Current Approaches to OCCUPATIONAL HEALTH 3

Edited by
A. WARD GARDNER
MD FRCPI FFOM DIH

WRIGHT
Bristol 1987

Published under the Wright imprint by
IOP Publishing Limited
Techno House, Redcliffe Way, Bristol BS1 6NX

British Library Cataloguing in Publication Data

Current approaches to occupational health 3.
1. Medicine, Industrial
I. Gardner, A. Ward
616.9'803 RC963

ISBN 0 7236 0739 7

Printed in Great Britain at The Bath Press, Avon

To Professor Richard Schilling
who has taught so many of us so much
about occupational health: with warmest
personal regards and all good wishes

Preface

The theme that runs through this book is a wish to illuminate some current problems, to give perspective and to review progress. It is, as I remarked in the prefaces to the previous volumes, like its predecessors 'a *potpourri*, the flavour of which is determined by the choice both of authors and of subjects.' I am grateful to reviewers and colleagues for suggesting topics for this book and would welcome further ideas for the future. The authors have laboured to produce their chapters within a short time span so that their contributions are topical and informative. I am grateful to them for their devotion to the task, for their toleration in matters editorial and have pleasure in thanking them all most warmly.

My colleagues, both in the Medical Research Council Environmental Epidemiology Unit and elsewhere, have been helpful and tolerant in discussing matters related to this book and I am most grateful to them all for giving their time and expertise.

Ward Gardner

Contributors

J. A. Bonnell MB BS FFOM

S. E. Brill MB ChB FRCP(Lond) FFOM DIH

D. Coggon MA BM PhD MRCP AFOM

D. G. Davies BTech (Hons)

A. Downie MSc MB ChB MFOM DIH

A. Ward Gardner MD FRCPI FFOM DIH

D. W. Gardner BSc MS

A. R. Hale MA PhD ABPsS MErgS MIOSH

Alun H. Jones MA MB BChir FFOM DIH OStJ

Tarek M. Khattab MB ChB DM DCv MPH MIH MFOM RCPI

J. A. Lunn MD FFCM FFOM DIH

G. Matthews MB FRCP FFOM

D. S. Minors BSc PhD

V. S. G. Murray MB BS MSc DIH AFOM

B. S. Ricketts MRCP MRCPsych

R. J. G. Rycroft MD MRCP MFOM DIH

A. R. Scott BSc MB MFOM

W. J. C. Scott VRD MB ChB FRCGP MFOM DIH

A. J. M. Slovak MB MFOM

Sheena A. Waitkins BSc PhD DipImmun MRCPath

J. M. Waterhouse BA DPhil

David T. H. Weir MA DipPSA

Contributors

John Bonnell qualified at King's College Hospital in 1948. He served his 'apprenticeship' in occupational medicine under Donald Hunter at the MRC Department for Research in Industrial Medicine at the London Hospital from 1950 to 1958.

He has worked in the Electricity Supply Industry since 1958, first as Deputy Chief Nuclear Health and Safety Officer of the Central Electricity Generating Board then as the Medical Adviser to the Board and to the Electricity Council since 1978.

He has contributed both nationally and internationally to occupational medicine on a wide range of topics. These have included papers on the hazards of ionizing radiation, radiation risk analysis, radiation protection procedures and the toxicology of metals notably mercury and cadmium. He has led a number of research projects within the electricity industry, particularly on asbestos-related diseases and, latterly, extensive studies on the health-related problems of electromagnetic fields.

He has a special interest in occupational mental health and has been deeply involved in the development and teaching of occupational medicine in the UK during the last 20 years.

Monty Brill is currently Chief Medical Officer of Ford Britain and President of the Society of Occupational Medicine (1985/6). He is a graduate of the University of Glasgow, spent 14 years in general practice before entering full-time Occupational Medicine in 1970. He is Chairman of the Special Advisory Committee on Occupational Medicine of the Joint Committee on Higher Medical Training. He has been a member of the Qualification Committee of the Faculty of Occupational Medicine and was a lecturer and examiner for the Occupational Health Nursing Certificate.

David Coggon read mathematics at Cambridge University before taking up a medical career. After SHO and registrar posts in general medicine he became a member of scientific staff at the Medical Research Council's Environmental Epidemiology Unit at Southampton in 1980. His main research interest is the epidemiology of diseases related to occupation, in particular cancer and musculoskeletal disorders.

David Davies is Executive Director of the Human Sciences and Advanced Technology (HUSAT) Research Centre at Loughborough University of Technology. HUSAT is widely regarded as the UK's foremost research institution in the field of ergonomics and human factors relating to the design and introduction of advanced technology.

On graduation from Loughborough University, he went into the steel industry to set up a new ergonomics unit. This grew from serving one steel works to a corporate unit with nationwide responsibilities. A major aspect of that unit's work was, and is, the health and safety of the workforce concerned, both on the shop floor and in the office.

He moved on from the steel industry to a wider consultancy role, carrying out assignments for the European Commission, Manpower Services Commission, Health and Safety Executive, ICI, Plessey and other private sector organizations. Again, a major aspect of much of his work was the health and safety of the people concerned.

In recent years he has been executive director of HUSAT. One of the particular interests of the research centre is the health and safety of office workers.

Archie Downie has recently been appointed a Consultant in Occupational Health with the Greater Glasgow Health Board after almost 14 years experience in both the upstream and downstream aspects of the oil industry. His particular interests have been hyperbaric medicine, hearing conservation programmes and alcohol-related disorders. His interest in teaching of occupational medicine led to him arranging the first overseas group in Saudi Arabia for the Distance Learning Course in Occupational Medicine centred in Manchester.

Ward Gardner is a graduate of Glasgow University and has worked for most of his life in occupational health/medicine and mainly in the oil and petrochemical industries, in maritime medicine and in safety. He has special interests in first aid, problem drinking and alcohol abuse at work, absenteeism, and writing about occupational medically related topics.

He now works as an independent consultant in the fields of occupational health, occupational medicine and in the medicolegal field in relation to occupational injuries and occupational diseases. Since 1985 he has been consultant in occupational medicine to the Medical Research Council's Environmental Epidemiology Unit at Southampton and teaches occupational medicine to undergraduates at Southampton University and to postgraduates in the Wessex Region.

He is a member of the examining board of the Faculties of Occupational Medicine both in Ireland and in London.

David Gardner works in the field of air pollution and is a doctoral student in the Plant Physiology Program at the university of Minnesota. His doctoral thesis is concerned with the long-term effects to soil and vegetation of emissions from a major coal-burning power-generating facility. He has extensive experience in air-pollution effects to vegetation and has been involved in a number of studies to determine the chemistry of rainfall and aerosols in Minnesota and other parts of the United States. He received his Master of Science degree in 1980 from the University of Minnesota in the Department of Plant Pathology where his research involved a detailed study of the chemical properties of rain in Minnesota. He obtained his Bachelor of Science degree in botany from the University of London in 1975.

Andrew Hale graduated in Natural Sciences from Cambridge in 1966 and joined the National Institute of Industrial Psychology to conduct research into occupational accidents and human error in the engineering and electricity distribution industries. He joined the newly established Safety and Hygiene Group (later the Department of Occupational Health and Safety) of the University of Aston in 1972 as lecturer and later senior lecturer. There he developed the Master's course and carried out research into hazard perception, safety training, the organization of health and safety, and the role of safety professionals and factory inspectors. He was a member of the National Examination Board in Occupational Safety and Health and lectured widely in Scandinavia, France, Canada and India on human factors' aspects of health and safety. In 1981 he led a multinational team in a project to develop training for health and safety inspectors in the petrochemical industry in Algeria. Since 1984 he has been Professor of General Safety Science at the Delft University of Technology in the Netherlands, heading a research and teaching group working in the areas of traffic and home, as well as occupational safety.

Alun Jones joined the Southern Region of British Rail in 1951 after service in the Royal Air Force. In 1966 he became assistant to the Chief Medical Officer of the British Railways Board specialising in environmental medicine over the whole range of the Board's activities. He developed a special interest in human factors, particularly in the field of passenger comfort and later in the special problems associated with the operation of the channel tunnel. He was general secretary of the Ergonomics Research Society (now the Ergonomics Society) for some years. He gained considerable experience of occupational medicine and human factors work in Europe during that period.

He joined the Employment Medical Advisory Service at its inception in late 1972 and became the Senior Employment Medical Advisor for Wales in 1974, a position which required considerable personal involvement in the application of statute law to occupational health. He experienced at first hand the transition from the era of the Factories Act to the enormously expanded activity of the Health and Safety at Work Act and more recently the increasing impact of the European Communities on British legislation. He has a special interest in the industrial and social history of the nineteenth century.

Dr Jones is now in private practice and is a consultant to the South Wales Electricity Board.

Tarek M. Khattab graduated from Cairo University in 1957, then served three years in the University Hospitals as a house officer and internal medicine registrar, as well as a year in the Ministry of Public Health. During that period, he obtained his diplomas of internal medicine (DM) and of cardiovascular disease (DCv). In 1962 he moved to Saudi Arabia, where he worked as a physician in ARAMCO clinics. He took a leave of absence to study at Johns Hopkins School of Hygiene, where he obtained his MPH in 1971, following which he was appointed as the corporate Occupational Health Physician. After another leave of absence at Harvard School of Public Health, he obtained his MIH in 1974. His efforts to create an occupational health organization were approved by corporate management and he was appointed as Chief of Occupational Medicine Division in 1977, the position which he is still holding. In 1980 he was admitted to the membership of the Faculty of Occupational Medicine, the Royal College of Physicians of Ireland. Presently, he is a Senior Associate at the Johns Hopkins School of Hygiene, and is a Member, WHO Expert Advisory Panel on Occupational Health.

John Lunn is district occupational health physician to the Paddington and North Kensington Health Authority, senior lecturer in occupational medicine and director of the student health service at St Mary's Hospital Medical School, London. His interest in occupational diseases was first developed when working at the Slough Industrial Health Service where he became deputy medical director. His special interest in occupational health problems of hospital workers began in 1970 when he established the occupational health unit at Northwick Park Hospital. Before taking up his present posts in 1980 he held a senior lectureship at St George's Hospital Medical School. He has written widely on occupational medicine subjects including a book on the health of staff in hospitals. He has a particular interest in alcohol-related health problems and is a regional adviser to the Medical Council on Alcoholism.

Geoffrey Matthews qualified from the London Hospital in 1953 after wartime service in the Navy. He spent some years in Canada and joined Esso Petroleum at Fawley in 1959. He joined BOC Ltd in 1972.

David Minors has studied circadian rhythms in humans since 1970 and completed a PhD thesis on this topic in 1975. His work in collaboration with Dr Jim Waterhouse, has concentrated largely upon changes in the human circadian system that result from imposed changes in sleep-wake schedules. These studies have been performed in an Isolation Unit, unique in this country and one of only a few in the world. In this unit, time-zone transitions may be simulated or subjects may live on non-24-hour days by providing them with a clock which runs at the wrong rate. More recently, his studies have concentrated upon the effects of irregular schedules of sleep and wakefulness. He has collaborated in writing numerous reviews on circadian rhythms and has co-authored the specialist textbook *Circadian Rhythms and the Human*. He has contributed chapters to physiology texts and is a managing editor of *Chronobiology International*.

Virginia Murray is the first occupational toxicologist in the National Health Service and has worked for the National Poisons Unit New Cross Hospital, London since 1980. She graduated in 1975 from the Royal Free Hospital School of Medicine and after three years, from houseman to registrar, at the Royal Free, took a MSc in Occupational Medicine at the TUC Centenary Institute of Occupational Medicine in 1979. She spent four years working in industry as the staff medical adviser to Harrods and during this time, carried out various research projects including taking part in a Health and Safety Executive funded literature search into the reported teratogenic effects of 50 chemicals. After being awarded the 1984 Esso Travelling Fellowship by the Faculty of Occupational Medicine to visit nine poison centres in Europe, she was given the task of being rapporteur for two international meetings, one in Geneva (the role and responsibilities of a poison control centre within a poison control programme, October 1985) and one in Munich (public health response to acute poisoning—poison control programmes, December 1985). She has now been asked to help to coordinate and co-author the guidelines on poison centres and poison control programmes from the International Programme on Chemical safety (WHO/UNEP/ILO). Within the UK, she has carried out various research work and is currently completing a major project on the epidemiology of acute occupational poisoning, funded by the Health and Safety Executive. She is also undertaking a study on the evaluation of the data in a poison centre and its use in relation to safety and health at work for the European Economic Community. She is married with two small children.

Bruce Ricketts is a consultant psychiatrist at the Royal South Hants Hospital, Southampton, and an honorary clinical teacher in the University of Southampton. He is a graduate of Cambridge University and Guy's Hospital Medical School. He trained in psychiatry at the Maudsley and Bethlem Royal Hospitals, London; as a senior registrar he rotated between those hospitals and St Mary's Hospital, Paddington. Since taking up his present post eight years ago he has worked in the general adult field taking a special interest in psychiatric day care. He has held a three-year appointment as tutor to the postgraduate psychiatric trainees in the district. More recently he has become chairman of the psychiatric departmental management team in his hospital and is involved in the establishment of a psychiatric service network independent of old institutional buildings. He is co-author of papers on long-term psychiatric out-patient care and the use of multiple choice questions in medicine.

Richard Rycroft graduated from Cambridge University and Guy's Hospital. He trained in both dermatology and occupational medicine, including a Dowling Travelling Fellowship at the Institute of Environmental Health of the University of Cincinnati. In 1977 he was appointed to a joint post in London as Consultant Dermatologist at St John's Hospital for Diseases of the Skin and Senior Employment Medical Adviser (Dermatology) to the Health and Safety Executive. He has since been practising occupational dermatology with a regular commitment to visiting places of work. He has published papers, contributed chapters to books and lectured widely on subjects within the field of occupational dermatology. His MD dissertation was based on detailed investigations of cutting oil dermatitis and was awarded the Raymond Horton-Smith Prize for 1982.

Alan Scott has researched into the epidemiology of occupational accidents for several years. This work has included the effects of shift work on accidents, and is orientated towards the workplace rather than the laboratory. He has just completed a large study of occupational accidents in the coal-mining industry and has published work on such varying topics as high-pressure injection injuries, shift work and health and cardiac disease and employment. He is a regular contributor of articles on occupational health to the British Medical Journal.

Jeffrey Scott graduated at the University of Glasgow and has worked in general practice and occupational health for over thirty years. For the past twelve years he has been a part-time Employment Medical Adviser with the Health and Safety Executive, while continuing to work in general practice. He was a member of the General Medical Council for eight years.

Andy Slovak is an occupational physician in the pharmaceutical and chemical industries in which he has now worked for over a decade.

He qualified from University College, London in 1970. His main special interests are in occupational asthma, reproductive effects of work and in the investigation of more efficient and effective ways of delivering occupational health care. He has contributed to the scientific literature on these and a number of other subjects. He is also a student of 20th century occupational medical history and is consequently a keen collector of books and other memorabilia of the period. He has a weakness for collecting and occasionally riding elderly and decrepit motorcycles.

Sheena Waitkins was born in Scotland in 1947, she was educated at St Patrick's High School, Coatbridge and graduated from Glasgow University with BSc Honours in Microbiology in 1970. From Glasgow she was awarded an MRC scholarship to study the pathogenesis and identification of *Neisseria gonorrhoeae* at the University of Sheffield, Department of Medical Microbiology. Her work established standard identification procedures used to identify gonococci in routine laboratories and she contributed to the knowledge of intracellular action of the Neisseria species on tissue culture cells and phagocytes in *in vitro* and *in vivo* experiments. In 1973 she was appointed lecturer in Medical Microbiology and was course tutor for BSc students.

In 1974 she was awarded a WHO Travelling Scholarship to work with the world expert on the gonococcus, Dr Alice Ryan at Copenhagen. During her stay in Denmark Dr Waitkins published several papers on the role of gonoccocal L-forms in the pathogenesis of non-specific urethritis. She also continued her interest in rapid identification procedures and established that penicillinase and non-penicillinase producing gonococci could be identified by using a fluorescent antibody technique.

On her return to the UK in 1975, Dr Waitkins was invited to work in the Streptococcal Reference Unit at PHL Colindale. Between the years 1975 and 1977 Dr Waitkins revolutionized the identification of viridans streptoccoci by introducing the use of the rapid API Enzyme system which she further developed in collaboration with API France to the now universally used and popular routine identification of pathogenic streptococci in all British laboratories.

While in London, Dr Waitkins attended the Chelsea College, part-time, and passed the Diploma in Immunology in 1976. She became a member of the Royal College of Pathologists in 1983.

During her training period for the MRC Path Dr Waitkins continued her research interests and published standard works on legionnaires' disease. She first managed to link the presence of legionella to the hotel water system at a Corby hotel, thereby establishing the risks of acquiring legionnaires' disease in Great Britain. While at Rhyl, she collaborated with clinical colleagues and reported the first isolation of metronidazole-resistant *Trichomonas*, thus explaining treatment failures noted in the area.

In 1982 Dr Waitkins was appointed the Director of the WHO Reference Laboratory and PHLS Reference Laboratory for the study of leptospirosis. During the years of her present appointment, Dr Waitkins has conducted several epidemiology studies on the incidence of leptospirosis in many risk-related occupations and hobbies. The current awareness of cattle-associated leptospirosis was due to the work carried out in the PHLS Reference Laboratory. Dr Waitkins has established research projects in the pathogenesis of leptospires, their interactions with antibiotics and the presence of leptospires in the environment and animal species for example coypu in East Anglia and the presence of leptospires in rivers throughout England and Wales.

She is currently studying the inter-taxonomic relationships of various cow strains of leptospires to establish the possibility of a human vaccine. New, rapid methods of detecting leptospirosis, both cultural and serological, are under investigation. The laboratory, under Dr Waitkins's directorship contributes both to the epidemiological and basic scientific understanding of the disease.

Dr Waitkins is a member of the WHO Consulting panel for Zoonotic Diseases and is a member of several related committees such as the International Taxonomic Sub-Committee on Leptospires and the National Subgroup of The Water Committee, to name but a few.

Jim Waterhouse has collaborated since 1969 with the late Professor John N Mills and Dr David Minors. His work has involved human volunteers in a variety of changed sleep-activity schedules which enable some properties of the circadian system to be investigated. His main interest is in the application of such studies to field situations and the dissemination of relevant advice to night workers and time-zone travellers. As well as collaborating in writing numerous reviews on circadian rhythms, he has co-authored the specialist text *Circadian Rhythms and the Human* and the more general textbook *Physiology and the Scientific Method*. He has also written chapters for basic physiology texts. He is a managing editor of *Chronobiology International* and frequently lectures on circadian rhythms to universities, scientific societies, drug firms and so on.

David Weir is the Head of the Department of Management Studies and Professor of Organisational Behaviour at the University of Glasgow. He graduated from Oxford University in 1960 and subsequently held posts in the Obstetric Medicine Research Unit, Medical Research Council, then at Leeds, Hull and Manchester Universities.

He was Dean of the Scottish Business School and a Member of the Finniston Committee on Engineering in Manufacturing Industry.

1. LOW-HUMIDITY OCCUPATIONAL DERMATOSES

R. J. G. Rycroft

An attractive young lady, who has had mild eczema since childhood, starts work as a receptionist in a large hotel. Her eczema gets gradually worse the longer she works there. A harassed-looking middle-aged businessman, involved in a take-over bid, itches all over after a hectic schedule of flights back-and-forth across the Atlantic. An ageing handyman in a modern electronics factory has a patchy facial eczema. It seems to have come up during a winter in which his redundancy has been on the cards. At first sight, the suggestion that there might be a common factor between all three of these skin problems seems improbable—unless perhaps it was stress. Yet these cases all resulted from the low humidity of indoor environments: hotel, aeroplane and factory. How could this be?

Below a water content of 10 per cent the outer tissue-paper-thin layer of the skin, the stratum corneum or horny layer, loses its softness and pliability.[1] The water content of the horny layer stays below 10 per cent when the relative humidity is less than around 50 per cent (relative humidity is the ratio of the actual vapour pressure of water to the saturated vapour pressure of water at the same temperature, expressed as a percentage). Drying out of the horny layer occurs most rapidly under conditions of high temperature, low humidity and fast-flowing air. The first two, and sometimes all three, of these conditions now occur in a large number of modern working environments.

It has therefore become necessary to recognize the effects of such working conditions on the skin. Those that are seen can roughly be divided into three: itching (pruritus), whealing (urticaria), and reddened flaky rashes (eczema). Dermatologists have long recognized similar, though not identical, skin problems arising from cold dry winter weather. Terms such as 'winter itch' have been coined to describe these related phenomena. The conditions arising from warm dry indoor working environments can conveniently be termed low-humidity occupational dermatoses.[2]

One point needs stressing at the outset. Low-humidity occupational dermatoses are more important than they might seem. There are at least three reasons for this. *First*, the annoyance and anxiety caused by itching can be out of all proportion to the visible signs of skin dryness. This can be highly destructive of good industrial relations. *Second*, the exigency of symptoms may lead to a large amount of time and money being spent on the investigation of other possible causes such as allergy. *Last*, the condition once diagnosed is usually amenable to remarkably simple and inexpensive measures. No one can afford to ignore this unstable borderland between skin physiology and pathology.

These skin problems are a prime example of the value in occupational medicine of making oneself as familiar as possible with the details of the working environment. It is far more difficult to make the diagnosis on an individual patient in a medical department than it is from looking at employees together as a group. But a warning is due. Time spent in reconnaissance is seldom if ever wasted: the problem should always be well considered from other angles before deciding on low humidity as the sole cause. It is not unusual, for example, for low humidity to be only one of two or more factors involved in an occupational dermatosis.

This chapter is intended to equip its readers with a realistic index of suspicion for low-humidity effects on the skin at work; to acquaint them with environments in which such problems may arise; and to indicate the ways in which the skin condition can be corrected. A number of other common disorders are discussed under differential diagnosis, since it is just as important not to over-diagnose this factor as it is to recognize it when it is there. From time to time I have imposed on the reader some point of general principle in occupational dermatology that seems to me to be illustrated by particular aspects of the outbreaks described. This is intended to broaden the usefulness of the discussion without, I hope, distracting from the subject at hand.

PRURITUS

The skin can itch following quite minor physiological alterations.[3] Individual itch thresholds vary widely. Atopics, for example, may relatively easily be stimulated to itch. A typical example of an outbreak of pruritus being caused by warm dry flowing air was provided by the introduction of hot unhumidified air downwards through a continuous vent around the periphery of an open-plan office. Complaints of itching came particularly from clerks sitting directly beneath this vent, though they came to a lesser extent from clerks working on the rest of the floor.

The worst case in this epidemic was that of a young female who already had atopic eczema, but whose skin was made very much worse by this environment. Beware of accepting such a case as typical: the remainder of the clerks had skin appearances varying from mild to scarcely detectable dryness (asteatosis or xerosis) and flaking (scaling).

It is not unusual for the worst case or cases in an outbreak of an occupational dermatosis to be atypical or even irrelevant to the more general problem. The occupational health practitioner should be reluctant to allow his clinical assessment of a widespread skin problem at work to rely too exclusively on the examination of the two or three worst cases, especially if they are self-selected or chosen by a representative without medical or nursing training. By all means examine such patients first, but then follow this up by talking to, and looking briefly at, as many other members of the working group as their cooperation—and your time—allows.

PRURITUS AND URTICARIA

Low-humidity occupational dermatoses do not only occur in offices. Ten years ago I was asked to investigate gathering complaints of severe pruritus and mild urticaria among men installing international telephone exchange equipment in a purpose-built new building.[2] The discrepancy between 'ten years ago' and the publication date of the reference just given, 1980, is a telling example of just how long it can take an unprepared mind to identify the real cause of an occupational skin problem. The problem eventually had to solve itself and, because all investigated causes had previously been rejected, the file on the problem had been kept open until that time. This at least allowed me to learn from my failure to identify the cause myself.

The alternative possible causes that I had previously investigated with essentially negative results are not uninstructive, since they adumbrate the range of conditions that may have to be considered in the differential diagnosis. The skin complaints comprised urticarial wheals and involved predominantly areas of skin covered by clothing, including the lower limbs.

Inhalable and ingestible allergens were considered. I well remember a searching examination that I made of the works canteen in pursuit of potentially allergenic foods. Glass fibre from filters within the air conditioning system was looked for in air samples: so near geographically, and yet still so far away mentally, did the investigation get to the source of the problem. Powder released from within electronic cables came under suspicion. Even psychological stress, such as might be induced by fear of redundancy after the contract was finished, was entertained as a possible explanation.

While these investigations continued, at a variable pace according to the pressure of other investigations, the situation progressively deteriorated. Much to my embarrassment the number of cases continued to rise. After a while, my secretary was scarcely able to bring herself to tell me that the company nurse was on the telephone again. During the summer of 1977 symptoms seemed to subside but such temporary alleviations can never be depended on: the autumn of 1977 saw a resurgence. I was fortunate enough to undertake a year's work in the USA at this juncture, though I must insist that the two events were totally unconnected.

By my return from the USA the problem was solved. This occurred when, following completion of the exchange equipment installation, the itching men no longer had to work for long periods in the carefully-controlled environment of the telephone exchange equipment. To prevent corrosion of metal contacts, temperature and humidity had been controlled within narrow limits by a double air-conditioning system: a central system supplying humidified air all year round, and a peripheral system supplying warm, unhumidified air during the winter months.

It was around the periphery of the telephone exchange floor that the worst-affected men had their temporary offices and it was in winter that their symptoms had been at their peak. The relative humidity around them had been approximately 35 per cent when they had been most symptomatic: their symptoms resolved when they eventually escaped into a relative humidity of 50 per cent and above.

Following this change, no further new cases presented and even previously persistent cases rapidly cleared up. The urticaria appeared to have been secondary to the scratching of itchy, dry skin, no doubt exacerbated by the men's natural frustration at the failure of their appointed investigator to work out the answer to their problem.

LOW-GRADE ECZEMA

In 1972 the Vancouver dermatologist, John Mitchell, justly renowned for his work in the great outdoors on plant dermatitis, reported a small but effective study that he had instigated in his own hospital.[4] He had noted that older male in-patients in the Shaughnessy Veterans Hospital frequently developed xerosis of the skin and mild eczemas. The relative humidity in the skin ward and two medical wards was measured during December and January. In all three wards the relative humidity varied between 20 and 40 per cent. The relative humidity of the air in one of the medical wards at 8.15 a.m. on 21 January was 16 per cent and the temperature 81 °F (27 °C). Mitchell commented, appropriately enough: 'These conditions may also prevail, I understand, in the Sahara desert.'

I have encountered at least 'semi-desert' conditions in factories manufacturing silicon chips and soft contact lenses. Since the first of these episodes, it has become clear that low humidity is a widespread problem in the new electronics industries where dry air is protective to the products during manufacture. In a silicon chip factory investigated early in 1979[2] the face was the sole site of complaints.

Four female process operators, all in their forties and fifties, had persistently complained of irritation and redness of their faces that winter. Their symptoms and signs had variably been ascribed by doctors to rosacea, the menopause, and even the combination of alcohol ingestion and trichloroethylene inhalation known as 'degreaser's flush',[5] because of their proximity at work to degreasing tanks. The latter two explanations, with their veiled implications respectively of sexism and alcoholism, were understandably not well accepted by the four women. Also, they had already found out for themselves that oral tetracycline did not cure their 'rosacea', whereas their own moisturizing creams had been surprisingly helpful. Their confidence in the medical profession had fallen to a low ebb.

Clinically, they all showed patchy redness and scaling over their cheeks and foreheads. Three had fair complexions and skins previously damaged by sun exposure (telangiectasia and elastosis). They all four worked at the same bench in one corner of the factory. They scored (scribed) lines between individual silicon chips in a block, separated (cracked) the chips along these scored lines, and finally subjected them to close visual inspection. This visually-demanding scribing and cracking required intense local illumination. In addition, vacuum pumps held chips in place for inspection; chips were stored in heated cabinets above the bench; and a diffusion furnace behind the bench was regularly checked for temperature through open doors. All these electrical devices were local sources of radiant or convective heat.

