eg, "seat belts save lives"). While there are some crashes in which the use of a seat belt would have changed the outcome, there are other crashes for which a seat belt would make no difference. As an example, while seat belt nonuse is highly specific for occupant ejection and the associated injuries (very few restrained occupants are ejected), it has low specificity for serious injury, as such injuries occur in both restrained and unrestrained occupants. Further, the relationship between seat belt use, serious injury risk, and crash severity is Gaussian (bell-shaped); and thus at very low crash speeds seat belt nonuse has little or no effect because injury is rare regardless of use, and at very high speeds seat belts have little or no effect because the injury is certain regardless of use. Therefore, while the conclusion that "seat belts save lives" is generally true, such a generalization sheds no light on the question of "would the decedent have survived the 50 mph (80 km/h) crash with a tree, on a more likely than not basis, if he had been using his seat belt?" Such evaluations cannot be performed without the appropriate analysis of relevant epidemiologic data.

Another source of disputed causation in traffic crash injury litigation is seen in claims for injuries that have a low degree of association with crash-related trauma. Two examples of such injuries are symptomatic derangement of spinal disks, and cervical artery (ie, carotid and/or vertebral) dissection and stroke. Both conditions often occur with an insidious onset, with no history of any distinct extraphysiologic trauma of any severity (Chandra et al., 2007; Freeman et al., 2009). Conversely, even the occupant forces occurring in relatively low-speed crashes can potentially act as a trigger for such injuries (Pettersson et al., 1997; Hauser et al., 2010). The lack of high degree of correlation between a lower energy collision and the wide variety of injuries that may be attributed to such crashes invites speculative and potentially fallacious expert opinion regarding causation based in probabilistic terminology. Commonly seen examples of such opinions are as follows:

> The vehicle damage was minimal, and thus the injury risk from the crash was correspondingly low, and therefore it is unlikely that the claimed injuries resulted from the crash.

This opinion relies on the fallacy of the transposed conditional, where $P(\text{injury}|\text{crash})$ is erroneously considered to be the equivalent of $P(\text{crash}|\text{injury})$ (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*). In other words, the correct assumption that the *absolute* risk of injury from the crash was low is confused for a *comparative* risk assessment

of the risk of injury from the crash versus the risk of the same injury occurring at the same time, but in the absence of the crash. Absolute risk alone is not a means of assessing individual causation, unless the absolute risk is reliably known to be 0 or 1.

CRASH INJURY CAUSATION METHODOLOGY

The three fundamental elements of an injury causation analysis are as follows (Freeman et al., 2009):

1. Determine whether the injury mechanism had the potential to cause the injury in question;
2. Evaluate the degree of temporal proximity between the injury mechanism and the onset of the symptoms reasonably indicating the presence of the injury; and
3. Assess whether there is a more likely alternative explanation for the occurrence of the injury at the same point in time.

The results of an injury causation analysis can be quantified in terms of a comparative risk ratio (CRR), described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*. The result of the analysis, given as either a CRR or probability of causation, meets the legal standard of what is "more likely true than not," when the CRR is $\geq$2.0 (95% CI $>$ 1.0 lower boundary), or the PC is $\geq$50%.

The CRR is often derived from a relative risk or odds ratio in FE investigations of crash injury causation, depending on whether the investigation was focused on the exposure (crash) or outcome (injury) of interest. As an example we can examine a commonly disputed issue in traffic crash litigation, which is the contribution that the failure of an occupant to use a seat belt made to the cause of a serious injury. For the example we can make the collision a frontal crash occurring at 20 mph (32 km/h) speed change, also known as "delta V" (this concept is discussed later in this chapter). The CRR applicable to the individual could be based upon the examination of the relative risk of serious injury in unrestrained occupants exposed to a 20 mph delta V frontal collision versus the frequency of serious injury in restrained occupants exposed to the same collision severity and type. If, for example, the frequency of serious injury in the group exposed to the presumptive hazard (failure to use a restraint) was 0.15 and the frequency in the unexposed (restrained) group was 0.05, then the CRR would be:

$$\mathrm{CRR} \approx \mathrm{RR} = \frac{0.15}{0.05} = 3.0$$

If the same issue were to be examined using a case–control approach (typically because the injury of interest was relatively rare), the study groups would be selected based on injury status and then examined for exposure. Using similar circumstances as in the previous example, a group of randomly selected patients with a unique injury after a traffic crash could be compared to a group of randomly selected patients with lesser injuries after a crash for restraint use. For this type of an investigation, a logistic regression analysis may be used to control for other predictive factors such as impact severity and direction, as well as patient characteristics, and then the frequency of restraint use in the two groups could be compared.

The result of such an analysis would be given as an odds ratio. Risk and odds are different, in that a risk is a frequency of occurrence (eg, 0.05/1.0), whereas an odds is the frequency of occurrence versus the frequency of nonoccurrence (eg, 0.05/0.95).

Risk and odds relate to each other mathematically as follows:

$$\text{Risk} = \frac{\text{Odds}}{1 + \text{Odds}}$$

$$\text{Odds} = \frac{\text{Risk}}{1 - \text{Risk}}$$

Thus, if 0.05 of the seriously injured patients in the example were restrained and 0.15 of the less injured patients were restrained, the odds ratio for lack of restraint among seriously injured occupants would be as follows:

$$\text{CRR} \approx \text{OR} = \frac{0.05/0.95}{0.15/0.85} = 0.3$$

This result would be interpreted as indicating that seriously injured occupants are 70% less likely to be restrained than less seriously injured occupants. The proportion can be inverted by dividing the reference odds of belt use of 1.0 in the less injured group by the odds in the seriously injured group of 0.3 to result in an odds ratio of harm for the failure to use a restraint of 3.3.

Note that using the same frequencies in the two previous examples resulted in a slightly greater CRR of causation for the odds ratio (case–control approach) than for the relative risk (cohort approach), 3.3 versus 3.0, respectively. When the frequency of the exposure is low the odds ratio approximates the relative risk, but when the frequency is higher the values can be substantially different. As an example of this relationship, a relative risk of 0.75 versus 0.25 would also result in a CRR of 3.0, as follows:

$$\text{CRR} \approx \text{RR} = \frac{0.75}{0.25} = 3.0$$

The application of the same frequencies within an odds ratio results in a substantially greater CRR, however:

$$\text{CRR} \approx \text{OR} = \frac{0.75/0.25}{0.25/0.75} = 9.0$$

In such a situation it would be more appropriate to convert the odds to risks prior to calculating the ratio, in order to arrive at a more accurate CRR that would be applicable to an investigation of specific cause.

In some investigations of crash injury causation the numerator and denominator of the CRR are derived from very different populations and thus neither a relative risk nor an odd ratio can be used (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* for additional discussion). This situation occurs when the numerator of the CRR is a per-event risk, and the denominator is a per-time risk (also known as a cumulative risk). An example of such an application would be a pulmonary embolism (PE) that occurred a week after a patient sustained a lower extremity fracture in a crash. Such complications often result from

blood clots forming in the legs and then traveling to the lungs. If the patient had a history of deep venous thrombosis (DVT) in the lower extremities prior to the crash, then the CRR approach could be used to assess the probability of PE given a lower extremity fracture (a per-event rate) versus the 1-week risk of PE in a patient with DVT (a time dependent probability). Both numerator and denominator can be described as fractions or probabilities despite the fact that they are derived differently. When the probabilities are very small, fractions tend to communicate the magnitude of a risk more readily to a lay fact-finder than a probability (ie, 1 in 8500 vs 0.000118).

In some crash injury causation investigations the CRR may be used to assess the individual risks associated with two competing theories of cause presented by opposing parties, with unique fact patterns that require disparate analysis. When this occurs the CRR stands as a unique metric without an obvious analog in population-based epidemiologic study. The following case serves as an exemplar of this type of application.

An unrestrained 35-year-old man was traveling on a highway on a winter evening in a 2008 Nissan Altima sedan. His vehicle struck the rear of a slow moving large truck at highway speed in the right lane, causing the air bags in the vehicle to deploy and disabling the vehicle. Within approximately 30 s following the first collision, the sedan was struck from behind by a semi-tractor/trailer traveling at highway speed. The man was pronounced dead at the scene, and later examination revealed extensive skull fractures and brain and spinal cord disruptions, along with severe chest, abdomen, and spine injuries. See Figs. 11.1 and 11.2.

The question addressed by the FE analysis was which of the two crashes was the cause of the death. Thus, the CRR was as follows:

FIGURE 11.1 The semi-tractor/trailer and sedan at final rest. The rear tires of the sedan are obscured by the front tires of the semi, which have ridden up over the rear of the car.

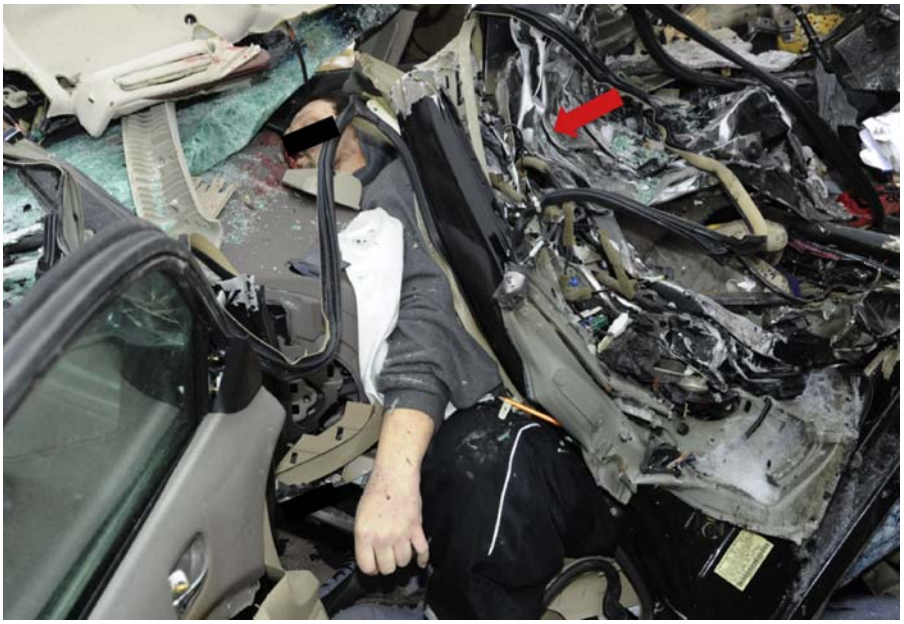


FIGURE 11.2 The position of the decedent after the semi has been towed off of the car. The arrow indicates where the front bumper of the semi was located prior to removal.

$$\text{CRR} = \frac{p(\text{death}|\text{first crash})}{p(\text{death}|\text{second crash})}$$

The analysis was simplified by the medical, biomechanical, and crash reconstruction evidence, as well as the common sense conclusion that the second crash was unsurvivable. Even if the decedent had not received the fatal head and chest injuries in this collision he would have died from compression asphyxia. An analysis of crash injury data from a crash injury database (the US National Automotive Sampling System-Crashworthiness Data System (NASS-CDS), described in more detail later in this chapter) indicated that the risk of death was very low for the initial frontal collision, despite the decedent's failure to use the seat belt, largely because of the air bag deployment. In a 17-year span of investigated crashes (1995–2011) there were an estimated 209,760 unrestrained drivers exposed to frontal collisions with an air bag deployment in the range reconstructed for the first collision (15–30 mph delta V (24–48 km/h)). The number of drivers with serious and greater injury was 11,108 or 5.3% (1 in 19) and the number of deaths was 181 or 0.09% (1 in 1159). The results of the CRR analysis thus indicated a 19 to 1 probability favoring the second collision as the source of the serious and greater injuries, and a 1159 to 1 probability favoring the second collision as the cause of the decedent's death.

As another example of the use of CRR where the numerator and denominator are from differing populations is with a low-speed collision that is temporally associated with a significant or persisting injury. As an example, we can consider an individual who was exposed to a rear impact traffic crash that was reconstructed to be a 6 mph

(9.6 km/h) delta V and shortly thereafter (within a day) was diagnosed with neck and arm pain that was ultimately attributed to a herniated cervical disk and operated on 3 months later. The exposure risk numerator of the CRR calculation would be related to the type and severity of the crash and would represent the frequency of injury per event. A previously published analysis of rear impact injury risk by delta V resulted in the risk curve depicted in Fig. 11.3 (Freeman, 2015a).

Thus, at a 6-mph (9.6 km/h) rear impact delta V the risk of a symptomatic cervical disk derangement is approximately 3.0% or 1 in 33 with a 95% confidence interval ranging from 0.9 to 9.3. The denominator of the CRR is an assessment of the competing theory proposed in the course of the litigation; for the example, it can be the theory that the onset of symptoms occurred at the same time as the crash but as a coincidence to, rather than as a result of the collision. The assessment of the cumulative risk of the spontaneous onset of the condition at the same time as the collision would be highly dependent on what is known about the individual.

If, for example, the individual had a history of a resolved neck injury from 3 years before the crash it can be reasonably assumed that this history would put him at greater risk of a spontaneous occurrence of neck pain than if there was no such history, due to a greater state of fragility in his neck. Age and gender are also important considerations, as well as other case-specific facts, potentially. One must keep in mind that, however, a characteristic that would tend to make an individual more likely to develop neck pain spontaneously would also tend to make the individual more likely to be injured as a result of the investigated crash.

One method for assessing the competing denominator risk of injury is to evaluate the rate at which the treatment for the injury occurs in the relevant population. For example, if the

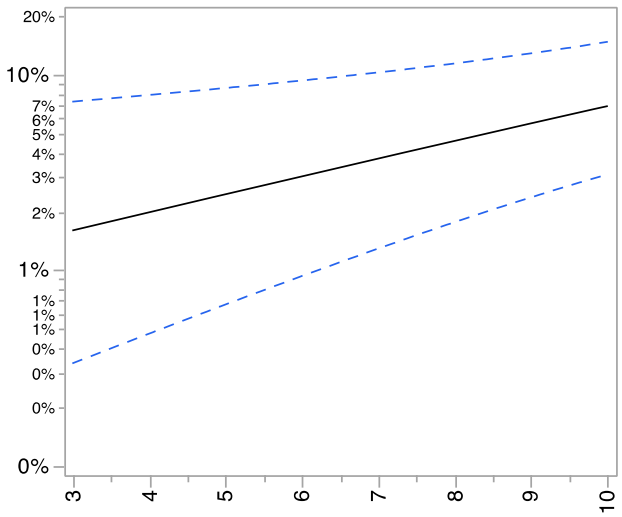


FIGURE 11.3 Log transformed risk (%) of cervical disk injury in a low-speed rear impact collision, by delta V (mph), with 95% CI (indicated by dashed lines).

individual in the example were a 40-year-old male with no prior history or other predilection to injury, it can be assumed that his precrash annual risk of neck surgery was no greater than that of the general population in his age and gender group. In fact, since many people who undergo neck surgery have symptoms that have been present for many months or years, the man in the example is likely to be at a much lower risk for surgery than the others in his age and gender group who will go on to have the surgery. Using an overestimated value for the denominator of the CRR decreases the risk of Type I error, however (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* for further discussion of error types). If it can be determined that, for example, the annual rate of neck surgery is less than 1 per 1000 population in the man's age group from population-based data, then the cumulative risk of surgery during the relevant time period between the crash and the onset of the signs and symptoms of injury can be estimated. If we assume that 1 in 1000 men like the investigated subject sustain a new onset of symptoms that lead to neck surgery in a year, and that the risk of this occurrence is uniform during the year, then the 1-day risk of such an occurrence (the cumulative risk) would be 1 in 365,000 (1 in 1000/365 days).

A CRR based on these values would be as follows:

$$\begin{aligned}\text{CRR} &= \frac{p(\text{cervical disk injury}|\text{6 mph rear impact crash})}{p(\text{spontaneous onset of disk injury}|\text{randomly selected day in the year})} \\ &= \frac{1 \text{ in } 33}{1 \text{ in } 365,000} = 11,060 \text{ to 1 in favor of the crash}\end{aligned}$$

Of course this example has many assumptions and room for disagreement over the values used; however the very large ratio favoring the collision as the cause of the investigated injury covers a very wide boundary of potential error without changing the net conclusion of the analysis. Even if the collision only presented 1/10th the risk of injury used for the calculation, and there was a 10 times greater risk of the spontaneous occurrence of the condition in the absence of the crash, the CRR would still favor the crash as the cause of the injury by more than 110 to 1.

CASE STUDY EXAMPLES

The following four unique case studies illustrate the concepts previously described in this chapter, as well as in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*.

In the first case the potential contribution of the failure to use a seat belt to a serious head injury is quantified using epidemiologic data. There was no dispute that the injury resulted from the crash, but the degree to which the injury was caused by the failure to use a seat belt was unclear. The results of the analysis are presented as a relative risk that was converted to a probability of causation via the estimated CRR.

In the second case, the cause of an indisputable lumbar spine fracture and other injuries associated with a low-speed traffic crash is examined. The central issue was the plausibility of the fracture given the nature of the collision. This case serves as in illustration of the logical application of epidemiologic concepts, rather than as a CRR calculation demonstration.

In the third case, the cause of the need for hip replacement surgery is examined in a middle-aged male following a low to moderate speed rear-impact collision. The primary disputed issue was concerning whether the trauma of the collision altered the natural course of the degenerative process that was present in the man's hip prior to the collision.

In the final case, the investigation of the cause and timing of the death of a young woman exposed to two crashes occurring in a very short span of time is described. This case study demonstrates a comprehensive FE analysis, incorporating crash reconstruction, biomechanical, epidemiologic, and forensic pathology features.

CASE STUDY #1: SEAT BELT EFFICACY ANALYSIS

There are several predictive factors to consider when performing an FE analysis of the efficacy of a seat belt in reducing the risk of an injury in a specific crash scenario:

1. The direction of the crash. Seat belts are primarily effective at reducing the risk of ejection from a vehicle in the event of a crash, most often involving a rollover. Secondarily, they reduce occupant motion, and thus injury risk, but only for certain types of crashes. As an example, seat belt use does not alter serious injury risk in side impacts, as 3-point restraints do not affect the frequency or severity of head contact with the side window/window frame/B-pillar in such collisions. On the other hand, seat belts provide the greatest reduction of injury risk for frontal collisions, because they have the greatest effect on reducing occupant movement toward and thus contact with vehicle interior structures in front of the occupant.
2. The severity of the crash. This parameter is quantified by the crash "delta V" or speed change, which is the near instantaneous change in speed exerted on the vehicle at the time of the crash (over approximately 1/10th of a second). It is important to understand that the ability of a seat belt to reduce injury risk is highly variable depending on the crash-related delta V. As an extreme example, in a 2 mph (3.2 km/h) delta V there would be no difference between the serious injury risk for a belted versus an unbelted occupant, as such injuries would be exceedingly rare for both occupants. Correspondingly, at a 100 mph (160 km/h) delta V there would also be no difference between the serious injury risk for a belted versus unbelted occupant, as virtually all occupants would sustain a serious or greater injury in such collisions regardless of belt use. Thus, there is a "sweet spot" of delta V in frontal collisions where seat belt use has its greatest effect on reducing injury risk (ie, the relationship has a bell-shaped distribution). Generally, seat belts have their greatest effect on reducing injury risk in the 10–30 mph (16–48 km/h) delta V range. Crashes with a speed change of more than 30 mph (48 km/h) are relatively unusual. See Fig. 11.4.
3. The seating position of the occupant. In comparison with front-seat occupants, rear-seat occupants do not have the advantage of an air bag in a frontal collision, and are often closer to the hard components of the vehicle, such as the roof rail, window frame, and B and C pillars.
4. The nature of the injury. Some injury types have a very high correlation with crash severity and seat belt nonuse; an example is femur fractures and head/facial injuries in

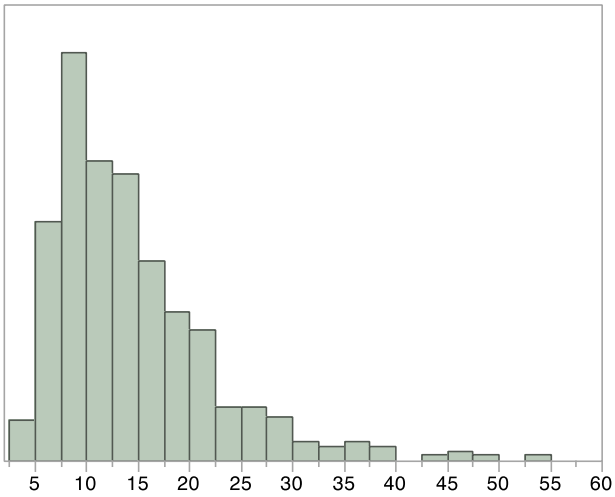


FIGURE 11.4 Distribution of frontal crashes (12 o'clock impact) by crash data recorder-based delta V in mph from NASS-CDS data for 2001–13. The average delta V is 14.5 mph, based on unweighted data gathered from 456 crashes.

frontal crashes, particularly when there has been no air bag deployment. Other injuries have a low correlation with either crash severity or seat belt use; an example is spinal disk injuries, which largely depend on the condition of an individual's spine, rather than the severity or orientation of a crash or seat belt use. Further, some injuries are very rare regardless of the severity of the crash or belt use, and there is simply insufficient information from which to draw any conclusions about the role of the seat belt on injury risk.

CASE FACTS

The case involved a high-speed impact between a Nissan sedan traveling at highway speed and a semi-tractor/trailer that was parked on the side of the highway. The collision was preceded by a collision between the sedan and another semi-tractor/trailer that made a lane change into the sedan's lane of travel, causing it to run off the road into the shoulder and strike the rear of the trailer (Figs. 11.5–11.7).

The speed at impact was reconstructed to approximately 45 mph (72 km/h); this was approximately the delta V as well. The restrained front passenger was killed in the crash. The driver's side rear-seat occupant sustained a serious head injury, including a diffuse axonal injury and subdural hemorrhage. It was the injuries of this occupant, who was not using an available seat belt, that were at issue. The central question in the investigation was whether the use of the seat belt would have prevented the serious head injuries sustained by the occupant, on a more probable than not basis.

FIGURE 11.5 Left front view of the sedan in full engagement with the trailer.

FIGURE 11.6 Right rear view of sedan, in full engagement with the trailer.

FIGURE 11.7 Right side view of sedan after it was towed away from the trailer, showing extensive damage to the front passenger compartment.

The defendant truck driver, whose actions caused the sedan to leave the road and collide with the rear of the trailer, retained an expert in biomechanical engineering to give an opinion on the effect that the use of a seat belt would have had on the injury risk for the rear-seat occupant. This expert opined that the results of crash testing for rear seated test dummies indicated that the use of a seat belt greatly decreased the risk of serious head injury for rear-seat occupants. The basis for the opinion was the head injury criterion (HIC) risk curve reproduced in Fig. 11.8 (Kleinberger et al., 1998).

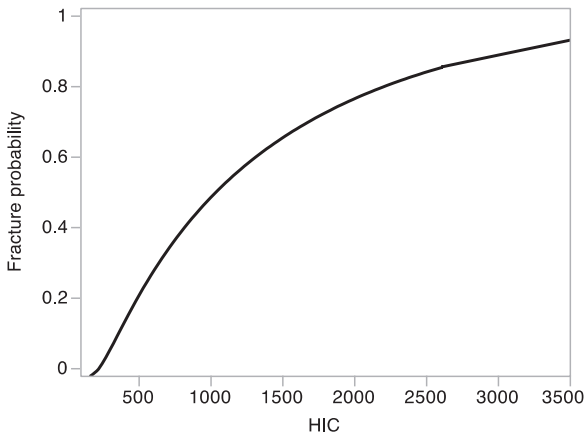


FIGURE 11.8 Head injury criterion risk curve *(adapted from Kleinberger, M., et al., 1998. Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems, NHTSA Technical Document).*

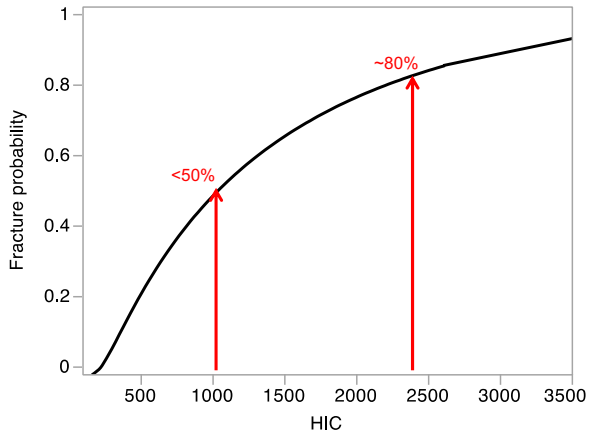


FIGURE 11.9 Illustration *(adapted from Kleinberger, M., et al., 1998. Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems, NHTSA Technical Document)* with the estimated points on the risk curve indicated by the biomechanical expert to correlate with the head injury criterion-based risk of serious injury for belted (arrow on the left) and unbelted (arrow on the right) rear-seat occupants.

The expert noted that 30 mph (48 km/h) barrier impact testing using unbelted crash test dummies resulted in a HIC of 2400, corresponding with a serious head injury (ie, skull fracture) risk of approximately 80%, and that the HIC for a belted dummy for a similar crash test was approximately 1000, corresponding with a serious head injury risk of less than 50%. The HIC is a function of linear acceleration and time. See Fig. 11.9.

ATTRIBUTABLE RISK METHODOLOGY FOR THE EVALUATION OF SEAT BELT EFFICACY

It could only be concluded that the nonuse of a seat belt is the cause of an injury if it can be shown that more than 50% of the frequency of an injury type and severity observed in an unbelted population is eliminated in a belted population in a comparative or attributable risk analysis.

As an example, if it was found that unbelted occupants sustain a particular injury in 30% of frontal crashes of >30 mph, and belted occupants sustain the same type and severity of injury only 10% of the time in the same type of crashes, it could be concluded that the relative risk of nonbelt use is 3.0, meaning that the injury occurs three times more often when a seat belt is not used. A relative risk of 3.0 also indicates that two out of three injuries that occur in an unbelted occupant would not have occurred if a seat belt had been used, and thus the risk of the injury that would be eliminated with seat belt use is 67%.

If the relative risk does not exceed 2.0, however, then it cannot be concluded *on a more probable than not basis* that the use of the seat belt would have prevented the injury. As an example, if the same injury described above occurs 15% of the time in unbelted occupants, and 10% of the time in belted occupants, the relative risk for injury given nonbelt use is 1.5. In this circumstance the attributable risk due to nonbelt use is 0.5 out of 1.5, which is 33% of the total risk. Thus, for every 15 crashes in which the injury occurs and no belt has been used, 10 (67%

of the total) would occur regardless of belt use. This is not to say that the use of the seat belt does not *reduce* the risk of injury in such a crash scenario, but rather that when looking at an individual case after the fact, the evidence does not support an opinion that the injury would have been *prevented* if a seat belt had been used, to a reasonable degree of probability (ie, more than 50% of the time).

CASE-SPECIFIC ATTRIBUTABLE RISK ANALYSIS OF SEAT BELT EFFICACY

For the purposes of the case described above a case-specific analysis of data accessed from the NASS-CDS database of the National Highway Traffic Safety Administration was performed. More information on this database, including how to access the data, can be found at http://www.nhtsa.gov/NASS.

The NASS-CDS investigates approximately 5000 crashes every year in 24 geographic primary sampling units (PSU) in the US. A record of over 800 variables including weather conditions, road conditions, injury to occupants or pedestrians, and vehicle damage is gathered for each crash by trained NASS crash investigators. In order for a collision to be recorded in the NASS-CDS it must meet several criteria: a police report was generated; it was located within a PSU; it involved at least one passenger car, van, or light truck; and at least one vehicle was towed from the crash scene. In turn, these data are weighted to provide a national estimate of all police-reported crashes occurring in the US and involving passenger cars, light trucks, and minivans that were towed due to damage.

The parameters of the NASS-CDS data query performed for the case described above were as follows: included were all outboard rear-seat passengers (no middle-seat occupants) of passenger vehicles including sedans, light trucks, and minivans, etc. that were exposed to a high-speed frontal collision of 30–60 mph delta V (48–96 km/h) with a direction of force ranging from 11 to 1 o'clock. Cases in which there was restraint use other than a 3-point seat belt, or the occupant was in a car seat, were excluded.

The variable of interest was 3-point seat belt use, and the outcome of interest was both overall injury severity and head injury severity. Injuries in the NASS-CDS are coded using a variant of the abbreviated injury scale (AIS), and range 1–6, as follows: 1 = minor, 2 = moderate, 3 = serious, 4 = severe, 5 = critical, and 6 = maximum. Only occupants with injuries with an assigned severity grade were included in the analysis (ie, occupants with no or an unknown injury grade were excluded from the analysis).

The results of the analysis were as follows: there were a weighted total of 6914 rear-seat occupants exposed to a frontal collision of 30–60 mph delta V severity in the 1995–2011 queried time frame (17 years, inclusive), based on a raw count of 106 occupants. Of these occupants, 50.2% (3469) were unrestrained and 49.8% (3445) were using a 3-point seat belt. The serious and greater injury rate in the unrestrained group was 25.6%, essentially the same as in the belted group (25.5%). These data indicate that there is no reduction in serious injury risk overall to rear seat occupants attributable to 3-point seat belt use in a high-speed frontal collision like the subject crash. There was a small difference between the rate of serious and greater head injury risk; there were 1477 such injuries among the unrestrained occupants (42.6%), and 1301 among the belted occupants (37.8%). This difference equated to a relative risk of 1.13

(95%CI; 1.06, 1.19), which means that 1.0 out of 1.13 (88.5%) of the serious head injuries in these crashes occurred regardless of seat belt use or nonuse, and 0.13 out of 1.13 (11.5%) of the injuries are attributable to nonuse of the seat belt. Because the CRR/probability of causation does not exceed 2.0/50%, it cannot be concluded that, for the rear-seat passenger, the use of a seat belt would have prevented his serious head injuries on a more probable than not basis. Indeed, there is an 88.5% probability that his injuries would have occurred regardless of seat belt use.

ANALYSIS OF THE DEFENDANT EXPERT'S METHODS AND CONCLUSIONS

The defendant expert's reliance on the theoretical injury risk derived from experimental cadaver testing data and two dummy crash tests lacks generalizability to real world, but there were other technical problems with the use of the experimental data, as well.

First, the expert failed to note that the relative risk between what he estimated to be "a little less than 50%" and "around 80%" from the graph reproduced previously in this chapter is 1.6 (80% is 1.6 times 50%), and this is insufficient to allow for the conclusion that the use of a belt would have resulted in a lesser injury to the occupant, on a more probable than not basis. Using these values, only 37% of the total injury risk, or 0.6 of the 1.6 relative risk, would be attributable to the failure to use the seat belt, and thus the legal standard for what is more probable than not would not be met by this evidence.

Further, the expert failed to note the error rate within the data that he relied upon. An examination of the data that the risk curve was based on revealed a total of 54 cadaver skull impact experiments. A reanalysis of the same data, including a 95% confidence interval, was performed, and is depicted in Fig. 11.10.

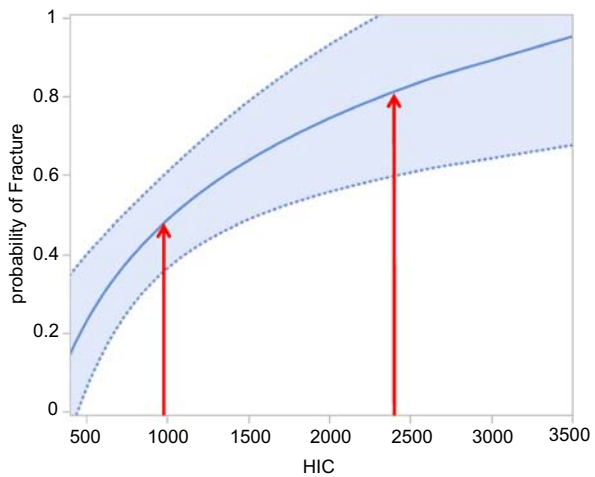


FIGURE 11.10 Reanalysis of data underlying the chart *(adapted from Kleinberger, M., et al., 1998. Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems, NHTSA Technical Document* with *arrows* indicating the level of head injury criterion and fracture risk for belted and unbelted rear-seat test dummies (left and right arrows, respectively), with 95% confidence interval.

It is apparent that the small number of specimens in the data resulted in a relatively wide confidence interval for the risk curve, such that the upper bound of the fracture risk for the belted occupants (0.60) overlaps with the lower bound for the fracture risk in the unbelted occupants (0.56). Thus, apart from other limitations of the analysis performed by this expert, the difference in risk between the belted and unbelted HIC values that he cited as a basis for his opinion cannot be said to be statistically significant at a *p*-value of 0.05.

CASE STUDY #2: LUMBAR SPINAL FRACTURE FOLLOWING A LOW-SPEED CRASH

A common approach utilized in the defense of low-speed crash-related injury claims is based on an analysis of the crash severity and associated biomechanical properties of the crash (Walz and Muser, 2000). Expert opinions stemming from such an analysis are often based on the (generally true) principle that the low severity of the crash indicates a low risk of significant injury, which is in turn used as corresponding evidence that there is a high probability that the injury did not occur in the crash (see the "conditional probability fallacy" described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*). In the case study below an unusual factual scenario made this approach impractical, as there was convincing evidence that the injury could only have resulted from the investigated collision, despite apparently low injury risk of the collision (Freeman, 2015b).

CASE FACTS

The subject was a 49-year-old previously healthy female, restrained driver of a 2006 Volkswagen Golf that was struck in the rear by a full-sized pickup truck. The only damage observed in the Volkswagen was to the license plate frame, and there was no apparent damage to the pickup truck.

The subject had an immediate onset of neck and back pain, and could not get out of the vehicle because of complaints of pressure and pain in the low back, as well as pain radiating into both legs. She was immobilized by emergency medical service providers and taken to the hospital for further evaluation. Plain radiographs taken at the hospital demonstrated an anterior superior vertebral body fracture at L5, an injury that was later confirmed as acute by MRI, which demonstrated edema of the adjacent marrow. Serial MRI examinations demonstrated progressive healing of the fracture over the months following the crash (see Figs. 11.11–11.14).

Because of persisting symptoms in the neck, right shoulder, and lumbar spine, the patient ultimately underwent five surgeries, including (1) an instrumented fusion for mechanical instability associated with bilateral facet fractures (confirmed intraoperatively) at L4-5, (2) removal of the L4-5 hardware for suspected pseudarthrosis and/or loose hardware, (3) a C3-4 anterior cervical discectomy and fusion with allograft, (4) a C5-6 discectomy and fusion with an intervertebral implant, and (5) an arthroscopic decompression of the subacromial space and repair of the glenoid labrum of the right shoulder.

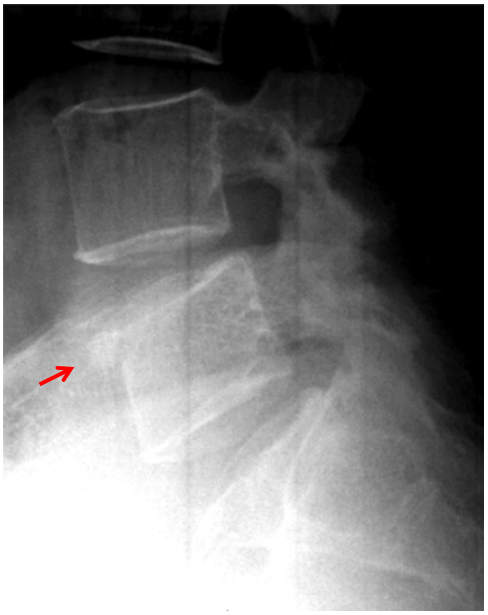


FIGURE 11.11 Lateral radiograph of the lumbar spine demonstrating the fracture at the anterior–superior aspect of the L5 vertebra, indicated by the arrow.

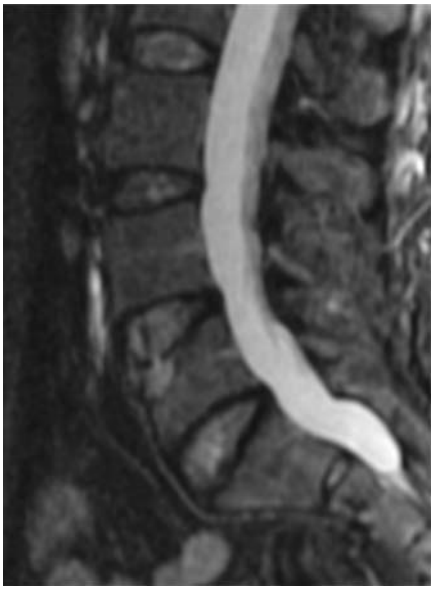


FIGURE 11.12 Lumbar MRI performed 9 days following the collision, demonstrating bright signal intensity in the anterior–superior aspect of the L5 vertebral body, consistent with edema and acute fracture.

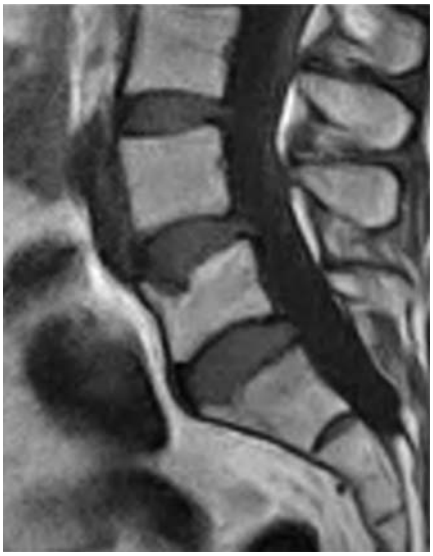


FIGURE 11.13 MRI from 5.5 months following the collision, demonstrating healing of the fracture.

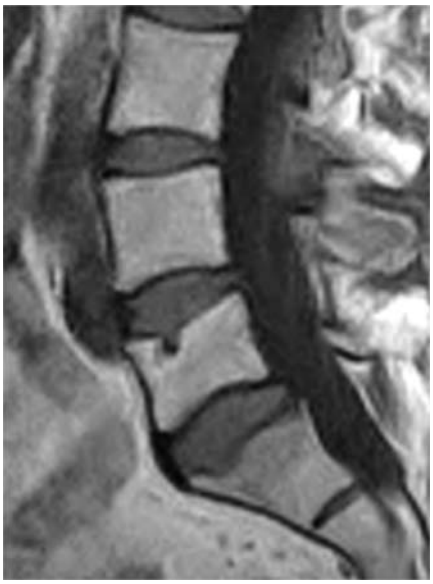


FIGURE 11.14 MRI performed 14 months after the collision showing complete healing of the fracture.

The patient made a legal claim against the insurer for her injuries and treatment following the collision, and the insurer obtained expert opinions that indicated the following:

1. The collision-related speed change of the Volkswagen was no more than 3.7 mph (6 km/h) (from an engineering expert);
2. A collision-related speed change of 3.7 mph (6 km/h) has been tolerated more than 500 times in experimental volunteer crash tests, and therefore the risk of injury to the patient in the crash was less than 1 in 500 (from a biomechanical engineering expert). The expert also opined that there was "no mechanism for injury" in the crash, essentially concluding that the observed injuries were impossible in the crash, as a means of denying that the injuries had resulted from the collision.

CAUSATION ANALYSIS

The case presented an unusual combination of conditions, in that a seemingly low-level trauma was closely associated with injuries with a high degree of specificity for a traumatic etiology; an acute superior endplate fracture of L5 (confirmed with serial imaging studies) and a bilateral facet fracture at the same level (observed intraoperatively). The injuries were conclusively dated to the time of the crash by the medical imaging, and no alternative explanations were considered plausible. For this reason, a comparative risk analysis became a moot task; mathematically, if the denominator base risk was essentially 0, then despite the very low risk of injury in the collision, the CRR would be so large that the attributable risk/probability of cause would be nearly 100%.

The opinion by the defendant's biomechanical expert regarding the results of volunteer crash testing as a basis for a less than one in 500 injury was erroneous on several levels; the absolute risk of injury from the collision has no context unless it is paired with the competing base risk, which, as mentioned above, was nearly 0. Further, the derivation of a less than 1 in 500 risk of injury in the general population based on the fact that 500 crash tests have not produced the same level of injury is erroneous in several regards. First, volunteers for crash tests are not representative of the range of injury susceptibility of the general population, as they are typically young, healthy, and robust males who are prepared for an impact. In contrast, real-world crashes involve occupants who are typically not prepared for the crash, often out of ideal seating position or in a rotated or awkward position, with prior history of injury or other health problems, and a variety of other factors that make them more susceptible to injury than a selected and prepared individual who is sitting in an ideal position in a vehicle waiting for an impact of a known severity (Freeman et al., 1999). The expert also failed to note that the 500 tests had been conducted on fewer than 100 mostly male young volunteers. As study of real-world collisions indicate that approximately 75% of the general population exposed to a 3.7 mph (6 km/h) rear-impact collision will not be injured to any degree (Freeman, 2015a), the accurate information that most occupants are not injured in similar collision is unhelpful in identifying the 25% of the population who are injured to some degree in such collisions, or the even smaller proportion of the population that is injured more significantly.

The expert's conclusion that, because injury was unlikely in the collision there was "no mechanism of injury," was a result of improper reasoning. The conclusion is akin to stating that "injury was impossible in the collision" without any supporting information or data, except the erroneous conclusion that 500 volunteer tests serve as an indication that the risk of injury was less than 1 in 500. The transmutation from "injury is unlikely" to "injury is impossible" with no additional information is semantic rather than material.

CASE STUDY #3: HIP REPLACEMENT SURGERY AFTER A TRAFFIC CRASH

A common and potentially contentious causation issue arises when a preexisting condition is worsened to the point that it requires treatment following a traffic crash of low or moderate severity. The question of whether the condition would have required treatment absent the influence of the crash becomes the central dispute. The following case study illustrates an analysis of the need for hip replacement surgery following a history of two crashes.

CASE FACTS

The subject was a 51-year-old male who was exposed to two moderate speed rear-impact collisions within 2 months. The circumstances of the collision were nearly identical; the subject was driving a 1993 Jeep Wrangler in both crashes, was using a 3-point seat belt, and his vehicle was struck from behind at 10–15 mph (16–25 km/h) by another passenger vehicle, resulting in an estimated delta V of 7–10 mph (11–16 km/h) and moderate vehicle damage. In both cases the subject had his right foot planted firmly on the brake. Following the first collision the subject developed low back and right thigh pain, and within a day of the second collision the pain in the hip worsened. He was diagnosed with symptomatic moderate severity degenerative joint disease of the right hip joint, and underwent hip replacement surgery 3.5 years after the first collision, although surgery had been recommended within 4 months of the second collision.

Approximately 6 months prior to the first crash the subject had sought evaluation for right hip pain following overuse during home activities. Following an X-ray of the hip he was diagnosed with mild degenerative joint disease. After a few weeks the hip pain resolved, and was asymptomatic until the time of the first crash.

CAUSATION ANALYSIS

An analysis of the nature and severity of the forces likely exerted on the subject's hip joint during the two collisions indicated that there was a plausible injury mechanism to his right hip. He gave a history of having heavily depressed the brake prior to the impact in both crashes, and thus preloading the hip joint. In the initial phase of the crash during the first 150–200 ms, the subject's body would have moved backward into the seat back. After this

time, his torso and hips would have loaded and deflected the seat back, which would have then sprung forward, propelling the subject's body forward. The braced right foot would then transmit force directly into the hip joint secondary to the axial loading from the forward body movement, and providing a plausible injury mechanism to the hip joint.

The first crash and the onset of symptoms in the hip were temporally proximate; the symptoms began in the hip approximately 1 month of the crash. Although the subject began limping after the second collision, he waited more than 3 years to have the hip replacement surgery.

This fact pattern raises the most important question to address in examining the cause of the subject's need for hip replacement surgery approximately 3.5 years after the rear impact collisions, which is whether there were any alternative explanations for the need for the surgery that are more likely than injury resulting from the collisions.

The comparative risk analysis for the crash was simplified by the competing theories for the cause of the hip surgery, as the surgeon who performed the total hip replacement asserted that the crash was the most probable cause of the need for surgery because of the symptom onset, and the surgeon who was retained by the insurer agreed that the collision caused the hip to become symptomatic and ultimately need surgery. He also opined, however, that the surgery would have been necessary even if the crash had not occurred because of the preexisting degenerative changes in the hip. Thus, the only facet of the CRR requiring investigation was the denominator risk of hip replacement surgery for the subject during the 3.5 years between the time of the collisions and when the surgery was performed.

In order to assess the frequency at which men in the subject's age group undergo partial or total hip replacement surgery (for any reason), an analysis of hospital data from the Nationwide Inpatient Sample (NIS) of the Healthcare Utilization Project (HCUP) of the Agency for Healthcare Research and Quality of the US Department of Health and Human Services was performed. The NIS is a publicly available database gathered and maintained by the US government, and it includes a 20% sample of all hospital discharges in the US. More information about the NIS, including how to access the databases comprising the NIS is available at http://hcup-us.ahrq.gov/nisoverview.jsp.

The results of the NIS database analysis were as follows: in 2010 (the most recent year available at the time of the analysis) there were an estimated 25,568 partial and total hip replacement surgeries among all US men aged 45–54 years of age. In the same year there were approximately 22,142,359 men living in the US in the same age group (http://www.census.gov/popest/data/datasets.html, accessed 10.05.15). Thus, the approximate annual frequency of hip replacement surgery in the subject's age bracket was 1 in 866.

It was further estimated from epidemiologic study that approximately 6.6% of the male population in the subject's age group has symptomatic hip osteoarthritis, which would equate to approximately 1,461,396 men in the US who were in a similar or worse condition as the subject prior to the crashes (Jordan et al., 2009). If the assumption is made that only those men who had symptomatic hip osteoarthritis underwent hip replacement surgery, then the 25,568 surgeries equate to 1 surgery per 57 men with symptomatic hip osteoarthritis in the subject's age group, per year. Assuming a uniform accumulated risk as a means of estimating the subject's precrash 3.5-year risk of hip surgery yields a 1 in 16 (6%) risk of conversion from nonsurgical to surgical hip during the time frame of interest, but absent the influence of the crashes (calculated by multiplying 1 in 57 by 3.5 years). Based on the

preceding analysis, it could not be concluded that the subject would have needed hip replacement surgery in the absence of the investigated collisions.

CASE STUDY #4: TIMING AND CAUSE OF DEATH

In the final case study of this chapter, the two questions of interest were (1) which of two crashes, occurring at nearly the same time, was the most probable cause of the death of the driver of a sedan, and (2) if the driver survived the first crash, how long did she survive? The second question related to whether the family of the decedent could recover for any amount of pain and suffering likely sustained by the decedent.

The investigated collision sequence involved a restrained 21-year-old female driver of a 2003 Saturn Ion sedan, traveling on a 2-lane highway. Her vehicle was stopped, waiting for opposing traffic in order to make a left turn into the parking lot of a high school. Behind her vehicle was a 2000 Peterbuilt 379 semi-tractor pulling a box trailer traveling at approximately 42 mph (67 km/h). The Peterbuilt struck the Saturn and propelled it forward and into oncoming traffic, where it was struck a second time. The overhead view of the crash scene is depicted in Fig. 11.15

Approximately 83 feet (25 m) behind the Saturn, the Peterbuilt began leaving tire marks from the rear dual tires. The impact speed was approximately 31 mph (50 km/h), slightly offset to right rear of the Saturn and approximately 5–10 degree clockwise to the long axis of the sedan. The impact resulted in a delta V of the Saturn of approximately 40 mph (64 km/h) causing the vehicle to rotate counterclockwise and travel approximately 77 feet (23 m) into the oncoming lane.

The extent of the crush to the rear of the sedan is apparent in Fig. 11.16.

Traveling in the oncoming traffic lane approximately 47 mph (75 km/h) was a 2008 Ford full-sized van. This vehicle left approximately 49 feet (15 m) of tire marks before impacting the right side of the Saturn. The van intruded into the passenger compartment of the Saturn by approximately 1.5–2 feet (0.5–0.7 m), and the impact produced an approximately 24 mph (38 km/h) delta V in the sedan.

FIGURE 11.15 Overhead view of the crash site indicating the approximate location of the first and second impacts.

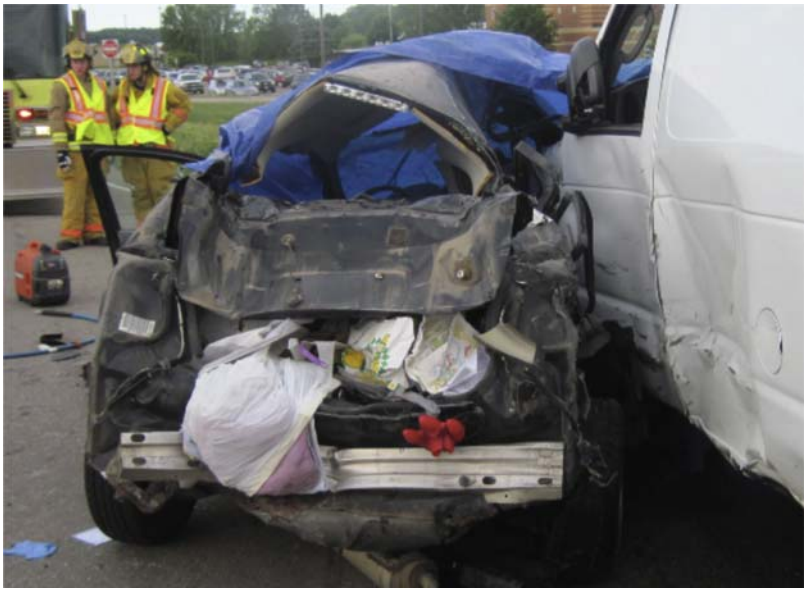


FIGURE 11.16 Rear view of decedent's vehicle at final rest, in contact with the van.

See Fig. 11.17, demonstrating the degree of intrusion from the right side of the Saturn as viewed from the rear (the postcrash position of the B-pillar is indicated by the "B_{post}" and the precrash position of the B-pillar is indicated by "B_{pre}," with the approximate distance between the two indicated by the arrow).

The vehicles continued to be engaged until coming to final rest approximately 56 feet (17 m) beyond the area of impact. The schematic depicted in Fig. 11.18 demonstrates the pre- and postcollision movement of the vehicles.

The driver of the sedan was pulseless at the scene and subsequently pronounced dead. She had obvious facial and skull injuries, including an open skull fracture at the back of her head. Upon autopsy extensive complex fractures of the skull were observed, along with subgaleal and subarachnoid hemorrhage, and complete dislocation of the base of the skull from the spine (atlanto-axial dissociation), with associated injury to the posterior proximal spinal cord just below the brainstem. These injuries would be appropriately categorized as at least critical (AIS 4) in severity. Also noted were multiple rib fractures, a liver laceration, and a fracture of the pubic symphysis (in the pelvis).

A reconstruction of the crash indicated an estimated range of time between the first and second impact of 3.1–5.1 s.

Although it occurred at a relatively high speed, the first collision would have been associated with a relatively low risk of a fatal head injury, as the decedent would have interacted primarily with her seat back, which failed (reclined) during the crash.

There was a reasonable question regarding the mechanism causing what appeared to be a puncture at the rear of the decedent's head, an injury that was associated with an open complex skull fracture on autopsy. On scene investigators opined that the injury

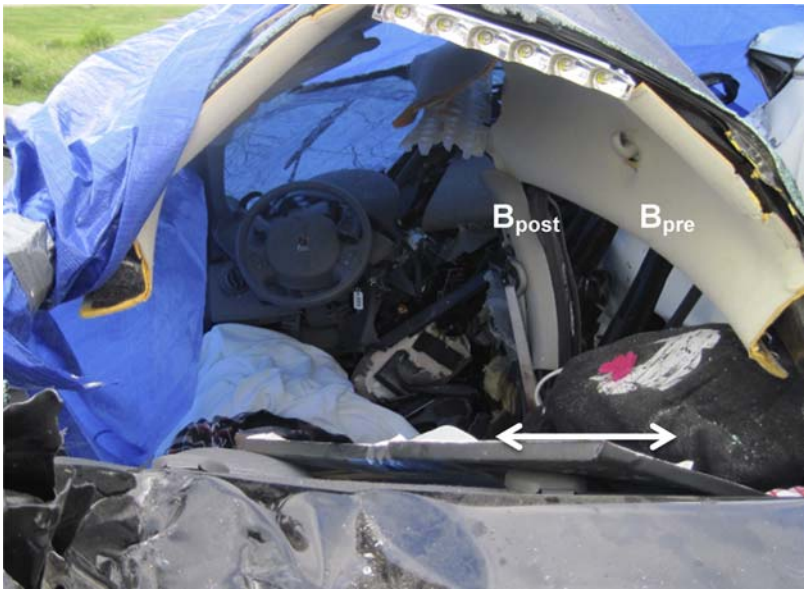


FIGURE 11.17 The view from the rear, demonstrating the left-ward displacement of the passenger side B-pillar (B_{post}) versus the approximately position of the B-pillar prior to the second collision with the van (B_{pre}).

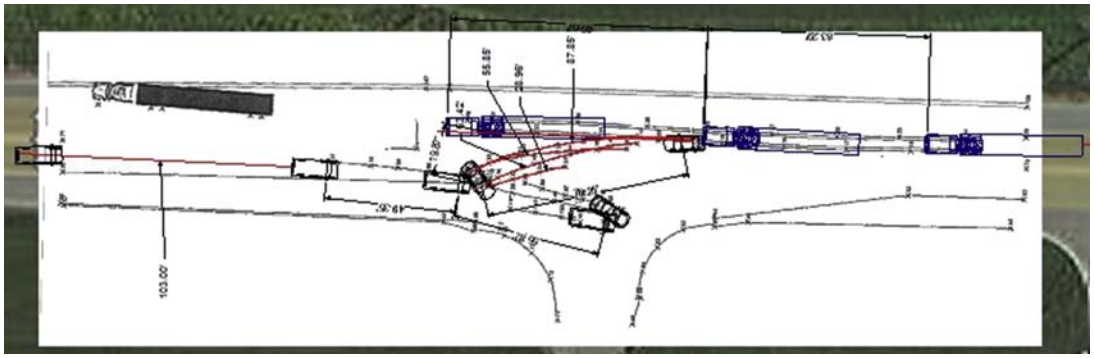


FIGURE 11.18 Diagram of the reconstructed travel paths of the three vehicles.

must have resulted from the occupant being struck by an object inside the vehicle during the rear impact. This explanation was unlikely, as the physics of a rear-impact collision causes all objects, including the occupant, to initially move rearward relative to the vehicle interior.

Further investigation of the injury and biomechanics of the two collisions indicated that head injury most likely occurred during the second collision. During this impact, as the passenger side B-pillar was crushed inward toward the driver, her shoulder belt would have provided no restraint to her corresponding movement to the right, as her seat back had already

collapsed backward during the first crash, moving her torso away from the belt. The blood on the B-pillar, and more particularly on the leading edge of the rear door gives certain evidence of contact with a hard and relatively pointed structure. Figs. 11.19 and 11.20 illustrate this finding.

Fig. 11.19 is for orientation; it shows the view of the interior passenger side from the left side of the vehicle after the doors and B-pillar have been removed. Fig. 11.20 is an enlargement of the area within the white box in Fig. 11.19.

Hard contact with the leading edge of the front door would have been the most likely cause of the skull injuries, including the atlanto-occipital dislocation, in which the skull was separated from the spine.

A population-based assessment of the fatal head injury risk of the two collisions was performed, in order to estimate the probability of cause attributable to the second versus the first collision. Data were accessed from the NASS-CDS database, described earlier in this chapter, for this part of the analysis.

The parameters of the case-specific search were as follows: First, an estimate of the rate of serious and greater (AIS 3+) and critical and greater (AIS 4+) head injuries in restrained front seat occupants of passenger vehicles exposed to rear impacts of 30–50 mph (48–80 km/h) delta V was determined. This range was bracketed around the approximately 40 mph (64 km/h) delta V resulting from the first collision. Next, this rate was compared to the rate at which the same severity of head injury was observed in restrained drivers exposed

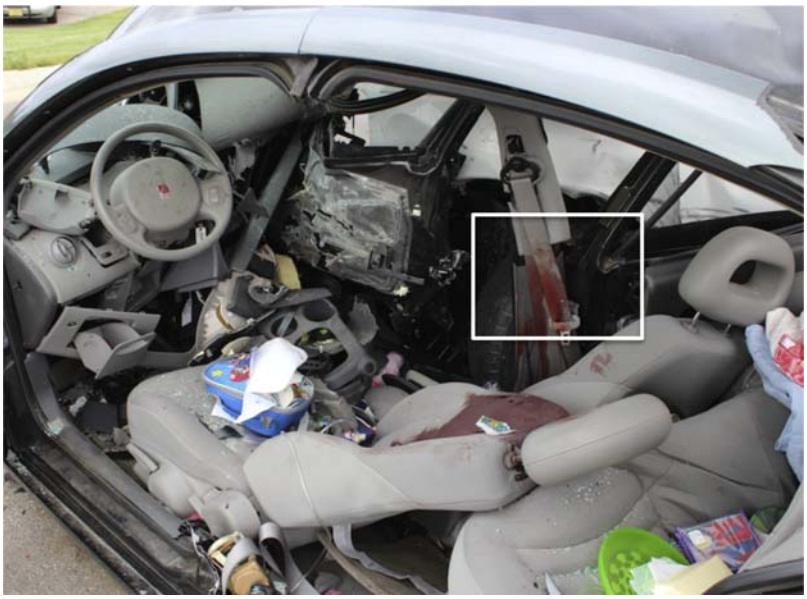


FIGURE 11.19 View of the decedent's vehicle from the driver's side, showing the point of head contact with the passenger side B-pillar (white square inset).

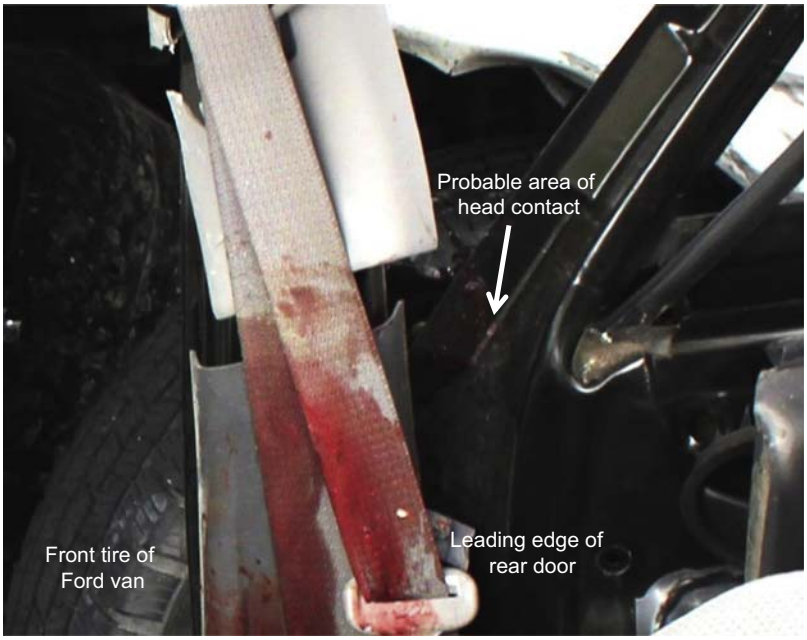


FIGURE 11.20 Enlarged view of the inset from Fig. 11.19.

to a right-side impact of 15–35 mph (bracketed around the 24 mph (38 km/h) delta V of the second crash) and in which there is also evidence of contact between the driver's head and a structure on the passenger side, including the door, B-pillar, window frame, etc. The years of the study included 1995–2011 (17 years inclusive).

The results of the analysis were as follows:

There were an estimated 34,804 front-seat occupants exposed to rear impacts of 30–50 mph delta V (48–80 km/h), and 3834 (11%) sustained serious and greater head injuries and 550 (1.6%) sustained critical and greater head injuries. In comparison, there were 8648 drivers exposed to right-side impacts where there was also any degree of head contact with a structure on the passenger side of the vehicle noted. Among these occupants, there were 2051 (23.7%) serious and greater injuries, and 397 (4.6%) critical and greater head injuries.

A CRR was calculated for these data, indicating a 2.2 times greater risk of serious and greater head injury for the side impact group (95% CI 2.1, 2.3), and a 2.9 times greater risk of a critical and greater severity head injury, also for the side impact occupants (95% CI 2.6, 3.4). Both CRRs would be associated with PCs of >50%.

As a result of the analysis it was opined that the most likely cause of the decedent's fatal head injuries was the second passenger side collision that resulted from the initial rear impact crash, and that it was likely that she survived for approximately 3.1–5.1 s after the first collision.

CONCLUSION

Causation in traffic crash–related injury litigation is a common source of dispute. Introduced in this chapter is a framework for evaluating causation in such cases, and suitable for the quantification of cause to meet the legal standard of what is more probable than not.

ENDNOTES

1. http://www.who.int/mediacentre/factsheets/fs358/en/, accessed 01.05.15.
2. http://www.law.harvard.edu/programs/olin_center/papers/pdf/Ramseyer_681.pdf, accessed 20.02.15.

References

Böstman, O., Varjonen, L., Vainionpää, S., Majola, A., Rokkanen, P., May 1989. Incidence of local complications after intramedullary nailing and after plate fixation of femoral shaft fractures. Journal of Trauma 29 (5), 639–645.

Crandall, C.S., Olson, L.M., Sklar, D.P., February 1, 2001. Mortality reduction with air bag and seat belt use in head-on passenger car collisions. American Journal of Epidemiology 153 (3), 219–224.

Chandra, A., Suliman, A., Angle, N., March 2007. Spontaneous dissection of the carotid and vertebral arteries: the 10-year UCSD experience. Annals of Vascular Surgery 21 (2), 178–185.

Freeman, M.D., Kohles, S.S., 2011. An evaluation of applied biomechanics as an adjunct to systematic specific causation in forensic medicine. Wiener Medizinische Wochenschrift 161, 1–11.

Freeman, M.D., Centeno, C.J., Kohles, S.S., 2009. A systematic approach to clinical determinations of causation in symptomatic spinal disc injury following motor vehicle crash trauma. PM R 1 (10), 951–956.

Freeman, M.D., 2015a. Biomechanical, mechanical, and epidemiologic characteristics of low speed rear impact collisions. In: Proceedings of 67th Annual Meeting of the American Academy of Forensic Sciences 2015 Feb 16–21: Orlando, FL, pp. 517–518. D11.

Freeman, M.D., 2015b. Medicolegal causation analysis of a lumbar spine fracture following a low speed rear impact traffic crash. Journal of Case Reports in Practice 3 (2), 23–29.

Freeman, M.D., Croft, A.C., Rossignol, A.M., Weaver, D.S., Reiser, M., 1999. A review and methodologic critique of the literature refuting whiplash syndrome. Spine (Philadelphia, Pa. 1976) 24 (1), 86–96.

Hauser, V., Zangger, P., Winter, Y., Oertel, W., Kesselring, J., 2010. Late sequelae of whiplash injury with dissection of cervical arteries. European Neurology 64 (4), 214–218.

Jacobs, Goff, Aeron-Thomas, Amy, Astrop, Angela, 2000. Estimating Global Road Fatalities. TRL.

Jordan, J.M., et al., 2009. Prevalence of hip symptoms and radiographic and symptomatic hip osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. Journal of Rheumatology 36 (4), 809–815.

Kleinberger, M., et al., 1998. Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems. NHTSA Technical Document.

McGwin Jr., G., Metzger, J., Porterfield, J.R., Moran, S.G., Rue 3rd, L.W., September 2003. Association between side air bags and risk of injury in motor vehicle collisions with near-side impact. Journal of Trauma 55 (3), 430–434.

Orsay, E.M., Dunne, M., Turnbull, T.L., Barrett, J.A., Langenberg, P., Orsay, C.P., March 1990. Prospective study of the effect of safety belts in motor vehicle crashes. Annals of Emergency Medicine 19 (3), 258–261.

Pettersson, K., Hildingsson, C., Toolanen, G., Fagerlund, M., Björnebrink, J., February 1, 1997. Disc pathology after whiplash injury. A prospective magnetic resonance imaging and clinical investigation. Spine (Philadelphia, Pa. 1976) 22 (3), 283–287.

Tencer, A.F., Kaufman, R., Ryan, K., Grossman, D.C., Henley, B.M., Mann, F., Mock, C., Rivara, F., Wang, S., Augenstein, J., Hoyt, D., Eastman, B., January 2002. Crash Injury Research and Engineering Network (CIREN). Femur fractures in relatively low speed frontal crashes: the possible role of muscle forces. Accident Analysis and Prevention 34 (1), 1–11.

Walz, F.H., Muser, M.H., 2000. Biomechanical assessment of soft tissue cervical spine disorders and expert opinion in low speed collisions. Accident Analysis and Prevention 32 (2), 161–165.

CHAPTER

12

Traffic Injury Investigation: Product Defects

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

OUTLINE

INTRODUCTION

In Chapter 11, *Traffic Injury Investigation*, the discussion and methods were focused on the cause of injuries resulting from an alleged act of negligence of a driver or occupant in traffic crashes. In some cases, however, the causation question relates to injuries resulting from a failure of a vehicle part that was due to an alleged defect in the manufacture of the vehicle or vehicle components. There are two causation questions that must be addressed in such cases. The first is whether the failure of the vehicle part was truly caused by a defect in

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00012-4

manufacturing. This question is most often addressed with engineering or other technical analyses, and often if the part does not meet a generally used industry or governmental standard, it may be deemed defective. In some situations the failure is due to the defect "per se"; this occurs when there are no alternative explanations for the failure associated with ordinary operation of the vehicle. An example is when an air bag deploys for no apparent reason, or an ignition switches off without being touched.

In other circumstances, the failure might be conditional. As an example, tires will fail if they have excessive wear after much use, or if they sustain significant damage, but a relatively unworn and undamaged tire should never fail. Thus, a tire failure in the absence of excessive wear or damage is much more likely to be due to a defect. It is important to keep in mind the fact that no vehicle component is impervious to failure if a crash is sufficiently severe, and thus a complete understanding of the conditions in which the failure occurred is important. As an example, a seat belt can fail during a collision by tearing. If the failure occurred in a 10 mph (16 km/h) delta V frontal collision, the suspicion that the belt was defective would be substantially greater than if the failure occurred in a 50 mph (80 km/h) delta V crash. In such circumstances, the assessment of whether the failure can be considered to be a result of a defect may require a population-based relative risk or odds ratio assessment, such that it can be concluded that there is an attributable fraction (AF) of failures that are purely due to a defective manufacturing process.

The second causation question concerns the injury, which in many cases can be assessed using a comparative risk ratio (CRR). How the CRR is calculated varies with the degree of specificity between the nature of the product failure and the injury. When characterizing the utility of causal "viewpoints," Hill described specificity as the degree of correlation between exposure and outcome, and used as an example deaths due to a particular disease that only occurred in certain workers, but not in others, to demonstrate when a high degree of specificity implied a causal relationship (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*). The concept of specificity has a unique role in the assessment of specific causation, but one that differs with how it is used in general causation (at least as Hill described it). As an obvious example, we can consider a vehicle with a defective gas tank that bursts into flames when it is struck from the rear at low speed. If an occupant were to perish from smoke inhalation after such an event, we would conclude, as readily apparent, that the death was highly *specific* to the fire, and the fire was highly *specific* to the defective gas tank. Had the fire not occurred the risk of death from the crash was exceedingly small, and the risk of death from smoke inhalation was zero. From a general causation perspective, however, most car fires are not due to a vehicle defect, and very few deaths are due to car fires. Thus, the concept of specificity in a general causal sense would have no utility and little meaning in the example, yet the specificity of the investigated relationships is the reason for the common sense conclusion that the death only resulted from the defect.

In contrast, the nature of traffic crashes is such that they are a dangerous event regardless of whether there is a known product defect. When there is a reasonable inference that an injury could have occurred in a crash with or without the contribution of a defective vehicle component, it can be said that there is a lack of an event-unique specificity between the component failure and the injury. In such a circumstance the only meaningful assessment of the probability of causation (a CRR), may be estimated using a relative risk analysis.

The following case study serves as an example of such an analysis:

CASE STUDY #1: AIRBAG FAILURE-RELATED COMPARATIVE DEATH RISK ANALYSIS

A 40-year-old female died as a result of blunt abdominal trauma following a single vehicle collision with a telephone pole, occurring at approximately 38 mph (61 km/h). The woman was unrestrained at the time of the collision, with a blood alcohol level of 201 mg/dL (versus a legal threshold of 80 mg/dL). Information stored in the air bag module (also called an event data recorder), was downloaded and demonstrated that no braking occurred prior to the impact, suggesting that the decedent was asleep at the time of the crash. The air bag in the vehicle did not deploy.