The reader will not be surprised to learn that the mean air temperature in this mini-environment was found to be a little over 77 °F (25 °C) and the relative humidity to average 35 per cent. While comparatively good conditions perhaps for the Shaughnessy Veterans Hospital male medical wards, they were still bad enough to dehydrate the facial skin of these scribers-and-crackers. The problem was solved when the relative humidity was raised with humidifiers so that it was at least 45 per cent in this corner and when the women made their use of moisturizing cream routine.

Complaints of facial itching in a soft contact lens factory reached epidemic proportions in the winter months of 1981.[6] By early April, 72 per cent of 78 mainly female employees examined showed patchy or diffuse superficial scaling of the facial skin, most noticeable on the cheeks, but also visible on the forehead, nose and neck. This scaling was frequently associated with a mild redness. Blondes were involved in a higher proportion than brunettes. Some had little in the way of signs, but much more in the way of symptoms. The shaven cheeks and chins of male employees tended to show a patchy redness without scaling, the scales presumably having been removed with the stubble.

Their large workroom was found to be air-conditioned rigorously, in order to protect the special acrylic polymer from which the soft contact lenses were machined. This hard polymer absorbs water and, when it does, swells and becomes soft, just like the horny layer of the skin. The ambient air was deliberately kept dry to prevent the polymer prematurely absorbing water from the atmosphere, and thus altering its dimensions during machining. Relative humidity and temperature in the workroom were continuously monitored. The answer to the problem had therefore already been plotted on a paper strip which was immediately available for inspection. The occupational hygienist is not often lucky enough to find his work already done for him.

At a peak period for symptoms during March, when the prevalence of symptoms had even led to threats of industrial action, the air temperature had varied between 68 and 74 °F (20–23 °C) and the relative humidity had remained largely between 30 and 35 per cent. The problem was solved by allowing the relative humidity to rise to 45–50 per cent and by the routine use of emollients.

Contributory factors

In both these outbreaks, but especially the second, there may have been additional factors tending to localize symptoms to the face. Much of the work carried out in both factories involved close proximity between the face and hot electrical machinery or inspection lamps. This almost certainly decreased the relative humidity of the facial micro-environment still further than in the mini-environment of the corner of the silicon chip factory or the more general indoor environment of the soft contact lens workroom.

In the second of these two factories, fine angular particles of hygroscopic acrylic polymer were recovered on swabs from the faces of employees. The percentage of free acrylic monomer (2-hydroxyethylmethacrylate) was below 0·3, so chemical irritation from this 'face powder' was probably not a factor. But rubbing the skin in response to pruritus would have ground these sharp particles into the already-compromised skin barrier. These skin lesions therefore probably resulted from the sum total of low humidity, high temperature, flowing air and mechanical microtrauma from plastic particles.[7]

When I presented this particular outbreak in the soft contact lens factory recently at a conference,[8] comments from other delegates indicated that skin symptoms frequently arose from a combination of low humidity and small particles, the latter being derived from sources such as paper, textiles or ceramics.[9] Many such particulates, being hygroscopic, would themselves be drying to the skin, further compounding the problem. It was also thought that the problem of low-humidity occupational dermatoses was better known to dermatologists than to ventilation engineers. Those who establish criteria for indoor environments needed to be made more aware of these problems.[10]

UNDER-REPORTING

It appears that the inherent difficulties of constructing convincing reports from outbreaks of low-humidity occupational dermatoses, with or without additional microtrauma, have conspired to make this group of disorders an under-reported phenomenon. Some readers, I hope, may by now have recognized an outbreak of skin trouble that they have seen in the past as possibly having involved low humidity. This surely underlines the benefit of occupational health journals sometimes publishing relatively anecdotal and unsubstantiated accounts of occupational skin problems of a non-allergic nature. Equally, investigators must be prepared to write them. It tends to be easier to put together a well-worked up case of a rare occupational contact allergen. Hence the literature on occupational dermatoses continues to give a misleading impression of the relative frequency of allergic and irritant contact dermatitis.

OTHER CONTRIBUTORY FACTORS

There are several other factors which may complicate a primarily low-humidity dermatosis, in addition to local heat sources and sharp or hygroscopic particles.

Dehydration of the horny layer not only makes it itch and break up, but also destroys its effectiveness as a barrier layer. This enhances the absorption of substances from the skin surface. Mild irritants, such as soap and detergents, can thus exert effects on a dehydrated skin that they would not achieve on a well-hydrated skin. Symptoms from low-humidity occupational dermatoses are often reported as exacerbated after showering or bathing. Because this usually occurs at home, such reports may give a spurious domestic relationship to symptoms. This can help to obscure the diagnosis.

A dry skin is also at an increased risk of sensitization from contact with ingredients of moisturizing creams used to combat the dryness. Such potential sensitizers include perfume, lanolin and preservatives, such as the parabens, Dowicil 200 (Quaternium 15) and Germall 115 (imidazolidinyl urea).[11] It is easy for the dermatologist to overlook an underlying low-humidity dermatosis in the professional excitement engendered by detecting such an allergy on patch testing. It is always germane to spare a thought as to whether the patient was having to increase the use of moisturizing creams for any reason prior to the onset of sensitization.

The relative humidity both in the home and outside must also be relevant to the speed with which a largely occupational low-humidity dermatosis takes hold. Most outbreaks of the problem that I have seen have peaked in the early months of the year. Spells of cold dry weather commonly occur at this time in the northern hemisphere and heating is turned well up in the home, both compounding any effect of the occupational environment. Often, occupational, domestic and general climatic conditions must all be working together towards the same end. But unless the occupational factor is recognized the problem may persist in susceptible skins well beyond the onset of spring.

OTHER OCCUPATIONS

The list of occupations in which I have seen cases of low-humidity occupational dermatoses is an ever-increasing one. Office workers, telephone exchange installers, patients in hospitals, and electronics and optical industry operatives have already been quoted. It remains for me to indicate the range of other occupations from which complaints commonly arise.

It should be clear that not only hospital patients but also the resident staff of hospitals, and not only the staff of hospitals but also the staff of hotels, restaurants, casinos and other leisure establishments, are liable to be affected by over-heated air-conditioned environments. Female croupiers in casinos tend to be less well-covered with clothing than their male counterparts, which is probably the reason why the croupiers whom I have seen with occupationally-related asteatotic eczema have so far all been female.

Main-frame computer rooms are kept dry to protect the equipment, which leads to problems in a similar way to that described in the soft contact lens factory. In a metal refinery, by contrast, the skin-drying effect was largely dependent on the very large number of air-changes per minute. This had been built into the environment as a safety factor should there be any leak of potentially-sensitizing precious metals.

The cabin crew of long-haul aircraft can also experience the drying effect of rapidly circulating air. A glance behind the scenes during a trans-atlantic flight may reveal a number of strategies employed by air-stewardesses to keep their skins in good condition, which even include the use of water sprays.

Down on the ground, travelling salesmen can dry out the skin of their lower legs in winter months from the enthusiastic use of their car-heaters to keep their feet warm while driving. A fully turned-up car heater provides optimal conditions for skin drying. The shins and calves take the brunt of the warm dry flowing air when the heat is directed downwards rather than up over the windscreen.

TREATMENT

The treatment of these conditions has already been indicated in most of the outbreaks discussed. It can be summarized as simply lowering the temperature, and thus indirectly raising the relative humidity; increasing the humidity directly, though this may be prohibitively expensive and also presents its own potential occupational health hazard, humidifier fever;[12] decreasing the frequency of air changes, as far as this is acceptable for other reasons; and the routine use of moisturizing (emollient) creams. The choice of emollient cream can usually safely and advantageously be left to individual employees. The routine use of an emollient is frequently all that is required, but it is almost as frequently overlooked as a solution, particularly by male employees.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of low-humidity occupational dermatoses should be kept in mind from the start if overdiagnosis of an attractively simple skin condition is to be avoided. A striking warning is provided by the case of a 68-year-old housewife whose home had a forced warm air heating system.[13] She developed itchy spots on the back and outside of her left arm and the back of her left thigh. The author of the case report elegantly demonstrated that it was not warm air that was causing these localized symptoms, but microscopic spines detached from her prickly pear cactus. These were drawn into the heating system, blown onto her clothing, and penetrated through it to her skin. The author does not relate whether the left side of her favourite armchair faced a warm air outlet but this is a tempting speculation.

There are at least two pseudo-occupational phenomena which should first of all be considered in the differential diagnosis. A chance cluster of non-occupational dermatoses in a workplace may result quite understandably in thoughts of occupational causation.[14] Pruritus may be a feature of many such dermatoses. Precise dermatological diagnosis of individual complaints should allow effective reassurance to be given. The background to such complaints is often worthy of consideration. Sometimes group anxiety or organisational conflicts may underlie the complaint.

It is worth remembering that the prevalence of skin lesions worthy of at least one visit to a physician was recently found, in a substantial general population survey in the United States, to be almost as high as one third. So point prevalences of skin complaints in small populations can be expected on occasions to be rather higher than one-third without there being any genuine occupational cause. Such clustering must of course remain a diagnosis of exclusion.

In 1978 Ann Maguire, a Lancashire dermatologist, whose qualifications unusually include a diploma in analytical psychology from the C G Jung Institute in Zurich, used the term 'psychic possession' to describe the different phenomenon of one or more genuine, but non-occupational, skin conditions generating an aura around them of imitative anxiety symptoms in fellow employees.[15] Again, an outbreak of occupational skin disease is simulated. I have since attempted to characterize this aetiology as follows: 'He has a rash... I wonder if I am going to get it... Wait a minute... I think I have'. Such a problem can easily present as a widespread pruritus. Accurate diagnosis of the central case (or cases) once again should allow effective reassurance to be given.

There is of course no reason why a genuine non-occupational skin disorder may not be lurking in the undergrowth of an otherwise perfectly genuine low-humidity occupational dermatosis. It may be precisely this case that is selected as 'the worst case' to represent the more general skin problem to the visiting 'expert': you have already been warned.

Office workers

Many low-humidity occupational dermatoses occur in offices. It might be supposed that office workers were safe from skin problems caused by their work itself, but their safety in fact is only relative. There are several entirely genuine occupational dermatoses that can affect office workers in quite large numbers at the same time, as in a low-humidity problem.

Pruritus of the facial skin may be part of the syndrome of symptoms experienced by office workers intensively handling carbonless copy paper.[16,17] Irritation, pricking and burning of the eyes, nose, mouth and, to a lesser extent, the skin, have now been reported in such groups from several different countries. Several different manufacturers' papers have been implicated.

There is usually nothing abnormal to be seen on clinical examination. Negative patch tests strongly suggest that the causal agent or agents act as an irritant rather than an allergen, though individual cases of allergic contact sensitization to carbonless copy paper have also been reported.[18]

Numerous solutions to the carbonless copy paper problem have been proposed, including in fact raising the ambient humidity, but no solution seems to have been universally effective, as might be expected in the absence of an accurately identified causal agent. Less intensive use is often sufficient to control the problem but some users have been forced back to their carbon paper.

Certain dry-process duplicating machines are capable of generating sufficient ozone, under conditions of intensive use and inadequate ventilation, to provoke irritation of the mucous membrances and, to a lesser extent, the adjoining facial skin. Such machines are sometimes stationed in rooms scarcely larger than broom cupboards and if employees are assigned full time to duplicating work they may present with minor symptoms such as a constant 'tickle' in the throat.

Other duplicating machines using a wet process can produce copies with enough ammonia still on them to irritate the skin of the hands and face of intensive users. Adjustment in the machine or working procedure may be necessary to make copies drier before they are handled. This irritant effect is to be distinguished from sensitization to the diazonium salt, which may be used in the same process.[19]

Reports have appeared in the last few years of various dermatological and systemic symptoms associated with measurable levels of formaldehyde inside buildings.[20] The usual source of this formaldehyde is urea-formaldehyde resin in insulation materials or fittings. The problem appears to be irritant rather than allergic and can therefore affect groups rather than individuals. Complaints include irritation of the eyes, nose and throat, and to a lesser extent the skin. There may be nothing abnormal to see on clinical examination.

Also capable of generation by insulation materials is glass fibre, and an outbreak of pruritus among office workers has recently been attributed to glass fibre derived from ceiling insulation.[21]

Currently the most controversial of causes of dermatological symptoms among office workers is the visual display unit. Reports of facial skin symptoms and rashes have come both from this country and Scandinavia.[22,23] No cause-and-effect relationship for such complaints has yet been established, but the similarity in signs and symptoms from one independent case to another seems to preclude the dismissal of a relationship without further research.

Several VDU operators in one office may make the complaint. Operators with symptoms are always intensive users of VDUs. Facial symptoms tend to be described in terms such as tingling, the feeling of being lightly stroked with a feather, or the feeling of a hair lying on the surface of the skin. The main areas of the face involved are the upper cheeks, forehead and chin. Sometimes a fine red papular eruption can be seen.

Numerous solutions to the VDU problem have been suggested, though, as with carbonless copy paper, no solution has yet been found to be universally effective. The electrostatic field that exists around VDUs has been incorporated into one possible explanation and measures to reduce this are the basis of several attempted solutions. The main differential diagnosis in an individual case is rosacea, but a number of cases in the same office might well prompt thoughts of a low-humidity dermatosis.

CONCLUSION

Low-humidity occupational dermatoses are a prevalent, if unspectacular, presence in today's world. Clinically they tend to be somewhat sparse in physical signs and not as easily identifiable as the 'chapping' that results from cold, dry, outdoor climates. Because of this paucity of physical signs, sufferers from these conditions run the risk of not being accorded sufficient medical attention. I hope that this chapter may have alerted its readers to a dermatological problem that they can both recognize and solve.

REFERENCES

1. Blank IH. Factors which influence the water content of the stratum corneum. *J. Invest. Dermatol.* 1952; **18**:433–40.
2. Rycroft RJG and Smith WDL. Low humidity occupational dermatoses. *Contact Dermatitis* 1980; **6**:488–92.
3. Rothman S. *Physiology and Biochemistry of the Skin*. Chicago: Chicago University Press, 1980.
4. Mitchell JC. Percentage humidity of the air in a dermatological ward, xerosis of the skin and irritant contact dermatitis. *Contact Dermatitis Newsletter* 1972; **11**:287–90.
5. Stewart RD, Hake CL and Peterson JE. Degreaser's flush. *Arch. Environ. Health* 1974; **29**:1–5.
6. White IR and Rycroft RJG. Low humidity occupational dermatosis—an epidemic. *Contact Dermatitis* 1982; **8**:287–90.
7. Rycroft RJG. Low humidity and microtrauma. *Am. J. Ind. Med.* 1985; **8**:371–3.
8. Kligman AM, Klemme JC and Susten AS. *The Chronic Effects of Repeated Mechanical Trauma to the Skin*. Proceedings of a Conference sponsored by the National Institute for Occupational Safety and Health, Cincinnati, Ohio, April 6–8, 1983. *Am. J. Ind. Med.* 1985; **8**:253–513.
9. Fregert S. In Questions and Answers. *Am. J. Ind. Med.* 1985; **8**:373.
10. Emmett EA. In Questions and Answers. *Am. J. Ind. Med.* 1985; **8**:373.
11. Nater JP and de Groot AC. *Unwanted Effects of Cosmetics and Drugs used in Dermatology*. 2nd ed. Amsterdam: Elsevier, 1985.

12. Newman Taylor A, Pickering CAC, Turner-Warwick M et al. Respiratory allergy to a factory humidifier contaminant presenting as a pyrexia of unknown origin. *Br. Med. J.* 1978; **2**:94–5.
13. Kirsch N. An unusual case of contact dermatitis. *Conn. Med.* 1958; **22**:756–9.
14. Rycroft RJG. Occupational dermatoses in perspective. *Lancet* 1980; **ii**:24–6.
15. Maguire A. Psychic possession among industrial workers. *Lancet* 1978; **i**:376–8.
16. Calnan CD. Carbon and carbonless copy paper. *Acta Derm. Venereol. [suppl] (Stockh.)*; 1979; **59**[85]:27–32.
17. Menne T, Asnaes G and Hjorth N. Skin and mucous membrane problems from 'no carbon required' paper. *Contact Dermatitis* 1981; **7**:72–6.
18. Marks JG. Allergic contact dermatitis from carbonless copy paper. *JAMA* 1981: **245**:2331–2.
19. Foussereau J, Benezra C and Maibach HI. *Occupational Contact Dermatitis*. Copenhagen: Munksgaard, 1982.
20. National Research Council. *Formaldehyde and other Aldehydes*. Washington: National Academic Press, 1981.
21. Verbeck SJA, Buise-van Unnick EMM and Malten KE. Itching in office workers from glass fibres. *Contact Dermatitis* 1981; **7**:354.
22. Rycroft RJG and Calnan CD. Facial rashes among visual display unit (VDU) operators. In *Health Hazards of VDTs?* Chichester: John Wiley, 1984.
23. Tjønn HH. Report of facial rashes among VDU operators in Norway. In *Health Hazards of VDTs?* Chichester: John Wiley, 1984.

2. OCCUPATIONAL ASTHMA

A. J. M. Slovak

WHAT IS ASTHMA?

In this chapter, no attempt is made to produce a stock overview of occupational asthma nor to provide a catalogue of causal agents. Many such already exist.[1,2,3] Instead, some of the conceptual and practical problems that interfere with our understanding of this set of disorders are highlighted, together with the recent developments which may help us to deal with these problems. The first of the problems lies less with occupation than with asthma, for asthma means different things to different people. To the clinician it is expiratory wheeze on auscultation, to the physiologist a temporary and variable decrement in a relevant respiratory parameter whilst the epidemiologist may describe asthma as that condition ascribed to a population in which given parameters of lung function show a statistically significant temporary decremental difference when compared to properly matched controls.

This series of different ways of perceiving asthma may seem laboured but is important to the way we choose to understand asthma now and how our perceptions may change in the future. In the last decade or two, we have come to accept a definition of asthma which incorporates physiological and classic, clinical concepts so that most of us would now be comfortable with a definition of asthma which stated it to be 'reversible airways obstruction sufficient to cause a stated degree of decrement in forced expiratory volume (FEV) or peak expiratory flow rate (PEFR)—usually 15 or 20 per cent'. However, this definition has shortcomings. These arise because the definition describes observable attributes of disease rather than more fundamental characteristics and also it sets arbitrary parametric cut-off points. A useful advance in understanding the nature of asthma is to incorporate the idea of airways reactivity. This permits us to view asthma as a circumscribed part of the spectrum of airways reactivity within the general population. It also allows us to perceive more easily that people may move in and out of that circumscribed part of the range according to age, season, level of challenge, concomitant illness and a number of other factors. When dealing with populations rather than individuals, the definition of asthma becomes even more important since different definitions may yield radically different results. Whilst different definitions may make comparisons between studies difficult they are not wholly to be deprecated for their use may be perceptually rewarding as may be seen later in this chapter.

OCCUPATIONAL ASTHMA

The prevalence of asthma in industrialized countries is believed to be about 2 per cent of which some 10 per cent or so is occupational. However, asthma tends to be under-reported and this is particularly so for occupational asthma. The main reason for this is stated to be fear of job loss but it is equally true that many persons with asthma may be genuinely unaware of their illness until it is drawn to their attention for some reason. Also in some regional societies, particularly in areas where chemicals or other asthmatogens have been worked for several generations, there may be an expectation built into the local culture that respiratory disorder is a 'natural' consequence of the work and it therefore passes unremarked.

The cases of occupational asthma which find their way to specialist clinics and thence into the literature are therefore the tips of a series of icebergs. Naturally enough, these cases tend to be the most severe and may be quite atypical of the totality of cases occurring in the population from which they derive. For this among other reasons, it is proper for industries in which occupational asthma occurs to survey their exposed workpeople to obtain a better understanding of the burden of disease and its nature. This has usually been done by means of cross-sectional prevalence studies.

Prevalence

Such studies have described prevalence rates of occupational asthma varying from a few percent to over 50 per cent in populations exposed to different causal agents. Cross-sectional studies capture the historical accumulation of developed disease in a population as well as disease that is only just developing. They fail to identify the disease experience of persons who have left the study population and those who have not yet developed disease but will do so. To this extent prevalence studies may still present a distorted picture and this is especially a problem where the disease is sufficiently severe to drive a substantial proportion of cases out of the particular population.

Incidence

These deficiencies may be overcome by studying the incidence of the disease in question prospectively following up cohorts and tracing leavers. Unfortunately, prospective studies are time and effort intensive and take a long time to yield results. It is tempting therefore to try to draw out conclusions before the results will support them. This is particularly so if the condition under study, as with some types of occupational asthma, takes a long time to develop fully.

Both prevalence and incidence studies tend to follow the classic, clinical model of disease investigation where the disease is first defined according to a series of characteristics. This syndrome is then sought in the population under-study. Such a 'case identification' system of investigation answers well to the needs of traditional scientific investigation but is less well-suited to answering questions arising out of modern society's relatively newfound preoccupation with the effect of general and working environments on health. The questions asked under these circumstances are whether there is a problem associated with a particular exposure or whether a particular problem is so associated. Further, information is sought on the characteristics and distribution of the disorder. There is also generally a need for speed in response which will seldom await the outcome of prospective developments. For these needs a different sort of cross-sectional survey is gaining popularity[4] where using physiological parameters and questionnaires for recording a wide range of relevant variables (e.g. symptoms, smoking etc.), occupational asthma is described only by the significance of the statistical differences in the findings between study and control populations.

Of course, there is an obvious, immediate quandary with this population-oriented statistical method of approach and that is how to reconcile it with traditional case identification. Perhaps this issue is best bypassed rather than reconciled, recognizing that these different ways of looking at occupational asthma both have useful perceptual benefits and are, in many ways, complementary. Nevertheless, attempts have been made at the reconciliation of objective and subjective views of occupational asthma, together with the incorporation of social and economic factors relating to the disease using decision analysis models.[5] These techniques, using Bayesian concepts, seek to set the parametric criteria for defining occupational asthma (and any other occupational disease) on a sliding scale according to the importance ascribed to different factors. This approach should appeal to the occupational health manager wishing to manipulate resources rationally and efficiently in diagnosing occupational disease.

SELECTION OF WORKPEOPLE

At a more mundanely practical level, the first management problem for the occupational practitioner is selection. There is little dispute that job candidates whose health would be seriously compromised by asthma should not be exposed to the risk of developing it. Most clearly, such a restriction would apply to those with serious pre-existing lung or heart disease. The case of those with asthma known to be caused by the particular substance they seek to work with, is also clear enough: they should usually be rejected. However, intrinsic asthma and asthma due to other extrinsic causes is not such a clear-cut contraindicator. At present there is little or no research evidence of occupational asthma being more likely to occur in people with pre-existing asthma, but certainly occupational asthma will worsen and be worsened by such pre-existing conditions. Nor would it seem that pre-existing bronchial reactivity, which has been much discussed lately, plays any great part in initiating occupational asthma. Another reason for exercising caution in admitting pre-existing asthmatics to work with asthmatogenic substances are the very real practical problems associated with differentiating the aetiology of the asthma which is subsequently observed.

Perhaps though, the most vexed question in relation to selection is the consideration which should be given to atopic status. The relationship of atopy to occupational asthma has not been well researched yet the practice of excluding atopic persons from work with substances causing occupational asthma is very common throughout the world. What research has been published shows that atopy is associated weakly or not at all with asthma caused by low molecular weight (MW) asthmatogens and moderately to well with higher MW substances. The association has usually only been expressed as sensitivity. Specificity and more importantly predictive value have seldom been examined.

Only in relation to asthma due to laboratory animal allergy (LAA) has there been any attempt to investigate this matter fully.[6,7] With this condition there is a strong association between asthma and atopy —defined by skin prick to common environmental allergens—with sensitivity in the range 70–80 per cent. However, specificity is mediocre and predictive value poor ($\simeq$ 80 per cent and $\simeq$ 35 per cent respectively). In different words, although there is a high proportion of atopics among those who develop LAA asthma (70–80 per cent) only about one-third of exposed atopics ever get the disease at all. Also, some people who are not atopic will also develop the disease. These comprise approximately one-fifth of those with LAA asthma. Exclusion of atopics is therefore a very crude selective tool. Since atopics comprise some 30 per cent of the population, it is also socially problematic and as occupational asthma is a largely manageable disorder, discrimination on grounds of atopy is questionable. Even more doubtful is the use of two other markers of atopy, family and past personal history. Skin prick testing is a fairly objective criterion whereas memory, especially at pre-employment examination, can be unreliable. Recently, I resurveyed a population involved in a study five years previously[8] and found that 30–40 per cent of those who had previously given a positive past or family history of allergy now gave a negative history. A similar switch was also observed from negative to positive.

CLINICAL ASPECTS

The traditional descriptive history of occupational asthma is one of chest-tightness occurring rapidly on exposure or after a symptom-free interval appearing in the evening or night and resolving by the next morning. These are the so-called immediate and late (Type 1 and Type 3) responses. Derived as they are from well-developed selected cases presenting in hospital clinics these symptom patterns do not necessarily reflect well the whole range of disease observed in populations at risk.

The low MW asthmas are usually associated with the rapid or sudden onset of vicious asthma which worsens in severity and frequency so that the period from inception to maturity of the disease will often be just a few weeks. Typically, the symptoms first occur some time after work although both immediate, late and dual response patterns of onset may be seen. Eventually the symptoms spread forwards and backwards so that they begin at work and do not remit, wholly or partly, before return to work the next day.

With the high MW asthmas the progress of the disease is frequently much more leisurely. Most low MW asthmas manifest themselves in the first year of exposure, the high MW asthmas more often take 2 or 3 years. A period of rhinitis may precede the development of overt asthma which may then take a year or two to develop the same severity as that which occurs so rapidly in low MW cases. The predominant pattern here is of immediate onset but prolonged persistence. In those asthmas where positive skin prick tests to specific antigen are accurate diagnostic markers, skin prick conversion from negative to positive may precede asthma by some appreciable time, not infrequently months.

These basic patterns of occupational asthma development vary substantially between individuals, but may be even more profoundly modified by individuals' work patterns. Shiftwork, especially that involving rapid cycling of shifts, may produce a very confusing clinical picture. Similar effects may also occur when exposures to the offending substance are short, intermittent or weak. Many of the most diagnostically difficult of the latter cases arise from so-called secondary exposures where the offending substance is carried by an exposed but unsensitized worker into another environment where a sensitized person is exposed. Typically this happens with secretaries and cleaners.

Asthma usually presents as chest tightness, wheezing or cough. There are regional, sexual and class differences in willingness to admit to certain symptoms. Thus working class British men in certain areas are reluctant to admit to wheezing because it has mental associations with sexual inadequacy. They are, however, quite happy to describe 'a tight cough'. Similarly, some women are not keen to admit to wheezing because they consider it a vulgar descriptive term. For those carrying out studies of occupational asthma in regions whose verbal sensitivities they are unfamiliar with, a pilot study of history taking or questionnaire validity is a prudent precaution.

SURVEILLANCE OF EXPOSED GROUPS

The management of occupational asthma is usually described in terms of the diagnosis and disposal of the individual case. However, the occupational practitioner more often has a population at risk in which only a few established cases are found. Within the population there will also be developing cases, people who will develop asthma at some time in the future and a large group who will never get asthma. This whole population needs to be monitored in such a way that the process is efficient and effective particularly with regard to obtaining maximal compliance.

The time-honoured process of screening in these cases is the periodic medical. The timing of this process is important and will vary in frequency if it is to effectively capture disease as it arises. Thus in the case of low MW asthma a 2- or 3-monthly medical would be indicated in the first year of employment reducing to some longer interval after this. The periodic medical has itself undergone something of a change in modern practice. The unstructured interview with doctor or nurse has tended to be replaced by a structural questionnaire so that the whole of the exposed population answers the same questions and the data obtained can be processed statistically to yield information about the development of disease in that population.