The photograph of the vehicle exterior indicates a high-speed frontal impact with the telephone pole, and the photograph of the vehicle interior demonstrates deformation of the steering wheel resulting from loading induced by the driver's chest and abdomen, as well as confirmation of the fact that the air bag failed to deploy (Figs. 12.1 and 12.2).

The decedent's estate made a claim against the vehicle manufacturer, asserting that the failure of the air bag to deploy was a contributing cause of her death. The manufacturer responded by acknowledging the defect, but also asserting that the legal standard that had to be met in order for the defect to be considered the cause of the death was "more likely

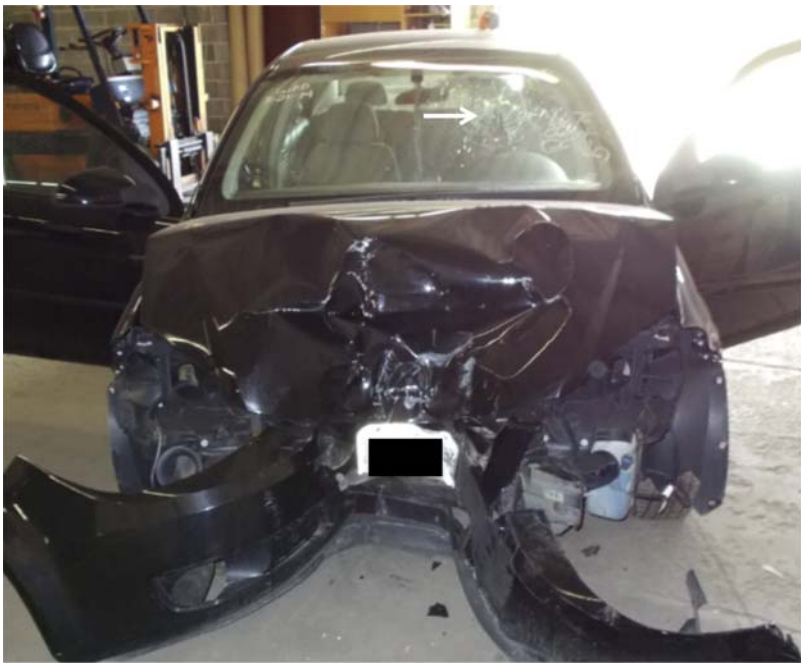


FIGURE 12.1 The front view of the Chevrolet, demonstrating the midline frontal configuration of the impact, as well as windshield damage consistent with a head or face impact (*white arrow*).

FIGURE 12.2 Interior view of the vehicle, showing the forward deformation of the steering wheel both above and below the air bag cover.

than not," or >50% probable. As described in the previous chapter and throughout this text, a CRR of >2.0 or a Probability of Causation of >50% typically satisfies the legal standard of what is more likely than not.

Comparative Risk Ratio Determination

The causation question for the investigation, in terms of CRR, is represented as follows:

$$\mathrm{CRR} = \frac{p(\text{death}|\text{airbag not deployed})}{p(\text{death}|\text{airbag deployed})}$$

As described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*, the CRR can be converted to an attributable fraction under the exposed, or probability of cause (PC), as follows:

$$\frac{(\mathrm{CRR}-1)}{\mathrm{CRR}} \times 100\% = \mathrm{AF_e} \approx \mathrm{PC}$$

In order to be applicable to the circumstances of the crash, however, the probabilities have to be conditioned by the crash-related features that are substantially predictive of fatal injury, and these predictive features have to be the same in both the numerator and the

denominator of the CRR, with the exception of the predictive factor under investigation (air bag nondeployment). The primary features of the collision that are predictive of fatal injury risk, independent of the air bag deployment status, are the type of vehicle, the severity of the collision in terms of crash-related speed change (delta V), the direction of the collision, the occupant's position as a driver, and the lack of seat belt use.

The manufacturer took the understandable position that the failure of the decedent to use a seat belt *may* have had a large influence on the fatality risk of the collision, but did not offer a quantitative appraisal of the relationship. As described in Chapter 11, *Traffic Injury Investigation*, seat belt nonuse can greatly increase injury for certain crash conditions, but has little to no influence in others.

From a biomechanical perspective, the primary function of the driver's side frontal air bag is to reduce the occupant force of an impact with the vehicle interior components in front of the driver (steering wheel, windshield, etc.) in the event of a frontal collision, typically of moderate or higher speed. The primary function of seat belts is to reduce the risk of ejection from a vehicle in the event of a crash. Secondarily, seat belts reduce occupant motion, and thus injury risk in certain types of crashes. In a frontal collision in which a frontal air bag has also deployed, much of the protection potentially afforded by the seat belt is already present from the air bag, and thus a seat belt has less effect in such collisions. The only way to accurately assess how the failure to use a seat belt in combination with an air bag nondeployment affected the death risk of the investigated collision is to examine a relevant and reliable data set. Because the purpose of the analysis is to examine the effect that a predictor variable had on the frequency of a specific outcome, the results of the analysis can be characterized in terms of a relative risk.

National Automotive Sampling System-Crashworthiness Data System Analysis

A case-specific analysis of data from the United States National Automotive Sampling System-Crashworthiness Data System (NASS-CDS) of the National Highway Traffic Safety Administration was undertaken. This database has been more thoroughly described in Chapter 11, *Traffic Injury Investigation*. Briefly, in order for a collision to be recorded in the NASS-CDS, it must meet several criteria, which includes that a police report was generated; that it was located within a primary sampling unit; it involved at least one passenger car, van, or light truck; and at least one vehicle was towed from the crash scene. In turn, these data are weighted to provide a national estimate of all police-reported crashes occurring in the United States that involved passenger vehicles.

The parameters of the search performed for the relative risk analysis were as follows: included were all unrestrained drivers of passenger vehicles that were exposed to a frontal crash of 32–42 mph (51–67 km/h) delta V (equally grouped around the actual delta V of 37 mph (61 km/h)), with no rollover and no ejection recorded, for the years 1990–2012 (the years available at the time of the analysis). The predictor variable of interest was frontal air bag deployment, and the outcome variable of interest was death. Only occupants with at least a minor injury (abrasions, strains) were included in the analysis to insure against the confusion of "uninjured" status with missing medical data. Given the severity of the collision, it is unlikely that excluding what were only possibly uninjured occupants would have been a source of bias for the analysis.

The results of the query and analysis were as follows: There were an estimated 11,763 unrestrained drivers with an air bag deployment and 19,885 without an air bag deployment. There were no significant differences between the groups with regard to type of passenger vehicle, average occupant age and gender distribution, or average delta V. Among the drivers with an air bag deployment, there were 492, or 4.2% of the drivers who were killed. Among the drivers with no air bag deployment, there were 2,720, or 13.7% of the drivers who were killed.

The CRR for the analysis was thus:

$$\mathrm{CRR} \approx \mathrm{RR} = \frac{p(\mathrm{death}|\mathrm{airbag\ not\ deployed})}{p(\mathrm{death}|\mathrm{airbag\ deployed})} = \frac{13.7\%}{4.2\%} = 3.27\ \ (95\%\ \mathrm{CI};\ \ 3.00,\ 3.59)$$

NB The confidence interval (CI) above was calculated based on the weighted sample, which was reasonable given the sample size and large risk ratio (RR). In some cases when using NASS-CDS data, it is necessary to incorporate the standard error associated with sampling methods.

The CRR was used to calculate a risk of death attributable to the nondeployment of the air bag as follows:

$$\mathrm{PC}_{\mathrm{non-deploy}} \approx \frac{(\mathrm{CRR}-1)}{\mathrm{CRR}} \times 100\% = \frac{(3.27-1)}{3.27} \times 100\% = 69.4\%$$

These results indicate a risk of death attributable to the nondeployment of an air bag for unbelted drivers of 69.4% of the total risk. The AF_e is approximately equal to the PC, and thus the air bag nondeployment PC specific to the circumstances of the decedent's death was also 69.4%. As a corollary, these findings also indicate that, on a more probable than not basis, the decedent would have survived the collision had the air bag deployed, despite the failure to use her seat belt.

Given the complexity of manufacture and operation, there are many components in a passenger vehicle that could fail prior to, or during, a crash. For this reason, prior to embarking on a forensic epidemiologic (FE) analysis of injury causation a multidisciplinary approach is required to understand the sequence and nature of events in a crash leading to an injury, and including application of principles based in crash reconstruction, injury biomechanics, and medicine.

In the following sections of this chapter are two case studies that illustrate the multidisciplinary nature of the FE investigation of causation associated with an alleged motor vehicle product defect and associated failure. The first case concerns the relationship between a serious spine injury and excessive roof crush following a rollover crash. The second case pertains to a fatal injury incurred during a rollover crash, in which the occupant was ejected from a vehicle in which the seat belt latch was known to be faulty.

CASE STUDY #2: ROOF CRUSH-RELATED NECK INJURY RISK ANALYSIS

A commonly alleged vehicle component failure associated with serious injury is excessive roof crush that can result from a rollover crash. In some vehicles, the failure results from an

inadequate or faulty design. The cause of a serious injury in a rollover crash can be difficult to determine, however, because, like the high-speed frontal crash described earlier in this chapter, rollover crashes are inherently dangerous events that can result in serious injury in the absence of any vehicle component failure or design defect.

In some cases with obvious fact patterns, there may be no need for an epidemiologic analysis of cause. If, for example, an occupant were to be killed in a vehicle with a known roof defect that rolled one revolution (landing on its wheels) and the roof was crushed to the door line (like the vehicle depicted in Fig. 12.3), any lay fact finder would understand the causal relationships with little expert assistance; it is readily apparent that the poor design caused the catastrophic failure roof, and that the catastrophic roof failure caused the death. One could say that the death was *specific* to the roof collapse, as it is difficult to understand how one could survive such a severe intrusion into the occupant compartment.

Compare the preceding scenario with the subject case study facts, which involve a serious spine injury in the front passenger seat occupant of a sport utility vehicle (SUV) that sustained two complete rolls. There was a maximum of 10 inches (25 cm) of vertical roof crush at the front seat passenger position (see Fig. 12.4). The occupant at that position was an unbelted (and unejected) 35-year-old male. He sustained a fracture-dislocation at the C5–C6 level, with an associated spinal cord injury. The occupant's injury was one that results from the axial compression and flexion of the cervical spine, which occurs in a rollover crash when the occupant is inverted and the weight of the torso loads the neck when the head contacts the roof. The load causes the spine to buckle, resulting in the fracture and dislocation, and often an associated spinal cord injury.

In contrast with the prior example of the sedan with the extreme roof crush, the fact scenario of the case study does not allow for a lay inference that the degree of roof crush was the most probable cause of the occupant's serious neck injury. Indeed, the knowledge that the occupant was unbelted during the rollover crash raises the question of whether

FIGURE 12.3 Example of vehicle with extensive roof crush following a rollover crash.

FIGURE 12.4 Passenger side view of Ford Explorer in rollover case study, showing moderate level of vertical intrusion from the roof crush into the passenger compartment.

the injury was in part or in whole due to his failure to use a seat belt. In providing an FE analysis of cause for this case, some background information on the controversies relating to rollover crashes is necessary.

Rollover Crashes

A rollover crash involves a vehicle that experiences at least two quarter turns ($\geq$180 degree) about its long axis. Although rollover crashes are less common than frontal, side, or rear impact collisions, they are associated with a higher rate of injury and fatality than any other crash type. The issue that has generated the most controversy in the literature is the association between roof crush (defined as vertical and lateral intrusion of the vehicle roof/ceiling into the occupant compartment) and how it may or may not relate to the risk of serious head and spine injury. The source of the controversy is largely due to the fact that the auto manufacturing industry has sponsored research and publications that indicate that roof crush is not related to head and neck injury, whereas nonindustry researchers, some of whom serve as consultants for plaintiffs who claim that a faulty roof design was the cause of their head and/or neck injury in a rollover crash, assert the opposite viewpoint.

Since 1975 automotive industry researchers have postulated a "diving theory" injury mechanism, in which injury to the head and neck is thought to result from the occupant moving toward the vehicle roof during a rollover, while the roof, which is in contact with the ground, temporarily remains stationary relative to the inverted occupant (Moffatt,

1975). The implication of the diving theory is that the degree of roof crush is unrelated to injury risk in rollover crashes, and thus a lack of roof strength is not a potential causal factor in the risk of serious head and neck injury in rollover crashes. In the competing explanation to the diving theory, called the "intrusion hypothesis," it is maintained that during roof-to-ground contact in a rollover that produces roof crush, the vehicle roof is momentarily stationary against the ground, while the rest of the vehicle continues to move downward, thus reducing occupant headroom and increasing head and neck injury risk. Implicit in the intrusion hypothesis is the tenet that increasing the strength of the roof, and thus its ability to resist intrusion into the occupant compartment is a key to reducing occupant head and neck injury risk (Friedman and Nash, 2001). A number of engineering and epidemiologic studies provide support the intrusion hypothesis versus the diving hypothesis; a 2009 study by the Insurance Institute for Highway Safety described roof crush testing on 11 midsize SUVs, resulting in roof crush of up to 10 inches (25 cm) (Brumbelow and Teoh, 2009). The authors also gathered and analyzed data from police-reported rollover crashes in 14 US states. Logistic regression was used to evaluate the association between occupant injury and roof crush, as well as other variables. The authors reported that a one-unit increase in the roof strength-to-weight ratio (ie, a stronger roof) was associated with a 24% (95% CI; 15, 33) reduction in the risk of fatal or incapacitating injury. Other authors have arrived at similar findings in support of the intrusion hypothesis using epidemiologic methods to evaluate crash injury data recorded in the NASS-CDS (Hu et al., 2007; Mandell et al., 2010; Dobbertin et al., 2013). Generally, the level at which roof crush is considered to be associated with excess injury risk is when it exceeds 6 inches (15 cm).

When designing an epidemiologic study of injury risk in rollover crashes for an FE analysis of causation, it is critical to keep in mind the variables that are predictive of injury, both generally in rollover crashes, and specific to the circumstances of the investigated crash, such that they can be controlled for in the study design. One of these important variables is occupant position relative to the rotation direction of the rollover. When a vehicle rolls toward its left (driver's) side, for example, the driver is termed the "leading" or "near-side" occupant, whereas the passenger (front-right seat occupant) is termed the "following" or "far-side" occupant. In a driver's side leading roll the passenger or far-side occupant has a significantly higher risk of injury and fatality, and vice-versa, likely due to the greater angular acceleration exerted on the occupant furthest from the momentary axis of rotation. Another important predictor of injury risk is the number of vehicle rolls, which makes intuitive sense; the more times a vehicle rolls, the more opportunity there is for occupants to make contact with vehicle components (roof, windshield, A pillar, B pillar, etc.), objects outside the vehicle, or to be ejected from the vehicle, particularly if they are not properly restrained. Additionally, vehicles that experience a greater number of rotations are likely traveling at a higher rate of speed, and thus there is an inherent increased potential for injury, irrespective of other roll characteristics.

Case-Specific Investigation

The analysis described here is directed at addressing the causal question of greatest interest for the case study; for an unbelted occupant who sustains a serious cervical spine injury in a rollover crash with ~10 inches of crush at his position (25 cm), what effect would

reducing the amount of crush to less than 3–6 inches (8–15 cm) have? The results ultimately have to be characterized in terms of a CRR or AF_e/PC such that they can be applied to the facts in the case example. Because there are multiple factors associated with injury risk in a rollover crash, the analysis for the case study was more complex than in the prior example of the air bag analysis.

Analysis

The ensuing analysis employed two different epidemiologic study methods that were applied to NASS-CDS data relating to rollover crashes and injuries to the head and neck. The first study was a general head and neck injury risk study, examining injury severity to this area of the body as a function of roof crush, while controlling for other important variables that might also dictate injury risk. Because this study employed a logistic regression analysis, the results were characterized in terms of odds ratios. The second analysis was a case–control study that was "nested" within the first study, in which occupants with serious neck injuries were identified as cases, and matched to a control without a serious neck injury for number of rolls and vehicle roof strength. The outcome of interest was the odds of roof crush exceeding 6 inches (15 cm).

Analysis 1

The first step in the analysis involved the identification of inclusion and exclusion criteria, including specification of the type of vehicle, age of the occupant, position of the occupant within the vehicle, and others. The methods for the analysis are more completely described in (Dobbertin et al., 2013). Rollovers that involved multiple vehicles, major vehicle fires, immersions, and end-over-end rollovers were excluded and only front-left and front-right occupants were included in the analysis. No center or second-row occupants were included although the results of the analysis should be applicable as a CRR for any outboard occupant, with a known amount of roof crush at their seating position.

Occupant ejection from a vehicle is a substantial source of injury in many rollover crashes, and thus controlling for this variable was important. Ejected occupants were not excluded from the study population, however, but controlled for by only including injuries that were determined to have resulted from contact with a structure that was over the occupant's seating position (roof, roof rail, header, etc.) and which could thus be affected by roof crush. The degree of roof crush that had occurred at the occupant's seating position had to be known, along with injury presence, location, type, severity, and source for all included occupants. As a convention, occupants in whom there was a head or neck injury attributable to contact with an overhead component were called "Type M" occupants, and those who did not sustain head or neck injury were called "Type O" occupants.

The results were as follows: of the 3088 vehicle occupants abstracted according to the inclusion and exclusion criteria listed previously, there are 1118 Type M (injured) and 1970 uninjured (type O). After elimination of injured occupants with missing information, there were 960 Type M occupants remaining. These occupants were used for an exposure study, in which the degree of roof crush was categorized, and then compared to head and neck injury presence and severity.

For this first analysis a single injury metric was used to describe the severity of injury to both the head and neck. The result was a modification of the New Injury Severity Score

(NISS) that only included injuries to the neck and head, and called the head-neck NISS (HN-NISS).

The analysis employed a generalized estimating equation logistic regression approach to determine the odds of HN-NISS $\geq$9 (the cut point at which injury was considered "serious"). This statistical model allows for the evaluation of the relative strength of the injury prediction variables that were considered for the model. First, the variables were examined in a univariable model in which the strength of the variable to predict injury was evaluated without controlling for the effect of other variables. The strongest and most statistically valid variables were then put into a multivariable model, which allowed for the evaluation of how the variables interact. The weakest variables were taken out and put back into the model in a "stepwise" fashion to see which ones had the greatest effect on the model, and then the final model of the strongest and most valid predictors of injury were arrived at.

After following the previously described process of variable elimination, the final model for predicting serious head/neck injury included roof crush, air bag deployment, seat belt use, and age. Although seat belt use was not statistically significant in the model ($p = 0.056$), it was kept in the final model because of the high odds ratio (OR) associated with the variable (1.90). The results of the model that were applicable to the subject case indicated a 3.5 times greater risk of serious head/neck injury for an occupant exposed to a rollover crash with roof crush of 25 cm (10 inches) versus less than 15 cm (6 inches) of roof crush (95% CI; 1.1, 11.0).

Analysis 2

The second analysis performed for the case was a nested case–control study, in which 155 serious neck-injured cases were "nested" among the Type M occupants, and compared to 155 uninjured controls "nested" among the Type O occupants. The criteria for matching the controls to the cases were that they had to be front seat occupants in the same vehicle, so that the severity of the crash, with regard to number of rolls and type of vehicle, was the same for both. The analytic technique used to model the odds of being an injured case versus an uninjured control was a conditional fixed-effects logistic regression. The advantage of the analysis was that variables that were vehicle specific and thus identical for both occupants could be omitted from the model.

The Abbreviated Injury Scale (AIS) ranking of serious (3+) neck injury served as the dependent variable in the analysis, and the proxy for the injury sustained by the subject in the investigated case, as cervical spinal cord injuries are deemed serious and greater in the AIS. The primary predictor variable in the study was the maximum amount of roof crush at the occupant's seating position, divided into two categories: $<$6 inches or $\geq$6 inches (15 cm). Other potential predictor variables were assessed first in a univariable analysis and then examined in the multivariable model using a stepwise approach, as with the previously described study.

A final multivariable model was constructed for serious neck injury using roof crush, roll arc side, and occupant height. The largest single predictor of serious neck injury was roof crush of 6 inches or more (15 cm), at an odds ratio of 8.2 (95% CI; 2.1, 32.0). Occupying the side of the vehicle that was following, rather than leading the roll, was also a source of increased serious neck injury risk (OR = 4.0 (95% CI; 1.4, 12.1)), and increased height (for every additional 1 cm of height there was a serious neck injury OR of 1.08 (95% CI; 1.03, 1.14)). The latter two variables were not considered of importance for the question of interest

in the investigated case, as the counterfactual causation scenario only involved the theoretical reduction of roof crush and not the alteration of the roll arc side or height of the subject.

Causation Conclusion

The results of the two studies described in the previous section can be used for a CRR estimate applicable for the circumstances of the investigated case, and as an indication that the risk of serious head and neck injury was 3.5–8.2 times greater given the observed roof crush of 10 inches (25 cm) in the crashed vehicle versus the alternative scenario of less than 6 inches (15 cm) of roof crush. Further, the results of the analysis indicated that the subject's failure to use the available seat belt was not causally related to his serious neck injury.

A CRR of 3.5–8.2 equates to an AF_e of 71–88%, meaning that 71–88% of the cause of the serious neck injury observed in the subject occupant can be attributed to the observed excessive (>6 inches (15 cm)) roof crush. (See the conversion calculations based on the two CRR estimates below;

$$PC_{\text{roof crush}>15\text{ cm}} \approx \frac{(CRR-1)}{CRR} \times 100\% = \frac{(3.5-1)}{3.5} \times 100\% = 71\%$$

$$PC_{\text{roof crush}>15\text{ cm}} \approx \frac{(CRR-1)}{CRR} \times 100\% = \frac{(8.2-1)}{8.2} \times 100\% = 88\%$$

CASE STUDY #3: SEAT BELT LATCH FAILURE-RELATED INJURY PATTERN RISK ANALYSIS

In the prior case study example, the disputed causal question of interest was "how much of the cause of the subject's serious neck injury would be eliminated if the roof crush had been less," and the answer was assessed using a counterfactual approach and quantified by the epidemiologic studies performed ad hoc for the investigation. In the following case study of an alleged seat belt latch failure, there was no dispute about an alternative or counterfactual scenario. The case involved a 16-year-old female driver of a 1999 Kia sedan who suffered fatal injuries following a rollover crash and subsequent ejection. Because of the ejection, the on-scene investigators concluded that the decedent had not been wearing her seat belt. Upon autopsy, however, the forensic pathologist noted a classic shoulder belt abrasion over the left shoulder, and as a result concluded that the decedent had indeed been using her seat belt at the time of the rollover crash, which was consistent with her history of customary use provided by her family (see Figs. 12.5 and 12.6 demonstrating the injury from autopsy photo and schematic of the orientation of the abrasion in the distribution of a driver's side shoulder belt).

Testing of the seat belt latching mechanism indicated that it was prone to a false latching scenario in which the latch plate makes a "clicking" sound when it is inserted into the receiver but does not completely latch, and thus can be unlatched with a relatively low load.

During the investigation, the question was raised as to whether the shoulder abrasion was a true seat belt abrasion, or whether it could also be explained by some other contact during

FIGURE 12.5 View of the left side of the head, neck, and left shoulder of the decedent demonstrating the conforming seat belt webbing-shaped abrasion.

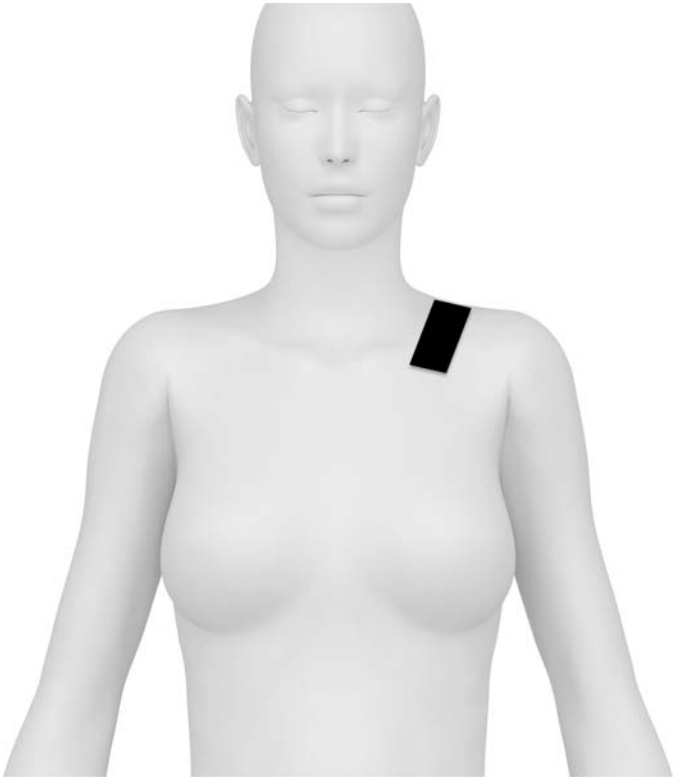


FIGURE 12.6 Schematic illustration of the location of the left shoulder abrasion.

the rollover or ejection. This alternative explanation was raised because of the lack of injury to the decedent's left arm. The auto manufacturer had retained several engineering experts who asserted the theory that when an occupant is ejected due to a seat belt latch failure that the retracting belt will cinch around the upper extremity closest to the shoulder belt anchor (the "outboard" upper extremity or OUE), invariably resulting in substantial injury, including amputation. The theory was based on several experimental studies using crash test dummies and several published case studies. The experts offered no epidemiologic data to substantiate their assertion of a 100% correlation between seat belt latch failure and serious upper extremity injury, but nonetheless concluded that the lack of the injury in the decedent corresponded to *certain* evidence that she was not using her seat belt at the time of the crash. Thus, despite the known defect in the latching mechanism, the defendant's experts asserted that this was not the cause of the ejection and associated death.

In order to test the assertion made by the defendant's experts, a unique analysis of the NASS-CDS database was undertaken. The question of interest was the reliability of serious OUE injury as an indication that an ejected occupant was using a seat belt that became unlatched at the time of the ejection. Seventeen years of data (1996–2011) were accessed for adult front seat occupants seated in the outboard position (driver or right-front passenger), for whom a 3-point manual seat belt was available, who were coded as completely or partially ejected, and for whom belt use status was known, including primarily cases in which there was a belt failure or no belt was used. Although injuries to all parts of the body were included in the analysis, the injury pattern of greatest interest was to the OUE, with an AIS injury severity rank of moderate or greater (2+). This categorization would capture more significant lacerations and most fractures of the upper extremity, along with the most serious injuries (ie, amputations and degloving) referred to by the defendant's experts. Confounding variables considered in the analysis were sex, age, body mass index (BMI), ejection status (complete or incomplete), and whether the collision involved a rollover.

A univariate analysis was performed as a means of identifying the predictor variables with greatest association to injury presence, and then summary statistics were tabulated for the two categories of seat belt status (belt failure and nonuse). Crude associations were tested using chi-square tests for categorical variables, and *t*-tests for continuous variables. Once the predictor variables were identified, logistic regression models were used to further investigate the confounder-adjusted associations between seat belt status and injury risk. As a final step, a matrix was constructed of posttest probabilities for seat belt status based on injury presence or absence and the precrash probability of seat belt use. In essence, the analysis treated the presence or absence of OUE injury as a test for preejection seat belt use, so that true positive rates (ie, sensitivity) and false positive rates (ie, (1-specificity)) could be used to construct a posttest probability matrix for a given pretest probability of seat belt use. The methods underlying this approach are more thoroughly described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* under Bayesian reasoning and analysis.

The results of the NASS-CDS analysis were as follows: there were a weighted total of 232,931 ejected occupants included in the analysis. Of these, 497 (0.21%) were coded as seat belt failures, and 232,434 (99.79%) were coded as not using their seat belt. There was no difference between the belt status groups with respect to age, sex, BMI, rollover status, mean maximum AIS, and the frequency of OUE, head, chest, abdominal, lower extremity, or upper extremity injury (there were no face or neck injuries among those with belt failure).

The only statistically significant difference was in the percentage of spinal injuries, with injuries observed in 1.2% of those with belt failure and 6.6% of those without a belt ($p = 0.05$).

Logistic models using injury status as the predictor and belt status as the outcome were used to identify significant adjusted associations. Models were adjusted for age, sex, BMI, and rollover and ejection status. Of the seven injury areas/types examined, only OUE and head injury were found to have a significant adjusted association (OR = 3.87 (95% CI; 1.2, 13.0) and 3.1 (95% CI; 1.0, 9.7)).

Finally, posttest probabilities of belt use were calculated for both the presence and absence of OUE and head injuries (based on the fact that both injuries were associated with belt failure in the adjusted model), based on the sensitivity and specificity values derived from the frequencies observed in the data set per the methods outlined in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*. The resulting posttest probability matrix is below:

	Outboard upper extremity		Head	
Sensitivity	0.42		0.49	
Specificity	0.85		0.78	
Precrash probability of belt use	**Injury present**	**Injury not present**	**Injury present**	**Injury not present**
0.9	0.96	0.86	0.95	0.86
0.8	0.92	0.73	0.90	0.72
0.7	0.87	0.61	0.84	0.61
0.6	0.81	0.51	0.77	0.50
0.5	0.73	0.41	0.69	0.40
0.4	0.65	0.31	0.60	0.30
0.3	0.54	0.23	0.49	0.22
0.2	0.41	0.15	0.36	0.14
0.1	0.23	0.07	0.20	0.07

This table provides the posttest probability of belt use given the presence or absence of an OUE or head injury, based on a range of assumed precrash probabilities of belt use. The sensitivity and specificity values were used in the following equation to transform the precrash probability that the occupant was using a seat belt into a postcrash probability that a belt was used, given the presence or absence of a head injury:

$$\text{Posttest probability} = \left[\frac{\text{pretest probability} \times \text{sensitivity}}{[\text{pretest probability} \times \text{sensitivity}] + [(1 - \text{pretest probability}) \times (1 - \text{specificity})]}\right]$$

The values in the table allowed for an assessment of the accuracy of the claim made by the defendant's engineers that the lack of a serious upper extremity injury observed in the decedent was a reliable basis for concluding that she was unbelted at the time of the crash. Given the fact that there was no direct contemporaneous evidence of belt use from any eyewitness, the precrash probability of belt use can be assumed to be indifferent (0.5). Thus going to the table, and finding the row with a pretest probability of 0.5, and the column under OUE with no injury present, we find that the postcrash probability of belt use given no OUE injury is 0.41. Alternately, if we were to use the additional information provided by the decedent's family that her practice was to always wear her seat belt, we can then use the 0.9 pretest probability to arrive at a 0.86 postcrash probability of belt use.

Conversely, if an OUE injury had been present in the decedent, the posttest probability of belt use would be increased from 0.5 to 0.73, and from 0.9 to 0.96.

Based on these results, we must conclude that the presence or absence of serious OUE injury, while predictive of belt use when an occupant has been ejected and belt failure is suspected, is not in and of itself a reliable basis for a determination of whether or not a seat belt was used. In the investigated case the common sense conclusion that the observed abrasion over the decedent's left shoulder was most probably caused by interaction with a shoulder belt, in combination with the knowledge that the seat belt latching mechanism was faulty, provided the strongest evidence of the cause of the ejection and ultimately the death of the decedent.

CONCLUSION

Most allegations of defective automotive products hinge on the ability to demonstrate that the defect was causally related to the injury outcome. An understanding of how to conduct an evidence-based approach to such analyses, using FE methodology and principles and valid data sources, is an important part of the assessment of any automotive product defect action.

References

Brumbelow, M.L., Teoh, E.R., 2009. Roof Strength and Injury Risk in Rollover Crashes of Passenger Cars and SUVs. Paper No. 09-0502. Insurance Institute for Highway Safety.

Dobbertin, K.M., Freeman, M.D., Lambert, W.E., Lasarev, M.R., Kohles, S.S., 2013. The relationship between vehicle roof crush and head, neck and spine injury in rollover crashes. Accident Analysis and Prevention 58, 46–52.

Friedman, D., Nash, C.E., 2001. Advanced roof design for rollover protection. In: 17th International Tehcnical Conference on the Enhanced Safety of Vehicles.

Hu, J., Chou, C.C., King, A.L., Yang, K.H., 2007. A weighted logistic regression analysis for predicting the odds of head/face and neck injuries during rollover crashes. In: Annual Proceedings Association for the Advancement of Automotive Engineering, vol. 51, pp. 363–379.

Moffatt, E., 1975. Occupant motion in rollover collisions. In: Proc. 19th Conference of the American Association of Automotive Medicine, pp. 49–59.

Mandell, S.P., Kaufman, R., Mack, C.D., Bulger, E.M., 2010. Mortality and injury patterns associated with roof crush in rollover crashes. Accident Analysis and Prevention 42, 1326–1331.

CHAPTER

13

Product Defect/Liability Investigation

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

F. Franklin

Oregon Health & Science University-Portland State University, School of Public Health, Portland, OR, United States; Morehouse School of Medicine, Atlanta, GA, United States; Thomas R. Kline School of Law, Drexel Universty, Philadelphia, PA, United States

OUTLINE

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00013-6

INTRODUCTION

Our lives are surrounded with man-made products designed for consumption and use. We encounter such products at home and work, during transportation, when undergoing medical care, and when recreating. While the mass majority of consumer products work safely and as they were designed, some products clearly do not. Some defects are self-evidently dangerous, and there is no need to surveil for increased injury frequency to determine that the product is unsafe; for example, a toy that is intended for use by infants that is made with small parts that can become detached and thus a choking hazard. Other defects are more difficult to detect, and they are not identified until a sufficient number of excess failures and associated injuries are apparent to indicate an association with a suspected defect. Medical device defects tend to fall into this category. As an example, inferior vena cava filters, which are designed to prevent blood clots from traveling from the legs and into the lungs, can dislodge from their implantation site in the vena cava or break even when they are not defective, sometimes with medically disastrous results. It is only after a sufficiently excessive number of patients have been injured by the failure of a specific filter that an alleged defect in manufacturing may be suspected.

Virtually any consumer product can cause injury if it is defective. While various national and international manufacturing standards and regulations are put in place and typically followed to ensure a minimal level of product safety, some consumer product defects are only investigated, publicized, and eventually eliminated through litigation. In this manner, the process of litigation causes the courts to function as a market-based consumer product regulatory agency. The assessment of excessive injury risk resulting from a defective product, via a relevant comparative risk ratio (CRR) assessment, can provide critical evidence in such an investigation.