Questionnaires may be interviewer administered which is time-consuming, or sent out by post. If posted, the response rate tends to drop and the omissions resulting from this loss of response have to be balanced against the time costs of alternative systems. More recently it has become feasible to call individuals to input their responses directly to a questionnaire on a computer linked screen. This obviates the need for an operator to input data from manually completed questionnaires before analysis and combines the advantage of the high response rate of interviewer administered questionnaires with the time saving obtained by self-completion. Questionnaires themselves are only as good as their design and ideally the quality of the responses obtained should be validated from time to time in a sample of the population.

LUNG FUNCTION STUDIES

The place of spirometry in the screening of occupational asthma is debatable. Certainly, it is a crude tool in the screening of the individual case, especially in the workplace. Pre- and post-work spirometry, daily or weekly will be insufficiently accurate to pick up changes in the majority of cases since asthma will not necessarily be present when the measurements are taken. However, spirometry does have a place both in the study of acute change in airways in populations and in observing long-term change. Where individual cases are being studied in a respiratory laboratory, spirometry can of course be used effectively but in the field less cumbersome methods are necessary and available.

A number of small, accurate devices are now on the market. These measure peak expiratory flow rate which has proved to be sufficiently sensitive to bronchial airflow changes to be comparable to spirometry recording forced expiratory volume. The pattern of asthma in occupational conditions is such that a 2-hourly self-recording of PEFR during the waking hours is sufficient to obtain an accurate picture of what is going on in the majority of cases. However, where there is an unusual pattern of change, hourly recordings may be needed. Compliance is seldom a problem in the well-motivated subject but obtaining motivation may be a time-consuming process of explanation, cajolery and continued support. In order to get a clear idea of what is going on, a consistent, gap-free record is to be preferred. Where there are doubts about compliance, a 4-hourly recording routine faithfully completed is superior to a 2-hourly schedule intermittently omitted.

The duration of the recording period will be dependant on the severity of the case and the complexity of the exposures. A typical case will need to be screened for a minimum of 4 exposure/non-exposure cycles and this will usually be work weeks and week-ends. It is useful to start the recording period with a prolonged period away from exposure and this can be most conveniently done after a holiday. This allows for recovery of long-term shut-down of airways which may occur in a proportion of more severely affected asthmatics. Occupational asthma is diagnosed from the size and pattern of decrements in PEFR caused by exposure and the effect of removal from exposure. The pattern of response varies according to a number of factors, among which are disease severity, severity of challenge and the immunological nature of the asthmatic response. A detailed description of these various patterns is beyond the scope of this chapter but is a rewarding subject for study. The review by Burge[9] is comprehensive.

The value of PEFR recording can be further enhanced if symptoms and location are recorded alongside the periodic readings. This is particularly useful with persons with a complex work pattern and also in discerning the symptomatic impact of the condition of the individual under study. These records can be designed more or less simply according to the needs of the study in question and the tolerance of those being tested. An example of a typical daily recording sheet is given in *Fig. 2.1*.

Observation Time	0200	0400	0600	0800	1000	1200	1400	1600	1800	2000	2200	2400
PEFR Record												
At work (Y or N)*												
Symptoms (Y or N)**												

Notes * If at work but off your normal job put Y and then NE (not exposed)
** put down Y for cough, wheeze or chest tightness only.
N for anything else.

Fig. 2.1. Sample simultaneous daily record for PEFR, exposure and symptoms.

Bronchial provocation or challenge testing is largely confined in its use to specialist respiratory physiology centres which may also be equipped with exposure chambers of varying complexity in order to be able to mimic occupational or environmental conditions. The main application of such testing is in the elucidation of new causes for asthma and in those cases where the diagnosis is doubtful or complex, for the studies are time-consuming and the resources costly. However, there is a half-way house between these sorts of complex resources and the other everyday methods of the practitioner interested in occupational asthma. This is the histamine or methacholine challenge test. Whilst this does require the availability of basic resuscitation facilities as a form of insurance, these and the test equipment are relatively simple and should be considered as a viable option by those occupational units with substantial populations exposed to agents causing occupational asthma. The test is administered by the delivery of serially increasing strengths of histamine or methacholine solutions by nebulizer. These solutions will provoke constriction according to the bronchial reactivity of the test subject. The bronchoconstriction obtained follows a well-defined dose response pattern and may be reversed rapidly by bronchodilators. The test is a simple way of quantifying bronchial reactivity and is particularly useful in cases of occupational asthma where the diagnosis is doubtful or where the condition is at an early stage of development. In these latter cases bronchial reactivity may be high when in exposure but will revert to 'normal' levels rapidly when exposure ceases. An example is shown in *Fig. 2.2.*

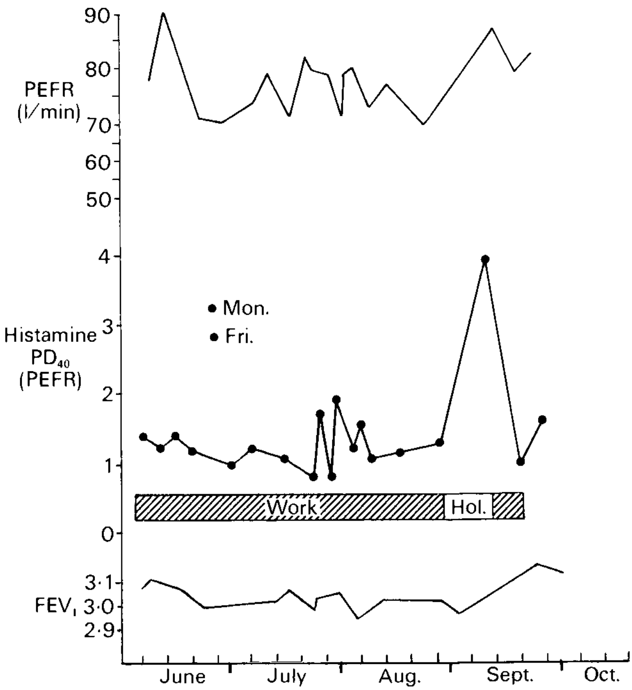


Fig. 2.2. Effect of histamine challenge on bronchial reactivity after a holiday period in a worker sensitized to laboratory animal antigens.

IMMUNOLOGICAL TESTS

The basis of much, although not all, of what is described as occupational asthma is immunological, mainly mediated by IgE. These immunological and other mechanisms are discussed below but first the contribution of immunological testing to diagnosis is outlined. The fundamental difference between the high and low MW substances causing asthma is that the high MW substances are independently, immunologically competent whereas the low have to form haptens before being able to induce antibody formation. From this it has been argued that it is mainly the difficulty of obtaining the appropriate hapten binding in *in vitro* experiments which prevents the accurate identification of specific IgE in persons with low MW asthmas.[3] However, it may be that the mechanism of genesis of asthma with low MW substances is more fundamentally different from that in high MW asthma than is supposed since the clinical manifestations of the diseases caused are markedly distinct.

In general, the diagnostic usefulness—as opposed to the theoretical and speculative benefit—of tests for substance-specific immunoglobins has been poor with regard to low MW asthmatogens and good with regard to high MW substances. Even with high MW substances it has been important to establish reliability on a case-by-case basis, paying special attention to the purity and strength of antigens. Where these have been well-developed, a high degree of reliance can be placed on the identification of specific IgE and this correlates well with skin prick results.

CASE MANAGEMENT AND PROGNOSIS

Until recently there has been little doubt expressed about the *a priori* view that occupational asthma resolves rapidly and fully upon removal from exposure. However, there is now increasing evidence that asthma may persist for a substantial period after the cessation of exposure and may become non-specific.[10] Much of the evidence for long-term effects is subjective, depending on reported symptoms or indirect evidence like persistent unemployment or failure to stay in work. Because both acute management and equitable compensation may be affected by changes in the perception of long-term prospects it is important that many series of occupational asthma cases be followed up for prolonged periods after diagnosis. If future research were to lead to a reappraisal of policy with regard to management and particularly to earlier relocation, it might conflict seriously with the social or economic interests of the individual concerned. Knowledge of a policy of early relocation can in some circumstances lead to concealment of symptoms until they are severe and manifestly obvious. The damage under these circumstances may be much worse than it would have been with a more conservative view of relocation. In order to avoid this problem, employers often produce a policy defining their actions and the provisions they will make for employees who develop occupational asthma.

Whilst exclusion and relocation removes the stimulus which maintains asthmatic symptoms, such a policy is only easy in circumstances of high employment and high labour turnover. These circumstances are not currently widespread. Further, there are certain types of work-people who are very reluctant to relocate. The most obvious of these categories are those who have invested substantial time and effort in acquiring job skills which are integrally bound up with the sensitizing exposure: among these are bakers, experimental research workers and industrial chemists. To these and those others with poor alternative job prospects, another strategy may seem more attractive and that is the use of preventive control measures to reduce exposure.

PREVENTION AND PROTECTION

The most successful attack upon a sensitizing problem has been achieved by the containment of the manufacturing process for proteolytic enzymes where new cases of occupational asthma are now exceedingly rare. There are special circumstances in this manufacture which should suggest caution about the likelihood of repeating the formula elsewhere. First the process is large scale, practised on only a few sites worldwide and, being of great economic importance, will bear substantial research and capital outlay. Second, the proteolytic enzymes are relatively mild sensitizers. With the much more potent platinum salt sensitizers very high levels of containment have failed to reduce the incidence of asthma significantly. One ploy, derived from the success with proteolytic enzymes, which is likely to have more universal application is the encapsulation of substances in 'inert' compounds to render them non-asthmatogenic or more simply too big to inhale. Another alternative is the substitution of pastes and solutions for powders.

However, even with the deployment of such ingenious subterfuges, the likelihood remains that containment at source will prove impractical for many asthma-inducing agents. This brings forward the issue of personal protection. Considering how widely masks and ventilated helmets are used, it is surprising that almost no research has been done in validating their efficacy. What little there is suggests that substantial benefits may be obtained but that the suppression of asthma is incomplete and consequently a high standard of monitoring is necessary. Conventional prophylactic or bronchodilator treatment is also effective in occupational asthma if circumstances warrant their use.

AETIOLOGY OF OCCUPATIONAL ASTHMA

Whilst a detailed analysis of the aetiology underlying the development of occupational asthma would be dauntingly long, a knowledge of the salient features and newer developments is beneficial in dealing with even the everyday problems of diagnosis and management. Three fundamental physiological mechanisms of varying importance are thought to be involved. In descending order of importance, these are immunological, pharmacological and irritant.

Immunological factors

The predominant class of antibodies involved in mediating occupational asthma is considered to be IgE which in the classic model of immune mediation is responsible for the immediate (Type 1) response. This model does not fit entirely well with the facts of occupational asthma where 'late', 'dual' and more complexly overlapping responses are frequently observed. Therefore a role in the late response has been hypothesized for IgE.[12]

IgG and IgM antibodies have also been implicated. Uncertainty about mechanisms has not been improved by the equivocal results from binding studies attempting to demonstrate specific IgE to low MW antigens.

Pharmacological factors

Better progress has been made with understanding the pharmacological bronchoconstrictive effect of some substances. The most easily understood of these are the β adrenoreceptor and H_2 blockers, propranolol and cimetidine, which have caused asthma in pharmaceutical workers exposed by inhalation. More importantly, with isocyanates, a β and prostaglandin receptor blocking role has now been demonstrated *in vitro*[13] although the *in vivo* importance of this finding is yet to be confirmed. Other direct pharmacological effects are hypothesized for agents mediated via cholinergic receptors or by inhibition of acetylcholinesterase.

Irritant factors

Irritance also plays a part in occupational asthma. A primary role in induction has been hypothesized especially for toluene di-isocyanate which is irritant at high concentrations and can act as a non-specific bronchoconstrictor. However, this is very hard to prove and it is much more likely that industrial irritants play a secondary role. Thus such agents as sulphur dioxide and diesel fumes have been shown to induce bronchoconstrictive episodes in asthmatic individuals. It is likely that the main role of the irritance associated with many industrial chemicals is to provide a major factor in the prolongation of non-specific bronchial hyperactivity even in those people who have been ostensibly removed from exposure, or at least from the specific causal agent that initiated the disorder.

CONCLUSION

The last few years have seen substantial progress in occupational asthma research on a number of fronts. The importance of these advances will vary according to the interests of the individuals concerned but, in practical terms, the most influential has certainly been the widespread adoption of serial recording of PEFR in diagnosis. Other promising leads in containment and protection strategy need to be developed. In experimental studies, the failure to identify the aetiological mechanisms of many low MW asthmas has been disappointing and more of these may yet turn out to have non-immunological origins. Occupational asthma will continue to be a productive target for many lines of research.

REFERENCES

1. Salvaggio J (ed.) In: *Occupational and Environmental Respiratory Disease.* (NIAID Task Force Report: Asthma and other allergic diseases. ed. Patterson R) USDHEW NIH pub. No. 79–387, May 1979.
2. Butcher BT and Hendrick DJ. Occupational asthma. In: Symposium on Immune Factors in Pulmonary Disease. *Clin. in Chest Med.* 1983; **4:**1.
3. Fink JN. Evaluation of the patient for occupational immunological disease. *J. Allergy Clin. Immunol.* 1982; **70:**11–14.
4. Field GB. Pulmonary function in aluminium smelters. *Thorax* 1984; **39:**743–51.
5. Harber P and Rappaport S. Clinical decision analysis in occupational medicine: Choosing the optimal FEV criterion for diagnosing occupational asthma. *J. Occup. Med.* 1985; **17,9:**651–8.
6. Cockcroft A, Edwards J and McCarthy P. The role of pre-employment allergy screening in animal workers. *Eur. J. Respir. Dis. (Suppl.)* 113, 1985; **62:**42–44.
7. Slovak AJM and Hill RN. Does atopy have any predictive value for laboratory animal allergy?: and a comparison of different concepts of atopy. Submitted to *Br. J. Ind. Med.*
8. Slovak AJM and Hill RN. Laboratory animal allergy: a clinical survey of an exposed population. *Br. J. Ind. Med.* 1981; **38:**38–41.
9. Burge PS. Advances in the diagnosis of occupational asthma. In: Slovak AJM (ed.) *Proc. Soc. Occup. Med.* Research Panel Symposium. 1981; 17–34.
10. Yeung M. Prognosis in occupational asthma. *Thorax* 1985; **40:**241–3.
11. Hendrick DJ, Jones RN and Weill H. Occupational lung disease. In: Simmons DH (ed.) *Current Pulmonology 3.* New York: John Wiley, 1981.
12. Slovak AJM, Orr R and Teasdale EL. Efficacy of the helmet respirator in occupational asthma due to laboratory animal allergy. *Am. Ind. Hyg. Assoc. J.* 1985: **4,6,8:**411–15.
13. Davies RJ, Butcher BT, O'Neil CE et al. The in vitro effect of TDI on lymphocyte cyclic AMP production by iso-proterenol, prostaglandin and histamine: A possible mode of action. *J Allergy Clin. Immunol.* 1977; **60:**233.

3. MONITORING FOR NEW OCCUPATIONAL RISKS OF CANCER

D. Coggon

INTRODUCTION

Cancer is one of the most serious health hazards which can arise from a working environment. Exactly how many cancers are caused by occupation is not known. A controversial report submitted to the US Occupational Safety and Health Administration (OSHA) in 1978 predicted that over the next few decades more than 20 per cent of cancer deaths in the United States would be occupational in origin.[1] However, the logic underlying this estimate has since been widely criticized, and Doll and Peto have suggested that a figure in the range of 2–8 per cent is more reasonable.[2] Even a proportion as low as 2 per cent would still imply a substantial health problem in absolute terms, accounting for approximately 2500 deaths per year in England and Wales. Moreover, industrial cancer arouses unique public concern. While workers may accept a hazard to their health from causes such as accidental injury, they will not tolerate the same level of risk from neoplastic disease. The identification of new occupational carcinogens is therefore an important area of research.

Table 3.1 summarizes the industrial products and processes which to date have been firmly associated with a risk of cancer at one or more sites. One remarkable feature of the list is that in almost every instance the hazard of cancer was first suspected on the basis of astute clinical observations. In comparison, laboratory investigations and formal population studies have been an unproductive source of hypotheses, although it is possible that some potential carcinogens identified through laboratory research have as a consequence not been produced commercially, and the opportunity to evaluate their carcinogenicity in man has therefore never arisen. Epidemiology has played an important role in confirming hazards once suggested, in pinpointing the responsible agents or processes, and in quantifying risks, but it has only provided the initial clue to four occupational cancers.

The circumstances surrounding these discoveries have been reviewed by Doll.[3] The high incidence of bladder carcinoma in coal

Table 3.1. Established occupational risks of cancer

Agent or process	Site of tumour
Combustion products of coal, shale oil, polycyclic hydrocarbons (chimney sweeps, makers of coal gas, cotton mule spinners)	*Scrotum and other parts of skin, bronchus
Ultraviolet light (farmers, fishermen)	Skin
2-naphthylamine, benzidine (dye manufacturers, rubber workers, makers of coal gas)	Bladder
Mustard gas (makers of mustard gas)	Bronchus, larynx, nasal sinuses
4-amino-diphenyl (chemical workers)	Bladder
Vinyl chloride (manufacture of PVC)	Liver (angiosarcoma)
Arsenic (sheep dip manufacturers, copper smelters)	Bronchus, skin
Asbestos (asbestos workers, insulation workers, gas mask manufacturers)	Bronchus, pleura, peritoneum
Benzene (workers with glues, varnishes etc.)	Leukaemia
Ionizing radiation (haematite miners, fluorspar miners, luminizers, radiologists)	*Bronchus, skin, bone
Bischloromethyl ether (makers of ion exchange resins)	Bronchus
Chromate manufacture	Bronchus
Nickel refining	Bronchus, nasal sinuses
Furniture manufacture	Nasal sinuses
Manufacture of isopropyl alcohol by strong acid process	Nasal sinuses
Boot and shoe manufacture	Nasal sinuses

*The site of the tumour depends upon the nature of the exposure. The occupations listed do not necessarily carry risks for all of these sites.

gas production workers was first noted by Henry, Kennaway and Kennaway in 1931 when they examined the death certificates of almost 6000 men who had died from the disease.[4] However, little importance was attached to their observation until it was finally confirmed nearly thirty years later. The finding that fluorspar miners are prone to bronchial carcinoma arose from an investigation into high death rates from lung cancer among the male population of St Lawrence, a centre of the fluorspar-mining industry in Newfoundland.[5] Two further hazards were revealed incidentally in the course of epidemiological studies directed to other ends. Case discovered the risk of bladder cancer in the rubber industry when he was looking through hospital records in Birmingham and noticed an excess of rubber workers with the tumour.[6] He had chosen Birmingham as a control area for an investigation of the dye industry. The association of nasal carcinoma with work in the boot and shoe industry was an incidental finding in an epidemiological study prompted by the high incidence of nasal cancer in furniture makers.[7] Acheson, then director of the Oxford

Record Linkage study, was investigating a focus of nasal carcinoma in High Wycombe, a centre of furniture manufacture, and in the course of his work he plotted the distribution of nasal cancer in the Oxford Health Region. This revealed a second, previously unrecognized cluster of cases in the area of Northampton, and subsequent investigation confirmed the link with the local boot and shoe trade.

Both of the last two discoveries owed much to chance. In choosing a control population Case might easily have selected a city which did not have a rubber industry. Similarly, if Northampton had not been situated in the same health region as High Wycombe, the risk of nasal cancer in boot and shoe operatives might well have gone undetected. The serendipity of these findings suggests that an appropriate systematic epidemiological search might well reveal further occupational carcinogens. In this chapter, I shall review the methods which epidemiologists have used in the pursuit of clues to occupational hazards of cancer, and discuss possible reasons for their relative lack of success to date. I shall also describe some newer approaches which are currently being explored and examine the prospects for the future.

THEORETICAL BACKGROUND

Before reviewing the epidemiological techniques which have been used to screen for occupational risks of cancer, it is helpful to consider those factors which determine the likelihood that an investigation will detect a carcinogenic effect if it is present. In any such study information about exposure to putative carcinogens is examined in relation to direct or indirect measures of cancer incidence. The probability that an association between carcinogen and cancer will be demonstrated with given statistical certainty depends upon the seven factors listed in *Table 3.2.*

Table 3.2. Factors determining the power of a study to detect a cancer hazard

1. The proportion of the study population exposed to the carcinogen
2. The size of the population studied and the time period over which disease is ascertained
3. The incidence of the cancer in the study population
4. The relative risk of the cancer due to the carcinogen
5. The accuracy with which exposure to the carcinogen is ascertained
6. The accuracy with which cancer incidence is ascertained
7. The influence of any confounding factors

Unless a sufficient number of the persons observed both develop the cancer and have previously been exposed to the carcinogen, the association will be missed—hence the importance of factors 1 to 3. Other things being equal, the higher the relative risk of cancer due to a carcinogen (factor 4), the more easily its effect can be detected. Inaccuracies in the ascertainment of exposure and cancer incidence (factors 5 and 6), provided that they are unbiased, reduce the chances that an association will be demonstrated.

Confounding (factor 7) occurs when the exposure of interest is associated with a second factor which independently influences the risk of cancer. Such confounding factors are not necessarily occupational in nature—for example smoking and alcohol. Positive confounding tends to exaggerate associations between carcinogens and the cancers which they cause, and may lead to spurious associations between occupation and cancer in the absence of any true occupational carcinogen. For example, an occupation associated with unusually high rates of smoking and alcohol consumption may carry an increased risk of oesophageal cancer, although the job itself is not directly harmful. Negative confounding can mask genuine hazards. For example, the effects of a lung carcinogen are less likely to be apparent if those who are exposed smoke less than the average.

With this theoretical framework in mind, we can now go on to examine in detail the various approaches that have been applied with the aim of generating hypotheses concerning possible unrecognized occupational risks of cancer.

METHODS USED TO SCREEN FOR NEW OCCUPATIONAL RISKS OF CANCER

Analyses of occupational mortality

The oldest and best known method of monitoring industrial health hazards is through occupational mortality. In England and Wales statistics of occupational mortality have been compiled since 1855.[8] Occupations are grouped according to a standard classification, and their mortality from specific causes is calculated for a period of one to five years surrounding a census. Information about the occupation and underlying cause of death of deceased persons is obtained from death certificates. This is then related to the total number of men and women in each occupational category derived from the census. Data on occupational mortality from cancer were first published in 1897.[9] A similar technique has been used to examine the mortality of men of working age in the United States.[10]

The method has the advantage of using information from large numbers of cases, but against this must be set several major weaknesses. There is the obvious difficulty that mortality is only an indirect measure of incidence. This does not present a problem for cancers with a high fatality rate, but limits the application of the technique to diseases such as carcinoma of the skin for which treatment is usually curative.

The inaccuracy of diagnoses reported on death certificates is a further limitation. Reports from several countries have suggested substantial rates of error in the certification of causes of death, misclassification being a particular problem with certain internal malignancies.[11-13] Even when diagnoses are correctly recorded they may lack adequate detail. Some occupational risks are specific for one histological type of tumour, but on death certificates information about histology is often lacking.

Another weakness is the poor quality of occupational information obtained from death records. In England and Wales where the method of analysis has been most widely used, only the most recent full-time occupation is recorded when a death is registered. However, many persons undertake more than one job in the course of their lifetime, and in the context of occupational cancer the work which they did ten or twenty years before death is likely to be the most relevant. Furthermore, the description of jobs on death certificates is frequently imprecise and may even be wrong. There is a tendency for informants when registering deaths to promote their deceased relatives to more prestigious jobs so that, for example, a scrap merchant with a small business might be described as a company director. Such errors can give rise to serious biases when the denominators for death rates are derived from more reliable census data, with the mortality of occupations such as company director being falsely inflated. These biases have been well documented by Heasman[14] and by Fox.[15]

The lack of information about confounding factors, in particular smoking, is also a drawback, although some attempt can be made to circumvent this problem by using data from other sources. In Great Britain information about the smoking habits of different occupational groups is available from the General Household Survey, and Fox has demonstrated how this can aid the interpretation of occupational mortality rates from lung cancer.[15]

Three other methods have been used to analyse occupational mortality, all of which avoid the biases that arise when different sources of occupational information are used to define the numerator and denominator of death rates.

The first approach employs proportional mortality as an alternative to true mortality rates. Only occupational data derived from death certificates are used in the calculations. For each occupational category the proportion of deaths attributed to a specific cause is compared with the proportion of deaths due to that cause in the total study population, allowance being made for the possible confounding effects of age and sex. This technique has been used by Milham to examine mortality patterns in Washington State.[16] Numerator/denominator biases are eliminated, but there is a risk that elevated death rates from individual causes may be masked if the overall mortality of an occupational group is high. Otherwise, the strengths and weaknesses of the techniques are similar to those of the traditional method of analysis.

The second method examines the occupations of deceased persons on a case-control basis. Men and women who have died from a specific cancer are compared with suitably selected controls who have died from other causes. Occupational data are again derived solely from death certificates. In general such studies are smaller and cheaper than proportional mortality analyses and concentrate on cancers for which an occupational aetiology is most likely, for example carcinoma of the bronchus and bladder.[17] In other respects their advantages and disadvantages are similar to those of proportional analyses.

In the third approach occupational information is obtained entirely from censuses. This is achieved by linking death certificates to one or more census records. Mortality rates are then calculated in the normal way. The method was first used in Scandinavia where the introduction of personal identification numbers during the 1960s facilitated record linkage,[18] and more recently has been applied to a 1 per cent sample of the population of England and Wales.[19] Apart from a reduction in bias, there is the advantage that mortality can be related to occupations held at intervals prior to death—thus allowing for any latency in the onset of cancer—and that by linking information from several censuses a more complete occupational history can be incorporated into the analysis for each individual. However, errors in the certified cause of death remain a problem.

The application of record linkage has in the past been limited by the cost of organizing and processing large quantities of data, but with the advent of modern computing facilities, it is likely that the method will be used on an increasing scale. Occasionally the method may be applicable to sources of occupational information other than censuses. For example, in the United States the mortality of a population of some 300 000 ex-servicemen over a 16-year period has recently been examined in relation to the occupations which they reported in 1954 in response to a postal questionnaire.[20]

The generation of hypotheses about industrial cancer is only one of many important applications of occupational mortality data. In the last few years several promising leads have emerged from surveys of the types described above, including for example, possible risks of stomach cancer in coal miners,[15] and of lung cancer among workers in aluminium smelters[16,21] and in butchers.[22] However, despite the long history of their use, studies of occupational mortality have so far only provided the initial clue to one proven industrial cancer hazard —the risk of bladder cancer in coal gas manufacture.

Analyses of cancer registration by occupation

Many countries now have systems for the registration of newly diagnosed cases of cancer, and often information is collected about the occupations of cancer patients. Such data can be used for analyses of occupational cancer incidence analogous to those of occupational mortality. For example, in Los Angeles County a record of the occupation at the time of diagnosis is obtained from the hospital notes of cancer patients.[23] By relating this information to census data on the distribution of occupations in the local population, cancer incidence rates by occupation can be derived. In Sweden and Finland the occupations of cancer patients are obtained by individually linking registrations to census records.[18] As with occupational mortality, the incidence rates derived by this method are less prone to bias.

As yet no definite industrial carcinogens have been identified through routine cancer registration statistics, although possible risks of carcinoma of the prostate in farmers[24] and of oesophageal cancer in men making vulcanized rubber[25] are among those which have been suggested. The discovery of nasal cancer in the boot and shoe industry was made using cancer registry data, but not on the basis of routinely generated statistics.