Injuries resulting from the use of commercially available products are most commonly investigated from a mechanistic perspective, which focuses on the physical characteristics of the alleged defect. This approach is quite reasonable when there is a high degree of association between the defect and the injury. In some cases, however, the association between the injury and the alleged defect is more difficult to quantify. This is particularly true when the product is used in an inherently dangerous activity that can cause serious injury even when no defect is present. An example is a bicycle helmet; the product is only used for the purpose for which it was designed when the rider has fallen from the bicycle, an event that can cause injury regardless of helmet use or efficacy of a helmet at preventing injury. A bicycle helmet that was manufactured with a defective strap, such that it was prone to come off during a fall from the bike, could be causally related to a head injury observed in a cyclist, if it were found that the helmet had come off during a fall from the bike. If, however, the fall resulted from a 60 mph (97 km/h) impact from an automobile, then it is obvious that the chance that the observed injury would have been prevented, on a more probable than not (>50%) basis, is quite low. In contrast, if the fall occurred at a walking speed then it is perfectly reasonable to question whether the injury may have been prevented if the helmet strap had not failed. In such an instance a counterfactual causation analysis is appropriate, and a CRR approach may be used to arrive at an attributable proportion in the exposed (AP_e) as an estimate of the probability that the helmet strap failure caused the victim's head injury (see Chapter 4, *Causation in Epidemiology and Law* for more information on these methods).

FIGURE 13.1 Exemplar multipassenger water tube. *Photograph accessed from http://www.waterskimag.com/files/2011/05/chreg_1.jpg.*

In product defect litigation, there is often no need to demonstrate the degree of association between the defect and the injury because there is such a highly specific relationship between the defect and the injury. As an illustrative example, we can use an inflatable water tube, designed to be ridden on while it is towed behind a ski boat (Fig. 13.1).

If there were an alleged defect that could cause the water tube to explode during ordinary use, then any injury that occurred as a result of the explosion would be causally attributed to the defect. No technical expertise is required to understand that the product should not explode if it is not defective, and the only investigation required in order to demonstrate the relationship between the alleged defect and the injury would pertain to the characteristics of the defect that led to the explosion (ie, design and manufacturing processes).

The nature of the injury might also be such that it is highly specific for the alleged manufacturing defect. For example, if one of the occupant handles at the front of the water tube was improperly designed so that it could fracture and leave a jagged surface, and an occupant sustained a laceration on a handle that had broken with ordinary use, then there would be little dispute as to the cause of the injury. This is because such an injury would be unique to the defect in the circumstances; a laceration is not an injury type ordinarily associated with the sport of water tubing, and if the defective handle had not broken, the injury could not have occurred. As was the case with the example of the exploding water tube, no particular expertise is required to assess the cause of the injury, as opposed to the manufacturing process that resulted in the defective product.

In some instances, the injury might result from an alleged defect that was thought to have increased the frequency of injury. With the example of the water tube, an injury may have resulted from the ejection of an occupant when the tube overturned. The postulated design defect in the water tube might be related to an allegedly increased propensity to overturn.

Unlike the prior examples of exploding tubes and fracturing handles, however, nondefective water tubes also overturn during both ordinary use and misuse. It would be incorrect to conclude, therefore, that absent the alleged defect that the water tube would not have overturned, and thus a CRR approach becomes necessary to quantify the probability that the defect caused the investigated injury. If it were determined that the water tube with the alleged defect overturned five times per hour of use, and other comparable water tubes overturn only one time per hour of use, then the resulting CRR of five would indicate a probability of causation (PC) of 80%, as

$$\mathrm{PC} = \frac{\mathrm{CRR} - 1}{\mathrm{CRR}} \times 100\% = \frac{(5-1)}{5} \times 100\% = 80\%.$$

If confounding and bias could be excluded as explanations for the excess overturn rate, *and* it was concluded that the only cause of the injury was the ejection, then the PC of 80% could be compared to the 50% "reasonable degree of probability" threshold to conclude that the most probable cause of the ejection and injury was the defect.

This last example using the hypothetical water tube defect illustrates the requirements of a valid causation analysis when an investigated defect is not exclusively associated with an injury mechanism or specific injury, as with the explosion and broken handle examples given earlier. In such a case, without a valid CRR analysis it can be difficult, if not impossible, to demonstrate a causal association between a well-established defect and an injury. If there were no data that demonstrated that the frequency of ejections were any greater in the defective water tube than in nondefective water tubes, then it would not matter how thoroughly the defect were described, or how well it was established that the defect was present in the investigated case. To return to basic principles described earlier in this book, unless it could be established that the defect accounted for more than 50% of the ejections and/or injuries in the exposed population, then the legal burden for establishing a causal link could not be met in most cases. Without adequate data even a lesser evidentiary standard, whether the defect was a "contributing factor to increasing the risk of the injury," could not be examined and quantified.

In the following sections of this chapter are two exemplar case studies of a forensic epidemiology (FE) analysis of causation of fatal asphyxial injuries suffered by infants in association with two different products. In the first case, the infant died while in an infant sleep positioner (ISP), and the investigation of CRR was directed at the incidence of asphyxial death in the sleep positioner versus other more common settings in which an infant sleeps.

In the second case the infant was strangled by interaction with a window blind cord, a well-established hazard. The manufacturer of the window blind retained an expert in statistics to provide a comparative analysis of fatal injury risk associated with a variety of household products that infants routinely come in contact with. The analysis allowed to the defending manufacturer to assert that the risk of death associated with window blinds was substantially less than with other household products that infants routinely come into contact with. In essence, the defense used the information from the expert to assert that a risk comparison so disfavored the risk of death from interaction with a window blind cord versus other much more common causes of death in the home that the product could not be considered relatively dangerous.

CASE STUDY #1: INFANT SLEEP POSITIONER DEATH INVESTIGATION

The case concerns the 2010 death of a 4-month-old female infant who was placed, for the night, in an ISP called a Nap Nanny Generation II. Over the course of the night, the harness of the device failed to secure the infant, and she was able to twist her body such that her head was hanging over the edge out of the device. The child was discovered in the early morning (3:00 am) with no signs of life. Immediate and continued attempts at resuscitation were unsuccessful and she was pronounced dead at the hospital. The cause of death was determined to be positional asphyxia.

The following is the account of the FE investigation of this death:

Epidemiology of Infant Suffocation Associated with Sleep Environment

The leading cause of injury deaths among infants less than 1 year of age is suffocation.[1] Over the past three decades there has been a fourfold increase in injury-related suffocation and strangulation in beds, a phenomenon that has been linked to an unsafe sleep environment.[2] Seventy percent of the suffocations among infants were attributable to mechanical asphyxia compared to respiratory obstruction.[3] The suffocation rate among infants less than 1 year of age is 10 times higher than among children 1−4 years of age, largely because of an inability to protect their airway.[3] Research that has examined the risk factors associated with sudden or unexpected infant deaths that occur in bassinets found anoxia or asphyxiation as the recorded cause of death in 85% of the victims.[4]

In the past decade, suffocation deaths among infants have been associated with the use of ISPs. The ISP devices are a response to the American Academy of Pediatrics' recommendation that infants sleep in a supine position. This recommendation was associated with the "Back to Sleep" campaign that was started to address the identification of the prone sleeping position as a high-risk factor for sudden infant death syndrome.[5] The ISP device is designed with the intention of keeping infants on their back while sleeping.

Notwithstanding the intent of the ISP, there have been nine deaths associated with the products reported between January 1, 1997, and August 20, 2009, prompting attention and concern from the US Consumer Product Safety Commission (CPSC).[6] An investigation of ISPs and their connection to infant asphyxiation concluded that the safest sleep environment for an infant is in a crib, supine, and without soft objects, loose bedding, or an ISP present.[2] By 2010 the FDA had issued an alert to the general public, indicating that there had been 12 deaths and dozens of reports of infants twisting into a hazardous position in ISPs (Fig. 13.2).

The Nap Nanny Infant Sleep Positioner

Three Nap Nanny ISP products were introduced to the market sequentially between 2009 and 2012: the Generation I, Generation II, and the Chill. There were approximately 5000 Gen I models sold during 2009, and no deaths associated with this product, prior to its recall.

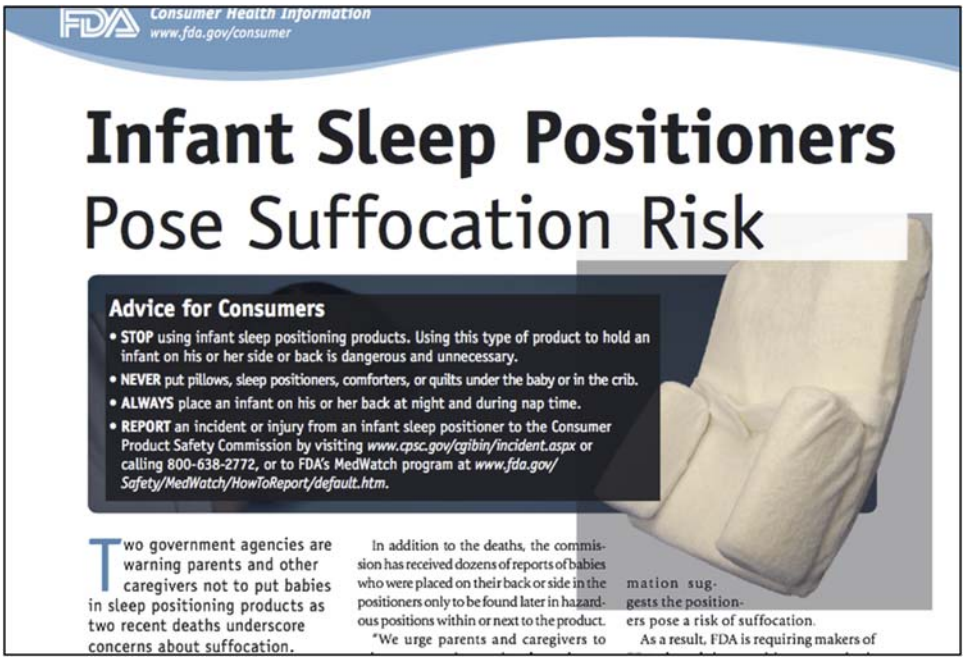


FIGURE 13.2 2010 FDA alert regarding infant sleep positioners. *Image accessed from http://www.fda.gov/downloads/ForConsumers/ConsumerUpdates/UCM227719.pdf.*

There were approximately 45,000 Gen II models sold between the end of 2009 and 2012 before the product was recalled by retailers and after the CPSC determined that the Nap Nanny was designed with defects that increased the likelihood of injury to infants. Specifically, the CPSC found that the Gen II device's 3-point harness failed to properly secure or restrain infants, which could allow the infant to turn sideways or twist out over the side of the product. The risk of serious injury, in particular asphyxiation/anoxic injury, was increased if the device was placed in a crib and the harness failed to prevent the infant from twisting out over the side of the device and getting wedged or otherwise trapped between the Nap Nanny and side rail or other structure inside the crib. Aside from the investigated death occurring on July 9, 2010, there were three other asphyxia/anoxia deaths associated with the Gen II Nap Nanny; on April 17, 2010, November 21, 2011, and April 12, 2012.

The third generation of the Nap Nanny that was introduced to the market was called the "Chill." The Chill Nap Nanny contained a number of modifications all designed to make it safer than the Gen II Nap Nanny. There were approximately 110,000 of these devices sold. There was a single asphyxial/anoxia death associated the use of the Chill which occurred on July 10, 2012.

An example of a Nap Nanny is depicted in Fig. 13.3.

In Fig. 13.4 is a Nap Nanny with a doll demonstrating a potential infant malposition and entrapment that could result in asphyxiation.

FIGURE 13.3 Nap Nanny infant sleep positioner. *Photograph accessed from http://www.mommyish.com/2012/12/06/nap-nanny-lawsuit-recall/.*

FIGURE 13.4 Demonstration of potential for infant malposition in Nap Nanny and associated positional asphyxia mechanism. *Photograph accessed from http://abcnews.go.com/US/sixth-baby-dies-recalled-infant-recliner/story?id=23970287.*

Analyses of CPSC Data for Comparative Risk Ratio of Asphyxia-Related Death Associated With the Nap Nanny

To assess the likelihood of asphyxia-related mortality in a sleep environment using the Nap Nanny ISP compared to similar sleep environments, an ad hoc FE analysis of asphyxia/anoxia-related deaths using data from the following three sources was performed:

1. The CPSC's National Electronic Injury Surveillance System (NEISS) database;
2. The Centers for Disease Control and Prevention WONDER database; and
3. The CPSC's administrative reporting site.[7]

The NEISS is a probability sample of US hospital emergency departments stratified by emergency department and size and geographic location. The analysis of the NEISS data was designed with the knowledge that death due to unintentional asphyxiation is the leading cause of death among infants less than 1 year of age[1] and that the deaths occur most frequently in the sleeping environments of cribs, beds, and couches.[3,8] For this reason, the analysis was limited to infants less than 1 year of age and to CPSC product categories that included beds, cribs, sofas/couches, and chairs. The purpose of the comparison was to assess the rate of asphyxia death associated with the use of the Nap Nanny versus the products that it was presumably intended to be safer.

Two analyses were performed; the first analysis was designed to be overly conservative, and grouped all three generations of Nap Nanny ISPs together, totaling approximately 165,000 of the ISPs assumed to have been in use over the 4-year period of 2009–12 that the products were on the market, and during which there were five asphyxia/anoxia-related deaths associated with product. The second analysis was more precise to the risk associated with the Nap Nanny Gen II product that was associated with the investigated death, and thus only included the 45,000 products that were on the market during the 3-year period of 2010–12 and the four asphyxia/anoxia-related deaths associated with them.

Analysis 1: Asphyxial Deaths Associated with All Nap Nanny Products Compared with Other Sleep-Related Products, 2009–12

Given that the Nap Nanny-related fatalities occurred in 2009 through 2012 time frame, this first analysis used an inclusive time period of 2009–12 for comparison with the NEISS data. The age of death for all of the Nap Nanny fatalities was less than 1 year, which is consistent with the age at which infants are most at risk for suffocation while sleeping due to an inability to fully protect their airway.[3] The comparison group of children exposed to other sleep products was thus limited to infants under the age of 1 year as well. The number of deaths used for the numerators for the asphyxia-related mortality rates for each product was derived from NEISS data, and the denominator population counts of exposed infants less than 1 year of age were taken from the census population data via the CDC WONDER database. The numerator and denominators for the Nap Nanny risk estimate were derived from the administrative records of the CPSC.

The first analysis was accomplished in two steps. The first step consisted of a comparison of the crude rate of asphyxia-related death associated with the various products, regardless of how much each one was used. Thus, the annual exposure to a Nap Nanny, bed, or crib, etc.,

for an infant less than 1 year of age was considered equal to all other products. In order to make an unbiased comparison between the products, however, the amount of time the infant would have been exposed to the product had to be accounted for. Thus, the second step of the analysis included a weighted denominator of time, based on the estimated use of the product. The hours of product exposure were based on sleep/wake patterns in infants during the first 12 months of life and CPSC documentation regarding ISP fatalities.[6,9,10]

The Nap Nanny exposure rate was based, in part, on the assumption that the total number of the products (165,000) were sold at a constant rate between 2009 and 2012, so that the market saturation ranged evenly from 0% to 100% during the time period.[11] Once the per-product rate of asphyxial death was determined for the first and second steps of the analysis, the risks between the products could be used in a CRR.

The results of the first step of the analysis are displayed in Fig. 13.5.

In this chart the Nap Nanny is associated with 3.0 deaths per 100,000 products, which yields a CRR of 2.4 relative to the rate of death associated with beds, 7.0 relative to sofas, and 30 relative to cribs.

Although these ratios indicate that there is substantially elevated risk of asphyxial death for the Nap Nanny relative to other infant sleeping environments, for a more accurate and less biased estimate of comparative risk it is necessary to adjust the analysis for the frequency and duration of the use of the product. For the following analysis, it was assumed that an infant spends 12 h/day sleeping in a bed or crib, and then naps for an additional 6 h on a chair, sofa, or couch. The analysis for the Nap Nanny was then conducted with the assumption of either 12 or 6 h of use per day, with both results presented as a form of a sensitivity analysis. While both values may be an overestimation of the average amount of time that a Nap Nanny was used, increasing the exposure denominator would have the effect of reducing the CRR for the Nap Nanny versus the compared products, and thus act as a form of a safety analysis (ie, the assumption favors the opposing party). The chart depicted in Fig. 13.6 illustrates the differences in rates between the Nap Nanny and other products using the modified exposure rates.

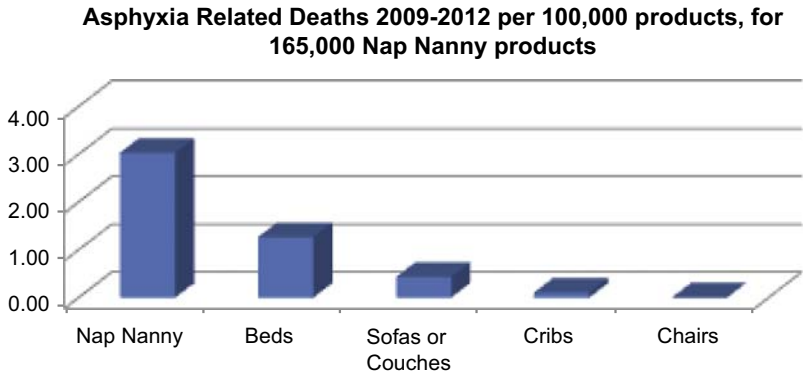


FIGURE 13.5 Rate of asphyxial deaths among infants <1 year of age for 2009–12, per 100,000 products (Nap Nanny) or households (for beds, sofas or couches, cribs, and chairs).

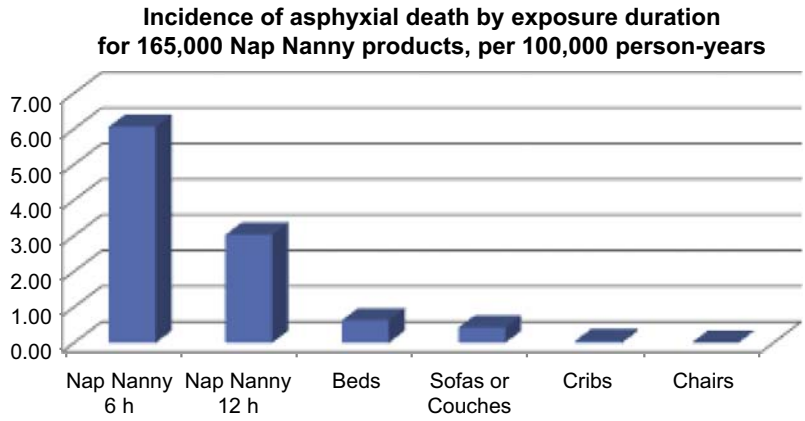


FIGURE 13.6 Rate of asphyxial deaths among infants <1 year of age for 2009–12, per 100,000 products (Nap Nanny) or households (for beds, sofas or couches, cribs, and chairs). Assumed exposure durations: Nap Nanny 6 h (6 h/day); Nap Nanny 12 h (12 h/day) Beds (12 h/day); Sofas/Couches (6 h/day); Cribs (12 h/hours day); Chairs (6 h/day).

The lowest CRR was observed between beds and the assumed 12-h daily use of the Nap Nanny, at 4.8. Even at the highest assumed rate of use of the Nap Nanny, the CRR was 59 versus cribs. Using the 6-h assumed daily exposure these CRRs were doubled to 9.6 for beds and 119 for cribs. In comparison with sofas and couches, the CRR for the Nap Nanny was 7.1 and 14.1 for 12- and 6-h daily exposure, respectively. All of these CRRs were statistically significant at the $p = 0.05$ level, as none of the associated 95% confidence intervals included 1.0 in their lower boundary.

Analysis 2: Asphyxial Deaths Associated with the Gen II Nap Nanny Compared with Other Sleep-Related Products for 2010–12

The same analyses described previous were performed only for the 45,000 Gen II Nap Nanny ISPs (the product associated with the investigated death), and the three deaths associated with the product during the 3 years that it was marketed. The results for all products, regardless of exposure duration, are depicted in Fig. 13.7.

The Gen II Nap Nanny was associated with 8.9 deaths per 100,000 products, versus the 3.0 per 100,000 rate for all of the Nap Nanny ISPs. This rate resulted in a CRR of 7.1 for beds, 20.7 for sofas or couches, and 89 for cribs.

The results of the 12- and 6-h Nap Nanny exposure rate comparison, using only the Gen II model rates, are depicted in the chart in Fig. 13.8.

The lowest CRR was again observed between beds and the 12-h use of the Gen II Nap Nanny, at 14.1. For sofas and couches the ratio was 20.7, and in comparison with cribs, the CRR for asphyxial deaths for the Gen II Nap Nanny was 174. As was the case with the prior analysis of all Nap Nanny products, these rates were doubled for the 6-h use CRR.

Based on the preceding FE analysis, it was concluded that there was sufficient evidence to provide an opinion that the Nap Nanny ISPs were substantially more dangerous than the sleeping products they were designed to replace. Further, given that the lowest CRR for

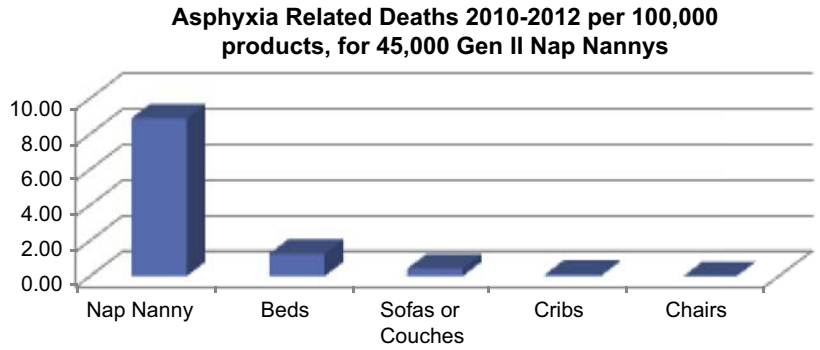


FIGURE 13.7 Rate of asphyxial deaths among infants <1 year of age for 2010–12, per 100,000 products (Nap Nanny II) or households (for beds, sofas or couches, cribs, and chairs).

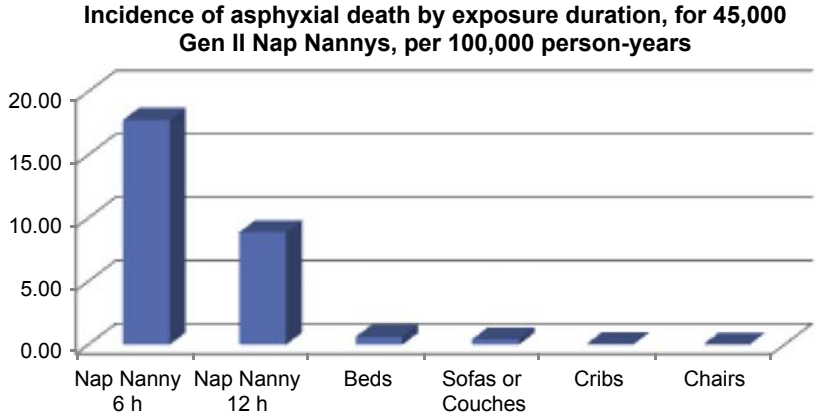


FIGURE 13.8 Rate of asphyxial deaths among infants <1 year of age for 2010–12, per 100,000 products (Nap Nanny II) or households (for beds, sofas or couches, cribs, and chairs). Assumed exposure durations: Nap Nanny 6 h (6 h/day); Nap Nanny 12 h (12 h/day) Beds (12 h/day); Sofas/Couches (6 h/day); Cribs (12 h/hours day); Chairs (6 h/day).

the Gen II Nap Nanny was 14.1 using the most conservative usage rate (ie, assuming the product was used at an equal rate as beds and cribs), and that a CRR of 14.1 equates to an AP_e and thus PC of 93%, all three of the deaths observed among the 45,000 products would have been prevented by substitution of a bed or crib for the ISP.

CASE STUDY #2: WINDOW BLIND STRANGULATION INVESTIGATION

This case concerns the death of a 29-month-old female toddler who was strangled to death while attempting to open a window blind in her bedroom. At some time during the early hours of the morning, she had climbed into a rocking chair that was situated below the

window and moved the window curtain to the side to access the cord of the blind. She pulled down on the operation cord and retracted the horizontal slats up to the top of the window blind's head rail, and, in doing so, her head was entangled in the looped pull cord. The child was discovered hanging from the blind cord 8:15 am; she had last been seen alive at 4:00 am for a diaper change. Following attempts at resuscitation, the infant was transported to a local hospital where she was pronounced dead shortly after arrival. Ligature marks were observed around the child's neck indicating the mechanism of fatal injury.

Epidemiology of Pediatric Window-Cord Strangulation

Corded window shades have been a well-recognized household hazard to small children for more than 30 years. A 1980 publication described 233 cases of childhood strangulation, using national data from the CPSC as well as data abstracted from medical examiner files.[12] These authors noted that strangulation by a rope or cord accounted for 26.6% of all pediatric strangulations. The authors concluded that while infants were more likely to strangle by being wedged between furniture or on a cord attached to a toy, toddlers were most often injured by entanglement with a hanging cord, like from a corded window blind.

A 1997 publication described an analysis of pediatric asphyxiations involving window blind cords using data from the CPSC for 1981–95.[13] The authors identified 183 cases of asphyxiation among children aged 0–3 in the United States, a rate of 0.14 per 100,000 children. The authors noted that the rate was probably low, since some states were not included in the CPSC database during the study period. The paper described the two most common injury scenarios; either an infant in a crib became entangled with a cord, or a toddler was suspended from a pull cord when playing on furniture located near a window.

The heightened risk of death by strangulation in toddlers has been observed consistently; in a study of 91 cases of unintentional mechanical asphyxia occurring in German children between 2000 and 2008, the authors describe 2 age peaks; one in toddlers and the other in adolescents aged 13–14.[14]

The Window Covering Manufacturer's Association (WCMA), in partnership with the CPSC, authored a report that described known pediatric window covering cord strangulations between 1996 and 2004.[15] Narrative descriptions in the CPSC incident reports, as well as scene photographs, were used to characterize the type and design elements of the window covering involved in the incident. Among 82 incidents, 50 were related to window treatments operated by cord lifts. Of these, horizontal blinds were responsible for 47 incidents. Sixteen cases involved a tassel loop, and 14 were related to the inner cord. Moreover, in the cases in which conformance with WCMA or other manufacturing standards could be determined, 76% of the blinds did not meet the standard.

The Defendant's Expert Opinion

Unlike the preceding case of the ISP death investigation, the epidemiology of window blind cord strangulation deaths was already relatively well described in the literature as summarized above. The defendant manufacturer retained an expert in statistics to dispute the assertion that window blinds with unsecured or looped cords present a greater hazard to children aged 0–3 than any other household product that infants and toddlers routinely

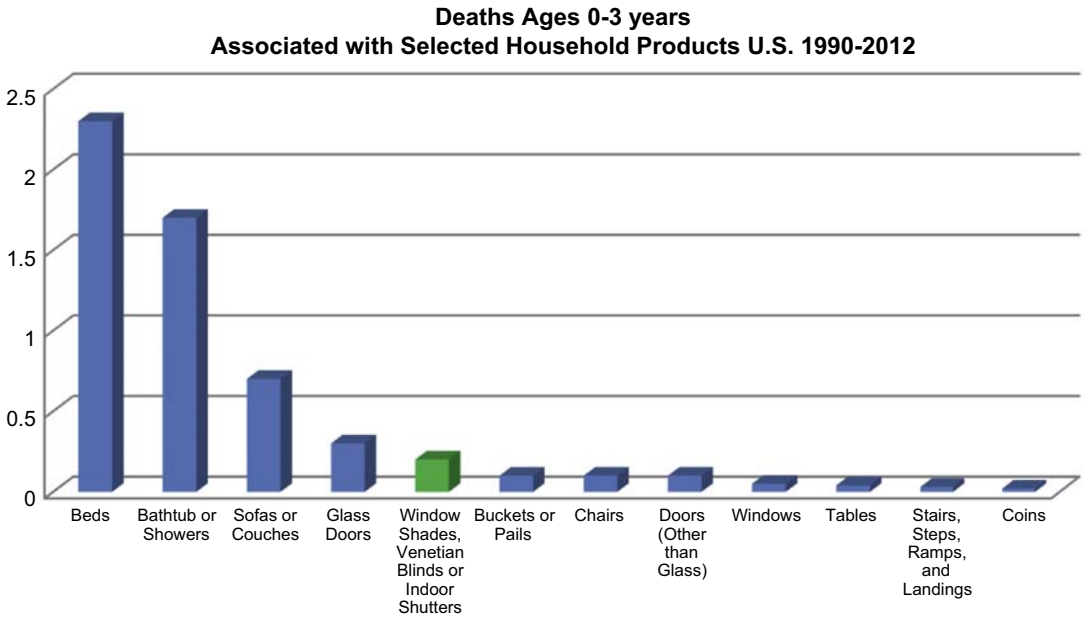


FIGURE 13.9 Analysis performed by defendant's expert, with all study periods (1990–93, 1990–2005, and 1990–2010) combined, per 100,000 households.

come into contact with. The expert described a review of four CPSC databases for a death rate comparison, including the NEISS, the Death Certificate Database (DTHS), the In-Depth Investigations (IDI), and the Injury and Potential Injury Incident (IPII) databases.

The expert relied primarily on data from the NEISS to provide both a hospitalized injury and a death count associated with window coverings and other household products. The injury and death counts were then matched to the populations of households where the products were likely to be in use, to arrive at estimated rates for the United States.

The expert's analysis was performed for three study periods: 1990–93, 1990–2005, and 1990–2010, and then combined and depicted in a bar graph, depicted in Fig. 13.9 (the rates are per 100,000 population).

Based on the analysis, the expert concluded that window coverings, including corded window blinds, were the fifth most common household item associated with death among 0- to 3-year-old children, and thus the product was not dangerous, relative to other household products. The manufacturer used the conclusions of the expert to argue that it was not the intrinsic danger of corded window blinds that resulted in the death of the infant, but rather that it was a rare, tragic, and largely unpreventable death that was even more likely to have occurred with other common household products that infants routinely come in contact with.

Analysis of the Defendant's Expert's Methods

There were such substantial methodologic flaws in the analysis provided by the defendant's expert that the resulting opinions were meaningless. The analysis was

designed in such a way that it diluted or obscured the true relationship between infant exposure to corded window blinds and the risk of injury. This was accomplished as follows.

Duration of Exposure

Per the approach described in the prior example of the ISP investigation, it should be readily apparent that the comparison of the risk of infant injury from exposure to window coverings to a range of products that included beds, bathtubs, stairs, buckets, and coins cannot be performed on a product-to-product basis without adjusting for the intensity and duration of exposure specific to each product. Every child sleeps in a bed, and depending on age will spend 12 h or more per day in bed.[16,17] Likewise, every child sits on the chair, and every child is bathed in a tub or shower of some kind. These are all products designed to be used by or for infants and small children. In comparison, corded blinds are designed to be used by adults and older children, but not infants and toddlers. It is reasonable to assume that in many if not most households, small children do not come into physical contact with window blinds at all, and that their average daily interaction with such products is brief, likely on the order of minutes. It is reasonable to consider even 10 min of daily interaction for an infant with corded blinds to be an overestimation of the likely duration of contact in the average household with both an infant and corded blinds. In comparison with corded window blinds, daily exposure to chairs and beds is 24 and 72 times greater (assuming 4 and 12 h of daily use, respectively). Thus, the defendant's expert's assumption of equivalent exposure time for all products that were compared with window blinds resulted in a substantially biased estimate of risk toward the null.

Age of At-Risk Infants

The defendant's expert pooled all fatal injuries occurring among children aged 0–3. In doing so, the expert ignored the rather obvious fact that corded window shades can only be used by children old enough to interact with these products, either as they were designed to be used, or as a toy. The muscular control and mobility required for a child to gain access to a corded window blind only develops in children that are typically >1 year of age. In comparison, beds and chairs are passively used products in which injury results from minimal interaction between the child and the product. Fatal injuries in children <1 year of age consist of relatively passive activities and mechanisms, such as being wedged between a mattress and bedframe, or from rolling off of a chair or sofa and sustaining a head injury. In large part, children <1 year of age are at risk of fatal injury because of the inability to protect their airway. In contrast, it is children aged 1–3 who are at greatest risk of death from corded window blinds, as demonstrated by the fact that nearly all of the reported deaths have been in children >1 year in age. Thus, the inclusion of infants age <1 in a study of deaths from corded window blinds versus all other household products by the defendant's expert had the effect of inflating the number of nonwindow blind deaths in a group at the lowest risk to the injury mechanism of interest and confounding the results of the analysis by age. This error had the effect of further biasing the estimate of risk from corded window blinds toward the null.

Inclusion of Noncomparable Products

The comparison of the fatal injury rate associated with bathtubs to window blinds was also inapposite, as bathtubs are only appropriately used by a caregiver and a child 0–3 years of age simultaneously. Thus, injury that occurs in a bathtub is almost always due to negligent use by the caregiver, most commonly from leaving the child unattended. There is little to no opportunity for an unsupervised infant to be injured in a bathtub without some degree of interaction of a caregiver. Since the investigated question of interest was whether a corded window shade is a comparatively dangerous product for unsupervised toddlers under the age of 3, it was improper to compare this product to a household item that is not designed or intended for use without caregiver supervision.

Additional Considerations

Nearly every child has a bed, chair/sofa, and bath, and in the analysis the defendant's expert assumed that 100% of children 0–3 years were at risk from injury from these products. For the comparison with corded window coverings, the expert assumed that 68% of children were exposed to a corded window covering, based on the results of a household survey that indicated this frequency of windows with blinds or shades. The assumption was another potential source of error in the defendant's analysis. There is no indication of the frequency at which blinds or shades are made with a cord pull, or whether the cord pull is accessible to a child. Further, it is reasonable to assume that in many households with both infants and a corded window blind that parents would keep cribs and beds away from such window coverings, and otherwise keep the cords tied or tethered and away from access to a child. The defendant expert's assumption that 100% of windows with shades or blinds both have cords and are accessible to infants and toddlers likely resulted in an overestimation of the frequency at which children are exposed to the hazard, and thus served as a further source of underestimation of the risk and bias toward the null.

Another potential source of error was the inclusion of all injury types and thus all causes of death in the analysis. The only potentially fatal injuries associated with window blinds are asphyxial in nature (strangulation), a unique mechanism of injury that is inevitably fatal if not interrupted. It has been previously found that most pediatric strangulations from corded window coverings were associated with window coverings that did not meet minimal industry safety standards (see the earlier discussion). Every pediatric strangulation death resulting from entanglement with a window covering cord could have been prevented if the products have been designed more safely. In contrast, a fall off of a bed or sofa is an inherent risk of all beds and sofas and not necessarily preventable (although most, if not all infant bed suffocation deaths are). Thus, limiting the analysis to asphyxial deaths is a reasonable further step to decrease bias in the analysis.

Reanalysis of National Electronic Injury Surveillance System Data

In order to provide a more accurate assessment of the danger of asphyxiation posed by corded window coverings versus other household products, a reanalysis of the data relied on by the defendant's expert was performed, with alterations to account for aforementioned

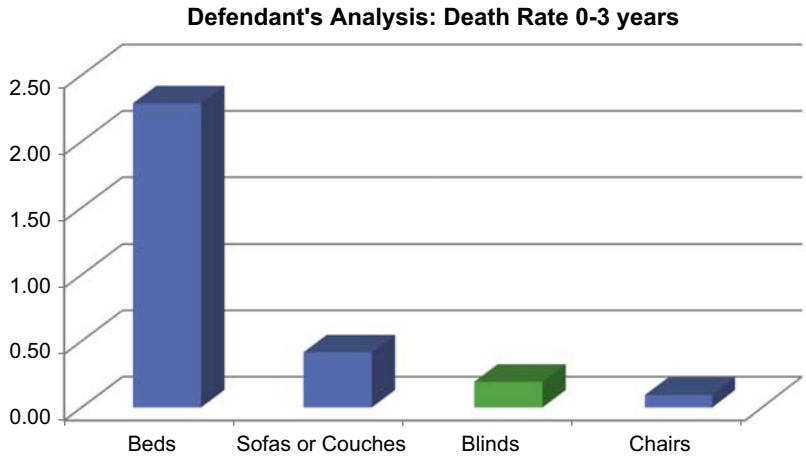


FIGURE 13.10 Simplified chart using the same rate data as the defendant's expert for the 1990–2010 time frame, but only including beds, sofas, chairs, and window blinds.

limitations of the original analysis presented sequentially. Fig.13.10 depicts a chart using the same rate data used by defendant's expert for the 1990–2010 time frame, but only including beds, sofas, chairs, and window blinds. Bathtubs and showers were excluded as deaths related to these products are nearly always due to negligence of the caregiver, as described earlier. With the exception of chairs, all of the products that the expert's analysis identified as having a lower fatal injury rate than window blinds were also eliminated to enhance the clarity of the chart.

In the chart shown in Fig. 13.11, the same analysis has been performed as in the prior chart, excepting that the 0–1 year age category has been removed.

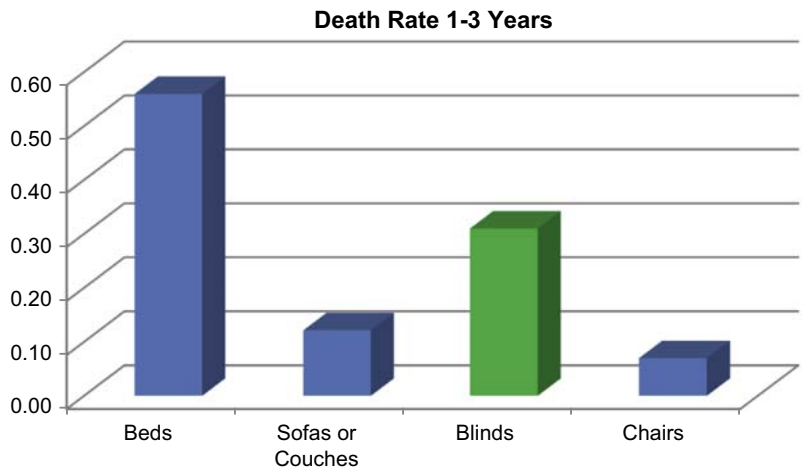


FIGURE 13.11 The same analysis depicted in the Fig. 13.10, excepting that the 0–1 year age category has been removed.

The rationale for removing the 0–1 age group from the analysis is immediately apparent in this chart. As expected, the elimination of infants at lowest risk from window blind cords and at highest risk from passive suffocation in bed and sofas resulted in a substantial increase in the rate of window blind-related deaths, relative to the other three products.

In Fig. 13.12 is a chart showing the same analysis as in Fig. 13.11 but excluding all deaths not due to asphyxia. Thus falls and other mechanisms of injury are eliminated from the analysis. The adjustment resulted in a decrease in the death rate associated with beds, and therefore resulted in a further relative increase in the death rate from blinds, as well as sofas and couches but not chairs.

The final adjustment in the analysis accounted for the adjusted time of exposure associated with the household product, as demonstrated in the chart in Fig. 13.13, based on an assumed daily exposure rate of 12 h for beds, 4 h for chairs or sofas, and 10 min for blinds.

As the chart illustrates, accounting for a more accurate daily duration of toddler exposure to the household products of interest provides a more realistic picture of the risk of asphyxia among 1- to 3-year-old children exposed to the various products. Corded window coverings carry 25 times the risk of asphyxial death in comparison with sofas and couches, and 67 times more than beds (40.4 per 100,000 versus 1.6 and 0.6 per 100,000, respectively). Thus, while the defendant's expert had initially assigned a death rate to beds that was 12 times greater than for window coverings (2.29 versus 0.19 per 100,000), this ratio underestimated the comparative risk of asphyxiation from window coverings versus beds by more than 800 times.

The progression of charts presented here provides a pictorial explanation of how the defendant's expert arrived at a counterintuitive and exceedingly biased conclusion that corded window coverings are no more hazardous than beds, chairs, or sofas. The FE analysis allowed for an evidence based and methodologically defensible rebuttal to the assertion that there is no elevated risk of death associated with corded window blinds in at-risk children, and demonstrated the lack of validity of the defendant's analysis. The expert

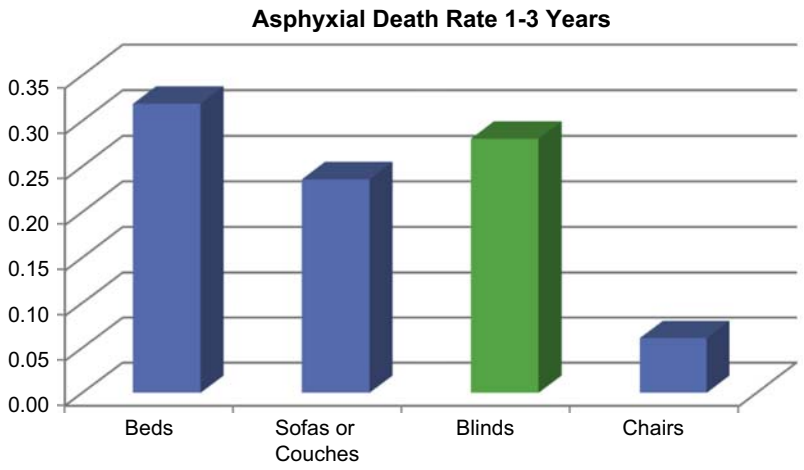


FIGURE 13.12 The same analysis depicted in the Fig. 13.11, excepting that the deaths are restricted to asphyxial causes.

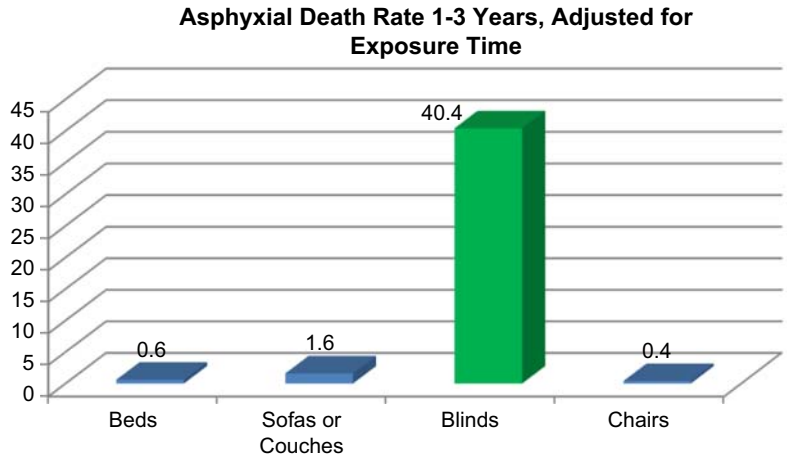


FIGURE 13.13 The same analysis depicted in the Fig. 13.10, excepting that the rates are adjusted by exposure time (a daily exposure duration of 12 h for beds, 4 h for chairs or sofas, and 10 min for blinds).

took a purely "statistical" approach to the analysis, without considering potentially confounding variables associated with the investigated mechanisms. The FE approach dictates that potential sources of confounding and bias must be examined and adjusted for, if necessary, prior to making a side-by-side comparison of dissimilar products, and the preceding case is an ideal demonstration of the application of these principles.

ENDNOTES

1. Centers for Disease Control and Prevention. 10 Leading Causes of Injury Deaths by Age Group Highlighting Unintentional Injury Deaths, United States-2010. National Center for Health Statistics (NCHS), National Vital Statistics System, 2010. US Department of Health and Human Services, CDC. 1-22-2014.
2. Lawrence et al. Suffocation deaths associated with use of infant sleep positioners-United States, 1997–2011. Reprinted from: MMWR 46, 933–937, 2012.
3. Drago, D.A., Dannenberg, A.L., 1999. Infant mechanical suffocation deaths in the United States, 1980 through 1997. Pediatrics 103 (5), e59.
4. Pike, J., Moon, R.Y., 2008. Bassinet use and sudden unexpected death in infancy. The Journal of Pediatrics 153 (4), 509–512.
5. Wanna-Nakamura, 2010. Unsafe Sleep Settings Hazards Associated with the Infant Sleep Environment and Unsafe Practices Used by Caregivers: a CPSC Staff Perspective. US Consumer Product Safety Commission.
6. Wanna-Nakamura. Infant Sleep Positioning Products and Wedges, 2009. National Electronic Injury Surveillance System (NEISS) 2009 Consumer Product Safety Commission 1-22-2014.
7. http://www.cpsc.gov/en/Newsroom/News-Releases/2013/Nap-Nanny-and-Chill-Infant-Recliners-Recalled-by-Baby-Matters-LLC-After-Five-Infant-Deaths-CPSC-Firm-Settle-Administrative-Litigation-/.
8. Pasquale-Styless, et al., 2007. Infant death scene investigation and the assessment of potential risk factors for asphyxia: a review of 209 sudden unexpected infant deaths. Journal of Forensic Sciences 52 (4), 924–929.
9. Burnham, et al., 2002. Nighttime sleep wake patterns and self soothing from birth to one year of age: a longitudinal intervention study. Journal of Child Psychology and Psychiatry 43 (6), 713–725.
10. So, et al., June 1, 2007. The use of actigraphy for assessment of the development of sleep/wake patterns in infants during the first 12-months of life. Journal of Sleep Research 16 (2), 181–187.

11. This assumption was not precisely correct, in that more Nap Nanny ISPs were sold in the last year that the product was on the market than in the prior 3 years combined, however, averaging the sales over the 4 years resulted in a more conservative estimate of the rate of death in the Nap Nanny products, and thus disfavored causation.
12. Feldman, K.W., Simms, R.J., 1980. Strangulation in childhood - epidemiology and clinical course. Pediatrics 65 (6), 1079–1085.
13. Rauchschwalbe, R., Mann, N.C., 1997. Pediatric window-cord strangulations in the United States, 1981–1995. JAMA 277 (21), 1696–1698.
14. Meyer, F.S., Trubner, K., Schopfer, J., Zimmer, G., Schmidt, E., Puschel, K., Vennemann, M., Bajanowski, T., 2012. Accidental mechanical asphyxia of children in germany between 2000 and 2008. International Journal of Legal Medicine 126 (5), 765–771.
15. American National Standard Institute/Window Covering Manufacturers Association. A study of design factors and other influences impacting window-covering cord strangulation deaths as recorded by the Consumer Product Safey Commission between 1996–2004, ANSWI/WCMA 100.1. 2005.
16. Galland, B.C., Taylor, B.J., Elder, D.E., Herbison, P., 2012. Normal sleep patterns in infants and children: a systematic review of observational studies. Sleep Medicine Reviews, 16 (3), 213–222.
17. Michelsson, K., Rinne, A., Paajanen, S., 1990. Crying, feeding and sleeping patterns in 1 to 12-month-old infants. Child: Care, Health and Development, 16 (2), 99–111.

This page intentionally left blank

CHAPTER

14

Medical Negligence Investigation

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

F. Franklin

Oregon Health & Science University-Portland State University, School of Public Health, Portland, OR, United States; Morehouse School of Medicine, Atlanta, GA, United States; Thomas R. Kline School of Law, Drexel University, Philadelphia, PA, United States

OUTLINE

INTRODUCTION

Causation plays a pivotal role in the evaluation of legal actions involving an allegation of medical negligence. Once it is established that a potentially negligent action (either commission or omission) has occurred and that an adverse health outcome has followed that action

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00014-8

in time, there are two questions that must be answered in order for the claim to advance legally. First, it must be determined that the allegedly hazardous action is plausibly related to the adverse outcome. Next, it must be demonstrated, on a more likely than not basis (>50% probability), that in the absence of exposure to the hazard, the outcome would not have occurred in the individual at the same point in time. In a tort action for personal injury, this is referred to as the "but-for" question; as in "but for the allegedly hazardous action of the defendant, would the plaintiff still have suffered the adverse outcome?" The process of answering this question is referred to variously as specific or medical causation.

Outside of a forensic or legal setting, causal evaluations are most commonly performed in a medical setting by clinicians. This is because the determination of the diagnosis of the condition for which the cause is sought is the responsibility of the physician, and not because clinicians are trained in causal methodology. In fact, there is no formal coursework on causality in medical school curricula. An exception is seen in the practice of forensic pathology, where the primary purpose of the postmortem examination is to determine the manner and cause of death. In this setting, when there is a high degree of association between the diagnosis and the cause of the death (for example, a gunshot wound to the head), the determination of causation is easily made as a matter of common sense. This is because the high degree of association of the causal relationship tends to rule out competing causes. In the example of a gunshot wound to the head, causation is obvious because such injuries are nearly always fatal, and the probability of an alternative cause of death coinciding with the time of the gunshot wound is exceedingly low in most circumstances. In contrast, the cause of death in a patient with pneumonia, an 80% occlusion of the left coronary artery, and who received an intravenous injection of a narcotic 30 min before going into respiratory arrest, cannot, and more importantly *should not* be determined as a matter of common sense. In such a circumstance the only causal analysis that can yield valid and repeatable results is the assessment and comparison of the risk of death associated with each of the plausible causes. This form of causal analysis is necessitated by a relatively low degree of association between the death and the cause.

A forensic epidemiology (FE) evaluation of specific causation for a medical negligence action differs from the clinical evaluation of causation in that the former focuses on a comparative risk ratio (CRR) analysis (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) based on a quantification of the risk of injury or disease associated with the investigated hazard versus the competing contemporaneous risk of the injury or disease in the absence of the exposure to the hazard, whereas the latter focuses on the differential diagnosis or etiology and patient history. In circumstances in which medical negligence is alleged as a cause of an adverse outcome, it is rare that there are not at least several alternative explanations for the outcome, including that it was a natural consequence of the disease or injury necessitating the medical care in the first place. In such cases, there will be differing opinions on causation, typically provided by clinicians on either side of the legal dispute. The differences of opinion often stem from disputes over the magnitude of the influence of the competing causes on the outcome (ie, the risks associated with the possible causes). The CRR approach is appropriately used when a differential etiology approach cannot be used to reliably estimate the risks associated with competing plausible causes.

In assessing the cause of an injury in a medical negligence action, the temporal relationships can play an important role. Temporality, as described by Hill in 1965 in his viewpoints

concerning general causation (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*), only pertained to sequence; that the "horse not come before the cart." While temporal sequence is the *sine qua non* of specific causation, in that it must be present in order to proceed with further analysis, there are two other parameters of temporality that are also important to consider in evaluating specific causation in a medical negligence action.

The first of these parameters is temporal *plausibility*. The outcome may not occur before or after the effect range of the hazard. As an example, some food-borne illnesses (ie, campylobacteriosis) only manifest after a matter of hours or days of incubation, and thus an individual who falls ill within minutes of eating undercooked chicken at a restaurant in which *Campylobacter jejuni* is found on the food preparation surfaces was not plausibly made ill by the consumption of the chicken, despite other collateral evidence suggesting causality. As another example, the otherwise unexplained death of a patient occurring 3 days after receiving an injection of an opiate with a half-life of a couple of hours (ie, hydromorphone) is not plausibly related to the injection, as the opiate would have cleared after approximately five half-lives and the death would be outside of the plausible temporal effect range of the drug.

The second parameter of temporality is latency. For an outcome that occurs within the effect range of the hazardous exposure, the quantification of the latency between the exposure and the first indication of disease or injury can be important in assessing the causal association. As an example, a death in a hospital patient that occurs within 20 min of an injection of hydromorphone is much more likely to be associated with the injection than one that occurs 6 h later. This is largely because the cumulative risk (ie, the risk over a given period of time) of competing causes of the death is a function of the latency period between the exposure to the hazard and the first sign of the adverse outcome. In other words, the greater the time between the exposure to the suspected hazard and the onset of the symptoms, the greater the opportunity for alternative risks to act as the cause of the injury.

In medical negligence actions the plausibility of the relationship is rarely challenged; even though an association may be generally agreed to be unusual, it is rare that an assertion of implausibility is encountered. Logical fallacy relating to the misuse of probability is sometimes encountered in opinions that are proffered in medical negligence claims, and an FE analysis can help identify such errors. Two fallacies of probability are the "middle ground fallacy" (argument to moderation) and the conditional probability fallacy. With the middle ground fallacy a compromise halfway between two diametrically opposing opinions is accepted as the most reasonable truth. The fallacy results from the failure to assess and compare the accuracy of each opinion. As a far-fetched example, in a case of postoperative infection resulting from the failure to use sterile practices, one expert might opine that sterile practices should be followed for all surgical procedures, and another expert could opine that sterile practices are never needed for surgery. The "middle ground" between the two opinions would be the absurd conclusion that sterile practices should be followed in only half of surgical procedures. A corollary to this fallacy is the assertion that one cannot choose between multiple differing explanations, as all explanations are possible (ie, $>0\%$ probable) and thus equal. The fallacy is exposed by quantification of the probabilities associated with the various proffered explanations.

Another probabilistic fallacy seen in medical negligence litigation expert opinion is the conditional probability fallacy, also known as the fallacy of the transposed conditional

(described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* and Chapter 15, *Criminal Investigation* as well). This fallacy results from the erroneous transposition of terms in a probability (eg, "all Germans speak German, therefore no non-Germans speak German). An obvious example from medical negligence litigation would be the case of patient who died from blood loss associated with a lacerated abdominal aorta after undergoing abdominal surgery, in which it was opined that because the frequency of aortic injury in such operations is less than 1 in 1000, the probability that the surgeon lacerated the aorta is similarly small. The obvious error of logic is that the wrong question is addressed. While it indeed may be true that there is only a 1 in 1000 chance the surgeon was going to injure the aorta *before* the surgery started, the correct causation question, now that the surgery is complete and the aorta has been injured is "what is the chance that the surgery was the cause of the injury"? The answer to this question is not 1 in 1000. This is the *absolute risk* of the injury before the surgery, but not the probability that the injury was caused by the surgeon versus all other potential causes. The analysis is "conditioned" by the additional information that an injury has occurred (thus the "conditional" probability fallacy).

As we have discussed in prior chapters, individual causation is assessed by examining the CRR, relevant to the known facts. Thus, for the example discussed earlier, the risk of injury from the investigated cause (surgeon error, estimated at 1 in 1000) would be compared to the risk of the same injury resulting from all causes other than surgeon error at the same point in time. The denominator of the CRR in this case would be the answer to the counterfactual question *when an aorta is lacerated during an operation how often is this* ***not*** *due to a surgical error?* Thus, while the risk of injury from a surgical mishap may only be 1 in 1000, the risk of the injury occurring at the same time due to all other causes may be less than 1 in 1,000,000. The conditional probability fallacy is avoided by the use of the comparative risk approach to assess causal relationships, rather than relying on absolute risk alone.

STEPS TO PERFORMING A COMPARATIVE RISK RATIO CAUSAL ASSESSMENT IN A MEDICAL NEGLIGENCE INVESTIGATION

Broadly speaking, there are two steps to a CRR in a medical negligence action; the first step is to determine if the association between the exposure and the outcome of interest is plausibly causal. Most causal evaluations that might involve an FE analysis concern relationships that are reasonably considered to be plausible, however, in some cases further investigation of plausibility is necessary. This first step of a CRR analysis, also considered to be the general causal relationship in legal settings, is not specific to the individual; it only addresses the question "*could the exposure have caused the injury?*"

The plausibility of a causal relationship is rarely an issue in FE assessment of causation in medical negligence cases, as the case would not have progressed to litigation without an adequately established link between the suspected cause of injury and the injury. Plausibility in this context refers to whether the observed association can be explained by known scientific principles. Hill put little stock in the viewpoint that he called plausibility (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*), asserting that it was a criterion "*that I am convinced we cannot demand.*" He noted that detailed scientific evidence describing an injury or disease mechanism may lag behind observational evidence of a consistently observed causal association.

A common error with plausibility assessments is to transpose low preevent probability of injury and *implausibility* of injury, a form of the previously described conditional probability fallacy. In the example, the laceration of the aorta during an abdominal surgery could be deemed "very rare." The rarity of such a complication does not make it *implausible* that the injury would occur in a surgery that requires the use of sharp instruments in the vicinity of the aorta, however. In the context of causation, plausibility and implausibility should not be considered as complements with no middle ground (as is the case with possibility (a probability of >0) and impossibility (probability of 0)). A causal relationship that may not be considered plausible by some (explainable by known scientific principles) is not by default implausible (impossible), as the mechanism by which a causal relationship may exist may simply be unknown at the present time. This is not to say, however, that when implausibility is well established that it should or can be ignored. Implausibility is present when a well-established biological principle must be violated in order to proceed with a causal assessment. An example of an implausible relationship would be the new onset of a Parkinson's disease–related tremor within hours of a biopsy performed under local anesthetic. The attribution of the tremor to the biopsy simply because it followed it closely in time, without consideration for the well-established implausibility of the relationship, results in the *post hoc ergo propter hoc* fallacy. This fallacy refers to the erroneous attribution of causality based purely on a close and sequentially appropriate temporal relationship, despite the fact that the relationship is implausible.

Plausibility can be assessed in one or both of two ways: (1) the relationship is widely accepted as generally plausible, a fact that is typically established via review of previously published biomedical literature or (2) via application of the Hill criteria that address the issues related to general causation, as described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*. In order of decreasing utility and/or availability of evidence, these criteria are as follows: coherence, analogy, consistency, specificity, biologic plausibility, experiment, and biologic gradient. General plausibility in the context of a causal evaluation refers to what is both *possible* (ie, not established as impossible) and *reasonable.* It is not the same as Hill's use of *plausibility,* which was more specific to the biologic mechanism by which the hazard acted in order to cause the outcome (and thus is sometimes referred to as *biologic* plausibility). A hypothetical example of a generally plausible relationship that cannot meet the Hill plausibility criterion would be an outbreak of gastroenteritis among independent patrons of a restaurant. Even if the microorganism responsible for the outbreak is not identified (and thus biologic plausibility cannot be examined), the general plausibility question is easily satisfied by other criteria, including coherence, analogy, and consistency. There is no set number of criteria that must be met to satisfy a conclusion of general plausibility; this is a judgment to be made by the investigator.

If the plausibility question can be answered affirmatively, then the analysis can progress to the second step, which is specific to the individual. This step is intended to answer the question "*what is the probability of that exposure having caused this injury in this individual?*" This part of the analysis incorporates predictive and relevant information about the nature and intensity of the exposure, the proximity of the temporal or spatial association between the exposure and the injury, and the medical and historical characteristics of the individual. As important as it is to quantify the risk of injury associated with the exposure, it is just as important to be able to answer the counterfactual question of "*what is the probability that the individual would have the injury at the same time if the exposure had not occurred?*"

Finally, the role of confounding and/or bias in the examined association must be assessed. If these factors are minimized then the analysis can proceed to an estimate of the CRR, as described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*.

The elements of the investigation leading to a CRR estimate of causal probability are as follows:

1. The risk of injury associated with the allegedly hazardous exposure is quantified via available epidemiologic data or study. This value is used for the numerator of the CRR. A reliable basis for this value may be derived from a previously published well-designed epidemiologic study, or it may come from analysis of information from existing data, similar to what has been described in Chapter 11, *Traffic Injury Investigation*, Chapter 12, *Traffic Injury Investigation: Product Defects*, and Chapter 13, *Product Defect/Liability Investigation* on other applications of FE. There are some situations in which there is no other need to quantify the risk of injury from a hazardous exposure because there is no reasonable dispute that the injury is certain when the hazard is present. An example would be death following the alleged failure to provide treatment for a cardiac arrest. Common sense, as well as cardiac physiology and medical experience, tells us that it is nearly certain (>99% probability) that an untreated cardiac arrest will result in death.
2. The temporal proximity between the exposure and the first indication of the injury outcome is quantified via the evidence specific to the case, typically gleaned from a careful review of the medical record, and/or interview of fact witnesses, if necessary. We can refer to this time frame as the *hazard period.*

   NB: The term "hazard" is being used here in its most common usage, as an event, exposure, action, or failure to take action that could potentially harm an individual's health or well-being. This should not be confused with meaning of hazard when used in "hazard ratio" as a means of describing the proportional difference between two explanatory variables in a survival analysis.
3. The cumulative risk of the adverse outcome occurring during the hazard period, but in the absence of the exposure to the hazardous exposure, is quantified via epidemiologic data or study (aka the "base risk"). This value is used for the denominator of the CRR. Two assumptions are typically inherent in assessing this value; first that the incidence is relatively consistent over time, and second that the risk posed by the hazardous exposure is independent of the competing risks. Like the numerator risk estimate of the CRR, the denominator may be derived from previously published well-designed epidemiologic study, or it may come from *ad hoc* analysis of information from an existing database. In many cases, the denominator risk for the CRR will be derived from annual incidence data, and thus, in order to derive an estimate of the daily or even hourly risk (depending on the hazard period), the annual rate must either be assumed to be relatively stable, or it must be adjusted to reflect the alteration of incidence over the hazard period.

As is the case with other FE applications suitable for presentation in a civil forensic setting, the result of the analysis, quantified as either a CRR or probability of causation (PC), is compared with the standard of what is "more likely true than not," and thus a CRR of $\geq$2.0 (95% CI > 1.0 lower boundary), or a PC of $\geq$50%. The results of the analysis are applicable to a specific individual to the extent that the predictive characteristics of both the

hazardous exposure and the plaintiff as he would have been preexposure, relative to the adverse outcome, are adequately accounted for in the study populations.

Case Presentations

In the following section of this chapter are four case studies in which serious injuries following alleged acts of medical negligence are described. The cases serve as illustrations of the methods for quantifying specific causation in medical negligence legal actions in which the primary legal dispute was the most probable cause of the adverse outcome, and the plausibility of the relationship between the alleged hazard and the adverse outcome was not disputed (at least, the cause and effect relationships were not deemed to be *impossible* by the experts for the adverse party).

The cases are described in the following format: (1) a brief history of the salient and undisputed facts is provided; (2) the alleged negligent act is described; (3) the opposing or defending theory is described; (4) the elements required to calculate the CRR elements are estimated; and (5) the CR is quantified and presented as a PC. In Table 14.1 are the relevant elements of the causal analysis performed for each of the presented cases.

In all of the cases, an *ad hoc* analysis of publicly available data was performed, along with an analysis of the relevant literature. In three of the cases the data were accessed from

TABLE 14.1 Causation Assessment Elements for the Four Described Case Studies

	Case 1	Case 2	Case 3	Case 4
Investigated hazard	Failure to timely treat ischemic stroke with thrombolytic agent	Cervical spine manipulation	Failure to timely diagnose and treat brain stem herniation after lumbar puncture	Treatment with cardiotoxic drug doxorubicin
Adverse outcome	Locked-in syndrome	Vertebral artery dissection and associated stroke	Upper cervical spinal cord infarct	Cardiomyopathy requiring heart transplant
Alternate explanation	Injury would have occurred regardless of treatment	Injury was of unknown cause and coincidental to manipulation	Injury is not predictable and occurs regardless of lumbar puncture	Cannot determine whether the cause was idiopathic or viral versus the drug exposure
Hazard period	1.5–2 h	2 h	2 h	6 months
Exposure hazard risk	Frequency of adverse outcome given no treatment	Frequency of adverse outcome given manipulation	Frequency of adverse outcome given lumbar puncture	Frequency of cardiac injury at drug dosage
Base risk	Frequency of adverse outcome given treatment	Frequency of adverse outcome given no treatment	Frequency of adverse outcome prior to lumbar puncture	Frequency of idiopathic or viral myocarditis/ cardiomyopathy
CRR	2.3–3.2 to 1	163 to 1	10.8 to 1	17.4 to 1
Probability of causation (%)	56–68	>99	91	94

the Nationwide Inpatient Sample (NIS) hospital discharge database, first described in Chapter 11, *Traffic Injury Investigation*. In the first case, however, the data used for the analysis were from a randomized controlled trial, also publicly available.

CASE STUDY #1: LOCKED-IN SYNDROME FOLLOWING THE ALLEGED FAILURE TO TREAT AN ACUTE ISCHEMIC STROKE WITH THROMBOLYTIC THERAPY (TISSUE PLASMINOGEN ACTIVATOR) IN A 28-YEAR-OLD MALE

A 28-year-old previously healthy and athletic man with no prior history of stroke or transient ischemic attack (TIA) developed a sudden onset of nausea, dizziness, and right-sided facial and eye weakness at approximately 9:00 pm. He was transported by ambulance to a hospital emergency department where he underwent a computed tomography (CT) scan of the head at approximately 10:00 pm. The examination was negative for any sign of intracranial mass or hemorrhage. No further action was taken until the man became unresponsive 4 h later. A neurology consultation was not obtained until the following morning at 7:30 am, at which time it was determined that the man had suffered an ischemic stroke in his posterior brain circulation. His condition worsened over the following day, and it was ultimately determined that he was "locked-in" due to an irreversible injury to his brain stem. Locked-in syndrome is a condition in which the patient has sensory awareness but no voluntary muscle movement or ability to communicate, with the exception of eye movement.

The family of the plaintiff alleged that the failure to treat him with thrombolytic therapy (tissue plasminogen activator or t-PA) resulted in the disastrous outcome from the stroke. The defense countered that, on a more probable than not basis, the patient would have suffered the permanent neurologic injury regardless of the thrombolytic therapy. The assertion was based on published literature indicating improvements of approximately 50% in efficacy trials of t-PA, rather than the doubling of benefit (ie, CRR 2.0 or greater) required to demonstrate specific causation. It was, however, agreed by both sides that the t-PA could have been administered within 90–120 min from the onset of the symptoms, as contraindications to administration had been ruled out by this time.

A pivotal issue in assessing the CRR for the FE investigation was to define the causal question of interest. The defense had asserted a correct but irrelevant fact that the relative increase in the probability of a good outcome for treated versus untreated patients, as reported in published clinical studies, is approximately 50%. Such studies have limited, and potentially misleading application to the investigation of specific causation of injury, as they report the i*ncrease in benefit* due to an intervention, when the question of interest is directed at the *decrease in adverse outcome* due to the same intervention. The alleged hazardous exposure that is of interest in the investigation is the failure to give the thrombolytic drug, and the investigated injury in the case dictated the outcome group the plaintiff belonged to; those who did *not* get treated and who had a bad outcome. Thus the defendant's citation to the results of t-PA efficacy studies, in which the goal was to describe the proportion of treated versus untreated patients with a good outcome, provided probabilities with minimal utility for a specific causation analysis. As an illustration of this principle, an untreated group could have a 60% probability of good outcome versus 90% in a group treated with a drug. The

disparity would equate to a relative increase of 50% and an absolute increase of 30% of good outcomes in the treated group. Using theses risks, the CRR, focusing on the benefit of the drug would be as follows:

$$\text{CRR} = \frac{0.9}{0.6} = 1.5$$

which would equate to a PC of only 33%:

$$\text{PC} = \frac{(1.5 - 1)}{1.5} \times 100\% = 33\%$$

Neither of the groups comprising the CRR include the patient who does not get treated and has a bad outcome, however.

Using these same figures, we can infer that the untreated group would have a 40% risk of a bad outcome and the treated group would have a 10% risk. The attributable proportion under the exposed (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) would be 30 of the 40%, or 75% of the total risk. The CRR for these values, which would be focused on the harm for failure to give the drug, would be:

$$\text{CRR} = \frac{0.4}{0.1} = 4.0$$

which would equate to a PC of 75%:

$$\text{PC} = \frac{(4.0 - 1)}{4.0} \times 100\% = 75\%$$

Thus, for an individual who did not receive the drug, in the alternative hypothetical scenario in which he did get the drug, 75% of his risk of a poor outcome would be eliminated, and it can be concluded that the failure to give the drug caused his poor outcome, on a more probable than not (>50%) basis.

An additional problem with using data from an efficacy study, which is highlighted by the facts of the case study described previously, is that the outcome-predictive characteristics of the individual, or the circumstances of the injury, may be substantially different than the average participant described in the study. Most stroke victims are of older age and have significant comorbidities, and many have a history of prior stroke or TIA. The average outcome of this population to stroke, both with and without treatment, is likely different than for a previously healthy and fit 28-year-old male.

Comparative Risk Ratio Analysis

A case-specific analysis of data from the National Institute for Neurological Disorders (NINDS) Recombinant Tissue Plasminogen Activator (t-PA) Stroke Trial was undertaken. The NINDS study was a randomized control trial conducted in 1995 in order to assess the efficacy of the emergency (within 3 h) use of t-PA in preventing long-term disability resulting from ischemic stroke. A total of 624 subjects were included in the trial, with 312 randomized to receive t-PA and the remaining 312 randomized to the control group. A vast amount of data were collected on each subject. The data, comprised of 336 variables,

cover information from a subject's age, gender, and blood pressure, to medical confounders such as the use of heparin, and hyperlipidemia. Initial stroke severity (baseline), as well as outcome at 3 months was assessed using several different scales, including the National Institute of Health Stroke Scale (NIHSS), the modified Rankin Scale, the Glasgow Outcome Scale, and the Barthel Index. All scales were coded such that greater value indicated better outcome. The NINDS data are publicly available, and a CD-ROM with the data can be obtained from www.ntis.gov order number PB2004-500031.