Studies incorporating more detailed occupational data

Most routine surveys of occupational mortality and cancer registration are founded on imprecise and incomplete occupational histories. There have, however, been several studies in which more detailed occupational information has been obtained by questioning cancer patients or their relatives.

At the Roswell Park Memorial Institute, a centre for cancer therapy in Buffalo, New York, all patients referred for diagnosis and treatment during the period 1956–65 were asked to complete a detailed questionnaire.[26] In addition to information about demographic and social factors (including smoking), a lifetime occupational history was requested. The occupational histories of patients with specific cancers were compared with those of controls diagnosed as having non-neoplastic conditions, allowance being made for the possible confounding effects of smoking.

As part of the US Third National Cancer Survey interviews were sought with a random 10 per cent sample of newly diagnosed cancer patients from eight geographical areas during the years 1969–71.[27,28] If patients had died before they could be contacted, their next of kin were approached for information. Questions were asked about possible risk factors for cancer including occupation, and histories were obtained for 7518 patients. The occupational data were analysed in two ways. Associations between each specific cancer site and main lifetime occupation and industry were examined using all other cancers as controls. If necessary allowance was made for potential confounding factors, such as tobacco and alcohol consumption. In addition, the industries in which patients were employed at the time of diagnosis were compared with data from the US 1970 census.

The occupational and diagnostic input to surveys such as those described above is superior to that of other surveillance systems, but the cost of collecting the data is correspondingly higher and restricts the size of such studies. The failure of the Roswell Park study and Third National Cancer Survey to reveal any new industrial carcinogens may stem from an absence of major hazards in the limited populations studied.

In view of the expense of large surveys, the idea of smaller investigations concentrating on those cancers which are most likely to be occupationally related is attractive. Several promising clues have emerged from such studies. For example, the possibility that cutting oils might cause carcinoma of the bladder was first suggested by Cartwright as a consequence of a case-control study in which occupational histories were obtained by interview.[29] The hypothesis is not yet proven, but is supported by the findings of other investigations.

Analyses by inferred occupational exposure

Ideally, all epidemiological studies of occupational cancer would use exact measures of exposure to potential hazards. Any inaccuracy in the measurement of exposure reduces the power of the investigation to detect a carcinogenic effect. When only one or two specific hypotheses are being tested it may be possible to ascertain exposure directly, if not with complete accuracy, for example from company records. However, such an approach will rarely be feasible in hypothesis-generating exercises where by definition a large number of putative hazards are examined simultaneously. Instead it is necessary to fall back on an indirect measure of exposure.

All the screening methods which have been described so far use job-title as a proxy for occupational exposure, and group jobs according to standard classifications of occupation and industry. However, the categories defined in such classifications do not always correspond closely with exposure to potential carcinogens. A given exposure may occur in many different job categories, while within a single job category only a proportion of workers is exposed. Consider for example the system used to code jobs in the analysis of occupational mortality for England and Wales around the 1971 census.[30] Within this classification exposure to diesel fumes occurs in occupational categories as diverse as underground miners, labourers in construction and merchant seamen. At the same time, not all men in these occupations are necessarily exposed to diesel fumes in the course of their work. Any carcinogenic effect of diesel fumes is therefore difficult to detect. Analyses of occupational mortality and cancer incidence might be improved if subjects could be classified in a way that corresponded more exactly with exposure to agents of interest.

Two approaches to this problem are currently being evaluated. The first uses job-exposure matrices as an adjunct to the conventional methods of analysis. In essence a job-exposure matrix is a table with job categories listed along one axis and exposures along the other. Whether or not a particular exposure occurs in a job is recorded in the appropriate cell of the table. Such a matrix can be used to group job categories which share exposure to a possible carcinogen and examine their combined cancer incidence. By extending the exposure axis a large number of potential hazards can be examined simultaneously.

The idea of using job-exposure matrices in the search for new occupational carcinogens was first put forward by Hoar and her colleagues in 1980,[31] and research teams from several countries have since taken up the concept. As yet, however, their efforts have not been rewarded by any new discoveries. In some studies the expected effects of known industrial carcinogens have been demonstrated, but even these have not been consistently apparent.[17,31,32] Perhaps the greatest limitation to the method is the poor specificity of many job categories as an index of exposure. The matrix method overcomes the difficulty that the same exposure can occur in many different jobs, but within a job category it does not distinguish those individuals who are exposed from those who are not.

The second new approach confronts this weakness by eliminating the use of a standard job classification. Instead, each individual occupational history is evaluated by a suitably trained hygienist or engineer who assesses the likelihood that the subject has been exposed to each of a range of possible carcinogens. Associations between cancer and likely exposure can then be examined. Clearly the accuracy of exposure estimates will depend upon the detail and reliability of occupational histories, and on the skill of the expert who carries out the evaluation. The pioneers of this approach are Siemiatycki and his colleagues in Montreal.[33] Since 1979 they have used probing interviews to collect detailed occupational histories from patients with newly diagnosed cancer and from controls in the general population. These histories are then translated by a team of chemists and engineers into lists of potential chemical and physical exposures. Among the possible hazards suggested by their work is a risk of lung cancer due to nickel, particularly in stainless-steel welders.[34] The analysis of their data is still at an early stage, however, and it remains to be seen whether further such leads will emerge. The method should in theory improve upon conventional analytical techniques. However, it is expensive in the time commitment of skilled personnel. Hence only limited populations can be covered.

Geographical analyses of mortality

Job title is not the only marker for occupational exposure which can be used in screening studies. Information about industrial hazards can sometimes be inferred from geographical differences in cancer incidence. Geographical patterns of morbidity and mortality have been studied for many years, but the past decade has seen new developments in this method of searching for environmental risks to health. The amalgamation of data for periods as long as 10 or 20 years has allowed analyses of mortality which are based upon relatively small population units, for example counties of the United States,[35,36] and local authority areas in England and Wales.[37] For a particular cause of death, areas with an unusually high mortality are picked out as promising places for further investigation. In addition, correlation of mortality rates with indices of environmental exposure may give direct clues to the causes of a disease. The method lends itself particularly to studies of industrial carcinogenesis since information on patterns of employment and industry within local areas is readily available from censuses. It has the advantage that hazards which are confined to one or two localized sections of an industry are not obscured by the absence of a risk elsewhere. Its main weakness is that risks which are not geographically localized cannot be detected. For example, the method would never reveal an occupational hazard specific to postmen.

In the United States, where this new approach to screening was first developed, an analysis of deaths from lung cancer revealed a previously unrecognized zone of high mortality along the eastern seaboard.[35,36] Subsequently this was shown to result, at least in part, from the use of asbestos in shipyards during World War II.[38] In England and Wales shipbuilding towns are prominent among the local authority areas with high mortality from mesothelioma in men, while high death rates from nasal cancer have been observed in centres of the furniture and shoe industries.[39] These findings were made at a time when the carcinogenic properties of asbestos, hardwood dust and leather dust were already well known. However, the fact that their effects are demonstrable by this method gives hope that it will reveal other unrecognized hazards in years to come.

SELECTION OF CLUES FOR FURTHER INVESTIGATION

In describing the various approaches which have been used to monitor for new occupational carcinogens, I have drawn attention to specific weaknesses which may have contributed to their failure to reveal more hazards. An additional problem which is common to all the screening methods is the difficulty of selecting leads for follow-up. The observant clinician who notices a cluster of disease in an occupational group has only one clue to investigate, but large epidemiological surveys generate a profusion of hypotheses. Usually it is the investigators who have carried out the study who are most strongly motivated to follow up the leads produced, but they do not have the resources to test more than one or two hypotheses in depth. It is possible that some carcinogens have been missed because leads arising from epidemiological studies have not been adequately pursued. An example of this phenomenon is the long delay between the initial report of Henry and colleagues on high bladder cancer rates among coal gas manufacturers and the final conclusive demonstration of the hazard.

Inevitably, if a study entails comparisons of cancer incidence between many different occupational groups or geographical areas, and for many different types of cancer, then a number of associations can be expected to achieve high levels of statistical significance simply by chance. This has led some critics to dismiss hypothesis-generating exercises as 'fishing expeditions'. Others have suggested that attention should only be given to associations which reach very high levels of statistical significance. However, there is no logical reason why the interpretation of a given association should be any different when it is found as one of a large series of observations rather than as a result of an isolated test. In both situations evaluation depends upon any biases inherent in the study method, the strength of the statistical association, and the weight of evidence from other sources supporting the association. What distinguishes the hypothesis-generating study is the low level of *a priori* evidence for most of the associations examined. However, this does not apply uniformly to all potential hazards. Given the weight of evidence which now exists for a risk of lung cancer in butchers and slaughtermen,[22] it would be foolish to dismiss an association of bronchial carcinoma with meat handling just because it was not highly significant statistically.

The most difficult aspect of selecting hypotheses for further investigation is the evaluation of evidence from other sources for or against the suspected hazard. Dubrow and Wegman have recently collated results from 12 large studies of occupational mortality and cancer incidence, and picked out the most consistent disease associations.[40] This seems a sensible idea which in years to come may improve our choice of clues for follow-up.

PROSPECTS FOR THE FUTURE

Given the fortuitous circumstances surrounding the discovery of some occupational cancers, it would not be surprising if other hazards were still lurking undetected. Moreover, the most stringent laboratory evaluation of new products and processes cannot guarantee against the further introduction of carcinogens into the occupational environment of the future. It is important therefore, that within economic constraints we adopt the most effective strategies of monitoring epidemiologically for such hazards.

In this section I shall argue the need for a diversity of screening approaches which should include the routine surveillance systems that are already in operation. Many of these screening methods could be improved, however, and I will suggest a number of modifications which might enhance their performance. In addition, I will propose the use of complementary *ad hoc* surveys, and discuss briefly the criteria by which populations might be selected for inclusion in such studies. Finally, I shall stress again the importance of adequately following up the leads generated by screening studies.

The need for a diversity of approaches

No single method of screening is ideally suited to the detection of all industrial carcinogens. Long and detailed occupational histories may not be necessary if the harmful exposure is restricted to a skilled occupation in which most workers remain throughout their professional lives—for example dentistry. If such an exposure causes a highly fatal cancer, the risk will be as well demonstrated in an analysis of occupational data obtained from death certificates as by any other method. If the tumour only rarely causes death, a study of cancer registration data will be more efficient. Studies using full lifetime occupational histories are more expensive, but would be the method of choice when risks occur in jobs with a high turnover of personnel. Geographical analyses of mortality are best suited to the detection of hazards in localized industries. Because the nature of new hazards cannot be predicted in advance, a combination of screening methods is indicated.

Established surveillance systems

Despite their poor track record to date, it would be unwise to abandon the established systems of surveillance. However, we should seek, where possible, to refine and augment these techniques. For example, most of the standard classifications of occupation and industry were not designed primarily for monitoring industrial health hazards. Analyses of occupational mortality and cancer incidence might be improved by the development of more specific classifications which relate more closely to patterns of exposure in the workplace. Job-exposure matrices are another possible means of enhancing such studies, but initial experience suggests that they may only be useful for the examination of widespread exposures.[32] The most informative method of analysing occupational mortality and cancer incidence in a large population is by the individual linkage of death and cancer registrations with census records. In the absence of a national scheme of identification numbers, however, such linkage is expensive, and it is likely that many countries will refrain from adopting this approach in the foreseeable future for reasons of cost and concern over the confidentiality of personal information. Where analyses continue to use occupational histories abstracted from death certificates there is often scope for improvement in the quality of information so obtained —for example by recording the deceased person's main lifetime occupation as well as his most recent job, and by regularly noting the industries in which occupations were carried out. Neither of these measures is currently practised in England and Wales where this analytical technique has been applied most often and on the largest scale.

The interpretation of geographical differences in mortality might be enhanced by concentrations on key areas in which a high proportion of the population is employed in a single industry. The effects of any hazard should be more clearly manifest in such areas than in places where only a very small section of the local populace is exposed to the carcinogens.

Complementary *ad hoc* surveys

In addition to the regular surveillance of large populations, there is a need for complementary *ad hoc* surveys of smaller population groups aimed at the type of hazard which would not be apparent in the routinely generated statistics. Siemiatycki's study in Montreal is an example of this type of investigation, and the results of his research are eagerly awaited. Even if his enterprise is unsuccessful, it does not necessarily follow that the method is invalid. It may be simply that there are no unrecognized hazards in the population which he has chosen to study. However, it is unlikely that anyone will repeat such an expensive and time-consuming exercise if no positive findings emerge from Montreal.

When, as in Montreal, financial considerations limit the size of an investigation, the choice of the study population can be of crucial importance. In the past, populations have usually been selected for study either because they have an abnormally high incidence of disease, or for the convenience of the investigator—data are often most readily available close to home. Sometimes, however, it may be more sensible to choose places for study because they are centres of suspect industries. For example, in a survey recently conducted by the Medical Research Council the study population was selected to include areas with heavy concentrations of chemical industry.[41] Many of the known occupational carcinogens are intermediates or products of chemical manufacture.

In addition to its geographical limitation, the above study was restricted to cancers occurring in men under the age of 55. The rationale for this decision was first that diagnostic and occupational information was likely to be more accurate for young persons; and second that for several known occupational carcinogens—bischloromethyl ether for example—the relative risk of cancer appears to be highest and therefore most readily detectable at young ages. It seemed possible that this approach might reveal hazards which were not obvious in routine statistics of occupational mortality.

Follow-up of clues

The other area of activity which offers scope for improvement in years to come is the evaluation and further investigation of clues arising from screening studies. Whatever the methods by which they are derived, it is vital that leads are properly assessed and appropriately followed up. Primary responsibility for this task must lie with the scientists who carry out the hypothesis generating exercises, since in the final analysis, the value of such studies can be determined only by following up the clues which they produce.

The improvements in methodology which I have discussed should in the future enhance the search for new industrial carcinogens. However, even the most effective epidemiological systems of surveillance cannot be certain of detecting all hazards. Thus, there will always be a need for vigilance among occupational health professionals who may still be the first to recognize a previously unsuspected problem.

REFERENCES

1. Bridbord K, Decoufle P, Fraumeni JF et al. *Estimates of the fraction of cancer in the United States related to occupational factors*. Prepared by NCI, NIEHS, NIOSH; draft of September 15 1978.
2. Doll R and Peto R. The causes of cancer. *JNCI* 1981; **66**:1191–1308.
3. Doll R. Pott and the prospects for prevention. *Br. J. Cancer* 1975; **32**:263–72.
4. Henry SA, Kennaway EL and Kennaway NM. The incidence of cancer of the bladder and prostate in certain occupations. *J. Hyg. Camb.* 1931; **31**:125–37.

5. De Villiers AJ and Windish JP. Lung cancer in a fluorspar mining community. *Br. J. Ind. Med.* 1964; **21**:94–109.
6. Case RAM and Hosker ME. Tumour of the urinary bladder as an occupational disease in the rubber industry in England and Wales. *Br. J. Prev. Soc. Med.* 1954; **8**:39–50.
7. Acheson ED, Cowdell RH and Jolles B. Nasal cancer in the Northamptonshire boot and shoe industry. *Br. Med. J.* 1970; **1**:385–93.
8. Registrar-General. *14th Annual Report of the Registrar-General of Births, Deaths and Marriages in England.* London: HMSO, 1855.
9. Registrar-General. *Supplement to the 55th Annual Report of the Registrar-General Part II.* London: HMSO, 1897.
10. Guralnick L. *Mortality by Occupation and Cause of Death among Men 20 to 64 years of age, United States 1950.* Washington DC: US Dept HEW, vital statistics —special report 53, No. 3, 1963.
11. Heasman MA and Lipworth L. *Accuracy of Certification of Cause of Death. (General Register Office Studies on Medical and Population Subjects No. 20).* London: HMSO, 1966.
12. Cameron HM and McGoogan E. A prospective study of 1152 hospital autopsies. I. Inaccuracies in death certification. *J. Pathol.* 1981; **133**:273–83.
13. Percy C, Stanek E and Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *Am. J. Public Health* 1981; **71**:242–50.
14. Heasman MA, Liddell FDK and Reid DD. The accuracy of occupational vital statistics. *Br. J. Ind. Med.* 1958; **15**:141–6.
15. Office of Population Censuses and Surveys. Occupational Mortality 1970–72. *Decennial Supplement for England and Wales (Series DS No. 1).* London: HMSO, 1978.
16. Milham S. *Occupational Mortality in Washington State 1950–79. (DHHS (NIOSH) Publ. No. 83–116).* Washington DC: US Govt Printing Office, 1983.
17. Coggon D, Pannett B and Acheson ED. Use of job-exposure matrix in an occupational analysis of lung and bladder cancers on the basis of death certificates. *JNCI*, 1984; **72**:61–5.
18. Jensen OM and Lynge E. Job exposure linkage systems for cancer research in Scandinavia. In: Acheson ED (ed.) *Job Exposure Matrices (Scientific Report No. 2).* Southampton: MRC Environmental Epidemiology Unit, 1983.
19. Fox AJ. The role of OPCS in occupational epidemiology: some examples. *Ann. Occup. Hyg.* 1979; **21**:393–403.
20. Walrath J, Rogot E, Murray J et al. *Mortality Patterns among US Veterans by Occupation and Smoking Status. (NIH Publication No. 85–2756).* Washington DC: US Govt Printing Office, 1985.
21. Milham S. Mortality in aluminum reduction plant workers. *J. Occup. Med.* 1979; **21**:475–80.
22. Fox AJ, Lynge E and Malker H. Lung cancer in butchers. *Lancet* 1982; **i**:165–6.
23. Mack TM. Cancer surveillance program in Los Angeles County. Natl Cancer Inst. Monogr. 1977; **47**: 99–101.
24. Coggon D and Acheson ED. Trends in cancer morbidity in England and Wales. *Health Trends* 1981; **13**:89–93.
25. Norell S, Ahlbom A, Lipping H et al. Oesophageal cancer and vulcanisation work. *Lancet* 1983; **i**:462–3.
26. Decoufle P, Stanislawczyk K, Houton L et al. *A Retrospective Survey of Cancer in Relation to Occupation. (US Dept Health Educ. Welfare Publ. No. (NIOSH) 77–178).* Washington DC: US Govt Printing Office, 1977.
27. Williams RR, Stegens NL and Horm JW. Patient interview study from the Third National Cancer Survey: overview of problems and potentials of these data. *JNCI* 1977; **58**:519–24.

28. Williams RR, Stegens NL and Goldsmith JR. Associations of cancer site and type with occupation and industry from the Third National Cancer Survey interview. *JNCI* 1977; **59:**1147–85.
29. Cartwright R. Occupational bladder cancer and cigarette smoking in West Yorkshire. *Scand. J. Work. Environ. Health* 1982; **8 (Suppl. 1):**79–82.
30. Office of Population Censuses and Surveys. *Classification of Occupations, 1970.* London: HMSO, 1970.
31. Hoar SK, Morrison AS, Cole P et al. An occupation and exposure linkage system for the study of occupational carcinogenesis. *J. Occup. Med.* 1980; **22:**722–66.
32. Pannett B, Coggon D and Acheson ED. A job-exposure matrix for use in population based studies in England and Wales. *Br. J. Ind. Med.* 1985; **42:**777–83.
33. Siemiatycki J, Day NE, Fabry J et al. Guest editorial. Discovering carcinogens in the occupational environment: A novel epidemiologic approach. *JNCI* 1981; **66:**217–25.
34. Gerin M, Siemiatycki J, Richardson L et al. Nickel and cancer associations from a multicancer occupation exposure case-referent study: preliminary findings. *IARC Sci. Publ.* 1984; **53:**105–15.
35. Mason TJ, Mckay FW, Hoover R et al. *Atlas of Cancer Mortality for US Counties 1950–1969.* Washington: US Govt Printing Office, 1975.
36. Blot WJ, Fraumeni JF, Mason TJ et al. Developing clues to environmental cancer: a stepwise approach with the use of cancer mortality data. *Environ. Health Perspect.* 1979; **32:**53–8.
37. Gardner MJ, Winter PD, Taylor CP et al. *Atlas of Cancer Mortality in England and Wales, 1968–78.* Chichester: Wiley, 1983.
38. Blot WJ, Harrington JM, Toledo A et al. Lung cancer after employment in shipyards during World War II. *New Eng. J. Med.* 1978; **299:**620–4.
39. Gardner MJ, Winter PD and Acheson ED. Variations in cancer mortality among local authority areas in England and Wales: relations with environmental factors and search for causes. *Br. Med. J.* 1982; **284:**784–7.
40. Dubrow R and Wegman DH. Setting priorities for occupational cancer research and control: synthesis of the results of occupational disease surveillance studies. *JNCI* 1983; **71:**1123–42.
41. Coggon D, Pannett B, Osmond C, et al. A survey of cancer and occupation in young and middle-aged men: I—Cancers of the respiratory tract. *Br. J. Ind. Med.* 1986; **4:**332–38.

4. ELECTROMAGNETIC FIELDS AND HEALTH EFFECTS

J. A. Bonnell

Man's interest in electricity dates from ancient times as depicted by the 470 BC bronze statuette in the Berlin museum of the *Zeus of Dodona* hurling his thunderbolt. The history of electropathology which is the study of the effects of the passage of an electric current through a living organism is, on the other hand, much more recent. An early experiment by Benjamin Franklin in 1751 is of interest, in which a chicken was rendered unconscious by an electric discharge to the head from a Leyden jar; the bird was revived by 'insufflation'. This must be the first account of the use of artificial respiration in the treatment of electric shock.[1]

These first studies were concerned with static electricity. Later studies looked at ventricular fibrillation thresholds, human body impedence and the differences due to the nature and types of current—direct current (DC) and alternating current (AC).

It is apparent that all forms of life on the planet have been subjected to electromagnetic fields from the beginning of time, for example from electric storms and from the earth's steady magnetic field which is 0·5 Gauss (G). However, during the past 50–100 years exposures from man-made sources have increased enormously. These increases have been due to the technological advances of the modern age which are associated with the increased use of short waves in radar and radio communication and the use of microwaves of different frequencies both for military and domestic purposes. Added to these is the growth of electrical power generation, transmission and distribution and the proliferation of electrical equipment in industry and in the home. Although significant increases have occurred in electromagnetic fields over a wide range of frequencies in the general environment, this chapter is concerned only with extremely low frequency (ELF) fields in the range 50–60 Hz, that is power frequency electric and magnetic fields. It is generally accepted in Europe that the frequency range 30–300 Hz is designated as 'extremely low frequency' (ELF).

Power transmission by high-voltage AC overhead lines is a well-established technology. The national electricity transmission grid in the UK came into full commercial operation in 1935. Many of the associated problems, such as noise, radiated radio and television interference, the danger of flashover, are well understood and suitable arrangements are made to minimize them.

ELECTRIC FIELDS

The electric field around the high-voltage conductors of an electric power transmission system depends both on the transmission voltage and on the shape and disposition of the conductors relative to their surroundings. Close to the conductors of a typical CEGB 400 kV line, the field is spatially non-uniform and intense (typically 1–2 MV/m) because the conductors have a relatively small cross-section and a high surface curvature (*Fig. 4.1a*). Near a flat ground surface beneath the transmission line, the field is essentially vertical, relatively uniform and much less intense.

Under 400 kV transmission lines in Great Britain, electric fields near a flat ground surface are usually below 5 kV/m but may rise in a very few places to a little above 10 kV/m. The field strength falls off to less than 1 kV/m at points more than about 25 metres from the centre line of a route. Electric fields in substations are of comparable strength, but in a few places rise to about 20 kV/m. The presence of electrically-conducting objects on or near the ground can considerably perturb the field pattern. In this context, most objects, including vegetation, buildings and people, are good electrical conductors, as is the ground itself. Above an object, the equipotential surfaces bunch together (*Fig. 4.1b*) and the electric field strength is increased to an extent which depends on the object's shape. At the top of the head of a person standing upright,[2] this increase would normally be by a factor somewhere between 10 and 20. To avoid ambiguity, in all instances electric-field strengths are expressed as 'unperturbed' values.

Increased fields above a conducting object are associated with reduced fields nearby and below. Trees and buildings, for example, provide substantial screening of the electric fields in their neighbourhood. Inside a house under overhead lines, measured fields have been found to be less than 10 per cent of their 'unperturbed' strengths some distance away.[3]

Induced currents and voltages

For alternating-current transmission, the electric field between the conductors and the ground is associated with surface charges which vary sinusoidally with time. The rate of change of charge represents

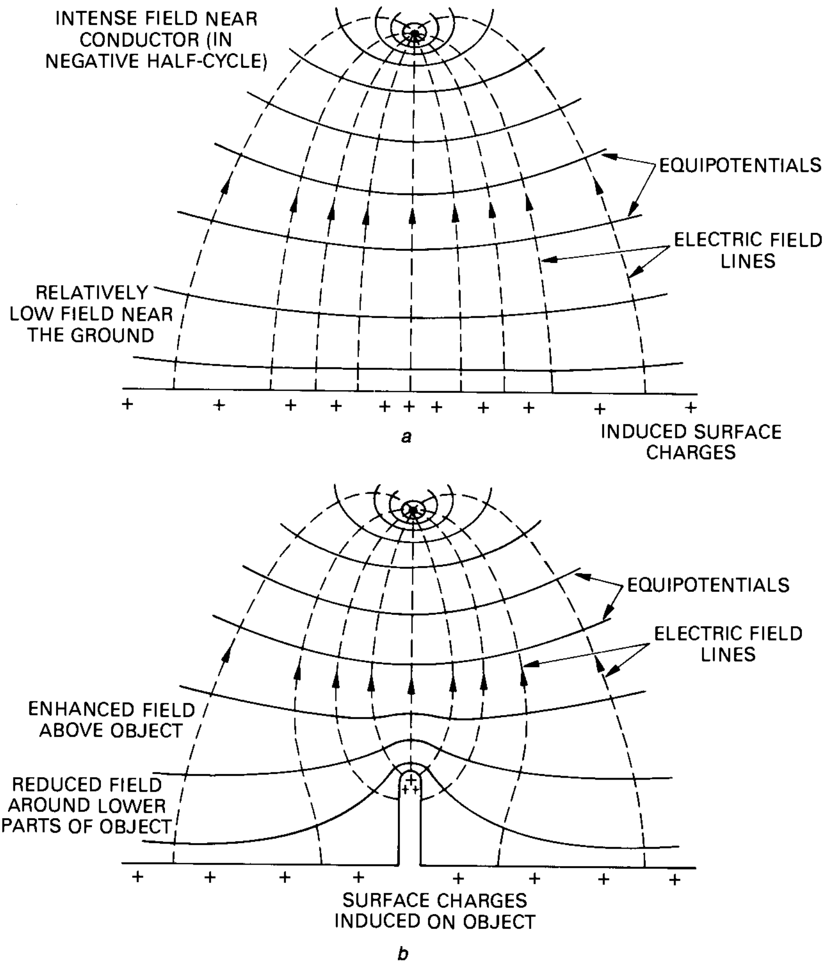


Fig. 4.1. a, Unperturbed electric field pattern below a high-voltage conductor (schematic). *b*, Perturbed field pattern introduced by a grounded object (schematic).

a current flowing from within the ground or the conductor to its surface.

A conducting object exposed to the alternating electric field has similar alternating surface charges induced on it and alternating currents flowing within it. The magnitude of the charges, and hence also of the internal currents, is proportional both to the unperturbed displacement-current density and to a factor involving the object's size, shape and position relative to the ground (*Fig. 4.2*). For a person

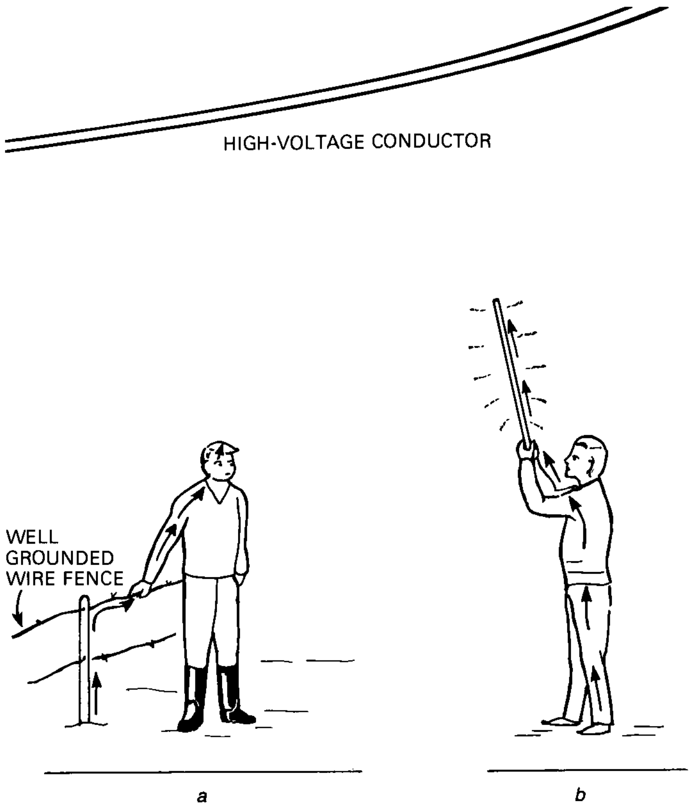


Fig. 4.2. a, If a person is in poor contact with the ground, current may flow through him via earthed objects. *b*, Current flow may be enough to light a gas discharge tube held in the air.

standing upright in a 50 Hz electric field of unperturbed strength 10 kV/m, the internal current would be about 140–150 μA (1 kV/m induces an internal current flow of about 14 μA).

Currents of this magnitude are well below the threshold of perception.[4] However, the alternating charges induced on the skin, hair or clothes are acted on by the electric fields to produce (100 Hz) mechanical forces. There will also be (50 Hz) forces on any static charge that the body may have accumulated. These forces are believed to account for the tingling sensations sometimes felt by people standing in alternating electric fields. A survey in the United States[5] found that 10 per cent of subjects could detect fields as small as 2 kV/m in this way. At the other extreme, 10 per cent were unable to detect fields of less than 30 kV/m.

Within the body, the voltage gradients are very much less than those outside, being determined by the current density and the largely resistive impedance of the tissues. The overall body resistance is about 1000 ohms (Ω), much of this being across the skin. The internal resistance is about 300 Ω. The capacitive coupling between the body and the overhead conductor has an impedance of about 10^9 Ω, depending on body size, posture and so on and may be considered as the controlling impedance for current flow through the body to ground.

Relatively large voltage differences may, however, appear between the body and ground, or between the body and other objects if the body is not in good electrical contact with the ground. If a person touches some object which is in better electrical contact with the ground, body currents may flow to ground via that object instead of taking the normal path through feet and shoes (*Fig. 4.3*). This

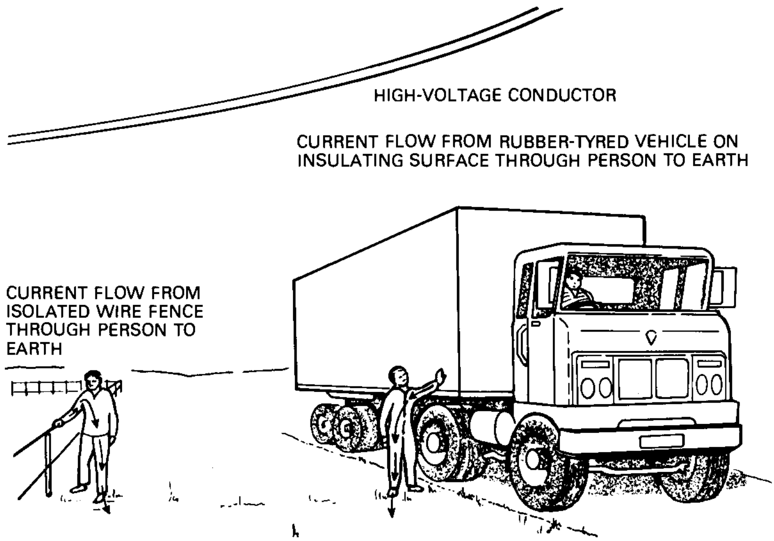


Fig. 4.3. A person in good contact with the ground may provide a route for current flow.

situation could occur, for example, when holding a metal-shafted walking stick, or when walking through long grass which brushes against the legs.

The human body may itself provide a path to ground for alternating charge induced on objects which are electrically isolated (*Fig. 4.3*). Umbrellas, fences or vehicles, for example, can be such a source of charge. Contact with very large vehicles may induce continuous body currents as high as a few milliamps—enough to be consciously, even strongly felt if the local current density is high. Sufficiently large currents may cause muscle seizure, (the 'no-let-go'* threshold is 10 mA) or heart fibrillation[4] (hand-to-hand threshold 50–100 mA), although such currents do not occur in any normal situation in which either employees or members of the public are likely to find themselves.

Normally, any sensation due to currents flowing via other objects is confined to the moment just before physical contact is made, when a momentary spark discharge may occur. 'Microshocks' of this kind can be compared with those commonly experienced in dry atmospheric conditions on touching a car door handle, or on touching an earthed object after walking across a nylon carpet. A survey of annoyance from sparks to the index finger caused by touching an earthed rod showed[5] that 10 per cent of subjects were annoyed in a field of 3 kV/m, whilst nearly all subjects were annoyed in a field 15 kV/m.

Measurement of electric fields

Although electric fields can, in some circumstances, be readily calculated,[6,7,8] measurement is often an easier way of finding what the field strength is at a given location, particularly for some of the more complex conductor configurations found in substations.

Alternating electric fields may be measured by an instrument which consists in essence of a microammeter connected between the halves of a split metal box whose size is normally a few centimetres on a side. The operator holds the instrument a metre or two away from him on the end of an insulating rod so that his presence does not perturb the measured field. Conducting objects significantly distort fields around them out to distances comparable with their own leading dimension. Distortion of the field by the meter itself is accounted for in the calibration, which is done between widely-spaced electrodes of a shape which allows the field strength to be computed from the voltage applied to them. In most cases, the vertical component of the field varies by only a few per cent within 2 m of a flat ground surface. CEGB practice is to make measurements at a height of 1·8 m above the ground (head height) so that the meter is well clear of surface irregularities, such as vegetation, and yet still in a convenient position to read.

A cruder measurement of field can be obtained by an observer wearing insulating footwear and carrying a metallic stick fitted with a microammeter to record the current flowing through the stick when it contacts the ground. The current is proportional to the unperturbed

* So called 'let-go' current is one which prevents the ability to release the hand grip—'no-let-go' is a better term.

field strength, but depends on the height and build of the observer as discussed above.

Assessment of exposure

A convenient unit of personal exposure is kilovolts per metre multiplied by hours of exposure (kv/m × h)—that is, the integrated product of field strength and time (measured in hours). While this unit of exposure is convenient, it should be stressed that there is little evidence for supposing that it has any significance in terms of biological effects—for example in suggesting that long exposure to low fields is in any way equivalent to short exposure to high fields. An approximate measure of exposure may be obtained by means of a simple integrating current meter—consisting of a pair of sensor plates, a rectifier and an electrochemical accumulator—attached at some point on the body surface such as the shoulder. The accumulator is interrogated electronically after exposure.

MAGNETIC FIELDS

Fields from overhead lines

As described above the *electric* field depends on the voltage—which remains virtually constant—while the *magnetic* field depends on the current which may vary widely according to demand. Since the ground is virtually non-magnetic and since its conductivity is too low for significant eddy currents to be induced in it, the power-frequency magnetic fields generated by the current in the line conductors are substantially unmodified by the ground. For similar reasons, most objects and buildings provide little or no screening.

The maximum field indicated for the highest rated line is about 0·4 G, assuming a current of 2 kA in each circuit and the normal arrangement of phases. As a comparison, the earth's steady magnetic field is about 0·5 G. At 25 m from the centre line, the field will usually be below 80 mG (0·08 G).

Fields from underground cables

High-voltage underground cables (400 kV, 275 kV and sometimes 132 kV) generally use separate phase conductors in a horizontal array. The field is then predominantly horizontal directly above the cable route, but further away it becomes predominantly vertical and falls off inversely as the square of the distance. For a 400 kV single circuit with a phase spacing of 0·4 m, buried at a depth of 1 m and carrying a current of 1000 A, the maximum field at ground level will be about

1·3 G, falling to about 200 mG at head height and to about 65 mG at a lateral distance of 5 m.

At lower voltages, all three phases are normally contained in one cable. Neutral and earth conductors (sometimes combined) may also be included, as for example in the 415 V mains supplies under most streets. For small loads, such as individual houses, 240 V single-phase service cables tee-off the street cable.

The individual conductors in a three-phase cable are laid helically to make the cable more flexible. This helical lay, combined with the small separation of the conductors, gives rise to relatively small external magnetic fields, which fall off very rapidly with distance away from the cable.

However, if the vector sum of all the currents in the cable is not zero, that is, there is some net unbalance or zero-sequence component with its return elsewhere, then the external field will be greater, even for quite small degrees of unbalance. It will be proportional to the unbalance, largely vertical and will fall off inversely as the distance. There are often several possible return paths for the out-of-balance currents, including other neutral conductors, various metallic service pipes, or the earth itself. Any net unbalance in any other three-phase system, underground or overhead, will have a similar effect on the field at a distance.[9]

Fields in the home

The magnetic fields from domestic appliances fall off quite rapidly with distance and at one or two metres have generally reached background level. Investigations made so far in the UK suggest that this background level is typically a few tenths of a milligauss and is often set by the field from unbalanced currents in various cables under the street.[10] Not only do the normal currents in such cables change markedly in response to daily and seasonal changes in the demand for electricity, but they also fluctuate from moment to moment as equipment is turned on and off.

A further element in the variability of the resulting fields probably arises because the degree and spatial pattern of unbalanced current fluctuate in a way not directly related to the load changes. *Fig. 4.4* illustrates the overall effect, which makes the characterization of the field at a given location very difficult. Because of this variability and because the field from the more obvious sources, such as overhead lines, falls off quite rapidly away from the source to become indistinguishable from the general background, caution is necessary in studies which seek an association between some feature of health and the fields from electrical plant, merely through proximity to the latter.

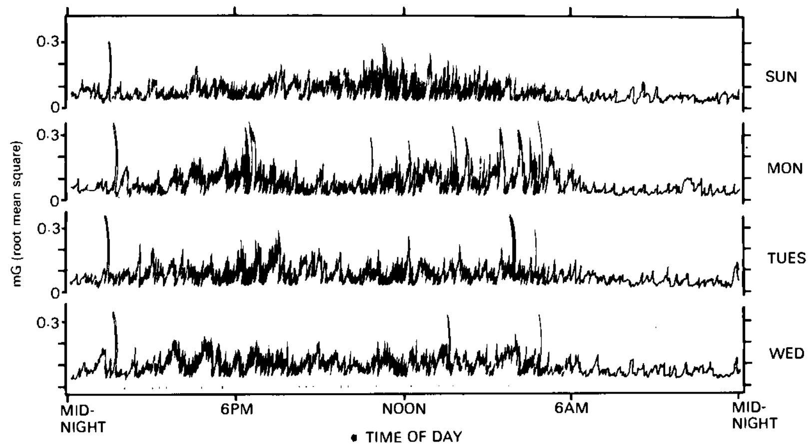


Fig. 4.4. Variations in magnetic field with time of day.

Induced currents

Alternating magnetic fields induce circulating currents in conducting bodies. The peripheral current density is an order of magnitude less than those induced by electric fields of comparable strength, thus it may be imprudent to study exposure to magnetic fields without taking note of any electric fields present.[9]

STUDIES OF HUMAN EXPOSURE

It was the study of complaints by Russian substation workers in the mid-1960s which triggered off the worldwide interest in high-voltage field effects.[11] The symptoms complained of were non-specific and included headache, fatigue and general gastrointestinal disturbances. The authors of the paper state that they gained the impression that changes in the cardiovascular system were encountered more frequently and were more marked in persons systematically subjected to electric fields—for example maintenance personnel—than those exposed sporadically such as signalmen and substation attendants. None of these statements was supported by quantitative data. There is no detailed account of the incidence of disorders in relation to the duration of work in electric fields, nor is the incidence of similar symptoms in unexposed workers discussed. No objective basis is given for asserting that the observed disorders were in any way related to electric fields.[12]

Further reports of positive findings in men occupationally exposed to electric fields have been made by a number of Soviet authorities since this time.[13,14,15] These tended to confirm the findings of Asanova and Rakov[11] and reported loss of libido in addition to the non-specific symptoms referable to the nervous, cardiovascular and digestive systems.

The fact that these disturbances were found mainly in men who had been exposed for long periods to strong fields led the Soviet authors to put forward the idea of a threshold effect. After a study with volunteers exposed to fields of 0–30 kV/m, Fillippov[15] reported that the haematological changes only occurred with fields of over 5 kV/m. He considered this to be the threshold above which it would be appropriate to limit either the length of exposure or the field strengths to which subjects were exposed.

On the other hand, Danilin et al.[16] considered that these symptoms could have been caused by other environmental agents and they suggested kerosene and gasoline vapour, materials which are frequently used by substation workers in Russia. It is also of interest to note that at a USA–Soviet symposium on extreme high voltage (EHV) AC power transmission held in Tashkent, USSR, B. M. Savin et al. (1978 unpublished) expressed reservations about the validity of some of the earlier Soviet reports of adverse effects due to electric fields. This view was endorsed by Bourgsdorf (1980)[17] at the 1980 Conference Internationale des Grand Reseaux Electriques a Haute Tension (CIGRE) conference in Paris who stated that operational experience with 750 kV lines confirmed that the expected dangerous biological effects of electric fields had been over-estimated.

Most of the research done in Europe and the USA has produced negative results. The first of these was by Kouwenhoven et al.[18] and Singewald et al.[19] who studied 10 linesmen over a 9-year period. Body currents calculated for these men were at times as great as those to be expected from exposure in unperturbed fields of up to 25 kV/m. The medical tests were extensive and no abnormality ascribable to electric fields was found although the number of subjects was too small for any but the most gross effects to emerge. There were no control subjects.

The only report of symptoms apparently associated with electric fields in Western Europe are those of Fole[20] and Fole and Dutrus.[21] They reported observations on 3 workers transferred from 200 kV to 400 kV substations who complained of vertigo, visual disturbances, nausea and lassitude. In the later paper[21] 6 subjects were exposed to fields of 15 kV/m for several hours and 2 subjects complained of loss of strength 'in the body', and changes in blood pressure and pulse rate were recorded. The brevity of these studies, the lack of controls, and the failure to confirm or substantiate any direct connection between the observed symptoms and the electric fields suggest that these reports are of doubtful validity.

In a study of families of Electricité de France employees, Strumza[22] reported on a four-year investigation of 70 men, 65 women and 132 children living within 25 m of 200 and 400 kV lines. His control group consisted of 74 men, 64 women and 120 children living more than 125 m from lines. He failed to discover any difference on the basis of medical records, frequency of visits to family doctors, or expenditure on pharmaceutical prescriptions.

Occupational health surveys

Malboysson[23] studied a group of 84 substation workers and 76 linesmen in Spain and compared these with 94 linesmen working on low-voltage systems. The men were examined over a four-year period by questionnaire, medical history, and medical haematological and retinal examinations. Biochemical tests consisted of blood glucose, serum cholesterol, triglycerides, urea and uric acid. The linesmen in both groups showed less sickness absence than the substation workers but there were no apparent adverse effects due to work in electric fields. No exposure measurements were taken and the data were not analysed statistically.

Roberge[24] studied 56 maintenance workers employed in 765 kV substations in Quebec; they were exposed to fields of up to 15 kV/m over a 4½-year period. There were no controls and the men who volunteered were asked to complete a questionnaire. There were no gross effects noted on the health of these workers. It should be noted, however, that the exposure measurements were poor, duration was not quantified and there were no control groups for assessment of the questionnaire responses. Additionally, the men taking part were volunteers. Two important suggestions emerged from this study: first, 22 of the men taking part expressed fear and anxiety of electric shocks; second, the ratio of sons to daughters of these men was disproportionately high (17:3).

A second Canadian study was reported by Stopps and Janischewsky[25] in which 30 high-voltage maintenance men from Ontario together with 30 employees, matched for age and educational level but not exposed to electric fields, were exhaustively studied in hospital. The investigation included EEG, ECG, blood biochemistry including tests of liver functions and serum electrolytes, and complete physical and psychological assessment with psychometric and personality tests. At the time of examination the subjects were not identified as being exposed or control. The exposed group consisted of: (*a*)

19 linesmen with an exposure experience calculated to be 7 kV/m hours per day—up to 8000 kV/m hours over 10 years; and (*b*) 11 substation workers with an average calculated exposure of 13 kV/m hours per day—up to 36 000 kV/m hours over 10 years.

It was concluded from this very exhaustive and thorough study that EHV work does not cause chronic ill health in substation staff in Ontario. Despite this it should be noted that the men taking part were all volunteers and represented a small number of total staff employed in these categories. The exposure expressed as kV/m hours was estimated and not measured, but some dosemeter checks were made. No acute effects were sought.

The fourth report is that of Knave et al.[26] in which 53 workers in 400 kV switching stations in Sweden were studied in association with 53 employees working on low-voltage systems: workers were matched by age, location and length of service, but not by educational level. Exposure details are incomplete in that they were assessed and not measured. Exposed and control groups were investigated thoroughly by questionnaire, EEG, ECG, medical and haematological examinations and various psychometric tests. There was no evidence of gross effects from working in 400 kV substations. It is noteworthy that the exposed personnel performed significantly better than the control group, but this may have been due to the better educational level of the exposed group. It was also noted that the exposed group had significantly fewer sons than daughters compared with the control group. This latter finding is the converse to that described by Roberge. These findings have no significance epidemiologically and will be discussed later.

Broadbent et al.[27] reported the results of questionnaire interviews with 390 electrical power transmission and distribution workers in the UK. In this study individual exposure measurements were obtained from 287 of 390 workers and attempts were made to relate these exposure measurements, together with estimates of exposure produced by local management, to several possible indicators of ill health as elicited by a questionnaire interview. The group chosen consisted of men responsible for the operation and maintenance of electrical transmission lines and substations operating at voltages of 132, 275 and 400 kV and distribution staff working on 11, 33, 66 and 132 kV equipment. From the nature of their duties it was expected that the transmission line staff would be exposed to higher electric fields than the distribution workers.

Broadbent et al.[27] looking for possible effects of exposure concentrated their attention on questions modified from the Middlesex Hospital Questionnaire (MHQ). The MHQ has been shown to give higher scores in people assessed as ill by more thorough medical examination. The questionnaire used had been slightly modified for use in industrial rather than hospital populations. It has been demonstrated that the MHQ gives significant scores for anxiety, depression, somatic symptoms and obsessional symptoms. No significant correlations of health indices were found with either measured or estimated exposure to electric fields. Although the general level of health was higher in electrical workers than for manual workers in other industries, there were significant differences in the health measures between different categories of job, different parts of the country and in association with factors such as overtime, working alone or frequently changing shift, thus confirming the sensitivity of the study method.

The most striking feature of the measured exposures reported in this study were the general low levels compared with prior expectations or estimates. These low exposure levels were confirmed by further studies of measured exposure elsewhere in England and in North America.[9] These findings are of importance when assessing the validity of reports purporting to show differences in the progeny of workers employed in 400 kV substations compared with those of 200 kV and 70 kV substation workers in Sweden, no measurements of exposure were taken.[28]

The European and US epidemiological studies are of a higher standard than the Russian reports in that detailed information on all examinations and investigations are provided. Even so, in some studies the medical examinations did not coincide with the period of exposure, the control groups were not always closely matched and the field strength and exposure were estimated and not measured. Despite this the evidence suggests overwhelmingly that occupational exposure to high voltage (ELF) fields does not give rise to chronic ill health.

LABORATORY STUDIES ON HUMAN VOLUNTEERS

Because of the shortcomings of some of the human studies, experimental exposure of people in laboratory tests are an important source of information on short-term effects.

The advantage of laboratory studies is that exposure conditions are much more readily controlled and monitored, although long exposures present practical problems. Hauf[29] exposed volunteers to fields ranging from 1 to 20 kV/m for periods of between 45 minutes and 5 hours. The experiments were not done double blind and it is possible that the subjects may have been able to perceive the fields in some instances. The main criteria used for comparisons were: (1) psychomotor tests—reaction times; (2) measurement of pulse rate, blood pressure, ECG and EEG; (3) peripheral blood counts (routine haematological tests); (4) serum electrolyte and blood biochemistry.

On the whole, these studies have produced negative results, apart from slight stimulation which delayed and reduced the effect of fatigue on reaction times. Slight changes were seen in the constituents of the blood, affecting the number of leucocytes, neutrophils and reticulocytes, compable with those reported by the Soviet authors.

In order to provide a more precise definition of these slight changes. Hauf[30] and his colleagues studied the effects of very low currents induced by low fields. These currents were close to the displacement currents which, from the physical point of view, are the only possible effect of exposure to an electric field. The negative results obtained from this new series of experiments have led Hauf to believe that the effects observed when subjects are exposed to an electric field are the effects of non-specific stimulation.

The work of Hauf[30] and his colleagues accords with earlier Swedish studies by Johansson et al.[31] who found that exposure to electric fields had no influence on performance in psychometric tests. The field strength was quoted as 100 kV/m 'at the head' and may be taken as indicating an unperturbed field strength of 5–10 kV/m.

It would seem that no effects occur at fields below 10 kV/m, but that if people are exposed to fields of 10 kV/m or more for periods of a few hours per day, then transitory physiological changes may sometimes be observed which may be due to apprehension or perception. It is very difficult in experiments of this kind to ensure that people do not perceive the electric field's presence through purely extraneous clues.

More recently, Graham et al.[32] exposed volunteers for 6-hour periods to simultaneous 60 Hz electric and magnetic fields of 9 kV/m and 16 A/m respectively. They found no effects of exposure on performance in a number of psychological tests, although visual and auditory evoked responses were apparently affected and cardiac interbeat intervals were transiently longer after exposure than after sham-exposure.

In a series of elegant experiments, Cabanes and Gary[33] carried out tests of direct perception of electric fields. By an arrangement of horizontal conductors placed at a height of 9 m above the ground, an electric field was produced which varied between 27 kV/m immediately under the conductors to 0·3 kV/m in other positions on the laboratory floor. Seventy-five volunteers were exposed to these field intensities in specified positions and they described their impressions and sensations at these different points, not knowing what the field intensity was at each spot. It was demonstrated that perception of electric fields in human beings was a function of the mechanical stimulation of the hair by electrostatic forces. The sensitivity of perception of the bare forearm depended upon whether the forearm had been shaved. Removal of the hair markedly reduced the sensitivity of the forearm. The vibrations of the hair on the forearm were recorded by high-speed photography; hair movements of 1 mm were noted in fields of 50 kV/m.

A parallel series of experiments[33] on animals confirmed these findings. Again, using high-speed photography, vibrations of as much as 10 mm of the whiskers of rats and mice in fields of 30–50 kV/m could be displayed, indicating clearly the extreme sensitivity of animals to the presence of electric fields.

In order to overcome the possible interfering effects of perception, experiments were carried out at Manchester University in which electrodes were placed on the body surface of volunteers to inject currents simulating exposure to high electric fields for 5½ hours while avoiding external stimulation.[34]

Pairs of male volunteer subjects were connected via electrodes on the scalp, upper arms and feet to a 500 μA, 50 Hz current source. This level of current was chosen as being the highest that could be passed with reasonable confidence that subjects would be unaware of it.[4] Current could be routed to only one subject at a time and the subject being sham-exposed was electrically isolated by a double-pole switch at the supply unit.

The electrodes were distributed so that 40 per cent of the total 500 μA current entered the body via the head and the remainder via the arms, thus crudely simulating the current distribution in a man of average build standing upright in a vertical electric field of 36 kV/m.

Before each test session, a check was made to establish the level of current that each subject could just perceive at a single arm electrode and, as a precaution, subjects were given a questionnaire form on which to record the nature and duration of any similar sensations they might experience during the session.

At the start and finish of each session, each subject completed a mood-adjective questionnaire[35] which probed subjective levels of stress and arousal.

Two of the performance tests probed verbal reasoning skills, the other two tested attentional skills. Seventy-six subjects took part in the study.

Despite the precautions taken to avoid perception, nearly two-thirds of the subjects did, at some time report itching, tingling or prickling sensations, mainly at arm electrode sites. More than half of these subjects reported sensations during sham-exposure. Performance was found to be related to perception in some of the tests.

Two of the measures analysed, however, showed an influence which was independent of perception. First, the arousal component of the mood-adjective checklist and, second, correct-response times for passive statements in the syntactic-reasoning test.

No general influence of exposure emerged in the remaining three performance tests. For both visual search and syntactic reasoning the above effects were equally clear in all four test sequences on the second day—that is there was no suggestion of any dependence on the duration of exposure to current.

It was found that current did not influence feelings of stress. This is an interesting finding in view of the strong correlation of such feelings with reported sensations and lends some support to the idea that stress responses in some earlier studies of the exposure of animals to electric fields[36] may relate to perceived effects, such as hair vibration or microshocks, rather than to effects of the associated induced body currents.

Caution should be exercised in interpreting the results of short-term exposure to current in the laboratory in terms of long-term occupational exposure to electric fields.

Non-occupationally exposed epidemiological studies

A relationship was claimed by Wertheimer and Leeper[37] between childhood cancer and residence in houses with high current flow due to external electrical wiring configurations—for example transformers —in the vicinity of these houses, compared with a control group. The finding was strongest for children who had spent their entire lives at the same address and it appeared to the authors to be dose-related. It did not seem to be an artefact of neighbourhood, street congestion, social class or family structure. They postulated that the correlation may have been due to the effects of current in the water pipes or of AC magnetic fields.

In a critique of this study, Miller[38] commented that a dose-response relationship was suggested, but no doses (magnetic field intensities) were given for any of the addresses. Furthermore, Miller provides evidence that the household magnetic field from electrical appliances in the home would be far in excess of any contributions from electrical wiring configurations in the environment outside the house. A similar study carried out by Fulton et al.[39] in Rhode Island failed to find any evidence to support the Wertheimer and Leeper hypothesis. Additional points of criticism of the Wertheimer and Leeper study are that the data were not collected blind and therefore there is observer bias; and the cases were ascertained after death and therefore no account was taken of cancer cases still alive. It is vital in case/control studies of this nature to ascertain whether birth or death addresses were used. It seemed from the published paper that both birth and death addresses were used, introducing a further strong bias in the data.

The relationship between childhood cancer and 50-Hz magnetic fields generated by overhead power transmission lines was studied in a case/control investigation in the Yorkshire Health Region by a team from Leeds University, the Yorkshire Regional Cancer Organization and the Central Electricity Generating Board (CEGB).[40] Particular care was taken to avoid observer bias and the magnetic field strengths were estimated on the basis of records of actual load currents rather than from spot measurements or from surrogates such as proximity to lines.

All cases of childhood cancer (aged under 15 years at diagnosis) born within the present Yorkshire Health Region and diagnosed over a finite period were identified. The search for these cases was aided by a childhood-cancer register which originated from Health Service records. The first complete year of records was for 1970 and it was, therefore, decided to select case children who were diagnosed in the decade 1970–1979.

Control children were selected from the local authority birth register which contained the record of the case child. These were children born as near as possible in time to the case child and with a nearby address.

Once case and control 'birth' addresses had been ascertained, a manual master index was constructed. Thereafter, a new record was used which contained no indication of which addresses belonged to a case and which to a control. This part of the work was done by an individual not otherwise connected with the study, who produced a total of 966 addresses (376 cases and 590 controls) which were then coded. Of these, 81 were within 100 m of an overhead power line and suitable for study. The coding of these 81 cases and controls was broken only after the relevant information about overhead lines had been assembled.