The first analysis performed on the NINDS data was an assessment of the long-term outcome in a cohort of ischemic stroke subjects who were treated with t-PA at various times from onset of symptoms to treatment, compared with those who were treated with a placebo. In order to reduce bias in the data, relative to the pre-stroke status of the plaintiff, only patients without a history of prior stroke or TIA were included in the analysis, which was adjusted for admitting NIHSS score, age, hyperlipidemia, and hypertension. The purpose of this initial analysis was to identify the proportional increase in good outcomes in the population subset, but not to estimate the CRR, which was accomplished with the second analysis, described below. A good outcome at 3 months poststroke was defined as a Barthel Index of 95 or more, a modified Rankin Scale of 1 or less, a Glasgow Outcome Scale of 1, or an NIHSS score of 1 or less. Crude odds ratios (ORs) were evaluated with Fisher's exact test and normal 95% confidence intervals, and adjusted ORs were calculated via logistic regression. All analyses were performed using SAS 9.4.

The results of this first analysis were as follows: there were a total of 429 subjects without a history of prior stroke or TIA. Of these subjects there were 215 were in the control group and 214 who were treated with t-PA. For all of the time of onset of symptoms to treatment categories combined (30 min−3 h), 84 (39.1%) controls were considered to have a good outcome at follow-up, whereas 120 (56.1%) in the t-PA-treated group had a good outcome. This resulted in a statistically significant adjusted OR of 2.6 (95% CI (1.6, 4.4)). *NB*: This OR may appear to be high given these values, as the probability of a good outcome is only 30% greater in the treated group, and thus the risk ratio is nowhere near double. Recall, from the discussion in Chapter 11, *Traffic Injury Investigation*, that an OR yields values that are substantially greater than a relative risk when the prevalence of the condition of interest is high.

The interaction between treatment group and time from stroke onset to treatment (OTT) was also examined. A logistic regression model was used to do this, which adjusted for baseline NIHSS score, age, hyperlipidemia, and hypertension. ORs and their associated 95% confidence intervals are presented in the graph depicted in Fig. 14.1. The values in black are all statistically significant. Thus, for example, treatment at 90 OTT resulted in a 3.76 OR of good outcome for treatment with t-PA versus no treatment, and treatment at 120 OTT resulted in an OR of 2.57.

A second analysis was performed in order to estimate a CRR given the circumstances of the plaintiff's case: where it was known that t-PA was not administered before the 90−120 min OTT window, and there was a poor outcome. The purpose of this second analysis was to examine the case-specific relative risk of a poor outcome in the hypothetical

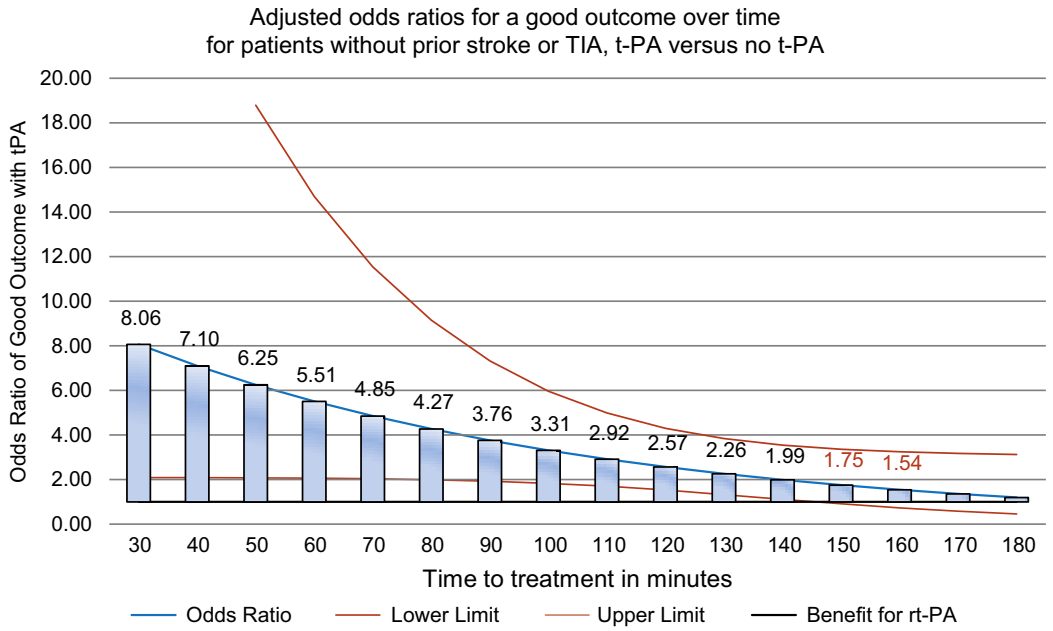


FIGURE 14.1 Adjusted odds ratios for a good outcome over time for patients without prior stroke or TIA, t-PA) versus no t-PA, based on National Institute for Neurological Disorders data analysis. *TIA*, transient ischemic attack; *t-PA*, tissue plasminogen activator.

scenario in which t-PA would have been administered in the 90–120 min time frame, and compare it to the risk of a poor outcome for patients who do not get t-PA. As noted previously, relative risk is a more accurate basis for an estimate of the CRR given the high proportion of the condition of interest in the examined data (poor outcomes).

A second logistic regression model, again adjusted appropriately for admitting NIHSS score, age, hyperlipidemia, and hypertension, was fit to the data and used to predict the probability of poor outcome over time for a 28-year-old patient with an assumed baseline NIHSS of 15 (estimated from the medical record), and without a history of hypertension or hyperlipidemia for both t-PA and non-t-PA groups. The graph depicted in Fig. 14.2 demonstrates the percentage of patients with a poor outcome at 90 and 120 min in the two groups.

Fig. 14.3 demonstrates the CRR (based on the relative risk) depicted in Fig. 14.2 over time, with a CRR of approximately 3.2 at 90 min and 2.3 at 120 min. A CRR of 3.2 is equivalent to a PC of 69%, and a CRR of 2.3 is equivalent to a PC of 56%. These values indicated that, for the plaintiff, more than 50% of the cause of his poor outcome can be attributed to the failure to administer t-PA during the 90–120 min window after the onset of the symptoms. The corollary of this conclusion is that there was a more than 50% probability that the disastrous stroke outcome seen in the plaintiff would have been avoided had he received t-PA during the 90–120 min OTT time frame.

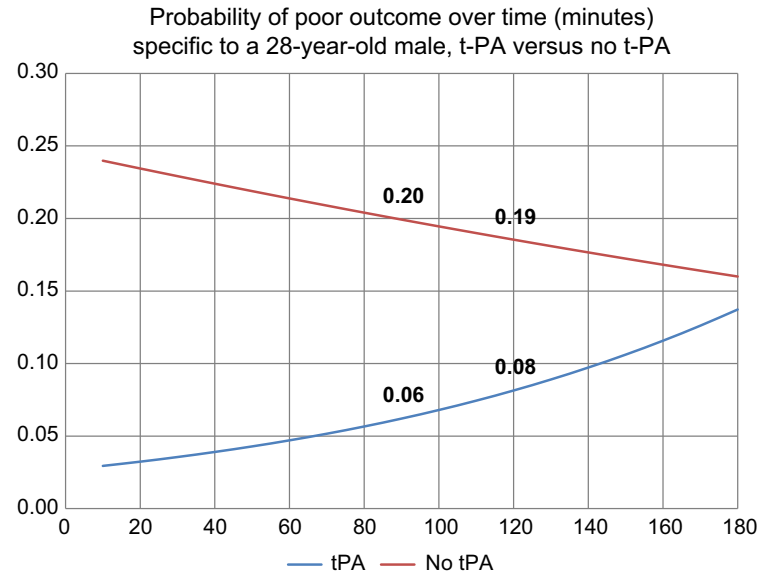


FIGURE 14.2 The probability of a poor outcome over time, by 20-min intervals of onset of symptoms to treatment, adjusted for a 28-year-old male, t-PA versus no t-PA. *t-PA*, tissue plasminogen activator.

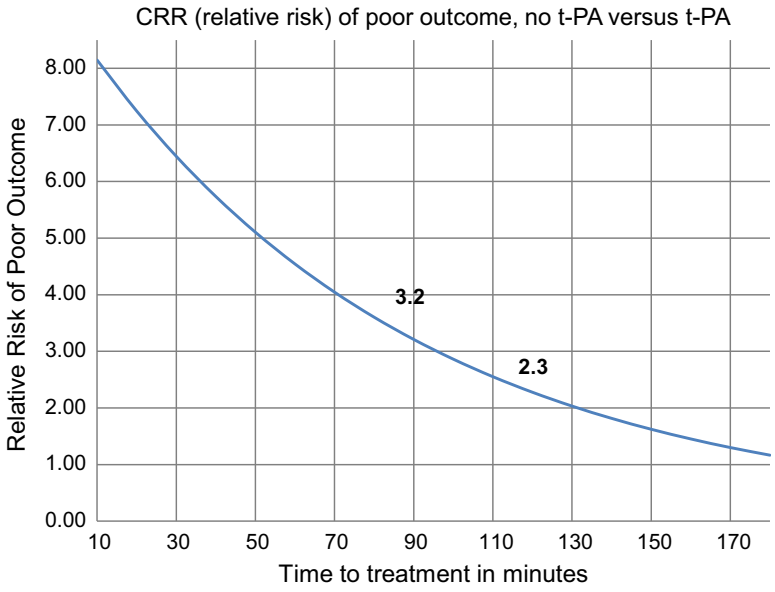


FIGURE 14.3 Comparative risk ratio (CRR) (based on relative risk estimate) of poor outcome, no t-PA versus t-PA, by onset of symptoms to time of treatment. *t-PA*, tissue plasminogen activator.

CASE STUDY #2: MANIPULATION OF THE CERVICAL SPINE FOLLOWED BY VERTEBRAL ARTERY DISSECTION AND STROKE RESULTING IN PERMANENT PARALYSIS

A 27-year-old previously healthy male presented for a first evaluation to a chiropractor for a recent onset of knee pain. As part of the therapy of the first visit, the patient underwent a manipulation of the cervical spine that included rapid rotation of the head and neck. Approximately 2 h following the manipulation, the patient began to feel that the left side of his body was numb and weak. The next morning his condition had worsened and he was unable to summon assistance. He was transported to an emergency department where he was found to have left hemiparesis, facial paresis, and dysarthria. A CT angiogram of the head and neck revealed a dissection of the right vertebral artery, and an MRI of the brain demonstrated an acute infarct of the right basal ganglia. Upon discharge from the hospital, the patient remained partially paralyzed. The patient had no known risk factors for stroke or arterial disease.

The plaintiff alleged that the rotational manipulation of the cervical spine was performed prior to an examination that demonstrated that the procedure could be performed safely on the patient (per generally accepted best practice), and that the ensuing improper manipulation resulted in a dissection of the right vertebral artery, which in turn resulted in the formation of a thrombus that embolized into the vertebrobasilar vasculature and caused the subsequent ischemic stroke.

The defense countered with the assertion that the stroke resulted from unknown factors, and the timing of the symptom onset in relationship to the cervical manipulation was purely coincidental.

In the ensuing FE analysis the relationship between the manipulation and the vertebral artery dissection and associated was deemed plausible. Although the plaintiff and defense standard of care experts disagreed about the risk of dissection associated with the manipulation, there was general agreement that the injury was plausibly associated with the manipulation. Rotation of the neck will produce strain on the vertebral artery, particularly where the artery has a highly tortuous configuration before entering the skull (see Figs. 14.4 and 14.5).

For the CRR assessment, the risk of dissection/stroke from a cervical manipulation was estimated from the literature. Such estimates range from as frequent as 1 in 20,000 patients (an almost certain overestimation) to as little as 1 in 5,846,381 cervical manipulations almost certainly an underestimation (Haldeman et al., 2002). Despite the likely underestimation of true risk, the lower risk figure was selected for the analysis as a means of minimizing the chance of Type I error.

In order to evaluate the competing base rate of *spontaneous* stroke during the 2-h hazard period between the manipulation and the first manifestation of symptoms, data from the NIS were accessed for men with vertebrobasilar stroke in the 25–29 age group for the same year in which the stroke occurred (2009). These values were compared with the number of men in the United States in the same year in the same age group in order to estimate an annual rate. The results of the analysis were as follows: in 2009, there were an estimated 42 cases of vertebrobasilar stroke among all men aged 25–29 who were admitted to US hospitals. Of note, there were only 6 cases that did not result from some external trauma and

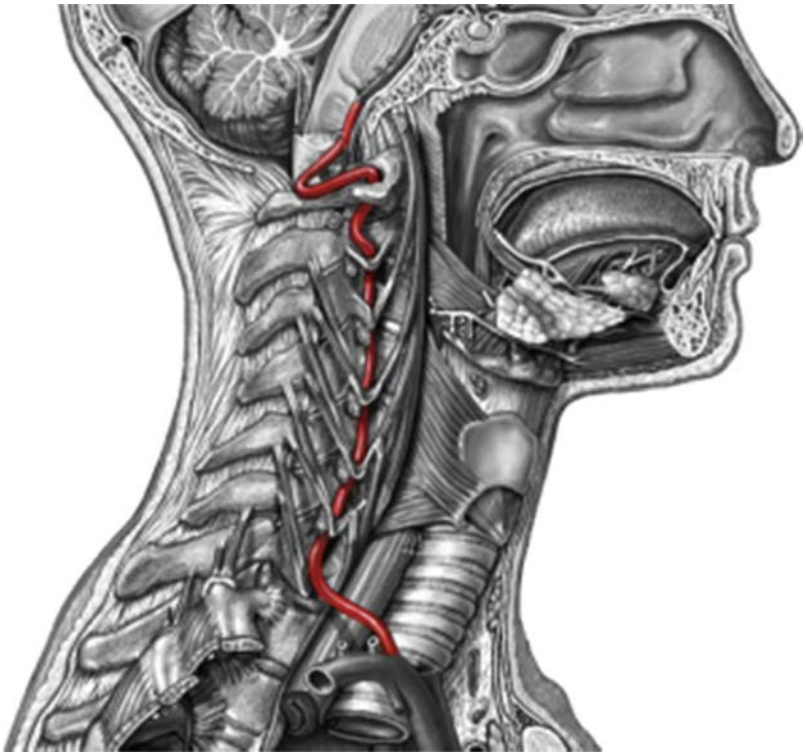


FIGURE 14.4 The course of the vertebral artery, traveling up through the neck and into the skull. *Accessed from* http://snyderphysicaltherapy.com/2013/02/11/cervical-manipulation-is-the-juice-worth-the-squeeze/.

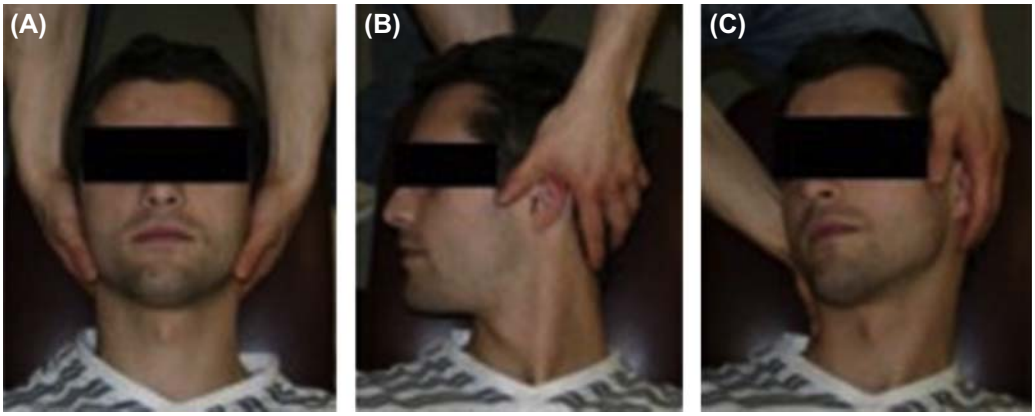


FIGURE 14.5 Demonstration of stages of a rotational manipulation maneuver, as performed by a physical therapist. *Accessed from* http://snyderphysicaltherapy.com/2013/02/11/cervical-manipulation-is-the-juice-worth-the-squeeze/.

thus could be considered spontaneous (20 were associated with a traffic crash, 11 due to assault with a firearm, and 5 were due to unarmed assault). In the same year, there were an estimated 9,744,000 men of the same age living in the United States. Thus, the annual incidence of all vertebrobasilar stroke among men aged 25–29 in the United States was approximately 1 stroke per 216,533. This figure, based on all strokes regardless of cause, was used for the denominator of the CRR estimate in order to again minimize the chance of Type I error. The cumulative risk for the 2-h hazard period derived from the annual incidence, assuming a constant rate over the year, was approximately 1 in 948 million.

The CRR estimate based on the preceding analysis was thus (values are rounded):

$$\mathrm{CRR} = \frac{1 \text{ in } 5,850,00}{1 \text{ in } 948,000,000} = 163$$

The CRR of 163 (95% CI 10, 2613) that favored the manipulation as the cause of the vertebral artery dissection and associated stroke was converted to a PC of more than 99%. As a result, it was concluded that, versus the alternative explanation that the injury was coincidental to the manipulation, the most probable cause of the plaintiff's vertebrobasilar artery dissection and associated stroke was the cervical spine manipulation.

CASE STUDY #3: FAILURE TO TIMELY DIAGNOSE AND TREAT A NEUROLOGIC COMPLICATION OF MENINGITIS RESULTING IN SPINAL CORD STROKE AND PARALYSIS

An 18-year-old previously health male college student fell ill with fever, chills, nausea, and vomiting and over a 24-h period became incoherent and combative. He was transported to a hospital and after evaluation was diagnosed with a suspected case of meningococcal meningitis. A head CT scan demonstrated edema in the brain. The following day, a lumbar puncture was performed on the patient in order to confirm the diagnosis, and 2 h later his condition deteriorated dramatically; he was agitated and combative despite sedation. After 6 h his condition worsened further; his pupils were unequal and he was not responding to painful stimuli. A subsequent CT scan demonstrated increased edema and herniation of the base of the brain through the foramen magnum (the opening at the base of the skull through which the spinal cord passes). Approximately 12 h after the lumbar puncture, intravenous mannitol therapy was initiated to reduce the intracranial pressure, but with no benefit. The patient became a complete upper cervical quadriplegic secondary to an infarct of the high spinal cord, with no sensation or movement from the chin down, and dependent upon mechanical ventilator for respiration.

The plaintiff alleged that the failure to rapidly recognize and reverse the brain stem herniation resulting from the combination of increased intracranial pressure (evidenced by the cerebral edema in the first CT scan), and the sudden decrease in spinal canal pressure following the lumbar puncture, was the cause of the high spinal cord injury. The defense asserted that brain stem herniation is a relatively common and unpredictable complication of meningitis, and as such an unpredictable and unpreventable complication, unrelated to the lumbar puncture. Further, the defense asserted that once the brain stem compression had occurred, the adverse outcome was unpreventable.

The plausibility of the relationship between a lumbar puncture in a patient with increased intracranial pressure and herniation of the brain stem through the foramen magnum is well recognized in medicine. The sudden disparity in the pressure gradient between the spinal canal (which drops at the time of the puncture) and the intracranial space makes the elevated risk of herniation a concern in the patient with evidence of elevated intracranial pressure.

The CRR analysis for the case was approached from two different perspectives; the first was from exposure hazard perspective; ie, if a patient with meningitis suffers from a brain stem herniation following a lumbar puncture how likely is it that the complication was due to the procedure rather than the natural course of the disease? The answer to this first question was found in a previously published study of the CT scans of 445 children with bacterial meningitis admitted to a large pediatric referral center hospital (Rennick et al., 1993). The authors documented time from lumbar puncture to herniation in 19 episodes of herniation. Out of 19, 12 herniations occurred in the first 10 h after the lumbar puncture, whereas the 7 others occurred over 6 other 10-h periods. The study data were used to calculated the odds of herniation in the first 6 h after a lumbar puncture, compared with the 6 h preceding the procedure, and this post hoc analysis indicated that, among those patients with herniation, the condition was 10.8 times more frequent in the 6 h following puncture (95% CI 1.4, 85.2). The relative risk of 10.8 was suitable to use as an approximation of the CRR in the subject case for the probability that the cause of the herniation was the lumbar puncture rather than a coincidental occurrence secondary to the disease process alone. A CRR of 10.8 equates to a PC of 91%.

A second question of interest in the case related to frequency of poor outcomes when a brain stem herniation occurs in a patient with meningitis. In the subject case it was alleged that it was the failure to rapidly recognize and treat the brain stem herniation that led to the spinal cord infarct and devastating sequelae.

In order to examine this question an analysis of NIS data for 2000−10 was undertaken. First, the relevant ICD-9 diagnostic codes for meningitis (all causes), brain compression, and serious adverse events, including stroke, paralysis, and coma, were identified, and then the corresponding data were accessed and analyzed for patients aged 30 and less.

The results of the analysis were as follows: there were a total of 684,654 hospitalized patients with a diagnosis of meningitis. Out of this group, there were 2991 (0.4%) who were diagnosed with brain stem compression/herniation. Among the patients with brain stem compression, there were 345 total cases of cerebral stroke (11.5%), with 168 cases of associated hemiplegia and no cases of spinal cord stroke. Based on the analysis, it was concluded that; (1) brain stem compression is a *rare* complication of meningitis; and (2) when brain stem compression/herniation occurs during hospitalization and the condition is presumably recognized and treated rapidly, in 88.5% of cases there is no serious adverse outcome.

As a result of the analysis it was concluded that the most likely cause of the brain stem herniation was the lumbar puncture. As a secondary conclusion, the defense theory that the spinal cord infarct was unpreventable given the presence of the brain stem compression associated with the meningitis was rejected, with the hospital data analysis indicating that in the majority of cases no serious neurologic injury results from the complication.

CASE STUDY #4: CARDIOMYOPATHY FOLLOWING EXPOSURE TO DOXORUBICIN

In 2007 a 26-year-old woman who was 30-weeks pregnant was diagnosed with invasive ductal carcinoma of the breast. Following successful cesarean section birth of the baby, the woman underwent a total mastectomy of the involved breast. Shortly thereafter she was started on a regimen of chemotherapy, including treatment with doxorubicin (prescribed at 240 mg/m^2), a drug with well-established cardiotoxic side effects. Six months after starting on the doxorubicin the woman developed symptoms of acute decompensated heart failure, a condition that was ultimately diagnosed as resulting from a dilated cardiomyopathy. Approximately 1.5 years later she underwent a heart transplant.

The allegation of medical negligence concerned the failure on the part of the physician prescribing the doxorubicin to assess the patient's heart function prior to the initiation of the therapy, or to adequately assess her family history of cardiomyopathy, which was positive.

The defense to the allegations was that competing causes of the cardiomyopathy, including an undiagnosed viral infection or idiopathic cause, as well as peripartum cardiomyopathy, could not be ruled out. The cardiology expert who provided this opinion acknowledged that the doxorubicin *could* have been the cause, but that it could not be determined which of the causes was more likely.

An FE analysis was undertaken to assess the CRR of the doxorubicin exposure versus all other causes asserted by the cardiology expert. Both the doxorubicin and the viral/idiopathic causes were deemed plausible, but the peripartum cardiomyopathy was rejected as a cause by a cardiologist serving as an expert for the plaintiff. A review of the diagnostic criteria for peripartum cardiomyopathy indicated that the onset of the symptoms had to be within 5 months of birth, and that there could not be any known competing causes present (Hibbard et al., 1999). Neither criterion was met for the case. In order to avoid Type I error, however, peripartum cardiomyopathy was included in the CRR analysis.

The CRR-numerator risk assigned to the doxorubicin exposure was estimated from an analysis of the dose received by the patient, which was compared to the dose-adjusted complication rate reported in a summary of clinical trials of the drug (Swain et al., 2003). A binomial logistic regression of the reported data resulted in an estimated risk of symptomatic cardiotoxicity at the dose received by the patient of 0.85% (95% CI 0.64%, 1.13%), or approximately 1 in 118 exposed patients. See Fig. 14.6.

In order to estimate a patient-specific CRR-denominator risk of the competing causes of the cardiomyopathy, relevant data from the NIS were accessed. Cases of viral or idiopathic cardiomyopathy diagnosed in women in the 5-year age block of the patient were identified in the NIS using appropriate diagnostic codes for data abstracted for the 2-year period during which the alleged negligence occurred. Following the exclusion of cases with significant comorbidities that the patient did not have, the number of cases was matched to census data to arrive at an age and gender-specific annual rate. The result of the analysis was an estimated annual rate of viral or idiopathic cardiomyopathy in women the age of the patient of 0.018%, or 1 in 5640 women. The cumulative risk during the 6 months from the date of the

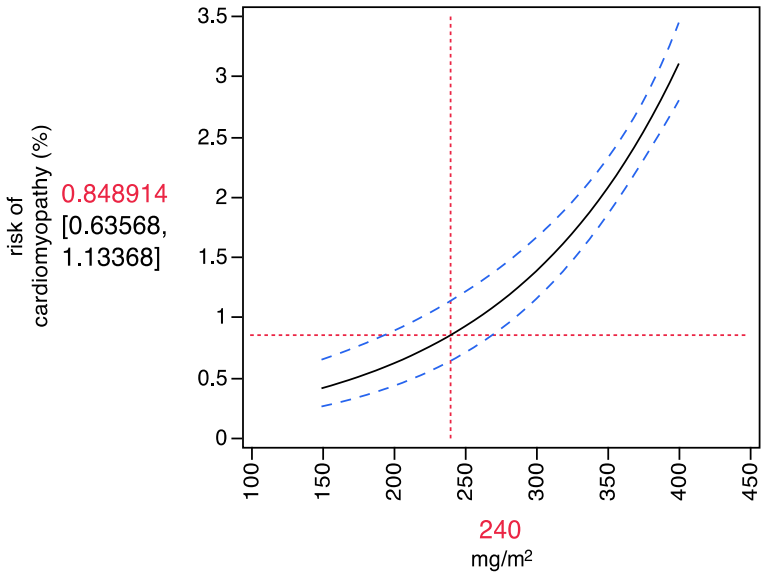


FIGURE 14.6 Profiler depicting the risk of cardiomyopathy at the dosage of doxorubicin received by the plaintiff, which was 240 mg/m^2 (JMP Version 11).

initiation of the therapy to the date of the first symptoms of heart failure was half of the annual risk; approximately 0.009%, or 1 in 11,280.

The rate of peripartum cardiomyopathy was derived from a review of the literature, with highest risk estimated at approximately 1 in 2500 (Givertz, 2013). The highest risk was used to reduce the risk of Type I error.

A CRR was calculated using the drug-related cardiotoxicity risk of 1 in 118 as a numerator versus the sum of the competing causes proposed by the defense in the denominator, including the 6-month risk of a randomly occurring cardiomyopathy of 1 in 11,280, and the risk of peripartum cardiomyopathy of 1 in 2500. Combined these two risks amounted to 1 in 2051.

The resulting CRR was thus 17.4 to 1 in favor of the drug exposure as the cause of the plaintiff's cardiomyopathy, equating to a PC of approximately 94%. The estimate was considered statistically significant, as the lower bound of the confidence interval did not include 1.0. As a result of the analysis, it was concluded that the most probable cause of the patient's cardiomyopathy was the doxorubicin exposure versus the competing causes postulated by the defense.

DISCUSSION

The four cases described in this chapter give a varied but limited view of the applicability of the methods described herein as a means of assessing specific causation in medical malpractice actions. In all of the four cases, causation was the primary contested issue, and in none of the cases was there a basis for a determining specific causation from the diagnosis alone. The medical experts on either side of the cases either opined regarding which of the

possible causes they deemed to be most likely (both plaintiff and defense experts), or they opined that there was no way to know which of the possible causes was the most likely (only defense experts, largely relying on the middle ground fallacy described earlier). Although in all of the four cases the FE analysis supported the plaintiff's theory of causation (that the allegedly negligent conduct was the cause of the adverse outcome), the methods used for a CRR analysis are unrelated to the side (plaintiff or defendant) for which the analysis is performed, and thus the outcome of the analysis is, by design, nonpartisan. If the underlying predictive facts of the case are accurately detailed, the confounding factors are identified and accounted for, and the data analysis is adequately matched to the relevant facts of the case, then the results of the CRR analysis should be the most accurate quantification of the true causal relationship between the alleged hazard and the adverse outcome available.

In these cases, and in the authors' experience generally, there is commonly a lack of scrutiny regarding how causal assessments are made in medicolegal settings. Most often, they are simply given by medical experts as a personally held belief as to what seems most *likely* without any quantification of *how* likely. This approach is often referred to incorrectly as the "differential diagnosis" approach to causation. The name is incorrect, as the medical expert is not differentiating between possible diagnoses to explain a set of signs and symptoms in a patient; rather, the expert is choosing between possible causes of a diagnosis based on an assessment of which cause presents the highest risk (what is more accurately called a "differential etiology" analysis). If this practice sounds like an intuitive or speculative approach to the evaluation of CRR as described in this chapter, it is because that is precisely what it is. In all four case studies there were medical experts who opined on the probability of cause based on either personal experience or their understanding of previously published epidemiologic studies. Courts tend to allow such testimony without question, in part because there are no widely known alternatives, and also because causal determinations are most commonly made by the same clinicians who diagnosed the condition in question. The lack of a systematic approach to causal determinations in medicolegal settings serves as an invitation for speculation, and even abuse, given the financial incentives to medical experts who provide causation testimony in court.

A common theme in all the four cases described here is the fact that the analysis was performed on behalf of the injured party bringing suit, and the result of the analysis favored the injured party. This should not be taken as a sign of a biased or unfair analysis, but rather bias in the selection process by which cases are accepted for analysis. In most instances, the initial impression of a demonstrable causal relationship gleaned from a summary of the case facts is born out by the subsequent analysis. Often, this is because of prior experience with similar fact patterns, or because the competing explanations are readily understood to be quite remote. In cases in which the initial impression is that there is not likely to be a demonstrable causal relationship, there is typically no subsequent analysis. Thus, the cases that undergo a full analysis are also those cases most likely to result in a conclusion that is aligned with the interests of the party requesting the analysis. This is not always the case, however, and in cases with unique causation questions for which no prior analysis has been performed and no literature exists, the results of the analysis may be disappointing to the retaining party.

As described in the FE application chapters, as well as in Chapter 3, *Methods Used in Forensic Epidemiology Analysis*, concept of the CRR is unique to the practice of FE. The CRR allows for a comparison between two or more plausible causes that are known to be present

in a specific case. As a practical matter in an FE analysis, common competing causes of injury can be eliminated by a review of the medical facts in a case, and it is often just the opposing theories of causation put forth by the plaintiff and defendant that require quantification and comparison.

This is not to say that a CRR assessment of individual causation is without potential error or weaknesses. An FE analysis of specific cause is Bayesian at its core; based on the conditioning of probabilities with relevant evidence that is specific to the investigated case. Possible causes must be considered or rejected in an unbiased and fair manner. A failure to consider relevant and predictive evidence may result in a fatally flawed analysis and incorrect conclusion. Conversely, the judgment as to what is relevant and predictive in the analysis of a specific case is by its nature a subjective process, based on the experience and knowledge of the forensic epidemiologist. A basic understanding of the physiologic, therapeutic, and pathologic processes at the center of an alleged act of medical malpractice is crucial prior to embarking on an analysis of a general or specific causation in such cases.

References

Givertz, M.M., 2013. Cardiology patient page: peripartum cardiomyopathy. Circulation 127 (20), e622–e626. Review.

Haldeman, S., Carey, P., Townsend, M., Papadopoulos, C., 2002. Clinical perceptions of the risk of vertebral artery dissection after cervical manipulation: the effect of referral bias. Spine J. 2 (2), 334–342.

Hibbard, J.U., Lindheimer, M., Lang, R.M., 1999. A modified definition for peripartum cardiomyopathy and prognosis based on echocardiography. Obstet. Gynecol. 94 (2), 311–316.

Rennick, G., Shann, F., de Campo, J., 1993. Cerebral herniation during bacterial meningitis in children. BMJ 306 (6883), 953–955.

Swain, S.M., Whaley, F.S., Ewer, M.S., 2003. Congestive heart failure in patients treated with doxorubicin: a retrospective analysis of three trials. Cancer 97 (11), 2869–2879.

CHAPTER

15

Criminal Investigation

M.D. Freeman

Maastricht University, Maastricht, The Netherlands; Oregon Health & Science University School of Medicine, Portland, OR, United States; Aarhus University, Aarhus, Denmark

F. Franklin

Oregon Health & Science University-Portland State University, School of Public Health, Portland, OR, United States; Morehouse School of Medicine, Atlanta, GA, United States; Thomas R. Kline School of Law, Drexel University, Philadelphia, PA, United States

OUTLINE

INTRODUCTION

In the previous chapters on forensic epidemiology (FE) applications the focus has been on disputes arising in the course of civil litigation, typically in which an allegation of negligent behavior is the basis for the legal action. In this chapter we focus on how FE concepts and methods are applied in the context of a criminal prosecution.

Forensic Epidemiology
http://dx.doi.org/10.1016/B978-0-12-404584-2.00015-X

Causality in criminal cases is often undisputed because of the high degree of association between the alleged exposure and the outcome of interest. The temporally proximate nature and high degree of lethality of the methods used to commit homicide (firearms, blunt trauma, sharp instruments) typically leaves little room for consideration of competing causes of injury and death. As an example, when death is the outcome and the exposure is a gunshot wound (GSW) to the head that was sustained moments before the exhibition of signs of injury (unconsciousness followed by cardiorespiratory arrest), there is no need for an expert forensic medical assessment of the cause of the death. The fact that penetrating trauma to the head is associated with a more than 90% death risk is widely understood and accepted (Siccardi et al., 1991). The chance that a competing cause of death acted on a decedent who died directly after sustaining a GSW to the head is so small that it is not worth considering in most circumstances. Even with a causal relationship that is obvious, however, we still have to keep in mind the basic underlying principle of the practice of FE, which is that *causation cannot be observed*. Thus, even in the prior example it is still *possible* that the decedent died due to an untraceable and 100% fatal poison that killed him just prior to sustaining a survivable GSW. A forensic pathologist who finds a bullet in the brain of the decedent will stop looking for a cause of the death, because it is, of course, impractical to consider an alternative cause of death that is so nearly (but not completely) implausible.

In some cases, however, death and injury investigations applicable to the prosecution or defense of a criminal action are aided by the use of epidemiologic data or concepts. In this chapter we present four case studies demonstrating a range of applications of FE methods to the investigation of probabilities associated with disputed issues in criminal cases. In all but the last case the application of physics (ie, biomechanics) plays an important adjunctive role in the assessment of the probabilities of interest.