The data were analysed both on the basis of proximity to the lines and for the calculated magnetic fields. No statistical correlations were seen with lymphoma/leukaemia, solid tumours, or all cancer diagnoses. This study which had been carefully designed to avoid the observer biases of the Wertheimer and Leeper study and to make accurate assessments of magnetic field strength, did not show any apparent relationship between the magnetic fields from overhead power lines and childhood cancer.

Domestic 50 Hz magnetic fields

In the home, contributions to the magnetic field may be produced by internal domestic electrical equipment and wiring arrangements, by external underground transmission and distribution cables or by overhead transmission and distribution lines. For a variety of reasons, it was not possible to investigate historical exposure to magnetic fields from many of these sources. Domestic equipment, for example, produces high local fields which fall off rapidly with distance and normally give rise to short-term exposure, which is difficult to estimate, especially retrospectively. Fields due to wiring arrangements could not be investigated either, because of the detailed and impractical on-the-spot investigation of the wiring this would have entailed, while the fields due to underground distribution cables could not be estimated with any certainty, partly because records of cable currents are not kept and partly because the relevant fields are due to net out-of-balance or zero-sequence currents in the cable, about which there is little information.[9]

The fields associated with overhead lines could be estimated with precision, because long-term records of loads were available and reliable calculations could be made.

Between 1982 and 1985[41,42,43,44,45,46] a number of letters were published in the *Lancet* and the *New England Journal of Medicine* in which it was reported that 'electrical workers' were at increased risk of developing leukaemia. The suggestion was based on either Proportional Mortality Ratio (PMR) studies or Proportional Incidence Ratios (PIR). The study populations were either cases from hospital records, cancer registries or death certificate data collected for occupational mortality studies. The occupations include radio and radar mechanics, telephone installers and repairers, telegraph radio operators, electronic engineers as well as electricians, linesmen, and persons of any occupation employed in the electrical and telecommunications industry. It was not possible from any of these studies to assess exposure to electric and magnetic fields. Indeed, from the described occupations it is most unlikely that the exposure of the majority to ELF fields was any different from that of the general public.[47] It was also apparent that there was exposure to solvents, fluxes and solders. Furthermore, the number of cases was small.[48]

Later studies by Coleman[49] and McDowell[50] failed to substantiate the hypothesis. It is difficult to postulate a pathogenic process whereby ELF fields from power distribution or transmission networks might affect the haemopoietic system. Because of the serious consequences of these proposals further investigation is required to examine the complete working environment of the occupational groups concerned.

A suggestion has been made of a link between electromagnetic field exposure and suicide.[51] The report lacks any biological hypothesis. Suicide is an event and not a disease and is frequently a symptom of a pre-existing psychotic illness. It is these diseases which need to be studied and classified. The paper is contradictory in its conclusions and is open to serious criticism for its incorrect use of epidemiological techniques.

ANIMAL EXPERIMENTATION

Numerous experiments have been carried out on the exposure of animals to electric fields. Reported effects have included reduced water consumption in rats, reduced size of offspring and high mortality in mice,[52] haematological changes in mice,[53] and slower bone healing in rats.[54] In all these experiments the animals were probably experiencing microshocks. Studies on bees in hives under power lines[55] suggest that the insects are affected only when they are subjected to microshocks on contact with the hive structure or with other insects.

At Battelle Pacific Northwest Laboratories, in the most careful and thorough work carried out so far, most of the physiological effects reported elsewhere have been sought but not found.[56] Some effects have, however, been confirmed in rats and mice exposed to fields up to 100 kV/m, fields which the animals could very probably perceive. Rats given the choice were found to spend more time out of fields greater than 90 kV/m than in them, and while in the fields were more active. At lower field strengths (25 and 50 kV/m) the rats spent the greater part of their time within the field zone.

Rats exposed from conception until 8 days after birth showed slight behavioural differences relative to sham-exposed controls, although these differences had disappeared at 21 days after birth. Results of a conditioned test response showed changes in the excitability of sympathetic ganglia in exposed rats and there were also changes in the numbers of red and white cells in the blood of exposed mice.

All the effects so far observed in the Battelle programme[56] are mild and within the normal range of variation for the animals concerned. A notable feature of the recent work is that as experimental techniques are refined so the number of observed effects has fallen.

Additionally, Le Bars and Andre[57] have carried out a comprehensive series of experiments on rats, rabbits and mice to study various haematological and biochemical parameters following exposure to controlled fields in the laboratory. No significant abnormalities have been noted. Reference has already been made to studies carried out on the direct perception of fields in animals.[33] Similarly, studies on the biological effects of electric fields on mice, rats, rabbits and dogs have been carried out in Italy within the 1000 kV project operated by Ente Nazionale per L'Energia Elettrica (ENEL).[58] These studies were concerned with basic cardiovascular, haematological and biochemical parameters and, in addition, growth, fertility and teratogenic effects were also investigated as well as changes in resistance to induced infections. Again no significant abnormalities have been noted to date but the work is continuing.

No animal experiments have yet given clear indications of what might happen to people exposed to electric fields, one of the difficulties being that because of differences in size, shape and orientation, the electric fields and currents do not scale in any simple way, either in magnitude or in distribution over the body.

The effects of electric and magnetic fields on *in vitro* systems have been studied in a few laboratories. This is an important source of information on possible mechanisms for the interaction of induced currents from either electric or magnetic fields on cell membranes.[59] There is an abundance of reports in the literature but essential questions remain unanswered. It is implied that ELF can influence calcium efflux from nerve cells and that direct ELF has the same effect as ELF modulation frequencies. The beneficial effects of ELF on bone healing also remains unproven. It is difficult to appreciate the direct relevance of these studies to human exposure at the levels met with either in the general or the occupational environment.

Epidemiological surveys and experiments with human subjects are essential to resolve satisfactorily the question whether subtle human health effects do exist.

CARDIAC PACEMAKERS

While the human body itself appears insensitive to relatively large induced currents, implanted electronic equipment may be less so and recent studies have established that currents of a few tens of microamperes can affect the performance of some models of cardiac pacemaker. The main effect is to cause the pacemaker to 'revert' unnecessarily to its designed fixed rate mode,[60] although occasionally it may pace in a slow or irregular fashion, which could be hazardous.

Butrous et al.[60] reported tests on 35 patients with various models of pacemaker. In the most sensitive models, reversion occurred at body currents as low as 26 μA, corresponding to a field of 2 kV/m at head height for the patient standing upright. At the other extreme, some models proved immune to all interference up to 20 kV/m, the highest field used. In all cases, the pacemakers resumed normal operation as soon as the field was removed.

It should be stressed that members of the public will rarely encounter power-line electric fields which are large enough to cause pacemaker interference and there has been no known instance of a patient suffering harm in this way.

REGULATION

The enormous amount of animal and other work which has been done over the past decade has given rise to no firm predictions of physiological changes that might be expected as a result of long- or short-term exposure of people to power line fields, either electric or magnetic.

There are three groups of reasons for considering the regulation of people's exposure to power frequency electric and magnetic fields. The first concerns the prevention of excessive irritation from perceived external effects of electric fields—for example microshocks. The second concerns the need to guard against discomfort due to the known perceptible effects of high induced currents in the body, such as visual phosphenes or muscle stimulation. The third concerns the possibility that relatively low and unperceived internal currents or magnetic fields may prove harmful in some way yet to be established.

High alternating electric fields in the environment are associated almost exclusively with large engineering works of the electricity supply utilities. The need to adopt measures to minimize perceived effects is something that all large utilities have coped with very satisfactorily for several decades.

From what has been said already, it will be apparent that the body is normally able to accommmodate internal currents much higher than those induced by exposure to alternating electric fields large enough to cause external perceived effects. Thus, if measures are taken to prevent excessive perception, they will simultaneously be more than adequate to prevent excessive internal currents.

It is tempting to suppress external perception simply by limiting the field strength at places to which people have access. However, such perception depends on other factors—for example the presence and shape of other objects, the scope for making contact, the degree of earthing, different conduction paths and humidity. Imposition of a limiting field strength is therefore not wholly satisfactory. It may be expensive in circumstances where it is not necessary and it may, in other circumstances, be inadequate.

Nevertheless, there is some suggestion that external perception (and possibly effects on pacemakers) should be the basis of a limit on the power frequency electric field strength in areas accessible to the general public. A level of 12 kV/m has been mooted for 50 Hz fields in the UK.[61]

As far as unrecognized effects of induced currents are concerned, the World Health Organization, following a comprehensive review,[48] concluded 'Whilst it would be prudent in the present state of knowledge not to make unqualified statements about the safety of intermittent exposure to electric fields, there is no need to limit access to regions where the field strength is below 10 kV/m'. Since induced currents are directly proportional to frequency, a level of 10 kV/m at 60 Hz (a power frequency used in the USA and Japan) is equivalent to 12 kV/m at 50 Hz. However, the WHO report also noted that

'It is not possible from present knowledge to make a definitive statement about the safety or hazard associated with long-term exposure to sinusoidal electric fields in the range of 1–10 kV/m. In the absence of specific evidence of particular risk or disease syndrome associated with such exposure, it is recommended that efforts be made to limit exposure, particularly for members of the general population, to levels as low as can be reasonably achievable.'

This report typifies the varying attitudes which may be taken in the face of evidence which is, at best, inconclusive. The imposition of arbitrary limits on electric (or magnetic) field strengths will inevitably be taken to imply the existence of public health hazard where none is, in fact, known. If the intention is to protect against perceived effects which constitute an irritation rather than a hazard, then the better approach may well be to address such problems directly as need arises, rather than to impose blanket field limits. Well-established screening and earthing procedures are available for obviating shocks. Simple precautions can be taken, for example, by pacemaker patients, to whom power frequency electromagnetic fields represent one of a number of potential sources of interference. Present practice in these matters has given rise to few problems and the consequences of change should be carefully considered.[62]

DISCUSSION

Apart from the obvious hazard of accidental electrocution, no occupational disease pattern has been reliably reported from among high voltage workers. Uncomfortable or unpleasant subjective sensations, such as induced voltage microshocks, do occur in workers exposed to high-voltage field radiance. It is of significance that as early as 1968 Krivova[63] was speculating on the importance of exposure to microshocks in inducing symptoms in substation staff. Roberge[24] described fear and anxiety of electric shocks amongst 22 of the 56 maintenance workers he investigated in Quebec. Takagi[64] studied the response of 40 subjects under a 500-kV test line in Japan, and in fields of 5 kV/m and upwards the microshocks gave rise to definite unease amongst the participants. Clearly, therefore, in the design of any laboratory experiment the elimination of microshocks is a vital prerequisite.

A constant theme of Soviet work is that problems arise only in substations where voltages are 500 kV or more. Krivova[63] provide data on field strengths in such substations. Maximum levels are between 20 and 25 kV/m which is high by Western standards. The incidence of annoyance from shocks would, on the basis of American work, probably also be high.

An observation which has given rise to some speculation has been made in two reports: Roberge[24] found an abnormally high male to female ratio of children born to high voltage field workers, and Knave[26] reported an abnormally high proportion of female children. However, Knave[26] emphasized that factors other than exposure to electric fields were probably responsible since the difference in the number of children was found to be present 10 to 15 years before the work in 400 kV stations began. The observation of a difference in the proportion of male and female children in small groups of people would be expected; in fact, even with large populations the proportion is extremely unlikely to be unity. It therefore has no significance epidemiologically but is purely and simply an observation in a cross-sectional study.

Some investigators have found transient minor physiological changes in people after several hours of exposure to fields of 10 kV/m or more. None of these investigators has claimed that such changes are inconsistent with causes such as apprehension or perception of the presence of the field. Tests in laboratory conditions show that people normally do show transient physiological responses to stimulation by small electric shocks such as they might experience in high electric fields near a transmission plant. Experiments on animals have failed to confirm any pathological effects after extended exposure to fields of up to 100 kV/m.

Physiological responses in animals have been confirmed only at fields high enough for perception to be probable. The responses appear to be mild and the results have no clear implications concerning the health of people near transmission plants.

Research programmes are under way in several countries: the topics under investigation range from fundamental problems of the interaction of electromagnetic fields with living material, to direct attempts to identify specific medical effects—a particular example in the latter category being the continuing work in Sweden on possible chromosome damage and congenital malformations.[29] This report included a retrospective study of the incidence of congenital malformations in the progeny of 542 male employees of the Swedish State Power Board and the incidence of chromosomal abnormalities in cultured lymphocytes from 20 workers employed in 400 kV substations.

A number of observations are relevant with regard to these studies. The increased frequency of malformations occurs evenly throughout the population studied irrespective of whether they worked in 400 kV, 130–200 kV or 70 kV substations. The control group is composed of the same individuals before they were exposed. The whole study is therefore invalidated because the exposed and control group are not matched for age. Parental age is a most important factor in any study of congenital malformations.

With reference to the chromosome study, the criteria for the selection of the 20 exposed and 20 control persons are not defined. Furthermore, it is assumed that persons who describe themselves as 'employees in occupation with 400 kV substations and transmission lines' are exposed to electric fields. There is evidence from recent UK studies that an assessment of exposure to electric fields may seriously overestimate exposure as measured by suitable monitoring instruments.[9] Occupational hygiene practice requires personal monitoring of environmental factors in the working environment; fixed monitors with calculated exposures based on time give notoriously inaccurate results.

At a fundamental level, there is the question of what physical mechanisms mediate the interaction of power frequency electric fields with the human body and what determines the thresholds of sensation, pain and muscle seizure. Various groups throughout the world are tackling this problem, but because of the complexities of human physiology it seems unlikely, in the short term, that evidence will emerge from this direction to cast light on the question of whether exposure to electric fields is harmful to health. It seems sensible, therefore, to take the more direct approach of measuring the response of people to electric fields, while keeping a watch on theoretical and mechanistic investigations as they advance.

CONCLUSIONS

A critical study of the physical and physiological phenomena associated with exposure of living organisms to electric fields, together with a review of the literature on the subject, was carried out by a group of experts from the industrially developed countries under the auspices of the World Health Organization[65] in 1978. It was concluded that electric fields were harmless up to transmission voltages of 400 kV. They were also of the opinion that this view was valid for electric fields associated with transmission voltages of up to 800 kV.

This view remains, with the additional statement that the Soviet authorities also believe that their earlier assessments of this problem were pessimistic.[17] Because of the overwhelming importance of the role of electricity in modern society, both industrially and domestically, further research work has been undertaken or placed in many countries, with strictly controlled conditions. There is, however, no new evidence to date to modify the view expressed by the World Health Organization group of experts.[66]

Acknowledgements

The technical assistance of Drs J. C. Male, W. T. Norris and Mr B. J. Maddock of the Central Electricity Research Laboratories is gratefully acknowledged.

REFERENCES

1. Cabanes J. Essay on electropathology history. *Revue generale de l'électricité* 1985; **11**:809.
2. Barnes HC, McElroy AJ and Charkov JH. National Analysis of electric fields in live line working. *IEEE Trans.* PAS 86 No. 4 1967.
3. Stringfellow GC. Electric fields near CEGB transmission plant. CERL note No. RD/L/N/174/80, 1980.
4. Dalziel CF. Electric shock hazard. *IEEE Spectrum*, Feb. 1972; 41–49.
5. GE/EPRI. Electrostatic and electro magnetic effects under ultra high voltage transmission lines. EPRI report No. EL-802, 1978.
6. Good AJ and Stringfellow GC. The calculation of electric fields in the vicinity of overhead lines. CERL note No. RD/L/N33/74, 1974.
7. Elson GD, PG Tran 3—a programme to evaluate the field gradient and probe voltage in the vicinity of overhead lines. CEGB computing programme, Report No. CC/P530, 1974.
8. Pickles JH. Monte Carlo calculations of potential gradients near boundary surfaces. CERL note No. RD/L/N/202/76, 1976.
9. Maddock BJ, Male JC and Norris WT. 50 Hz electric and magnetic fields near power transmission circuits and some associated exposure and health studies. International Conference on Electric and Magnetic Fields in Medicine and Biology. *IEE London* 1985; 112–116.
10. Norris WT, Male JC and Maddock BJ. Magnetic and electric power frequency fields in the home. 23rd Annual Hanford Life Sciences Symposium, Technical Information Centre, US Dept of Energy, 1984. In press.
11. Asanova TP and Rakov AI. The state of health of persons working in the electric field of outdoor 400 and 500 kV switchyards, *Gig. Tr. Prof. Zabol.* 1966; **10**:50–52.
12. Michaelson SM. Analysis of studies related to biologic effects and health implications of exposure to power frequencies. *The Environmental Professional* 1979; **1**:217–232.
13. Sazanova TE. Nauchnye Raboty Institutov. *Okhrana Truda TSSPS* 1967; **46**:34–39.
14. Korobkova VP et al. In: Large high voltage electric systems (Paper 23–06), Conference International des Grands Reseaux a Haute Tension (CIGRE) Paris 1972.
15. Fillippov V. Second International Colloquium on Prevention of occupational risks due to Electricity. International Social Security Association, Cologne 1972; 70–80.
16. Danilin VA, Voronin AK and Modorskii VS. *Gig. Tr. Prof. Zabol.* 1969; **13**:51–52.
17. Bourgsdorf V. In: UHV transmission feasibility. Conference Internationale des Grands Reseaux Electriques a Haute Tension (CIGRE) Paris 1980.
18. Kouwenhoven WB, Langworthy OR, Singewald ML et al. Transaction on Power Apparatus and Systems. *Institute of Electrical and Electronics Engineers, London* 1967; **86**:507–511.
19. Singewald ML, Langworthy OR and Kouwenhoven WB. Transactions on Power Apparatus and Systems. *Institute of Electrical and Electronics Engineers, London* 1973; **92**:1307–1309.
20. Fole FF. Proceedings of the Second International Colloquium on Prevention of Occupational Risks due to Electricity. International Social Security Association. Cologne: 100–109. Reprinted in *Medicina y seguridad del trabajo* 1973; **21**:15–18.
21. Fole FF and Dutrus E. *Medicina y seguridad del trabajo* 1974; **22**:25–44.
22. Strumza MV. *Archives des Maladies Professionnelles de médecine du travail et de securite sociale* 1970; **31**:269–276.
23. Malboysson E. *Revue generale de l'électricité* 1976; **7**:75–80.

24. Roberge PF. Final Report: Study on the state of health of electrical maintenance workers on Hydro-Quebec's 735 kV power transmission systems. Hydro-Quebec Health Department, Montreal 1976.
25. Stopps GJ and Janischewsky W. An epidemiological study of workers maintaining HV equipment and transmission lines in Ontario. Canadian Electrical Association, Vancouver 1979.
26. Knave B, Gamberale F, Bergstrom S et al. *Electra* 1979; **65**:41–54.
27. Broadbent DE, Broadbent MHP, Male JC et al. Health of workers exposed to electric fields. *Br. J. Ind. Med.* 1985; **42**:75–84.
28. Norstrom S and Birke E. Investigation of possible genetic risks among employees of Vattenfall/Sydkraft who had been exposed to 400 kV: summary of preliminary results. (Press release.) Department of Medical Genetics. University of Umea, Umea, Sweden 1979.
29. Hauf R. Elektro Technische Zeitschrift-B **26**:318–320. (Central Electricity Generating Board, London, Translation CE 6689) 1974.
30. Hauf R. *Revue generale de l'électricité* 1976; **7**:31–49.
31. Johansson R, Lundguist AG, Lundquist S et al. Is there a connection between electricity in the atmosphere and the function of man? Part 3: Alternating Fields. Swedish National Defense Research Institute, Stockholm (FOA C2627H5) 1973.
32. Graham C, Cohen HD, Cook MR et al. A double blind evaluation of 60 Hz field effects on human performance, physiology and subjective state. 23rd Annual Hanford Life Sciences Symposium, Technical Information Centre, US Department of Energy 1985. In press.
33. Cabanes J and Gary C. In: Transmission lines and the environment (Paper 233–08). Conference Internationale des Grands Reseaux Electriques a Haute Tension (CIGRE), Stockholm 1981.
34. Bonnell JA, Broadbent DE, Lee WR et al. Can induced 50 Hz body currents affect mental functions? International Conference on Electric and Magnetic Fields in Medicine and Biology. *IEE London* 1985; 117.
35. Mckay CS, Cox T, Burrows G et al. An inventory for the measurement of self-reported stress and arousal. *Br. J. Soc. Clin. Psychol.* 1978; **17**:283.
36. Marino AA, Cullen JM, Reichmanis M et al. Power frequency electric fields and biological stress: a cause and effect relationship. 18th Annual Hanford Life Sciences Symposium, Technical Information Centre, US Department of Energy, 1979.
37. Wertheimer N, Leeper E. *Am. J. Epidemiol.* 1979; **109**:273–84.
38. Miller WM. *Am. J. Epidemiol.* 1980; **112**:165.
39. Fulton JP, Cobb S, Preble L et al. *Am. J. Epidemiol.* 1980; **111**:292.
40. Myers A, Cartwright RA, Bonnell JA et al. Overhead power lines and childhood cancer. International Conference on Electric and Magnetic Fields in Medicine and Biology 1985, *IEE London* 1985; 126–130.
41. Wright WE, Peters JM and Mack TM. Leukaemia in workers exposed to electrical and magnetic fields. *Lancet* 1982; **ii**:1160–1.
42. Milham S Jr. Mortality from leukaemia in workers exposed to electrical and magnetic fields. *N. Engl. J. Med.* 1982; **307**:249.
43. McDowall ME. Leukaemia mortality in electrical workers in England and Wales. *Lancet* 1983; **i**:246.
44. Coleman M, Bell J and Skeet R. Leukaemia incidence in electrical workers. *Lancet* 1983; **i**:982.
45. Milham S. Leukaemia mortality in amateur radio operators. *Lancet* 1985; **i**:812.
46. Pearce NE, Sheppard RA, Howard JK et al. Leukaemia in electrical workers in New Zealand. *Lancet* 1985; **i**:811.
47. Bonnell JA. Leukaemia in electrical workers. *Lancet* 1983; **i**:1168.
48. Environmental Health Criteria, 35, Extremely low frequencye (ELF) fields. World Health Organization, Geneva, 1984.

49. Coleman M, Bell CMJ, Taylor HL et al. Leukaemia and electro magnetic fields: A case control study. International Conference on Electric and Magnetic Fields in Medicine and Biology. *IEE London* 1985; 122.
50. McDowell ME. Mortality of persons resident in the vicinity of electricity transmission facilities. *Br. J. Cancer* 1986; **53**:271–279.
51. Reichmanis M, Perry FS, Marino AA et al. *Physiol. Chem. Physics* 1979; **11**:35–413.
52. Marino AA and Becker RO. *Physiol. Chem. Physics* 1977; **9**:131–147.
53. Blanchi D, Cedrini L, Ceria F et al. *Arch. Fisiol.* 1973; **70**:30–32.
54. Marino AA, Cullen JM and Becker RO. Proceedings of the 18th Annual Hanford Life Sciences Symposium, Richland, Washington. Ed. RD Phillips et al. Technical Information Centre, US Department of Energy, 1978; 258–276.
55. Greenberg B, Kunich JC and Bindokas VP. Final Report: Research Project 9341 (EPRI EA-841). Electric Power Research Institute, Palo Alto, California, 1979.
56. Phillips RD. Symposium on Energy and Human Health, Pittsburgh, Electric Power Research Institute, Palo Alto, California and University of Pittsburgh, 1979; 19–21.
57. Le Bars H and Andre G. *Revue generale de l'électricité* 1976; **7**:91–97.
58. Cerretelli P, Veicsteinas A, Margonato V, et al. Proceedings of the 18th Annual Hanford Life Sciences Symposium, Richland, Washington. Ed. RD Phillips et al. Technical Information Centre, US Department of Energy, 1979; 241–257.
59. Adey WR. Tissue interactions with non-ionising electro magnetic fields. *Physiol. Rev.* 1981; **61**:435–514.
60. Butrous GS, Male JC and Webber RS. 'The effect of power-frequency high intensity electric fields on implanted cardiac pacemakers', 1983. *PACE* **6**:1282–92.
61. Dennis JA. Towards the development of exposure limits for low-frequency electromagnetic fields. International Conference on Electric and Magnetic Fields in Medicine and Biology, London, 4–5 December 1985.
62. Male JC, Norris WT, Maddock BJ et al. Alternating electric and magnetic fields near electric power equipment: Are they a health hazard? *J. Soc. Radiol. Prot.* 1985; **5**:4.
63. Krivova TI, Lukovkin VV and Morozov YuA. Nauchnye Raboty Institutov. *Okhrana Truda VTSSPS* 1977; **108**:33–39. (Central Electricity Generating Board, London, Translation CE 7496.)
64. Takagi T. Document R7604. Japan Institute of Electrical and Radio Engineers Council, Tokyo, 1976.
65. Bonnell JA, Cabanes J, Hauf R et al. Electric and Magnetic Fields and Man. *J. Soc. Occup. Med.* 1980; **30**:135–7.
66. Bonnell JA. Effects of electric fields near power transmission plant. *J. Roy. Soc. Med.* 1982; **75**:933–41.

5. GASES

G. Matthews

DEFINITIONS

A *gas* can be defined as a substance in a physical state such that its constituent molecules move freely, allowing it to expand indefinitely.

A *vapour* is a gas below its critical temperature: it can therefore be liquefied by an appropriate increase in pressure.

A *fume* is a cloud of particles which arises from a condensation of vapour or from a chemical reaction.

RANGE OF INJURY CAUSED BY GASES

The number of industrial employees who die as a result of industrial gassing incidents in their plants is very small; it is minute compared with the number of deaths in such disasters as Bhopal where in 1985 about 2500 people died of the effects of methyl isocyanate which escaped into the environment, or as a result of a large leak of liquefied petroleum gas which caught fire in Mexico City in 1984 and burned to death more than 600.

In the United Kingdom the Office of Population Censuses and Surveys[1] in 1981 recorded 194 poisonings from gases and vapours and 189 in 1982—these included the few industrial cases, and were nearly all domestic—carbon monoxide being the main cause. By contrast, gassing injuries notified to the Factory Inspectorate[2] were as follows:

	1980	*1981*	*1982*
Total	157	93	69
Fatal	4	4	0

Chlorine, sulphur oxides and again, carbon monoxide, caused the largest number of injuries specifically due to gas; fumes however, constituted a greater hazard, being responsible for 31 of the 69 incidents in 1982. These figures demonstrate clearly that most gassing incidents arise from exposures to gases which are either commonly used or commonly met with, or from exposure to fume which usually occurs from its explosive release or from a failure of its entrapment.

The old adage still applies—common things commonly occur—and attention to day-to-day procedures pays the biggest dividends.

Fume and smoke arising from the combustion products of solid plastics and plastic foams are intensely toxic due mainly to nitrogen oxides and acrolein and present a particular occupational risk to fire and rescue services; the risk in general industry is low.

DISTRIBUTION OF GASES

The historical use of gases in industry is largely in metal cutting and welding, using acetylene and oxygen, and steel making, using oxygen. Hospitals use large quantities of oxygen as well as anaesthetic gases and vapours. Small volumes of more esoteric gases have been used for research purposes for years and the quantities and range is increasing. Microchip manufacture in the semi-conductor industry has required a great expansion in the use of erstwhile unusual gases like silane and germane, and of highly toxic but better known gases like phosphine and arsine.

The volume of established industrial gases used has increased markedly in the last four years due to wider applications—for example, liquid nitrogen is used much more extensively as an inerting agent and for food freezing and refrigerant processes, as well as for a multitude of purposes where extreme cold is needed. It is also used to freeze sections of earth prior to tunnelling or boring in civil engineering and to destroy a wart by cryoprobe. Oxygen is used increasingly to augment heat yields from various fuels in combustion processes.

Gaseous by-products are common in chemical processes: hydrogen sulphide in oil refineries, carbon dioxide in breweries. By-products occur more commonly in non-industrial situations: methane (fire damp) occurs in mines and marshy caves, mixed with carbon dioxide; in slurry tanks and silage towers; carbon dioxide in chalky pits. The most lethal exposures are to carbon monoxide which arises from incomplete combustion of fuel gas, usually from domestic gas boilers or water heaters.

Because of the flexibility of gases, their compressibility and the comparative ease of evolving continuous process systems, their use is certain to increase in both scope and volume.

ANATOMY OF THE RESPIRATORY TRACT

The upper respiratory tract comprises the nasal cavity and the nasal, oral and laryngeal parts of the pharynx. The epiglottis rises to protect the vestibule of the larynx, leaving the lower pharynx to continue into the oesophagus. The smaller laryngeal cartilages are shielded by the thyroid cartilage behind which is the narrowest part of the respiratory tract—the slit bounded by the vocal chords. It is here where oedema can be fatal. Beneath the chords the infra-glottic cavity expands into the trachea which descends, divides into the two main bronchi and thence ultimately to the smallest bronchioles. These small airways, many of whose diameters are less than the breadth of human hair, end in the alveoli. A single layer of epithelial cells separates the gas in the alveoli from the blood and gas in the pulmonary capillaries. The junctions between these cells can be split apart by inflammation and by irritant gases and fume, so allowing leakage of fluid and blood fractions from the vascular compartment into the alveolus.

PHYSIOLOGY OF THE RESPIRATORY TRACT

The function of the respiratory tract is to enable oxygen to be taken into the body and carbon dioxide to be blown out.

It is also involved with the regulation of hydrogen ion concentration in the blood. The processes concerned are:

Ventilation —the movement of air in and out of the lungs.
Distribution—the distribution of the inspired air throughout the lungs.
Diffusion —the alveolar and pulmonary capillary gaseous exchanges.

The distribution and dynamics of pulmonary blood flow as well as the transport of blood gases, can also be affected by lung changes.

The total area available for gas transfer is about seventy-five square metres (75 m^2).

The time lapses from the inhalation of a soluble molecule of gas are:

Nose to alveolus	7 seconds
Alveolus to capillary	1 second
Capillary to brain	7 seconds

The pressure of oxygen in the alveolar space is normally 100 mm mercury. Unconsciousness occurs rapidly when the pressure falls to 38 mm mercury.

Respiratory changes which occur during exercise are simply a response to the increased oxygen demands of the active musculature (*Table 5.1*).

It is plain that a toxic gaseous environment which might be physiologically tolerable for a given time during slight exertion would be intolerable during heavy work.

Table 5.1. Respiratory changes on exercise (Taken from Campbell, Dickinson and Slater. *Clinical Physiology*. Oxford: Blackwell, 1963)

	At rest (l/min)	*Violent exercise* (l/min)
Ventilation	8	120
O_2 intake	0·25	3–4
CO_2 output	0·25	3–4
Pulmonary blood flow	5	30

Response of the lung to gas exposures

The important characteristics of the lung which determine the physical response to the effects of toxic gases are:

—the huge absorptive area which allows soluble gas to enter the vascular compartment of the body extremely quickly.

—the production of inflammatory exudate by the epithelium of the upper respiratory tract and by the bronchial epithelium in response to increasing irritation.

—the disintegration of the cell junctions of the alveolar epithelium in response to irritation, leading to transudation from the pulmonary capillaries.

GENERAL EFFECTS OF GASES

Gases have great potential to interfere with physiological functions, even if they are not intrinsically toxic, because of the ease with which they enter the body. The effect depends mainly upon the solubility and type of reactivity of the molecule. A soluble gas is distributed swiftly. If it is inert, nothing will happen unless the alveolar concentration is so high that the oxygen level falls.

A highly reactive gas like chlorine or fluorine will in low concentrations irritate, in high concentrations corrode the tissues with which it comes in contact. In very high concentrations the upper respiratory tract may be attacked very severely—this is even more likely with a soluble corrosive gas like ammonia which produces rapid laryngeal oedema.

Principles of prevention

It is a truism, and the figures for deaths given earlier lend more credence to it than any words, that the more overtly dangerous the environment, the more detailed care is taken. Most gassing or asphyxiation incidents occur in confined spaces or in situations where a heavy gas displaces air—they are always avoidable.

It is beyond the remit of this book to go into great detail about matters which are properly the province of safety and engineering departments but a brief summary of the necessary safety considerations is useful:

—appraisal of the work by competent supervision.
—knowledge of gases which are to be used or which may be generated or encountered and their significant physical and chemical characteristics, i.e. if they are heavy or light, give adequate warning of their presence, inert or toxic.
—especial care if flames or sparks will arise in confined spaces or onto pipework; their effect in creating possible explosive or chemical reactions.
—those who may have to wear respirators should have had them fitted and be trained on their use.
—necessity for automatic warning systems, or oxygen or other analysers.
—permit to work system in operation, which considers the above factors plus any civil, mechanical or electrical engineering hazards.

Principles of treatment

1. With few exceptions, there is no specific antidote against the effect of toxic and corrosive gases.

2. Victims of gassing incidents require general supportive treatment.

3. Hypoxia needs rapid assessment; it is a very common accompaniment of gassing. The hypoxia may be hypoxic hypoxia when the oxygen in air has been displaced by an inert gas; or the anaemic hypoxia of carbon monoxide poisoning; or, rarely, the histotoxic anoxia of cytochrome oxidase blocking by a gaseous cyanide.

4. Pulmonary oedema due to severe acute smoke exposure or delayed subacute insults from any of several common gases will produce severe hypoxia as well as other systemic and pulmonary disturbances.

5. The effective treatment of acute gas exposures consists of applying specific remedies where possible and ensuring that adequate oxygenation of the blood is secured. Both these aims can often be achieved at the place of work by competent first aid. A volume controlled oxygen resuscitator should be available where toxic gases are in use.

Where lung damage has occurred from the effects of smoke or corrosive gas, more energetic and sophisticated treatment in hospital is needed. Very seriously ill patients with pulmonary oedema require specialised techniques for assessment of alveolar permeability, for reduction of fluid in the alveoli and for forced ventilation. The aim of the treatment is to restore a dry alveolus; this is achieved by lowering blood pressure and reducing inflammation and damage. In addition, adequate hydration and oxygenation are maintained.

VERY COLD LIQUEFIED GASES

The danger of these gases is due mainly to their coldness and it is convenient to describe the treatment here. The common gases with their boiling points are:

Helium	−268 °C	Argon	−185 °C	Ethylene	−103 °C
Nitrogen	−195 °C	Oxygen	−183 °C	Propane	−42 °C

A fall in core body temperature to approximately 25 °C will result in death while the skin temperature may fall to 0 °C without harm. Should the peripheral temperature fall further, tissue damage will occur from the formation of ice crystals and consequent disruption of cell membranes. Injuries to the hands and feet are not uncommon.

The injury becomes apparent on thawing when extravasation of fluid from damaged cells and vessels gives rise to oedema; thrombosis of small arteries leads to ischaemia, and gangrene may follow. This is the classic description of frostbite, of course, and the degree of damage is related to the extent and to the temperature of the liquefied gas.

Immediate first aid is to warm the part slowly by immersing in lukewarm water at a temperature no higher than 45 °C. Dry heat must never be used—the part is anaesthetic and undue heat will simply impose a burn on the cold tissue. If warm water is not available, water from a cold tap is useful. The pallid, waxy looking injured part will change colour slowly through blue to red; it will become painful and swell.

Should the damaged area be at all large or functionally significant, it should be covered with dry sterile dressings and the patient removed to hospital for specialised care.

Hypothermia is common among the elderly in northern winters, but is very rare in industry.

Asthmatics

It is well known that asthmatics may respond to cold weather with an attack of bronchospasm, and the response is more pronounced on exertion. The bronchi of normal subjects respond similarly but to a lesser extent. Anyone who has lived in temperatures below −20 °C will know the feeling. In asthmatics the cause is neural and local; in normal subjects it is purely neural.[3] Asthmatic subjects should therefore be advised not to work in very cold atmospheres—chill rooms, freezer depots and the like. Exposures of those dispensing liquefied gases should be properly assessed.

Spillages of liquefied gas

The following are volumes arising from the expansion of a unit volume of liquefied gas at 15 °C and 1 bar:[4]

Argon	835	Helium	748	Oxygen	854
Ethylene	482	Nitrogen	691	Propane	311

The evaporation of liquid inert gas in confined spaces may rapidly diminish the oxygen content. All evaporating gases are cold, dense and heavy and will find low-lying places and sink into pits. The risk of anoxia from spillages of inert gases and the risk of fire with oxygen enrichment are evident. The best illustration of what can happen in an atmosphere of oxygen enrichment, although in this instance not caused by evaporation of liquid oxygen, is contained in the Factory Inspectorate's report *The Fire on HMS* Glasgow *23 September 1976*[5] which describes the catastrophe in which 8 men died. Ionized oxygen's high redox potential—its greed for electrons—makes it terrifyingly dangerous in high concentrations: in an atmosphere of pure oxygen it is said that a man's body can be consumed by fire in a few seconds.

RECENT DEVELOPMENTS

Two gases and a new industry

The two gases are ethylene oxide and nitrous oxide. The story of the evaluation of hazards associated with their use is useful in that it illustrates how long it can take to appreciate a potential risk although the lights are shining for he or she who would see. The new industry is the semiconductor or microchip industry.

Ethylene oxide and nitrous oxide

These gases are not commonly encountered in industry: ethylene oxide is used as an intermediate in the chemical industry and as a sterilizing agent; nitrous oxide as an analgesic and as an anaesthetic agent for a hundred years.

There are two lessons:

1. A specific use may in itself suggest a hazard. Ethylene oxide sterilizes: it is therefore not unreasonable to suppose that it could have an effect on other biological systems.

2. Communication between various academic disciplines is often poor.

A seminal paper published in a chemical journal in 1968 on a reaction of nitrous oxide with transition metal complexes did not attract the attention of medical workers until the mid-seventies. At this time a good deal of interest was being taken in the possible hazards of anaesthetics: once the biological significance of the nitrous oxide reaction with cobalt—an integral part of the vitamin B_{12} molecule—was realised, fundamental research work took a precisely orientated direction.

Ethylene oxide

Ethylene oxide is highly reactive, flammable and subject to explosive decomposition. It is used as an intermediate in the chemical industry within sealed process plants and also as a sterilizing agent for materials which could be damaged by other sterilizing techniques. For sterilizing, a chamber is filled with items requiring sterilization, the compartment sealed and ethylene oxide piped in. After an appropriate period, the compartment is force-ventilated. Ethylene oxide has a propensity to cling to materials, especially plastics, and so the sterilized items are then usually taken into a non-occupied, closed, well-ventilated waiting area where the remaining ethylene oxide slowly diffuses out.

Local effects

In high concentrations the gas is irritating to the mucous membrane and lungs. Prolonged heavy exposure affects the central and peripheral nervous systems. The boiling point is 10·5 °C and so the gas is easily liquefied—liquid on the skin is deceptively painless, but unless it is washed off very quickly, will create a deep burn. Ethylene oxide in gumboots is asking for skin grafting.

Systemic effects

In 1979, Hogstedt et al.[6] described two cases of leukaemia and one of macroglobinaemia among workers engaged in ethylene oxide sterilization. It is perhaps not unreasonable in any case to postulate that a chemical used to kill micro-organisms might have some effect on other biological systems.

Much investigative work was performed; the most important findings follow. Ethylene oxide produced a positive Ames test. Snellings et al.[7] conducted a long-term inhalation study in rats and found a higher incidence of tumours than expected. Chromosome aberrations were found in exposed workers.[8] An increased abortion rate was observed in women occupationally exposed.[9]

It is now accepted that ethylene oxide is a potential carcinogen and exposure to it should be as low as possible. The TLV in the United States and exposure limit in the United Kingdom until recently were 50 ppm (90 mg m^{-3}). The US Federal Register 49 25734 published a permitted limit of 1 ppm (1·80 mg m^{-3}) in 1985: the limit in the United Kingdom is 5 ppm.

Nitrous oxide

Nitrous oxide (N_2O) is a colourless non-irritating, sweet smelling gas. Its main use is in general anaesthesia. In combination with oxygen, as a 50 per cent N_2O/O_2 mixture, it is also an excellent inhalation analgesic in cases of severe immediate pain—fractures, crushed limbs, severe burns, heart attacks—and acts very quickly. It has a mild addictive potential.

In the late sixties, Vaisman in Russia[10] reported a high incidence of subjective symptoms, spontaneous abortions and premature births among female anaesthetists. Following Vaisman, many further studies among operative theatre personnel, conducted largely by questionnaire, provoked much anxiety as worrying effects—cancer, congenital abnormalities—were alleged to be associated with exposure to anaesthetics in general and to nitrous oxide in particular. Several experimental studies were undertaken to determine the effects of individual anaesthetics and combinations of anaesthetics upon reproduction processes in small animals. It was suggested that trace concentrations affected human performance.[11,12] This was not confirmed elsewhere.[13] NIOSH produced a criteria document in 1977[14] suggesting an exposure limit of 25 ppm. This level has not been adopted.

Some objective observations had however been overlooked. In 1956 Lassen[15] had described bone marrow depression after usage of nitrous oxide in severe tetanus. In 1968 Banks et al.[16] described the reaction of monovalent cobalt with nitrous oxide. Transposing this reaction to the univalent cobalt in vitamin B_{12}:

$$2Co^{I} + N_2O = 2Co^{II} + N_2$$

This action inactivates vitamin B_{12}. The vitamin is necessary for the activity of methionine synthase. Methionine synthase is a necessary part of folate metabolism.

Deacon et al.[17] showed that a few hours exposure to nitrous oxide inactived methionine synthase, with the ultimate result that DNA synthesis was impaired and normal haemopoiesis gave place to a megaloblastic bone marrow.

In 1983 Sharer et al.,[18] in an important series of controlled observations, established a range of concentrations of toxic significance in rats. It was shown that an exposure of 450 ppm for 24 hours or longer was without effect. The concentration at which inhibition of methionine synthase significantly occurred was a little under 1000 ppm (*Figs. 5.1, 5.2*). An effective dose 50% (ED_{50}) of 5400 ppm was established.

Sweeney et al.[19] in a study of dentists showed a conclusive relationship between exposure levels and adverse effects and suggested that a practicable exposure level would be 400 ppm.

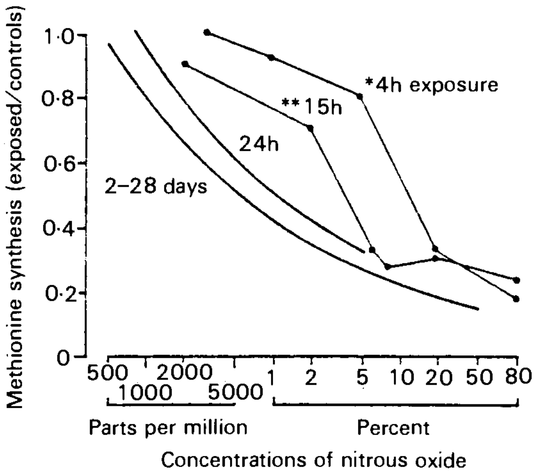


Fig. 5.1. Dose–response curves for inhibition of methionine synthase activity for different durations of exposure. (Reproduced with permission of J. F. Nunn and the editor of *Br. J. Anaesthesia.*)

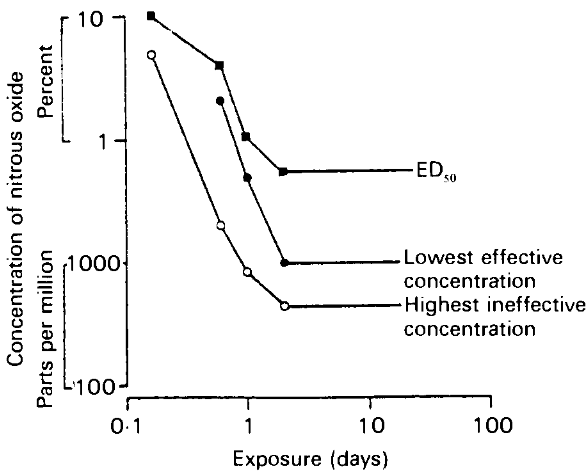


Fig. 5.2. ED_{50}, lowest effective and highest ineffective concentrations for inhibition of methionine synthase activity shown as a function of time. (Reproduced with permission of J. F. Nunn and the editor of *Br. J. Anaesthesia.*)

The doubts about nitrous oxide seem to be clearing and its effects understood. There are aspects of the biochemistry which are worthy of further comment and which are still in process of elucidation but are really beyond the scope of this book. These are the essential points which matter in preventing harm:

1. The effect of nitrous oxide is to inactivate vitamin B_{12}.
2. Exposures at anaesthetic or analgesic concentrations more frequently than every four days pose a risk of megaloblastic marrow change. In very ill people the risk is greater.
3. Repeated high exposures, as in nitrous oxide sniffers or addicts can present the picture of subacute combined degeneration of the cord and megaloblastic anaemia.

A new industry

There have been two developments in the last fifteen years which are significant engines of industrial and social change—microchip manufacture and genetic engineering. The first is transforming science, communications and manufacturing; the second, a little more slowly perhaps, is opening the doors to hardier and better yielding crops, more precisely patterned pharmaceuticals and deliverance from congenital disease. Genetic engineering need not concern us further here.

Microchip manufacturing takes a wafer of silicon and turns it into a complex electrical system. A typical chip is about 6 mm square and 0·3 mm thick. It contains about 7000 components—transistors, resistors, capacitators and diodes. The process of manufacture is theoretically simple but technically demanding. Silicon is a poor conductor of electricity and its crystalline structure serves as a bed for the insertion of negative or positively charged atoms.

A disc of silicon is covered with a layer of light-sensitive materials. Ultraviolet light is passed through a glass screen which has within it the pattern required. The pattern, now marked on the surface of the disc, is etched by chlorine or fluorine compounds. The remaining light-sensitive material is removed, leaving the pattern of excised segments. These are then doped with either boron, phosphorus or arsenic compounds at high temperatures. The process is repeated up to ten times so that layers of complex circuitry are created.

Manufacturing is a continuous process. Potential exposures of the workforce will be to the customary range of industrial solvents, to the toxic gases arsine, phosphine and diborane which are used as doping gases, and to inorganic acids, especially hydrogen fluoride and hydrogen chloride.

The hazards of the materials used in the industry are reasonably well known in themselves but as development occurs, there is a clear need for wide dissemination of the knowledge gained. It has been suggested by La Dou[20] that some increase in occupational illness has occurred in the electronics industry. The essential risks associated with the gases used are those of exposure to the gases themselves and to their reaction products within the process—very largely acidic residues.

ACUTE TOXICITY AND EXPOSURE LIMITS

Details of considerations governing appraisal of chemicals, in order to arrive at limits to which workers may be exposed for an 8-hour day, 5 days a week, without adverse effect, are well described in the Threshold Limit Values and Biological Exposure Indices[21] published in the United States.

Exposure limits may be governed by several factors, the main ones being industrial and experimental evidence. As well as pure acute toxicity, other criteria for a very low limit are an appalling odour, irritation, chronic effects or carcinogenic potential; mercaptans, ammonia, nitrogen oxides, vinyl chloride are examples of each respectively.

Especially for common materials, many exposure limits were established years ago and if a comparison is made between the concentration of a gas accepted as being dangerous to life after one hour and its exposure limit, a wide variation is apparent. If the dangerous concentration is divided by the exposure limit a figure can be reached which could be termed an injury index. This has been done as shown:

Gas	Concentration (C) in ppm dangerous to life after 1 hour	Exposure limit (EL) in ppm	Injury index (C/EL)
Phosphine	100	0·3	330
Hydrogen cyanide	about 100	10·0	10
Hydrogen sulphide	about 500	10·0	50
Phosgene	10	0·1	100
Arsine	6	0·05	120
Nitrogen dioxide	100	3·0	33

It can be seen that according to the established limits it takes 330 times the limit to risk a life with phosphine, but only 10 times the limit in the case of hydrogen cyanide.

Death of course is not the same as toxicity although it may result from it: both nitrogen dioxide and phosgene cause delayed pulmonary symptoms, arsine is a powerful haemolytic agent, cyanide is a biological poison whose acute effect is accurately known by dose, and both phosphine and hydrogen sulphide are acute systemic poisons. Were exposure limits related to acute injury the injury index would be the same: it is other factors which determine the case. In the case of phosphine it may be the possibility of chronic effects akin to those of phosphorus although none have been recorded.

What is clear, and although stated often, not entirely believed, is that an exposure limit does not define toxicity.

CLASSIFICATION OF TOXIC GASES

The simplest classification is one based upon physiological action. The intrinsic chemical and physical characteristics of the gas are the determinants of toxicity. At normal temperatures and pressures gases can cause:

Acute	local injury
Chronic	local injury
Acute	systemic injury
Chronic	systemic injury

In addition, in unusual circumstances of high pressure, all gases will have their effects enhanced and many which are regarded generally as non-toxic, like oxygen or nitrogen, will exert toxic effects. Liquefied gases carry their own risk of cold contact burns and frostbite.

There follows a listing of the most important gases within the general classification given above. The relative density (RD) is to air (=1) and is approximate at 20 °C and atmospheric pressure; the RD figure is not intended to be precisely accurate but is meant to indicate the degree of heaviness or lightness. Many gases hydrolyse readily; where the hydrolysis product is toxicologically important, this is indicated. If a control limit is not yet established, this is indicated by NE (*Tables 5.2 and 5.3*).

Table 5.2. Acutely severely irritant or corrosive gases

Gas	Formula	Relative density to air	Exposure limit	Hydrolysis product
Ammonia	NH_3	0·59	25 ppm	Nil
Boron trichloride	BCl_3	4·12	NE	(HCl)
trifluoride	BF_3	2·4	1 ppm	(HF)
Bromine pentafluoride	BrF_5	(liquid)	0·1 ppm	(HF)
trifluoride	BrF_3	(liquid)	0·1 ppm	(HF)
Carbonyl fluoride	COF_2	2·29	2 ppm	(HF)
Chlorine	Cl_2	2·43	1 ppm	
Chlorine trifluoride	ClF_3	3·21	0·1 ppm	(HF)
Cyanogen chloride (also acute systemic)	ClCN	2·16	0·3 ppm	
Dichlorosilane	SiH_2Cl_2	3·52	NE (10 ppm)	(HCl)
Dimethylamine	$(CH_3)_2NH$	1·56	10 ppm	
Fluorine	F_2	1·5	1 ppm	
Hydrogen bromide	HBr	2·8	3 ppm	
Hydrogen chloride	HCl	1·26	5 ppm	
Hydrogen fluoride	HF	1·85	3 ppm	

Table 5.2. (*cont.*)

Gas	Formula	Relative density to air	Exposure limit	Hydrolysis product
Hydrogen iodide	HI	4·46	NE	
Hydrogen selenide (also acute systemic)	H_2Se	2·8	0·05 ppm	
Hydrogen sulphide (also acute systemic)	H_2S	1·18	10 ppm	
Iodine pentafluoride	IF_5	(Liquid. VP=29·1 mb)		
Monoethylamine	$C_2H_5NH_2$	1·6	10 ppm	
Monomethylamine	CH_3NH_2	1·1	10 ppm	
Nitric oxide	NO	1·04	25 ppm	
Nitrogen dioxide	NO_2	2·6	3 ppm	
Nitrosyl chloride	NOCl	2·31	NE	
Oxygen difluoride	OF_2	1·9	0·05 ppm	
Ozone	O_3	1·7	0·1 ppm	
Perchloryl fluoride	ClO_3F	3·6	3 ppm	
Phosgene	$COCl_2$	3·5	0·1 ppm	
Phosphorus pentafluoride	PF_5	4·5	NE	(HF)
Silicon tetrafluoride	SiF_4	3·6	NE	
Sulphur dioxide	SO_2	2·3	2 ppm	
Sulphur tetrafluoride	SF_4	3·8	0·1 ppm	(HF)
Tetrafluorohydrazine	N_2F_4	3·6	NE	(HF)
Trimethylamine	$(CH_3)_3N$	2·1	10 ppm	

Table 5.3. Mildly or moderately irritant gases

Gas	Formula	Relative density to air	Exposure limit
Ethylene oxide	C_2H_4O	1·5	5 ppm*
Bromotrifluoroethylene	$F_2C{=}CBrF$	5·6	NE
Methyl bromide (also acute systemic)	CH_3Br	3·35	5 ppm
Methyl chloride (also acute systemic)	CH_3Cl	1·7	50 ppm
Methyl mercaptan (also acute systemic)	CH_3SH	1·66	0·5 ppm
Nickel carbonyl	$Ni(CO)_4$	5·95	0·05 ppm

*Because of systemic action.

Chronically irritant gases

Many of the acutely irritant gases have been associated with chronic changes in the respiratory tract; the exposures are usually mixed with dust and fume. The best known are:

Ammonia
Chlorine
Nitrogen dioxide
Organic acids
Sulphur oxides

Table 5.4. Acute systemically toxic gases

Gas	Formula	Relative density to air	Exposure limit
Acrylonitrile (also carcinogenic)	CH_2CHCN	Liquid	2 ppm
Arsine	AsH_3	2·69	0·05 ppm
Carbon monoxide	CO	0·967	50 ppm
Carbonyl sulphide	COS_2 (hydrolyses to H_2S)	2·10	10 ppm
Cyanogen	C_2N_2	0·99	10 ppm
Cyanogen chloride	ClCN	2·16	0·3 ppm
Diborane	B_2H_6	0·965	0·1 ppm
Dibromodifluoromethane	CBr_2F_2	7·2	100 ppm
Germane	GeH_4	2·6	0·2 ppm
Hexafluoroacetone	F_3CCOCF_3	5·7	0·1 ppm
Hydrogen cyanide	HCN	0·97	10 ppm
Hydrogen selenide	H_2Se	2·8	0·05 ppm
Hydrogen sulphide	H_2S	1·18	10 ppm
Methyl bromide	CH_3Br	3·35	5 ppm
Methyl chloride	CH_3Cl	1·7	50 ppm
Methyl fluoride	CH_3F	1·2	NE
Methyl mercaptan	CH_2SH	1·66	0·5 ppm
Nickle carbonyl	$Ni(Co)_4$	5·95	0·05 ppm
Nitrogen trifluoride	NF_3	2·46	10 ppm
Perfluoro-2-butene	$CF_2CF{:}CFCF_3$	7·0	NE
Phosphine	PH_3	1·2	0·3 ppm
Phosphorus trifluoride	PF_3	3·0	NE
Silane	SiH_4	1·1	0·5 ppm
Sulphuryl fluoride	SO_2F_2	3·7	5 ppm
Vinyl bromide	C_2H_3Br	3·8	5 ppm

Table 5.5. Chronic systemically toxic gases

Gas	Formula	Relative density to air	Exposure limit
Acrylonitrile	CH_2CHCN	Liquid	2 ppm
Ethylene oxide	C_2H_4O	1·5	5 ppm
Nitrous oxide	N_2O	1·5	NE
Vinyl chloride	C_2H_3Cl	2·2	3 ppm

Table 5.6. Inert asphyxiants

Gas	Formula	Relative density to air	Exposure limit
Argon	Ar	1·38	NE
†Carbon dioxide	CO_2	1·53	5000
*Carbon tetrafluoride	CF_4	3·05	NE
*Chlorodifluoromethane	$CHClF_2$	3·11	1000
*Chloropentafluoroethane	C_2ClF_5	5·54	NE
*Chlorotrifluoroethylene	C_2ClF_3	4·13	NE
*Chlorotrifluoromethane	$CClF_3$	3·6	NE
Deuterium	D_2 or $^2H_2$	0·14	NE
*Difluoroethylene	$H_2C{:}CF_2$	2·21	NE
Helium	He	0·14	NE
*Hexafluoroethane	C_2F_6	4·8	NE
Hydrogen	H_2	0·07	NE
Krypton	Kr	2·89	NE
Methane	CH_4	0·55	NE
Neon	Ne	0·7	NE
Nitrogen	N_2	0·97	NE
Perfluorobutane	C_4F_{10}	8·4	NE
Perfluoropropane	$CF_3CF_2 \quad CF_3$	6·7	NE
Sulphur hexafluoride	SF_6	5·1	NE
Tetrafluoroethylene	$F_2C{:}CF_2$	3·5	NE
*Trichlorofluoromethane Liquid. VP 0·92 bar at 21 °C	CCl_3F		1000
Xenon	Xe	4·5	NE

*These are all fluorohydrocarbons. Many are used as refrigerants and known under trade names.