In the first case study, the circumstances surrounding a crash-related death are described. The issue investigated with FE methods was the position in the vehicle (driver or passenger) of the surviving and intoxicated occupant. The second case involved the death of a pedestrian struck by a motorcycle, with the investigation focused on the speed of the motorcycle at impact. The third case concerns the investigation of the most likely cause of a skull fracture observed in an infant, in which it was alleged that the injury history provided by the father was so improbable that the likely alternative explanation was that the injury was the result of intentional violence. In the last case an epidemiologic investigation of hospital data was undertaken to estimate the probability of causation attributable to maternal gestational cocaine exposure in a full-term delivery of a stillborn baby. Although these cases are varied in nature and the type of analysis performed, there are many other circumstances in which an FE analysis may provide reliable insight into an important question arising in a criminal matter. Like the civil cases described in Chapter 11, *Traffic Injury Investigation*, Chapter 12, *Traffic Injury Investigation: Product Defects*, Chapter 13, *Product Defect/Liability Investigation*, and Chapter 14, *Medical Negligence Investigations*, the analyses are primarily dictated at the assessment of causal relationships, although this is a bit difficult to see in some of the case studies.

CASE STUDY #1: IDENTIFICATION OF THE SEATING POSITION (DRIVER vs PASSENGER) OF AN EJECTED OCCUPANT IN A VEHICULAR HOMICIDE INVESTIGATION

A potential difficulty for fact finders in vehicular homicide prosecutions arises from the lack of reliability and precision associated with injury pattern evidence. Occupant injury patterns are often used in a vehicular homicide investigation to help determine where an occupant was seated during a collision, as some injuries are more commonly associated with a driver's position than a passenger's position, and vice versa (Freeman and Nelson, 2004). The difficulty occurs when there are differing expert interpretations of the significance of the injuries. As an example, one expert may claim that a chest abrasion observed in a decedent could have been caused only by contact with a steering wheel and therefore the surviving defendant must have been in the driver's seat at the time of the crash. In contrast, another expert may interpret the abrasion as having no such meaning. Thus, the evidence of injury, the presence of which both experts agree on, can be characterized as a *positive test* for steering wheel contact. The first expert interprets the positive test as having a high-positive predictive value or PPV (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) for steering wheel contact, and thus reaches the conclusion that the evidence serves as a reliable indication of the precrash position of the occupant as being in the driver's seat. In contrast, the second expert can agree with the first expert that a chest injury is *associated* with a steering wheel impact, but at the same time reject the assertion that the association excludes alternative explanations for the injury (ie, the finding has low specificity for steering wheel contact). While both experts essentially agree that the injury serves as a positive test for steering wheel contact, the second expert infers a low PPV for the finding and rejects the inference that it is reliably associated with steering wheel contact. A fact finder is thus left with two differing interpretations of the meaning of piece of evidence, and no means of quantitatively comparing the accuracy of one interpretation to the other. In such a manner, epidemiologic concepts that are crucial to understanding the meaning of evidence are hidden in plain sight in many criminal investigations.

Injury pattern analysis (IPA) is the method, used primarily in crash injury and death investigation, in which injury patterns observed from postmortem or medical evaluation of decedents and survivors can be systematically paired with crash reconstruction, biomechanics, and epidemiologic data in order to draw inferences regarding the seating position, restraint use, ejection route, and other parameters of occupant status in a fatal crash investigation (Smock et al., 1989; Freeman and Nelson, 2004). IPA is an exemplar of FE methods, as the technique requires the probabilistic interpretation of evidence via the application of knowledge from multiple adjunctive disciplines. Presented in the following case study is the account of an IPA analysis that employed a Bayesian evaluation of the posttest probabilities associated with multiple pieces of evidence relating to the seating position of a surviving occupant of a fatal crash (Freeman et al., 2009). The use of the Bayesian posttest probability formula, described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* allowed for the consideration and relative weighting of the evidence so that it could be presented to lay fact finders in a meaningful form.

The investigated collision consisted of a high-speed frontal impact of a pickup truck with a tree followed by a passenger side leading $^1/_4$ turn rollover in which the surviving occupant

was ejected and the decedent was trapped in the vehicle, and subsequently died in an ensuing fire. There was no definitive evidence regarding which of the occupants was driving, such as an eyewitness account. Because the surviving occupant was found to have a blood alcohol concentration that was three times the legal limit, the death was investigated as a homicide.

The following undisputed evidence was used to construct a posttest probability calculation that the surviving occupant was the driver:

1. The ejected occupant was found to have high-energy (comminuted and/or open) fractures of the right femur, tibia, fibula, and foot.
2. There was extensive crush to the front end of the vehicle on the driver's side, and the driver's side foot well was obliterated (see Figs. 15.1 and 15.2).
3. The decedent was found to have no lower extremity fractures upon autopsy.
4. There was little crush to the front end of the vehicle on the passenger side, and the passenger's side toe pan was preserved. See Fig. 15.3.
5. The only apparent opening allowing for the ejection of an occupant was on the driver's side, between the driver's side door and the A-pillar (see Fig. 15.4).
6. The deployment of the airbags would have made a passenger ejection through the windshield improbable during the initial collision with the tree, and the subsequent $^1/_4$ rollover to the right would have had the effect of trapping rather than ejecting an occupant in the passenger's seat.

This evidence was used to develop true positive and false positive rates for four "diagnostic" tests pertaining to (1) whether the ejected surviving occupant was in the driver's

FIGURE 15.1 The vehicle at final rest on its passenger side. The *arrow* indicates the extensive crush to the left front of the vehicle.

FIGURE 15.2 Photograph of the driver's side foot well with a measurement depicting the distance from the front of the seat frame (*white arrow*) to the end of the foot well. There are approximately 7 inches (0.2 m) of space for the legs of the occupant.

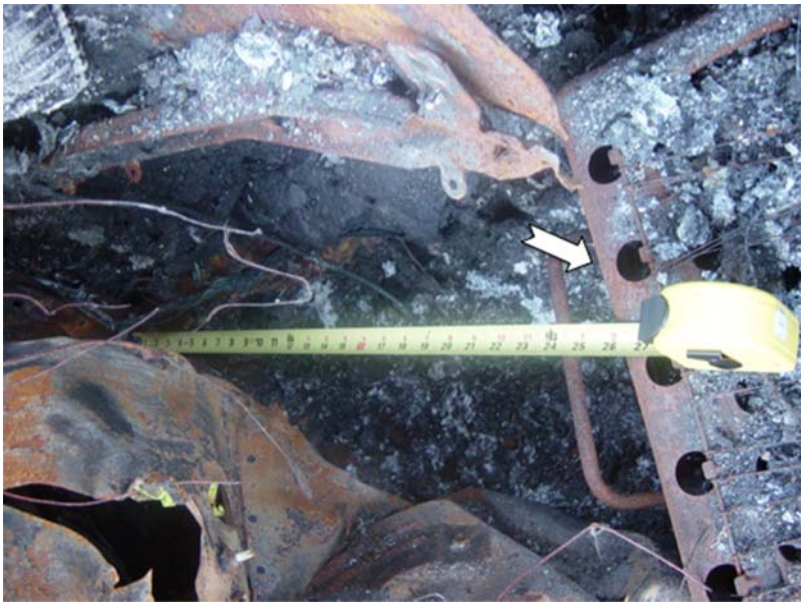


FIGURE 15.3 Photograph of the passenger's side foot well depicting the distance from the front of the seat frame (*white arrow*) to the front of the foot well; approximately 26 inches (0.66 m) of occupant leg space.

FIGURE 15.4 The induced crush resulting in outward bowing of the driver's side door (*upper arrow*) and driver's side windshield pillar (*lower arrow*) formed a large opening and potential ejection route for the driver between the windshield and the door. The space indicated by the double arrow is ~2 feet (0.7 m).

seat, or (2) whether the decedent was in the front passenger seat. The true and false positive rates were used arrive at an estimate of the posttest probability that the survivor was the driver using the following equation for posttest probability:

$$\text{P(driver|tests)} = \frac{(\text{P(driver)} \times (\text{true positives}))}{(\text{P(driver)} \times (\text{true positives})) + (\text{P(passenger)} \times (\text{false positives}))}$$

The equation was simplified by the fact that the pretest probability of driver versus passenger seat position for the survivor was assigned an "indifferent" value of 0.5, and thus (P(driver)) and (P(passenger)) were the same and canceled each other out. The resulting equation was for positive predictive value:

$$\text{P(driver|tests)} = \frac{(\text{true positives})}{(\text{true positives}) + (\text{false positives})}$$

In the same way that the comparative risk ratio (CRR) can be converted from a ratio to a percentage probability, the posttest odds are converted as follows:

$$\text{Posttest odds} = \frac{(\text{posttest probability})}{(1 - \text{posttest probability})}$$

When viewed as tests of the precrash position of survivor, the true and false positive rates for the following four pieces of undisputed evidence from the fatal crash investigation were as follows:

TEST #1—PRESENCE OF A FRACTURED LOWER EXTREMITY OF THE EJECTEE

This was considered a test for the probability the ejected survivor was in the driver's seat, based on the high degree of crush to the foot well at this position and corresponding high risk of lower extremity fracture. This probability was estimated to range from 0.85 to 0.95 based on previously published epidemiologic data (Augenstein et al., 2005). The false positive rate used for Test #1 was the probability that the survivor would have suffered the same fracture if he had been occupying the passenger seat, given the lack of crush at this position. This probability was estimated to range from 0.56 to 0.63 based on an analysis of National Automotive Sampling System-Crashworthiness Data Sample (NASS-CDS) data (see further description of this database in Chapter 11, *Traffic Injury Investigation*).

TEST #2—THE LACK OF A FRACTURED LOWER EXTREMITY IN THE DECEDENT

Test #2 is the mirror image of Test #1, but because of the mutually exclusive nature of the "who was driving" scenario the evidence can, in essence, be counted twice. The true positive rate of Test #2 is equal to the probability that the decedent would not have sustained a fracture had he been seated in the passenger seat position. This value is the complement of the false positive rate (ie, the specificity) for Test #1, or a range of 0.37 to 0.44 (derived from $(1 - 0.63)$ to $(1 - 0.56)$). The false positive rate of Test #2 was based on the probability of no lower extremity fracture had the decedent been in the driver's seat. This range of values was the complement of the true positive rate of Test #1, or 0.05 to 0.15 (derived from $(1 - 0.95)$ to $(1 - 0.85)$).

TEST #3—THE DEFENDANT WAS EJECTED

The estimated true positive rate used for Test #3 was 0.5–0.75 based on the investigation findings that indicated the driver's side as the most probable ejection route (see Fig. 15.2). The false positive rate used for Test #3 was the probability that the ejectee could have been ejected from the passenger seat of the vehicle, estimated to be 0.05–0.15. This probability is given a very low value because the reconstruction of the collision events produced no identifiable route through which the passenger could have been ejected. It could reasonably be argued that it is even lower, if not outright implausible.

TEST #4—THE DECEDENT WAS NOT EJECTED

As was the case with Test #2 relative to Test #1, Test #4 is the mirror image of Test #3. Thus, the true positive rate used for Test #4 was the complement of the false positive rate (0.85–0.95), which represented an estimate of the probability of no ejection given passenger seat position. The false positive rate was the complement of true positive rate (0.25–0.75), which was the probability of no ejection if the defendant had been in the driver's seat.

From the discussion above the true positive rates of the tests were estimated be in the following ranges:

Test #1 = 0.85–0.95
Test #2 = 0.37–0.44
Test #3 = 0.5–0.75
Test #4 = 0.85–0.95

The false positive rates were estimated as follows:

Test #1 = 0.05– 0.15
Test #2 = 0.56–0.63
Test #3 = 0.25–0.5
Test #4 = 0.05–0.15

For the posttest probability calculation, only the lowest true positive and highest false positive values were used to in order to minimize the probability of a Type I error (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) and maximally favor the defendant, as the analysis was performed for the prosecution. The posterior probability that the defendant was the driver which calculated from all of the probabilities as follows:

$$\mathrm{P(driver|tests)} = \frac{(0.85 \times 0.37 \times 0.5 \times 0.85)}{(0.85 \times 0.37 \times 0.5 \times 0.85) + (0.15 \times 0.63 \times 0.5 \times 0.15)} = 0.949645$$

Thus, the odds that the decedent was the driver were:

$$\mathrm{Odds(driver|tests)} = \frac{0.949645}{(1 - 0.949645)} = 19$$

These posterior odds indicate that, using the data most favorable to the defense, the ejectee was at least 19 times more likely to have been the driver versus the passenger.

The sequence of the lower extremity injury mechanism and ejection are illustrated in Figs. 15.5–15.9:

FIGURE 15.5 Right side view of the pickup and tree just prior to impact. Preimpact speed was reconstructed to 55 mph (88 km/h).

FIGURE 15.6 The point of maximum engagement with the tree and maximum crush to the front of the pickup. Both airbags have deployed. It is at this point that the foot well of the driver is crushed toward the driver (see Fig. 15.2), presenting increased risk of lower extremity fracture. The foot well on the passenger side is preserved (see Fig. 15.3).

While the preceding analysis was limited somewhat by the fact that some of the data were adapted from a previously published paper, the paper did provide a description of vehicle damage and thus could reasonably serve as a basis for comparison to the investigated crash. Most importantly, the posttest probability calculation was performed using the true and false positive values that least favored a correlation between occupant position and ejection and lower extremity injury risk (also known as a "safety" analysis), reducing the chance of Type I error.

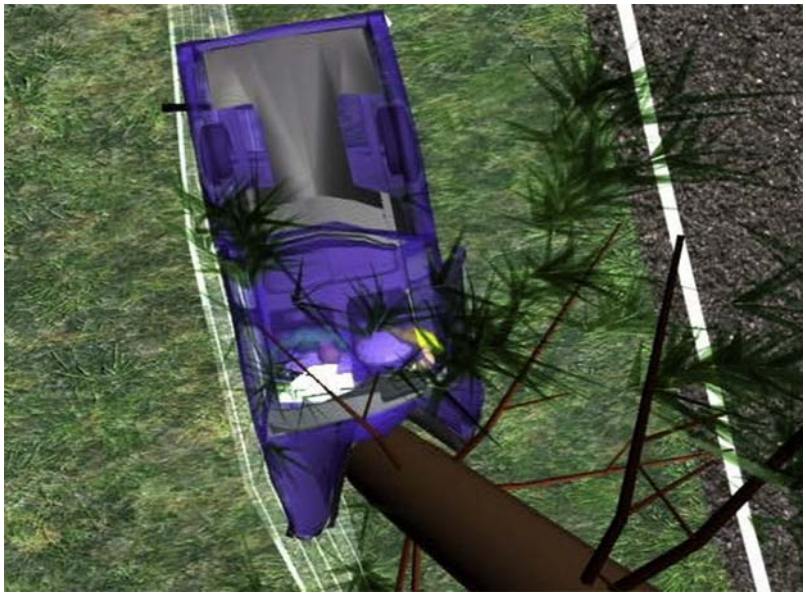


FIGURE 15.7 Top view of the same point in the crash sequence shown in Fig. 15.6. Both occupants have shifted to the left of the vehicle interior because the truck is beginning to roll toward the passenger side.

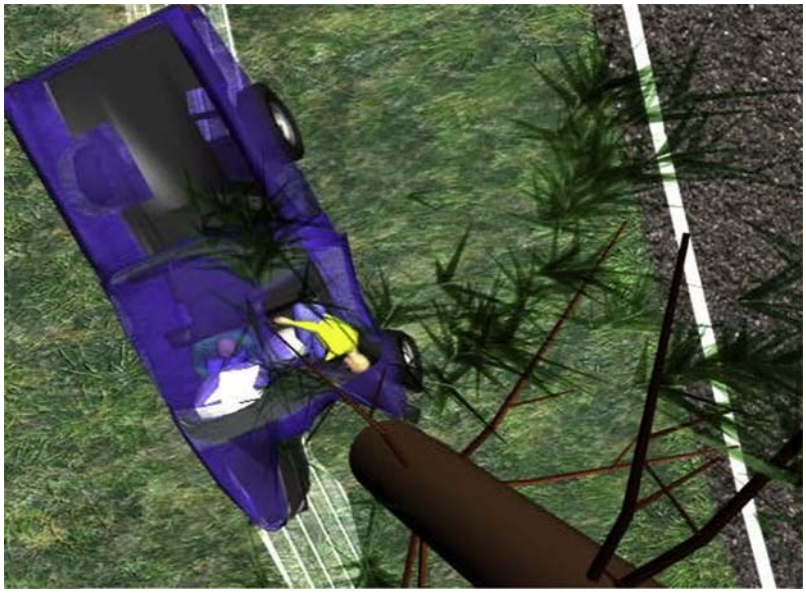


FIGURE 15.8 Top view as the truck continues to roll toward the passenger side and the driver is ejected from the opening between the door frame and the A-pillar.

FIGURE 15.9 Top view depicting the continued ejection route of the driver as the vehicle continues to final rest on the passenger side.

CASE STUDY #2: MOTORCYCLE VERSUS PEDESTRIAN: SPEED AT IMPACT INVESTIGATION

Like the first case study this second case illustrates the application of epidemiologic methods and data to medical, biomechanical, and crash reconstruction investigation findings from a fatal crash in order to assess a discrete issue. The case concerned the death of a 13-year-old female pedestrian who was walking on the sidewalk in a small village with three friends. The decedent was the furthest from the road, and to the right of her friends. She was struck from behind and killed by a Yamaha dirt bike operated by a teenaged male. The motorcycle was a very powerful two-stroke motorcycle that was not legal for street use, and thus not equipped with a headlight or turn signals.

Fig. 15.10 is a figure of the crash scene, with the preimpact approach direction of the motorcycle indicated in the lower right-hand aspect of the picture. The two blue X's indicate the general area in which the decedent was struck, and the red X is the estimated point of final rest for the decedent. The motorcycle is depicted in Fig. 15.11.

The witnesses to the crash (the three friends) recalled that they heard the motorcycle behind them as it entered the sidewalk area, and they turned to see what was making the noise. They turned back and kept walking, assuming that the motorcyclist was entering the driveway to the parking lot, which was approximately 100 feet (30 m) behind them. It was noted by the investigating officer that the driver of the motorcycle lived in the village where the crash occurred, and he had been seen riding the bike illegally on the street previously (the motorcycle was not street legal).

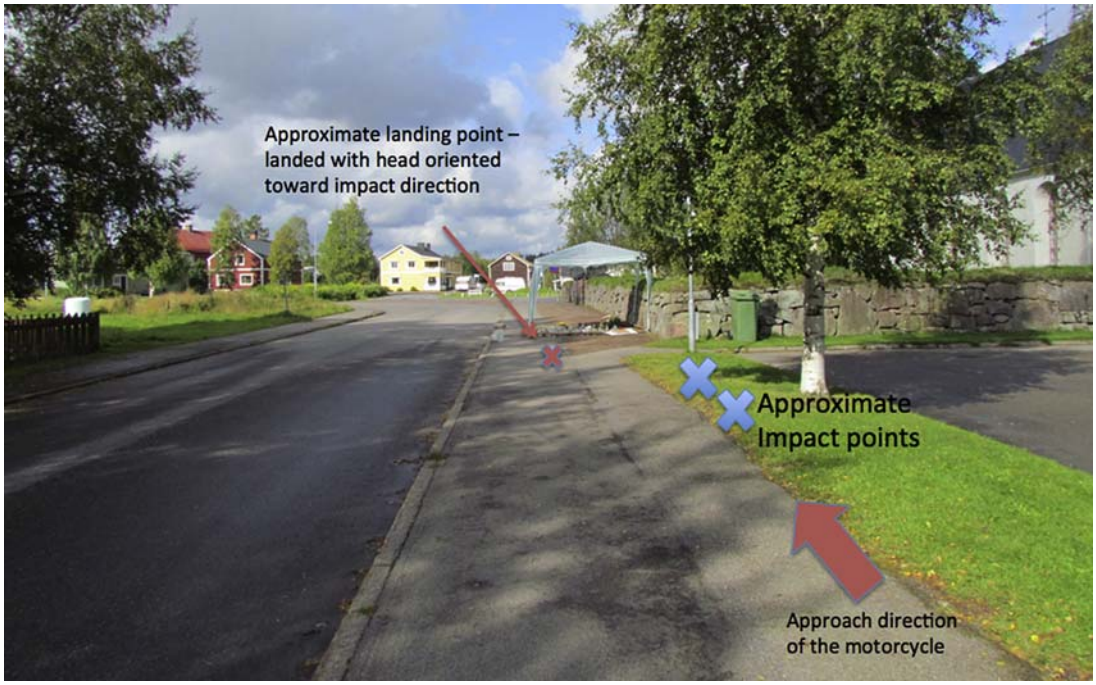


FIGURE 15.10 See text for description of the annotations.

FIGURE 15.11 The involved motorcycle, a Yamaha 450 cc two-stroke dirt bike.

There was no dispute as to how the death occurred. The motorcyclist initially claimed that prior to the collision he was driving on the road and traveling in the same direction that the girls were walking, when he lost control of the motorcycle, resulting in the bike traveling up onto the sidewalk and the subsequent collision. He claimed to have been traveling at 12 mph (20 km/h), which was an important issue for how the defendant might be charged criminally. If it could be proven that the defendant was, in fact speeding, then the level of criminally negligent behavior would warrant a more serious criminal charge.

The three witnesses indicated that after the impact the driver stayed upright for a few meters before falling off of the bike. The bike fell over and slid approximately 100 feet (30 m) to final rest, leaving gouge marks in the sidewalk for the last 23 feet (7 m). The decedent was projected at least 10 feet (3 m) forward by the impact.

Among other injuries, the decedent sustained a massive skull base fracture (ring fracture—see Chapter 6, *Forensic Pathology*) with associated fatal central nervous system injury, along with a right femoral neck fracture and an upper thoracic spine fracture. The fractures to the skull base and the femur fracture are depicted in the 3D CT scan reconstructions depicted in Figs. 15.12–15.14. These reconstructions illustrate the utility of this technology for clearly showing the orientation of bony injuries resulting from high-energy trauma.

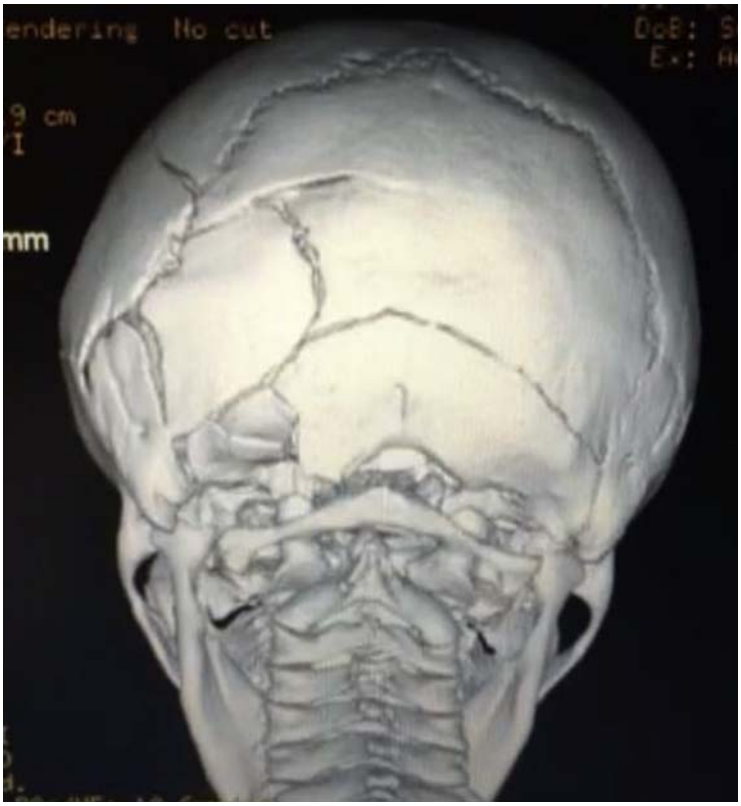


FIGURE 15.12 Postero-inferior view of the skull demonstrating complex fractures of the skull base.

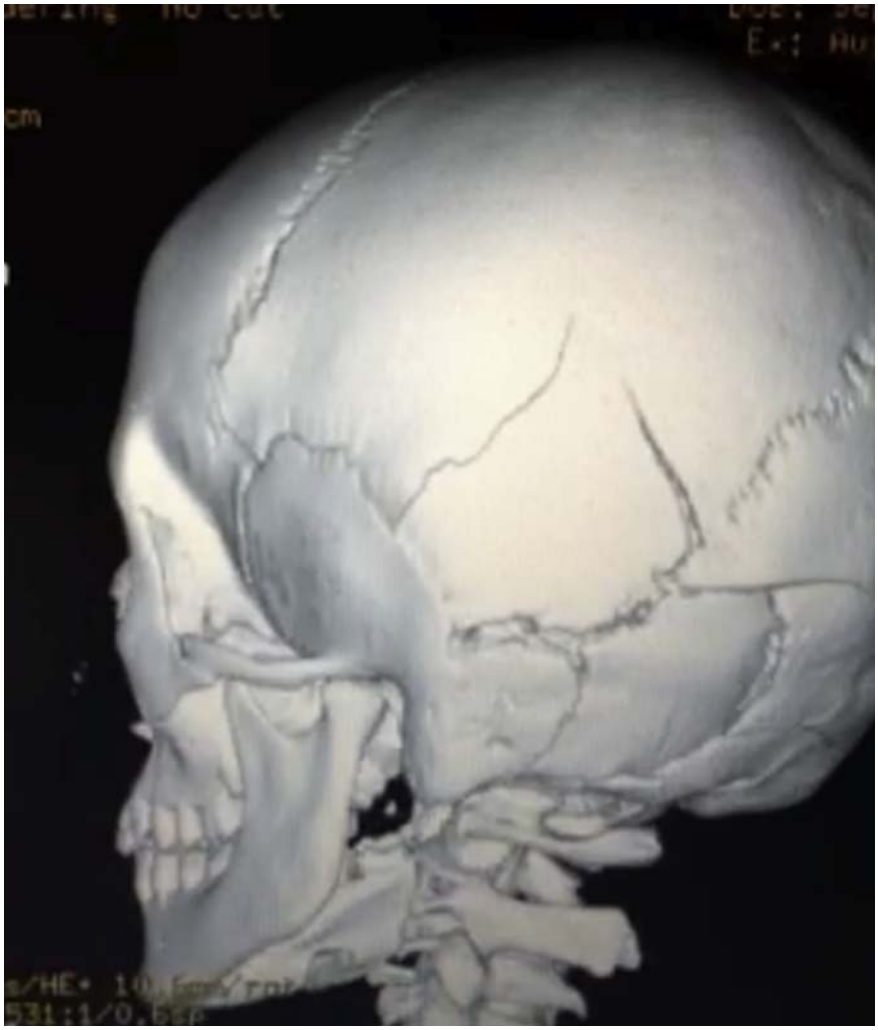


FIGURE 15.13 Left posterior view of skull.

As described in the first case study, an IPA investigation was helpful to understand the injury biomechanics of the collision. The cutaneous injury pattern observed in the right lower quadrant of the decedent's back was matched to the structure, geometry, and height of the right front brake lever and housing. See Figs. 15.15 and 15.16.

The smooth and contiguous injury pattern almost certainly resulting from impact by the brake lever suggests that the driver was not using the front brake when the motorcycle struck the decedent, as the pattern from the two fingers on the brake lever would likely have been evident (the brake type was designed for two-finger use).

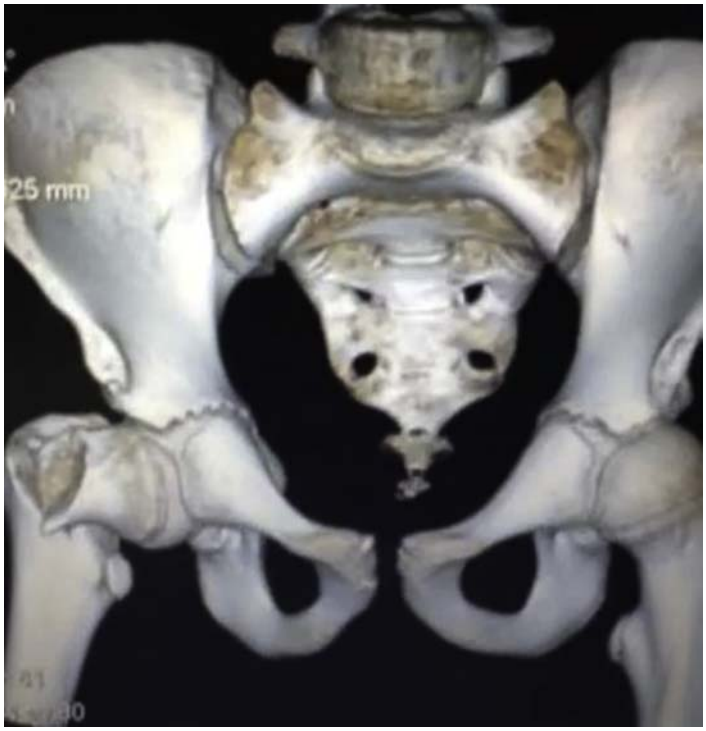


FIGURE 15.14 Anterior view of pelvis demonstrating right femoral neck fracture.

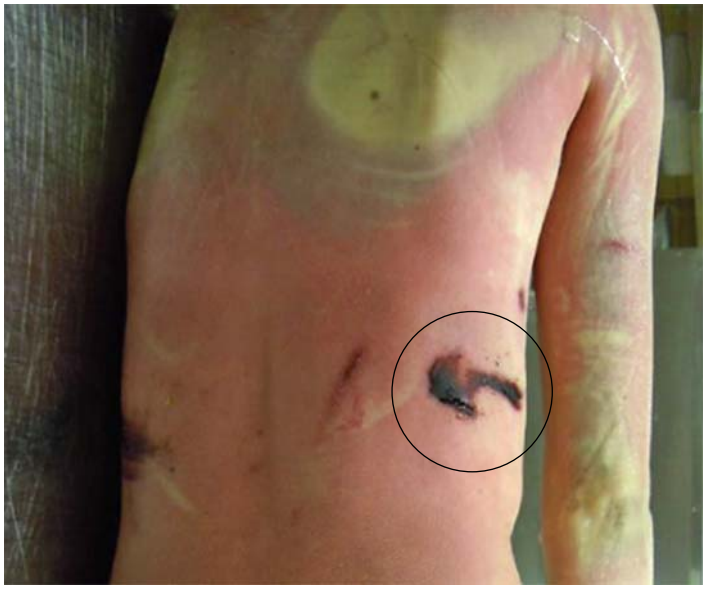


FIGURE 15.15 Posterior view of the decedent's torso and right upper extremity. The widely distributed reddening is postmortem livor (see Chapter 6, *Forensic Pathology*). The injury pattern seen in the circled area is a near-perfect match for the right front brake lever and housing (see Fig. 15.16).

FIGURE 15.16 Right front brake lever and housing (*circle*).

IMPACT SPEED ANALYSIS

There were two types of evidence used to assess the impact speed of the motorcycle. The first and most reliable was the distance traveled by the motorcycle after impact, which was approximately 100 feet (30 m). A motorcycle sliding on pavement will lose speed at a relatively consistent rate because of the friction between the bike components and the roadway (known as the friction coefficient or drag factor) (McNally). If the distance that the motorcycle has slid is known then a standard "slide-to-stop" calculation can be used to estimate the preslide speed of bike.

The formula for this calculation, assuming that the final speed of the bike was 0, is Pre-skid speed $= \sqrt{2 \times (\text{road drag factor}) \times (\text{distance of skid})}$. The result is in feet or meters per second, which can be converted to mph or km/h.

Taking into account the distance that the motorcycle traveled during the skid (100 feet (30 m)), and using a middle value for the coefficient of friction for the motorcycle on the paved sidewalk (0.5), the calculation yielded a preimpact speed of 38 mph (62 km/h), or around three times the speed claimed by the driver of the motorcycle. In the calculation there was no accounting for the energy lost (ie, slowing) when the motorcycle struck the decedent, and thus the preimpact speed was likely greater than just the skid to stop calculation result. Depending on the degree of engagement between the bike and the decedent, the impact could have added an additional 3 mph or more (5 km/h) to the impact speed.

In keeping with the safety analysis approach described in the first case study, assumptions most favorable to the defendant were further examined. Had the motorcycle began to slide while traveling only 12 mph (20 km/h) as claimed by the defendant, and using the lowest published drag factor for a sliding motorcycle, the postimpact travel would have been only approximately 13 feet (4 m), nearly 90 feet (27 m) less than what was observed. This is such a large disparity from what was observed at the scene that it is reasonable to conclude, even assuming a relatively high degree of inaccuracy or error in the measurements or reported distance estimations and an unrealistically low coefficient of friction for the entire slide distance that it is a physical impossibility for the subject collision to have occurred at less than 30 mph (50 km/h).

The second analysis of the impact speed of the motorcycle (made somewhat moot by the reconstruction findings matched to the evidence) was based on an analysis of epidemiologic data, relative to the injuries observed in the decedent. As a demonstration of the forces typically associated with just the decedent's skull fracture (ie, ignoring the additional presence of the femur fracture), data were accessed from the NASS-CDS database (described in Chapter 11, *Traffic Injury Investigation*).

The parameters of the search performed for the present case were as follows: all data were accessed for all occupants 10–25 years of age with a skull base fracture coded as "severe" (AIS 4+) or greater, injured in a passenger vehicle that was involved in a crash, but with no vehicle rollover and no ejection of the occupant. The crash also had to have been reconstructed for impact-related speed change to be included in the analysis. The purpose of the analysis was to find circumstances in which a skull base fracture resulted from a quantified intravehicular impact, as the speed change of the vehicle would be the approximate speed at which the occupant would collide with components in the vehicle interior, and provides the best estimate of impact severity associated with the injury.

Ideally the analysis would have been performed on pedestrian impacts, but no such database containing information on impact speed versus fracture risk was available.

The results of the analysis were as follows:

There were an estimated 7285 occupants with the injury of interest during the years queried (1995–2012). The average speed change at which the injury occurred was 32 mph (53 km/h). Only 2.5% of the injuries occurred in crashes with a less than 12 mph (20 km/h) speed change.

From these data it could be concluded that skull base fractures like the one seen in the decedent are unusual at occupant impact speeds of 12 mph (20 km/h) or less. If the additional impact energy required to cause both the skull fracture and the femur fracture were to be accounted for, it is likely that there would be an even smaller proportion of injuries occurring at similar speed changes.

As a final point to be included in the investigation of the cause of the fatal crash, the fact that the witnesses turned and looked at the motorcycle and then turned back and kept walking suggests that initially the bike was not perceived as a threat. This fact pattern suggests that the driver was traveling slowly when he mounted the sidewalk and the girls saw him, and then he accelerated toward the girls to reach the impact speed. A bike as powerful as the subject Yamaha would be able to accelerate to the likely impact speed within a very short distance. This fact pattern tends to cast doubt on the driver's initial explanation that he lost control of the bike prior to it leaving the roadway.

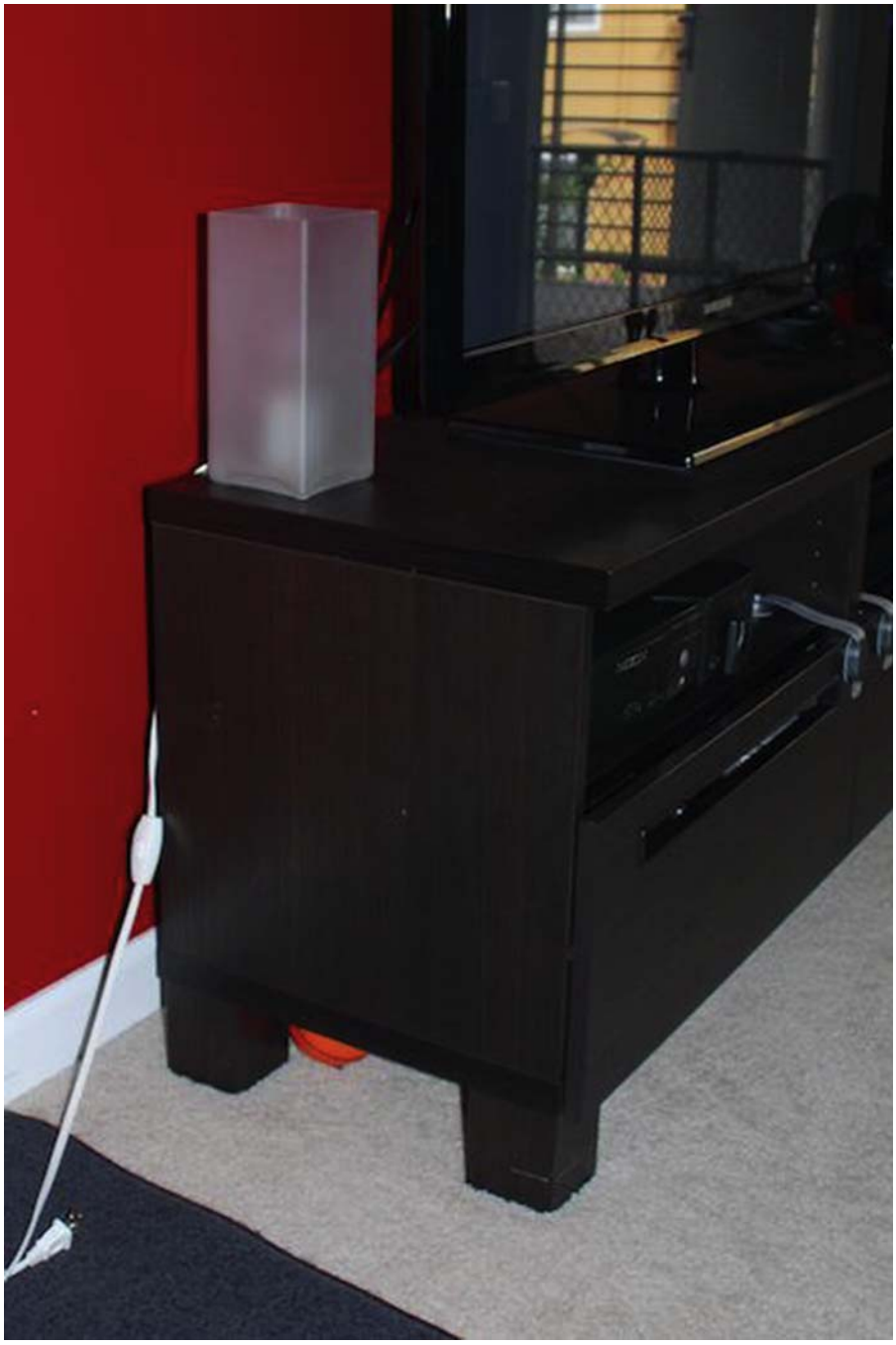


FIGURE 15.17 The lamp and the table that it was ordinarily situated on in the father's home.

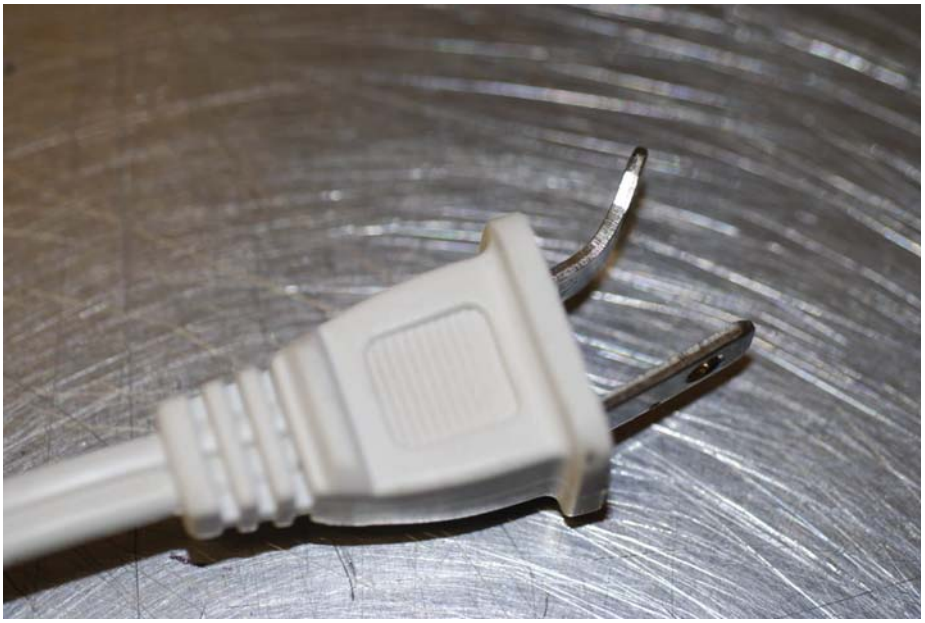


FIGURE 15.18 Photograph of the plug, demonstrating the bent prong.

CASE STUDY #3: ACCIDENTAL VERSUS INTENTIONAL HEAD INJURY IN A TODDLER

This case study concerns a skull fracture sustained by an 8-month-old male infant who was in the care of his father. The child was crawling around the family room of the father's home, and the father left the room for a short time. He heard the child begin to cry and rushed back to the room to see what had happened. He found the child in obvious distress, with a small mark at the back of his head and swelling that was rapidly forming. He also found a small glass lamp lying next to the child, unplugged. The position of the lamp on the table where it is normally situated is depicted in Fig. 15.17. Figure 15.18 is the appearance of the plug when the father examined it.

The father took the child to the emergency department of the nearest hospital where a physical examination was conducted and a CT scan of the child's head was performed. The scan demonstrated a comminuted skull fracture to the left parietal aspect of the child's skull with subgaleal hemorrhage (bleeding between the scald and the skull), subdural bleeding over the surface of the brain, and bleeding within the brain from a contusion, with associated subarachnoid bleeding. Images from the CT scan are reproduced in Figs. 15.19 and 15.20.

A pediatrician who examined the child at the emergency department was shown photographs of the lamp and the table the lamp ordinarily sat on. Additionally, the pediatrician was shown a photograph of the lamp from a cell phone (investigators later took

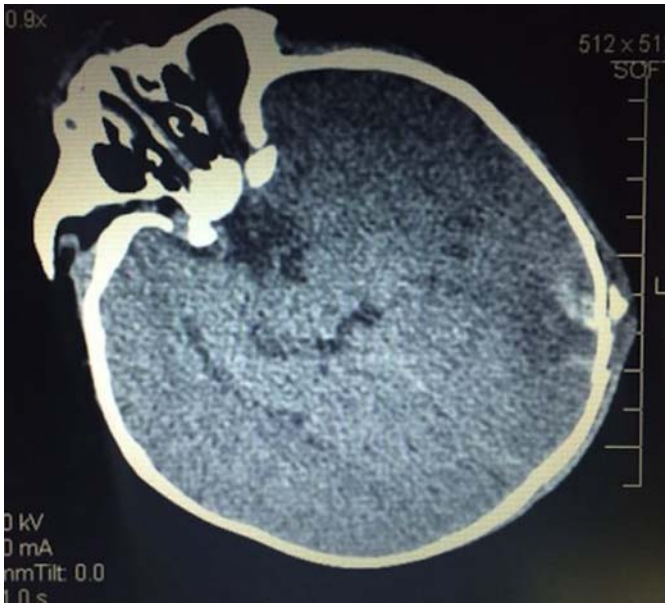


FIGURE 15.19 Image from CT scan performed on the infant at the emergency room.

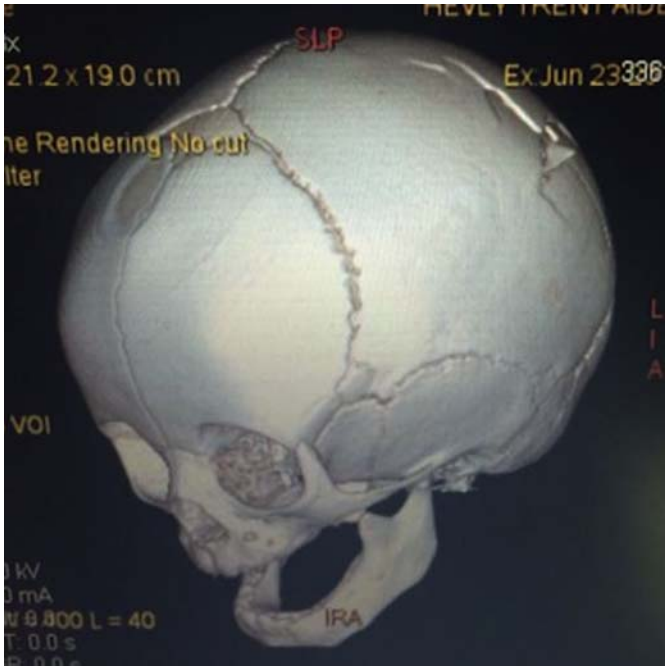


FIGURE 15.20 A 3D reconstruction of the CT scan examination performed in the emergency room, demonstrating in the discrete pattern of the fracture consistent with an impact from a pointed object, with a 3 mm depressed and comminuted fracture noted.

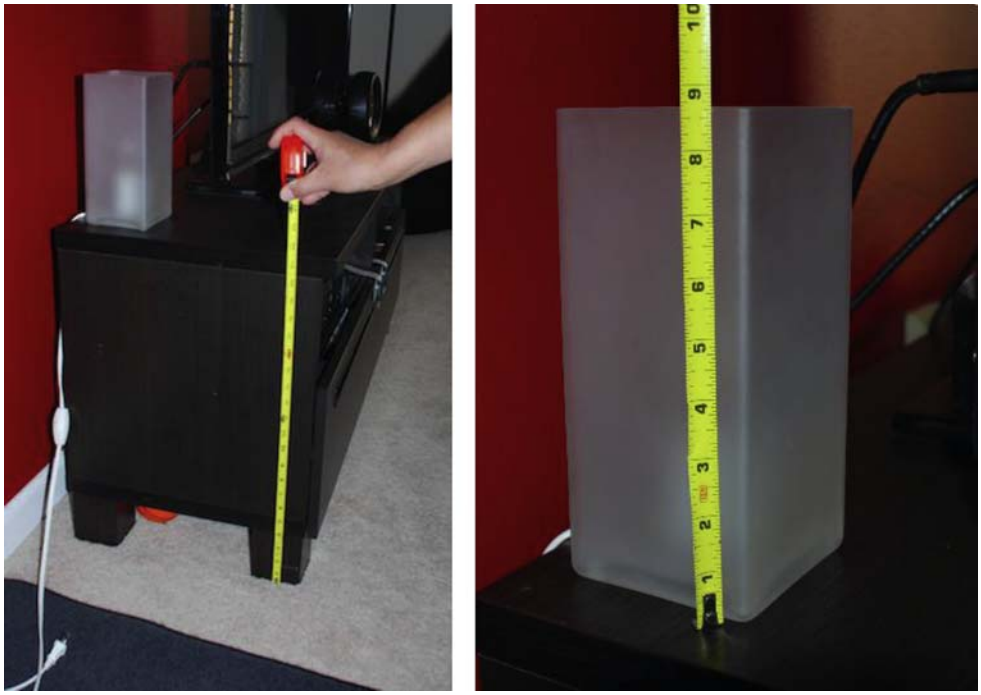


FIGURE 15.21 Photographs of measurements of the table and lamp performed by investigators for law enforcement after it was determined that the injury could not have resulted from an impact from the lamp.

measurements at the scene and from the lamp (see Fig. 15.21), and the weight of the lamp was measured at 2.4 lb (1.1 kg)). Based on this information the pediatrician concluded that the history of how the injury occurred was "not consistent with this kind of injury" and that consequently "this kind of injury is *highly concerning* for inflicted injury (emphasis added)." Based on this conclusion the father was charged with child abuse. The child ultimately had a complete recovery from the injury.

The necessity of an FE analysis of causality was made imperative by the speculative assertion by the pediatrician regarding the cause of the injury. The pediatrician's conclusion was in the form of a logical fallacy called the conditional probability fallacy, described in Chapter 3, *Methods Used in Forensic Epidemiologic Analysis* (also known as the fallacy of the transposed conditional). The pediatrician erroneously concluded that the common sense conclusion that the risk of the injury from the history given by the father was "low" (ie, it "makes sense" that such a small lamp falling such a short distance would be unlikely to cause such a significant skull fracture and intracranial hemorrhage) was used as evidence to reach the seemingly complementary conclusion that the alternative explanation, that the injury was intentionally inflicted, was high. In essence the pediatrician used the presence of the injury in the context of the history of how it occurred as a *test* of intentional injury in the same way the injury pattern evidence described in the first two case studies was used. The flaw in the approach is that there is no indication of the false positive rate of the test for the circumstances in which it was used.

The fallacy is avoided by first identifying the elements of an *actual* investigation of injury risk from a falling object, which would include a biomechanical investigation of the range of forces potentially resulting from the known physics of the event, combined with an estimation of the injury threshold of the involved tissue (if known) (see Chapter 9, *Biomechanical, Epidemiologic, and Forensic Considerations of Pediatric Head Injuries* for additional discussion of methods). The next step is to examine the probabilities associated with the alternative explanations, such that a posttest probability of unintentional (or intentional) injury could be estimated. An appropriate analysis of the probability of intentional injury would require substantially more information than the pediatrician possessed at the time that the opinion was proffered (which resulted in the arrest and charging of the father). As an example of additional information that would have been useful, the preevent probability that the injury was intentionally inflicted could have been enhanced by the knowledge that the father had a previous history of child abuse, or that he was known to be violent. Conversely, if the father had a completely absent history of violence this would tend to decrease the probability that he had suddenly become violent, absent contemporaneous evidence that this was the case.

As a practical matter, the risk of injury from the history provided by the father would be very difficult to assess. This difficulty could be overcome with a counterfactual approach; ie, by assessing the implication that the injury could not have resulted from the history provided by the father.

An investigation of the biomechanical aspects of the event began with an assessment of the physics of the falling lamp. Measurements taken at the scene and from the lamp were used to assess the kinetic energy of the lamp at the time of head impact. Assuming a minimal fall distance from the table top to the child's skull of 18 inches (0.5 m), and 2.4 lb (1.1 kg) in weight, and using the fall height formula described in Chapter 9, *Biomechanical, Epidemiologic, and Forensic Considerations of Pediatric Head Injuries*, the head impact speed of the lamp would have been approximately 7.0 mph (11.2 km/h). The associated kinetic energy of the falling lamp would have been approximately 4 ft-lb (5.4 J). An important aspect of the analysis is to understand that it is not just the weight and speed of the falling lamp, but also the geometry and stiffness of the lamp that predicts injury. It is not difficult to understand that if the child was struck in the head by the pointed corner of the glass lamp that the injury potential of the impact would be greater than if the child was struck by a flat side or edge of the lamp. A calculation of the pressure exerted on the child's skull by one of the corners of the falling lamp, assuming a 0.25 inch (6 mm) stopping distance, indicated 147 MPa (megapascals) of mean pressure from the impact. A comparison with the failure thresholds demonstrated in experimental study of infant skull fractures indicated that the impact had the potential to exceed the fracture tolerance of an infant skull (Margulies and Thibault, 2000).

A further factor to be considered in the analysis was the deformation of the electrical plug for the lamp (Fig. 15.18). If the plug was bent because the child pulled on the cord, which then resulted in the lamp striking him in the head, the fall energy analysis would likely be an underestimation of the upper bound of the force, and thus injury risk, of the impacting lamp as the lamp could have accelerated toward the infant's head at a rate greater than the pull of gravity.

As a result of this analysis it was concluded that the pediatrician's assertion that the injury risk from the lamp impact was so low that intentional injury should be considered as the most likely explanation for the child's skull fracture and associated injuries was deemed speculative and inaccurate. Moreover, further investigation demonstrated no collateral evidence that supported a determination that the injury resulted from abuse. There were no additional injury findings indicating abuse in the child, no history of abuse or violence on the part of the father or any other caregiver, and no injury mechanism that explained the discrete skull injury other than the stated history. As a test for intentional abuse, the uninformed assertion that "it did not seem like the lamp could have caused the diagnosed injury" was neither precise nor reliable.

As an endnote to this case study the prosecution dropped the charges against the father of the child following review of the results of the aforementioned analysis.

CASE STUDY #4: FETAL DEATH FOLLOWING MATERNAL COCAINE INGESTION

In this final case study we describe another criminal prosecution in which the pivotal issue was one of the probabilistic assessments of evidence. The underlying facts of the case were as follows.

In 2006 a 15-year-old crack cocaine-using African-American female gave birth to a stillborn fetus at 37 weeks gestation (full term). A toxicologic examination of fetal blood indicated the presence of a small and nonlethal amount of benzoylecgonine, a cocaine metabolite. Based solely on this finding the pathologist who performed the autopsy on the fetus determined that the manner of death was homicide. Based on the laws in the US state where the birth occurred, the mother was charged with first-degree murder.

An FE analysis of causal probability was undertaken, in order to assess the reliability of the inference by the pathologist that the presence of a nonlethal level of cocaine in the fetal blood was the "most probable" cause of the fetal demise. The assumption by the pathologist was that the finding of cocaine metabolite in the fetal blood served as the sole explanation for the stillbirth. The assumption ignored the well-established fact that stillbirth occurs both with and without maternal–fetal cocaine exposure, and maternal–fetal cocaine exposure occurs both with and without stillbirth. Further, stillbirth occurs disproportionately among disadvantaged and women of color (Stillbirth Collaborative Research Network Writing Group, 2011), characterizations that both are accurate descriptions for the defendant mother. Although the etiology of stillbirth in individual cases is often unclear, a number of associated factors, including poverty, single motherhood, inadequate prenatal care, maternal age, infection, obesity, diabetes, thrombophilia, fetal genetic or structural abnormalities, and umbilical cord abnormalities have been identified. Notably, the mother was also diagnosed with a thrombophilia (a tendency to form blood clots).

An analysis of the epidemiologic literature, performed as the initial step of the case analysis, indicated nonsignificant elevation of risk for stillbirth secondary to maternal–fetal cocaine exposure (Miller et al., 1995; Wolfe et al., 2005).

In the FE analysis, the relationship between maternal–fetal cocaine exposure and stillbirth was considered to be plausibly causal but potentially *confounded* (see Chapter 3, *Methods Used in Forensic Epidemiologic Analysis*) by some of the previously mentioned factors. To further

quantify the relationship a case-specific analysis of hospital inpatient birth data was performed. Data from the Nationwide Inpatient Sample Database (NIS) of the Healthcare Utilization Project of the Agency for Healthcare Research and Quality of the US Department of Health were accessed. This database is described in more detail in Chapter 11, *Traffic Injury Investigation*, Chapter 12, *Traffic Injury Investigation: Product Defects*, Chapter 13, *Product Defect/Liability Investigation*, and Chapter 14, *Medical Negligence Investigations*.

Initially, a univariate analysis of the contribution of maternal cocaine presence to stillbirth risk, along with other known risk factors, was conducted. These findings were used to construct an adjusted model of the relationship between cocaine exposure and of stillbirth, using binomial logistic regression. The results of the analysis resulted in an odds ratio of 1.58 (95% CI 1.02, 2.45). This value was used as a CRR for the analysis and was converted to a probability of causation of 37%.

As a result of the FE analysis it was concluded that nonlethal maternal–fetal cocaine exposure in a case of stillbirth does not account for more than 50% of the cause of the stillbirth. The assumption by the pathologist that the presence of fetal cocaine was highly specific for the stillbirth in the individual case, and thus the manner of death was homicide rather than due to natural causes was rejected as erroneous. While the cocaine exposure *could* have caused the stillbirth, it could not be concluded that the exposure was the most probable cause of the stillbirth, much less that a homicide had been committed *beyond a reasonable doubt*, which was the relevant standard of proof for a criminal conviction in the jurisdiction where the crime was charged.

As a final note, the charges against the mother were ultimately dismissed.

References

Augenstein, J., Perdeck, E., Mostafa, K., Digges, K., Bahouth, G., Morgan, R., 2005. The Role of Intrusion in Injury Causation in Frontal Crashes. SAE Paper Number 2005-01-1376.

Freeman, M.D., Nelson, C., 2004. Injury pattern analysis as a means of driver identification. Laboratory Medicine 35 (8), 502–505.

Freeman, M.D., Hand, M.L., Rossignol, A.M., 2009. Applied forensic epidemiology: a Bayesian evaluation of forensic evidence in a vehicular homicide investigation. Journal of Forensic and Legal Medicine 16 (2), 83–92.

McNally, B. Summary of Motorcycle Friction Tests. http://mcnallyassociates.com/techpapers2/MCFrictionSummary.pdf (accessed 3.03.14.).

Margulies, S.S., Thibault, K.L., 2000. Infant skull and suture properties: measurements and implications for mechanisms of pediatric brain injury. Journal of Biomechanical Engineering 122 (4), 364–371.

Miller Jr., J.M., Boudreaux, M.C., Regan, F.A., 1995. A case-control study of cocaine use in pregnancy. American Journal of Obstetrics and Gynecology 172 (1 Pt 1), 180–185.

Siccardi, D., Cavaliere, R., Pau, A., Lubinu, F., Turtas, S., Viale, G.L., 1991. Penetrating craniocerebral missile injuries in civilians: a retrospective analysis of 314 cases. Surgical Neurology 35 (6), 455–460.

Smock, W.S., Nichols 2nd, G.R., Fuller, P.M., Weakley-Jones, B., 1989. The forensic pathologist and the determination of driver versus passenger in motor vehicle collisions. The need to examine injury mechanisms, occupant kinematics, vehicle dynamics, and trace evidence. American Journal of Forensic Medicine and Pathology 10 (2), 105–114.

Stillbirth Collaborative Research Network Writing Group, 2011. Causes of death among stillbirths. JAMA 306 (22), 2459–2468.

Wolfe, E.L., Davis, T., Guydish, J., Delucchi, K.L., 2005. Mortality risk associated with perinatal drug and alcohol use in California. Journal of Perinatology 25 (2), 93–100.

Glossary

AIS Abbreviated Injury Scale.

Association (Syn: correlation) The statistical relationship between events or variables. If the events occur more or less frequently together than one would expect by random chance, then they are considered to be associated. Associations are not necessarily causal.

ATD Anthropomorphic Test Device; a crash test dummy.

Attributable fraction (exposed), AF_e (Syn: attributable proportion (exposed), attributable risk, etiological fraction (exposed), relative attributable risk, probability of causation).

For a causal association, the proportion of the investigated condition that can be attributed to the exposure of interest. The AF_e is also defined as the proportion by which the incidence of the investigated condition among the exposed would be reduced if the exposure were eliminated.

Background Level, Rate The frequency of occurrence of an event during a specific time in the absence of an investigated hazard.

Bayes' Theorem (Syn: Bayes' Law) A method of revising or "conditioning" the probability of the occurrence of an event given the occurrence or nonoccurrence of an associated event or events.

Bias In epidemiology, bias refers to a form of error that may threaten the validity of a study by producing results that are systematically different than the true results. Two main categories of bias in epidemiologic studies are selection bias, which occurs when study subjects are selected as a result of another unmeasured variable that is associated with both the exposure and outcome of interest, and information bias, which is systematic error in the assessment of a variable.

Biomechanics The field of study pertaining to how force affects tissue.

BMI Body Mass Index, calculated by (weight in kg)/(height in m^2).

Case–Control Study A retrospective study design that starts with the identification of persons with a particular disease or injury and compares them with a control group of persons without the same disease or injury for exposures of interest. The results are presented in the form of odds ratios.

Causal Criteria Considerations that help to guide case-specific judgments about causality. Most commonly referring to the "Hill viewpoints" but often incorporating other factors and considerations, depending on the circumstances.

Causation The relationship between an antecedent event, condition, characteristic, or agent that produces a disease or injury outcome. General causation is concerned with the cause of disease and injury in populations and the proportion of the ill or injured population attributable to the exposure. Specific causation is concerned with the cause of disease and injury in individuals.

Cohort Study (Syn: concurrent, follow-up, incidence, longitudinal, panel, prospective study) A study that starts with the identification of persons who have been exposed to a suspected cause of injury or disease and compares them to an unexposed group of persons for rate of occurrence of the disease or injury. The results are presented in the form of risk ratios.

Complement For the probability of an event or outcome or measurement A, the complement is (not A). Calculated by $[1 - A]$.

Component Causes Refers to a situation in which multiple factors must act jointly to result in a given outcome, as none of the factors can result in the outcome alone. The model is useful for analyzing the effect of antecedent causes, the sequence and timing of causal events, and injury or disease-producing processes that operate at different levels.

Conditional Probability The probability of one event given that another event has occurred.

Confidence Interval (CI) A range of values constructed around a point estimate of effect (typically an odds ratio or relative risk) that indicates, at a specified level of "confidence," that the true value falls within the range. When a 95% confidence interval does not include 1.0 this is typically interpreted, in terms of statistical significance, as the equivalent of a p-value that is <0.05.

Confounding A situation in which a noncausal extraneous or "nuisance" variable is associated with both the exposure and outcome of interest.

Death Rate [Syn: mortality rate] The proportion of a specified population that has died during a certain period of time. Death rate may also refer to a per-event metric that ignores time (eg, the death rate associated with a gun shot wound to the head).

Dependent Variable The outcome or measured variable of interest, and related to the value of the independent or predictor variable. In a cohort study the dependent variable is the disease or injury outcome, and in a case–control study it is the exposure of interest.

Diagnosis The process and result of assessing the presence of disease in an individual, family, group, or community.

Disease Any abnormal affliction in an individual.

Epidemiology The scientific study of the distribution, determinants, and deterrents of injury and disease in populations of people.

False Negative An incorrect negative test result in an individual with the condition of interest.

False Positive An incorrect positive test result in an individual without the condition of interest.

Forensic Epidemiology (FE) A hybrid discipline of epidemiology and forensic medicine, concerned with the application of epidemiologic methods and data as a means of investigating specific causation in a legal or forensic context.

Forensic Medicine (FM) The branch of medicine concerned with the interpretation of medical issues and evidence in a civil or criminal legal matter, such that the interpretation can be presented as a factual opinion suitable for the relevant standard of admissibility in a court of law.

Forensic Pathology (FP) The branch of forensic medicine devoted to the investigation of the manner and cause of death in deceased persons.

HN-NISS Head-Neck Injury Severity Score. The sum of the squares of the three highest AIS severities, only from the head or cervical spine regions.

Incidence The number of cases of disease or injury occurring during a given period in a specified population. Commonly presented as the number of cases occurring per a multiple of 10 (1000, 10,000, or 100,000, etc.) of the at-risk population per year.

Independent Variable The predictor variable under study, which is related to the occurrence or value of the dependent variable. In a cohort study the independent variable is the exposure, and in a case–control study it is the presence of the disease or injury of interest.

Injury Physical harm resulting, most commonly, from the external application of energy at a level that exceeds the tolerance of the exposed tissue. Injury can also result from the deprivation of an agent that would normally prevent or ameliorate a pathological process (eg, oxygen, pharmaceutical agent, safety device).

Injury Pattern Analysis (IPA) A method of matching medical evidence of injury with crash reconstruction, occupant kinematics, and epidemiologic data in order to draw inferences regarding the seating position, restraint use, ejection route, and other parameters of occupant status, typically in the circumstances of a fatal crash investigation.

ISS The Injury Severity Score. A composite injury score that is comprised of the sum of the squares of the three highest AIS severity scores from three different body regions.

Kinematics The study of occupant movement for a particular crash scenario. Based on the disciplines of both crash reconstruction (in order to establish the vehicle kinetics) and biomechanics.

Life Expectancy (Syn: survival projection) The average or median number of years an individual of a given age is expected to live if current mortality rates continue to apply. In a forensic setting the value is most accurately described as a range falling within what is considered to be likely to occur on a "more probable than not (>50%)" basis.

Meta-Analysis A statistical analysis of the results from separate or independent studies, typically including an examination of the differences, commonalities, and variance among the results and leading to a quantitative summary of the synthesized results.

NASS-CDS The National Automotive Sampling System-Crashworthiness Data System, a data gathering branch of the NHTSA.

NASS-GES The National Automotive Sampling System-General Estimates System, a data gathering branch of the NHTSA.

NEISS The National Electronic Injury Surveillance System.

NHTSA The National Highway Traffic Safety Administration, the United States governmental agency dedicated to road traffic.

NINDS The National Institute of Neurological Disorders and Stroke.

NISS The New Injury Severity Score, a composite injury score that is comprised of the sum of the squares of the three highest AIS severity scores from three different body regions, regardless of the body region.

Positive Predictive Value (PPV) The probability the condition is present given a positive test.

Posttest Probability A Bayesian method of combining a preevent prevalence of a condition with a PPV in order to arrive at the postevent probability of the condition or outcome.

***P*-Value (probability)** The probability that future testing will result in a value that is equal to or more extreme than what was observed in a test, assuming that the tested hypothesis is true. Before the test is performed, a threshold value is chosen, called the significance level of the test, most commonly 5%, but occasionally 1% or 10%, and denoted as the α (alpha) of the test.

If the *p*-value is equal to or smaller than the chosen alpha level, then the null hypothesis should be rejected, as the result is deemed "statistically significant." The meaning of *p*-values in hypothesis evaluation is vigorously debated in epidemiology and statistics, and it is improper to place too much weight on this measure without considering all of the elements of an investigated hypothesis. This is particularly true for the use of the *p*-value in a forensic setting, where often too much reliance on this test parameter may lead to Type I or Type II error.

Prevalence The total number of individuals who have a disease or injury during a specified time divided by the total relevant population at risk for the disease or injury.

Random Error Errors in measurement that lead to inconsistency in the result, which are randomly scattered about the true value. All measurements are prone to some degree of random error, the degree of which is indirectly related to the sample size.

Relative Risk (RR) (Syn: Risk Ratio) Risk ratios that quantify disease frequency differences between groups with different exposure levels.

Review, Systematic The use of methods to limit bias in the critical assessment and synthesis of all relevant epidemiologic studies on a focused topic.

Risk A probability that an event will occur (eg, that an individual will be ill or die within a specified period of time or will be injured due to a certain exposure).

Risk Difference (RD) (Syn: absolute risk reduction) The absolute difference between two risks.

Sensitivity (Syn: true-positive rate) The probability of a positive test given the presence of the condition of interest.

Specificity (Syn: true-negative probability) The probability of a negative test given the absence of the condition of interest. The false positive rate is the complement of specificity (false positive = (1 − specificity)).

This page intentionally left blank

Author Index

'*Note:* Page numbers followed by "f" indicate figures, "t" indicate tables, and "b" indicate boxes.'

Subject Index

'*Note:* Page numbers followed by "f" indicate figures, "t" indicate tables, and "b" indicate boxes.'

C

D

E

F

G

H

O

P

T

U

V

W

Y

Z

FORENSIC EPIDEMIOLOGY

Principles and Practice

Edited by Michael D. Freeman and Maurice P. Zeegers

It is an inescapable fact that causation, both generally (in populations), and specifically (in individuals), cannot be observed. Rather, causation is determined when it can be inferred that the *risk* of an observed injury or disease from a plausible cause is greater than the *risk* from other plausible causes. While many causal evaluations performed in forensic medicine are simplified by the fact that the circumstances surrounding the onset of an injury or disease clearly rules out competing causes (eg, a death following a fall), there are many cases that present a more complicated picture. It is these types of investigations, in which an analysis of comparative levels of risk from competing causes is needed to arrive at a reliable and accurate determination of the most likely cause, that forensic epidemiology (FE) is directed at.

In *Forensic Epidemiology*, the authors present the legal and scientific theories underlying the methods by which risk is used in the investigation of individual causation. Methods and principles from epidemiology are combined with those from a multitude of other disciplines, including general medicine, pharmacology, forensic pathology, biostatistics, and biomechanics, *inter alia*, as a basis for investigating the plausibility of injury and disease exposures and mechanisms. The ultimate determination of the probability of causation (PC) results from an assessment of the strength of association of the investigated relationship in the individual, based on a comparison between the risk of disease or injury from the investigated exposure versus the risk of the same disease or injury occurring at the same point in time in the individual, but absent the exposure.

The principles and methods described in *Forensic Epidemiology* will be of interest to those who work and study in the fields of forensic medicine, epidemiology, and the law.

Key Features

- Historical perspective on how epidemiologic evidence of causation has been used in courts in the US and Europe
- Theory and science underlying the use of risk to assess individual causation
- Primer on epidemiologic methods, and various measures used to arrive at individualized comparative risk assessments and PC
- The use of statistical methods applied to publicly available data for *ad hoc* analysis of PC applicable to the specific circumstances of a case
- Background on complementary disciplines, including forensic pathology, death investigation, biomechanics, and survival analysis
- Examples of applied FE in the investigation of traffic injury and death, automotive and other product defect litigation, medical negligence, and criminal prosecution and defense

LAW / Forensic Science,
MEDICAL / Epidemiology,
MEDICAL / Public Health

ISBN 978-0-12-404584-2

9 780124 045842

ELSEVIER

ACADEMIC PRESS

An imprint of Elsevier
elsevier.com