†Carbon dioxide is not strictly speaking inert. Its physiological action is to increase the breathing rate: over 10 per cent there is increasing distress and 25 per cent is fatal.

Table 5.7. Anaesthetically active asphyxiants

Gas	Formula	Relative density to air	Exposure limit
Acetylene	C_2H_2	0·98	NE
Allene (Propadiene)	CH_2:C:CH_2	1·4	NE
*Bromotrifluoromethane	$CBrF_3$	5·31	1000
1,3 Butadiene	C_4H_6	1·878	1000
n-Butane	C_4H_{10}	2·11	600
1-Butene	C_4H_8	2·0	NE
CIS-2-Butene	C_4H_8	2·0	NE
trans-2-Butene	C_4H_8	2·0	NE
Cyclobutane	C_4H_8	2·0	NE
†Cyclopropane	C_3H_6	1·45	400
Dimethylether	$(CH_3)_2O$	1·6	NE
*Dibromotetrafluoroethane	$C_2Br_2F_4$	8·97	NE
*Dichlorofluoromethane	$CHCl_2F$	3·5	NE
*Dichlorotetrafluoroethane	$C_2Cl_2F_4$	5·9	1000
*Difluoro-1-chloroethane	$C_2H_3ClF_3$	3·5	NE
*1,1-Difluoroethane	$C_2H_4F_2$	2·3	NE
2,2-Dimethyl propane	$C(CH_3)_4$	2·5	NE
Ethane	C_2H_6	1·05	NE
Ethylacetylene	CH_3CH_2C:CH	1·9	NE
Ethyl chloride	C_2H_5Cl	2·23	1000
Ethylene	C_2H_4F	0·95	NE
Fluoroform	CHF_3	2·4	NE
Isobutane	$CH(CH_3)_3$	2·06	NE
Isobutylene	$(CH_3)_2$C:CH_2	1·9	NE
Methylacetylene	C_3H_4	1·4	NE
3-Methyl-1-butene	$(CH_3)_2$CHCH:CH_2	2·5	NE
Methyl vinyl ether	CH_3OCH:CH_2	2·0	NE
†Nitrous oxide	N_2O	1·5	NE
Propane	C_3H_8	1·5	NE
Propylene	C_3H_6	1·5	NE
*1,1,2-Trichloro-1,2,2-trifluoroethane Liquid: VP 0·38 bar at 21 °C			1000
Vinyl fluoride	CH_2:CHF	1·6	NE

*These are all fluorohydrocarbons. Many are used as refrigerants and known under trade names. †A general anaesthetic.

CONCLUSION

Although gases are not commonly the cause of poisonings, only care can maintain and improve this. The commonly encountered, unsuspected gases present the greatest risk. Some aphorisms:

1. Know the industry's processes;
2. Know the industrial gases used;
3. Remember that many gases are solvents and may be corrosive. This will affect the choice of protective clothing for those working with them;
4. Ensure that the hazards are known to the right people;
5. Work with the right people to get the handling and controls right;

6. Ensure the first-aiders are aware of the specific hazards;
7. Meet your local casualty department doctors or nurses and let them have a list of your potential poisons, especially if they are rarities;
8. Ensure that a data sheet of the gas accompanies any victim to the hospital.

REFERENCES

1. Office of Population Censuses and Surveys. *Deaths by Cause.* Ref. DH2 83/5 4 October 1983. London: HMSO.
2. Health and Safety Executive. *Statistics: Health and Safety* 1981–82. London: HMSO, 1985.
3. Heaton RS, Henderson AF, Gray BJ et al. The bronchial response to cold air challenge: evidence for different mechanisms in normal and asthmatic subjects. *Thorax* 1983; **38**:506–11.
4. L'Air Liquide Division Scientifique. *Gas Encyclopoaedia* (English and French). Amsterdam: Elsevier, 1976.
5. Health and Safety Executive. *The Fire on HMS Glasgow 23 September 1976.* London: HMSO, 1978.
6. Hogstedt C, Malmqvist N and Wadman B. Leukaemia in workers exposed to ethylene oxide. *JAMA* 1979: **241**:1132–3.
7. Snellings WM, Weil CS and Maronpot RR. (1981) Ethylene oxide: *Two year inhalation study on rats: Final Report 44–20* 1981. Bushy Run Research Center, Pittsburgh, USA.
8. Hogstedt B, Gullberg B, Hedner K et al. Chromosome aberrations and micronuclei in bone marrow cells and peripheral blood lymphocytes in humans exposed to ethylene oxide. *Heriditas* 1983; **98**:105–13.
9. Hemminki K, Mutaanen P, Saloniemi I et al. Spontaneous abortions in hospital staff engaged in sterilizing instruments with chemical agents. *Br. Med. J.* 1982; **285**:1461–3.
10. Vaisman AI. Working conditions in surgery and their effect on the health of anaesthetists. *Eksp Khir Anesteziol* 1967; **3**:44–9 (Rus).
11. Bruce DL, Bach MJ and Arbit J. Trace anesthetic effects on perceptual, cognitive and motor skills. *Anesthesiology* 1974; **40**:453–8.
12. Bruce DL and Bach MJ. Effects of trace anaesthetic gases on performance of volunteers. *Br. J. Anaesth.* 1976; **48**:871.
13. Smith G and Shirley AW. A review of the effects of trace concentrations of anaesthetics on performance. *Br. J. Anaesth.* 1978; **50**:701.
14. NIOSH. Centre for Disease Control. *Criteria for a recommended standard... occupational exposure to waste anesthetic gases and vapors.* US Government Printing Office, Washington, DC. 1977.
15. Lassen HCA, Henriksen E and Nevkirch F. Treatment of tetanus. Severe bone marrow depression after prolonged nitrous oxide anaesthesia. *Lancet* 1956; **i**:527–30.
16. Banks R, Henderson RJ and Pratt JM. Reaction of gases in solution. Part III. Some reactions of nitrous oxide with transition-metal complexes. *J. Chem. Soc.* 1968; **A**:2886.
17. Deacon R, Lumb M, Perry JA et al. Selective inactivation of vitamin B_{12} in rats by nitrous oxide. *Lancet* 1978; **ii**:1023–4.
18. Sharer NM, Nunn JF, Royston JP et al. Effects of chronic exposure to nitrous oxide on methionine synthase activity. *Br. J. Anaesth.* 1983; **55**:693.
19. Sweeney B, Bingham RM, Amos RJ et al. Toxicity of bone marrow in dentists exposed to nitrous oxide. *Br. Med. J.* 1985; **291**:567.
20. LaDou J. Potential occupational health hazards in the microelectronics industry. *Scand. J. Work Environ. Health* 1983; **9**:42–6.
21. ACGIH. *Threshold Limit Values and Biological Exposure Indices for 1985–6.* 6500 Glenway Ave., Bldg D-7, Cincinnati.

6. LEPTOSPIROSIS

Sheena A. Waitkins

Leptospirosis is a zoonotic disease. A wide variety of animals act as a maintenance host but man is the end-accidental host. The organism that causes leptospirosis is a spirochaetal bacterium, *Leptospira*, belonging to the family *Spirochaetaceae* which is morphologically distinct from all other spirochaetes. The leptospira has a characteristic spiral morphology with typical hooked ends. Their length may vary between 6 μm and 20 μm and their width is very narrow—about 0·1 μm in diameter. These fine, delicate spirochaetes can be seen only by dark-ground, phase contrast or electron microscopy. They do not stain well with normal bacteriological stains but are considered on biochemical and morphological grounds to be Gram-negative. They have a characteristic dancing movement which is often misdiagnosed in blood cultures from suspect patient cases. This pseudoleptospiral appearance is due to membranes of red blood corpuscles which look deceivingly like real leptospires by dark-ground microscopy. It is therefore, very important that proper isolation and identification of suspect leptospires are made before the disease is confirmed.

CLASSIFICATION *(Table 6.1)*

The strains of leptospira are arranged into two species, *Leptospira interrogans*, comprising the parasitic and potentially pathogenic organisms, and *Leptospira biflexa* whose members are saprophytic and non-pathogenic, free-living. The two species can be differentiated by biological and physiological tests. These tests are routinely used to separate and identify unknown strains of newly isolated leptospira.[1] *Leptospira interrogans* can be further subdivided into serovars according to their antigen inter-relationships revealed by microscopic agglutination and absorption reactions with polyvalent rabbit sera. Like-serovars are then grouped together to form serogroups. Currently there are 26 serogroups comprising 188 serovars (*Table 6.2*).

In Great Britain there are three main important serogroups in human leptospirosis: *Icterohaemorrhagiae*, which has 21 different serovars, *Canicola* with 13 serovars and *Sejroe* with 20 serovars

Table 6.1. Classification of leptospira

Strain	Growth at 13 °C	Growth 8-azaguanine (200 μg per ml)	Hamster inoculation
Pathogenic	No growth	No growth	Death/clinical symptoms
L. interrogans	−	−	
Saprophytic	Growth	Growth	No symptoms
L. biflexa	+	+	

Table 6.2. The serogroups of the *L. interrogans* species

Icterohaemorrhagia	Hebdomadis
Canicola	Huanuco
Sejroe	Javanica
Australis	Kenya
Autumnalis	Louisiana
Ballum	Manhao
Bataviae	Mini
Bufonis	Panama
Butembo	Pomona
Celledoni	Pyrogenes
Cynopteri	Ranarum
Djasiman	Shermani
Grippotyphosa	Tarassovi

Table 6.3

Serogroups—serovars reference strains of *L. interrogans*

Serogroups	Serovars	Reference strains
26	188	188

Serogroups found in Great Britain

Icterohaemorrhagiae	21 serovars
Canicola	13 serovars
Sejroe	20 serovars

(*Table* 6.3). The most common serovars occurring are *Icterohaemorrhagiae* var. *copenhageni* and *icterohaemorrhagiae, Canicola* var. *canicola* and *schueffneri* and *Sejroe* var. *hardjo*.

Classification is based on cross-absorption methods which are tedious, time consuming and technically difficult. Newer methods employing bacterial restriction endonuclease DNA analysis

(BRENDA)[2,3,4] has recently been used. BRENDA is accurate and easy to do and should prove invaluable in studying the epidemiological distribution of strains.

GROWTH

Leptospira interrogans are obligate aerobes requiring demanding growth factors for multiplication, optimal pH range 7·2–7·6, temperature 28–30 °C, essential vitamins B_1 and B_{12}. Their source of energy is not carbohydrate but long-chained fatty acids. Media used for the cultivation of leptospires are usually enriched with either fresh rabbit serum or bovine serum albumin.[5] Various media such as Fletchers,[6], Korthoffs[7] and Stuarts[8] have been successfully used in the past; all are difficult to make and depend on fresh animal products which are not always easy to obtain. Semi-synthetic media such as Ellinghausen and McCullough[9] (EM) containing bovine serum albumin and Tween 80 and its modification by Johnson and Harris[10] (EMJH) are now widely used and can be obtained commercially. Although leptospires will survive in routine bacteriological media, such as blood cultures,[11] further growth in specialized media such as EMJH is essential for successful isolation of leptospires for biological fluids.

It is therefore not surprising that the routine isolation of leptospires from human tissues and fluids is not often attempted. Although serological diagnosis of the disease is easier, cultivation of leptospires is still difficult, but despite these difficulties serological diagnosis of leptospirosis can be attempted by all routine microbiology laboratories.

SEROLOGY

There are four established serological methods used in Great Britain. Two are specifically reference tests, these are the macroscopic slide agglutination test which is a rapid indicator of the presence of leptospiral antibodies and the microscopic agglutination test (MAT) which is a serogroup specific confirmatory test. The other two methods are used primarily as screening methods for early diagnosis of leptospirosis. Both the complement fixation test and the ELISA methods detect genus-specific antibodies and therefore can be used routinely by all microbiology laboratories as screening tests.

Routine laboratories detect the possibility of leptospirosis in patients. The serogroup is confirmed by the MAT reference tests carried out by the Reference Laboratory. Understanding of the clinical diagnosis and epidemiology of the disease has been hampered because the organism is difficult to grow, requires specialized identification procedures and depends on serological diagnosis of the disease.

EPIDEMIOLOGY AND CLINICAL PRESENTATION

The natural reservoir of leptospires is wild animals, particularly the rodent family such as rats, mice, voles and coypu[12] which carry the causative agent of Weil's disease, *Leptospira interrogans* var. *icterohaemorrhagiae*. Domestic animals may also become infected either directly with a leptospiral serogroup specific to their species, for example *Leptospira interrogans* var. *canicola* in dogs or *Leptospira interrogans* var. *hardjo* in cattle (these are examples of maintenance hosts) or indirectly by contact with rat's urine in water, dogs may become severely ill due to *Leptospira interrogans* var. *icterohaemorrhagiae* (accidental host).

Man is always an accidental host. He is the endline of the transmission of the disease from maintenance to accidental host (*Fig. 6.1*).

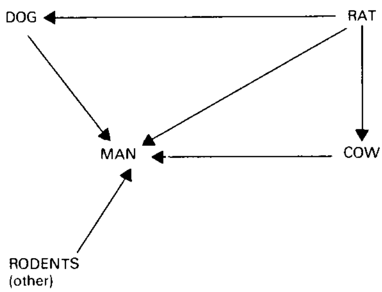


Fig. 6.1.

The host animals are often asymptomatic but will carry leptospires in large numbers in their kidneys and excrete the organisms to the surrounding environment via their urine, thus contaminating canals, rivers and animal foodstuff on farms.

Human infection may result from direct exposure to urine as in the case of cattle-associated leptospirosis (hardjo infections) or indirect exposure to contaminated environment, for example when swimming in infected canal or river waters (icterohaemorrhagiae infections).

Leptospires enter the body through cuts and abrasions of the skin or by contact with mucous membranes of the nose, mouth or eyes.

Clinical presentation

Leptospirosis is *not* synonymous with Weil's disease. Other febrile illnesses are more commonly associated with leptospiral infection. The degree of severity may vary from the characteristic severe presentation of Weil's disease with hepatorenal failure and meningitis to the milder flu-like illness with severe headaches associated with cattle acquired leptospirosis (*Table 6.4*). There is considerable overlap in the presentation of the clinical syndrome known as leptospirosis.

Table 6.4. Classical symptoms

Icterohaemorrhagiae	Canicola	Hebdomadis (hardjo)
Severe hepatorenal failure	Jaundice	Flu-like illness
Conjunctivitis	Conjunctivitis	Severe headache
Meningitis	Meningitis	Meningitis
Death	Death	

Weil's disease is caused by *Leptospira interrogans* var. *icterohaemorrhagiae.* Two distinct phases of the illness are noted, the first phase is essentially leptospir*aemia* while the second is predominantly a leptospir*uria*. The bacteriaemic phase corresponds to the incubation period and varies from 7 to 12 days with subsequent spread to tissues and organs such as liver, kidney and occasionally brain. This accounts for the varied presenting symptoms often noted in leptospirosis. Blood cultures taken during the first week of the illness may give positive growth of leptospires. During the second phase of the illness the leptospires have been removed from the blood and tissues by phagocytes and increasing concentrations of antibodies. The kidney is the only organ that retains the leptospires which migrate to the convoluted tubules where they may multiply producing congestion of the tubules and renal failure. Because the kidney is involved, leptospires may be excreted via the urine and can persist for several months even after clinical cure has been achieved.

During the second week of the illness, increasing amounts of antibodies may be detected and by the tenth day of illness sufficient IgM has been produced to confirm the original clinical diagnosis. When meningitis is a clinical feature, leptospirosis may be isolated from cerebrospinal fluid and antibodies demonstrated.

Although there is severe jaundice and renal failure, biochemical findings are only transitory, the liver function tests are usually normal with raised total bilirubin, while renal function tests are barely above normal.

Milder forms of leptospirosis occur with cattle-associated leptospirosis and there is usually a history of working with dairy cattle. The illness presents with flu-like symptoms including fever and severe headaches—often with mental confusion. The recovery is prolonged, usually 8–10 weeks if antibiotics are not given. Lethargy is common. In a small minority of cases the symptoms may proceed to meningitis, renal and hepatic failure.

Clinical management of all types of leptospiral infection is the same. It involves first taking the correct specimens for laboratory diagnosis (blood cultures and serum) and then *immediately* administering high doses of penicillin. At least 8–10 mega units of penicillin per day should be given for 10 days. In cases of confirmed hardjo infections amoxycillin 500 mg three times daily for 5 days is an appropriate therapy. Alternatively, if the patient is hypersensitive to penicillin, tetracycline at an initial dose of 500 mg followed by 250 mg 8-hourly intravenously for 24 hours and 500 mg 6-hourly orally for 10 days may be used if there is no renal failure or, preferably, erythromycin may be given in 250 mg doses 6-hourly for 10 days. Delay in treatment may mean the death of the patient. In severely ill patients with hepatorenal failure, supportive medical therapy is also required.

HUMAN EPIDEMIOLOGY OF LEPTOSPIROSIS

The predominating infecting serogroups found in this country are *Leptospira interrogans* var. *icterohaemorrhagiae*, var. *canicola*, var. *hardjo* (*Table 6.5*).

In the past, occupations associated with water or sewage were at particular risk from Weil's disease. Personnel often worked in rat-infested conditions and in water polluted with leptospires-infected urine. Not surprisingly leptospirosis was quite common in sewer workers and miners (*Table 6.6*). Today with modern pest control measures, the use of protective clothing and the presence of detergents in waste waters which rapidly destroy leptospires, the epidemiological pattern of leptospirosis has changed dramatically. During the eight years 1978–1985, only five sewer workers and two miners were reported to have suffered from leptospirosis (*Table 6.6*).

The major occupational risk today is among farmworkers. Of the 243 cases of icterohaemorrhagiae infections diagnosed during this period, 68 occurred in the farming industry, including 6 fish farm workers. As in the sewer workers and miners, protective clothing and so on cannot be applied easily to current farming practices. Nevertheless, awareness of the risks is often helpful in itself.

Water-associated sporting activities have also been shown to be major risk factors for leptospirosis—71 of the 243 cases of icterohaemorrhagiae infections between 1978–1985 were linked to water exposure (*Table 6.6*). Most were bathers in fresh-water ponds and streams and some were canoeists practising capsizing drills such as eskimo rolls—14 of these were reported in 1984. Infections resulting from water exposure are usually found in the younger age groups ranging from 12 to 17 years, partaking in adventure sporting activities although properly supervised, completely lacking in the knowledge of the risks

Table 6.5. Human leptospirosis in the British Isles 1978–1985; cases confirmed by Leptospira Reference Unit. Results and numbers (and percentages)

	1978	1979	1980	1981	1982	1983	1984	1985	Total
Icterohaemorrhagiae	26	23	27	39	23	39	37	29	243
	(40%)	(41%)	(56%)	(54%)	(37%)	(32%)	(41%)	(27%)	(39%)
Hardjo	33	18	13	18	13	55	32	54	236
	(51%)	(33%)	(27%)	(25%)	(21%)	(46%)	(36%)	(51%)	(38%)
Canicola	5	5	2	4	7	8	6	7	44
	(8%)	(9%)	(4%)	(6%)	(11%)	(7%)	(7%)	(7%)	(7%)
Others	1	9	6	11	19	18	15	17	96
	(1%)	(17%)	(13%)	(15%)	(31%)	(15%)	(16%)	(15%)	(16%)
Total	65	55	48	72	62	120	90	107	619

Table 6.6. Human leptospirosis—British Isles

	1933–1948	1978–1985*			1978–1985†
Occupation or type of contract	Ictero-haemorrhagiae	Ictero-haemorrhagiae	Hardjo	Canicola	Total
Farmer	45	62	187	1	250
Livestock contact	21	5	13	0	18
Veterinarian	0	1	6	1	8
Coal miner	139	2	0	0	2
Sewer worker	79	5	0	0	5
Water worker	16	1	0	0	1
Fish worker/farmer	216	6	0	0	6
Armed forces personnel	97	5	1	0	6
Bathing and water sports	48	71	6	0	77
Rat bite or contact	11	32	0	2	34
Dog contact	2	0	0	20	20
Miscellaneous or not stated	309	53	23	20	96
Total	983	243	236	44	523

* 1985 provisional figures.
† Other infecting leptospiral serogroups *not* included (96) in total.

being taken, result in late diagnosis and usually the more severe presentation of icterohaemorrhagiae infections. Indeed, leisure activities account for the major proportion of Weil's disease in Great Britain.[13]

Leptospiral infections due to the serogroup canicola are usually isolated incidents and do not contribute greatly to the total number of infections in this country. Dogs, of course can become infected, often without any clinical signs but compared to the dog population of Great Britain, the numbers recorded make canicola infections insignificant. The low incidence of leptospirosis in dogs is probably due to the vigorous vaccination programmes which exist in Great Britain.

CATTLE-ASSOCIATED LEPTOSPIROSIS

The new emerging hazard is undoubtedly those infections associated with cattle-associated leptospirosis. This particular infection usually presents with flu-like illness and is due to the serogroup *Sejroe* serovar *hardjo*.

During the years 1968–1972, one-third of all bovine serum submitted to the Central Veterinary Investigation Centre, Weybridge, had serological evidence of leptospirosis using the microscopic agglutination test as a screening method. Over 90 per cent of these were due to the *Sejroe* serogroup.[14] It soon became apparent to veterinarians that approximately 10 per cent of all cattle abortions were associated with high levels of leptospiral antibodies. They found that 96 per cent of the recorded positive cases could be attributed to a single leptospiral serogroup—*hardjo*.[15] Supporting work from Ellis and his colleagues in Northern Ireland showed that 45 per cent of randomly selected aborted bovine fetuses, and 69 per cent of aborted fetuses from farms with known abortion problems could be shown to be *Sejroe* var. *hardjo*.[16] The same pattern of disease has been described in Australia, New Zealand and Israel. Indeed, the New Zealand survey by Hellstrom and Blackmore[17] estimated that nearly 90 per cent of all cattle herds in that country showed serological evidence of infection. However, figures of about 30 per cent have been reported in the UK by Little and his colleagues[15] and Ellis and his associates.[18]

It is therefore not surprising that such evidence of epidemic proportions of leptospirosis in cattle should also be reflected in the incidence of cattle-associated leptospirosis in man.

This high prevalence of infection in cattle has, presumably, led to an increase in the disease in human *hardjo* infections, first reported in 1968 (although first diagnosed in 1961) and diagnosed fairly commonly in recent years (*Fig. 6.2*).

In 1985, most patients were cowmen or farmers who regularly milked their cows (*Table 6.7*).

Table 6.7. Human leptospirosis and occupational/risk factors, 1985

Occupation/risk factor	Ictero-haemorrhagiae	Hardjo	Canicola	Other or not determined	Total
Dairy or cattle farming	—	42	—	—	42
Fish farmer	1	—	—	—	1
Farmwork (other/unspecified)	4†	2	—	—	6
Other livestock contact*	2†	7	1	2	12
Contact with natural waters	13†††	1	—	6	20
Rats	2	—	—	1	3
Soldiers	1	—	—	1	2
Abroad	4	—	—	4	8
Dogs	—	—	4	—	4
Sewer workers	1	—	—	—	1
Miscellaneous/not stated	1	2	2	3	8
Total	29	54	7	17	107

* *See* Table 6.8.
† Deaths.

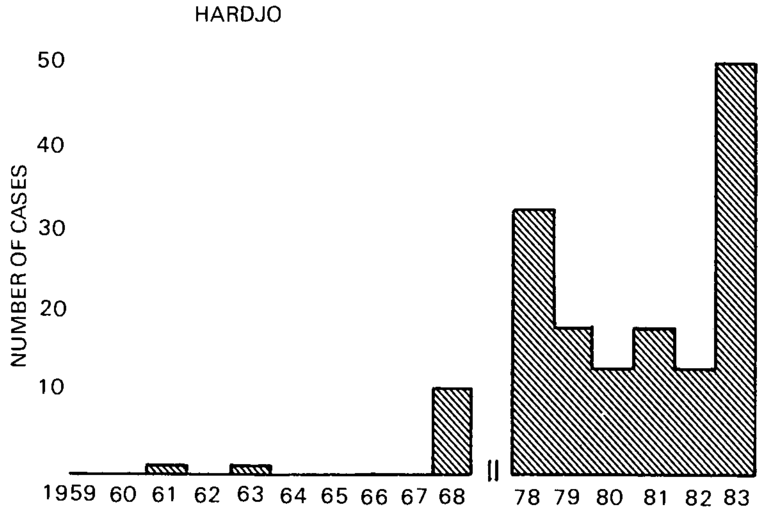


Fig. 6.2. The emergence of hardjo infections from 1959.

During 1984–1985, 500 sera from dairymen were examined retrospectively for leptospirosis (*Table 6.8*).

Table 6.9 shows that in dairy farming areas, 11·5 per cent of dairy farmers had serological evidence indicating past infections. A much lower percentage of positive cases was recorded in arable areas such as Derbyshire (1·8 per cent) and in beef cattle areas such as Herefordshire (3·7 per cent). The overall incidence was about 6 per cent.

In a previous study undertaken in Worcestershire, only one case of leptospirosis (icterohaemorrhagiae) was found in 800 sera tested by the microscopic agglutination test.[18] Assuming that the procedures used were similar, this suggested that cattle-associated leptospirosis is underdiagnosed in many parts of the country, is a recent phenomenon and that there is a lack of general awareness of the condition in dairymen.

In conclusion, the epidemiology of leptospirosis in Great Britain, although complex, is easily differentiated into three major groups, namely *icterohaemorrhagiae, canicola* and *hardjo*. The major hazard in cattle-associated leptospirosis is in those working in the dairy industry. Water-associated leisure pursuits are now the most common risk-related occupations with icterohaemorrhagiae infections.

Table 6.8. Livestock contact occupations associated with leptospiral infections during 1985

Livestock contact occupations	Ictero-haemorrhagiae	Hardjo	Canicola	Others not determined	Total
Farm engineer	—	1	—	—	1
Abbatoir worker	—	1	—	—	1
Slaughterman	1†	1	—	—	2
Butcher	—	1	—	—	1
Veterinarian	—	2	1	—	3
Artificial inseminator	—	1	—	—	1
Total	1†	7	1	—	9

† Deaths.

Table 6.9. The incidence of hardjo infection in dairymen

	Number of sera tested	Number of positive with MAT at 1:80
Herefordshire (1983–1984)	400	15 (3·7%)
Derbyshire (1984–1985)	125	2 (1·6%)
Cheshire (1984–1985)	184	21 (11·5%)
Total	709	38 (5%)
Control (urban)	500	0 (0%)

REFERENCES

1. Turner LH. Leptospirosis Part III Maintenance, isolation and demonstration of Leptospires. *Trans. R. Soc. Trop. Med. Hyg.* 1976; **64:**623–46.
2. Marshall RB, Wilton BE, and Robinson AJ. Identification of leptospira serovars by restriction-endonuclease analysis. *J. Med. Microbiol.* 1981; **14:**163–6.
3. Robinson AJ, Ramadass R, Lee Alison et al. Differentiation of subtypes within Leptospira interrogans serovars hardjo, balcanica and tarassovi by bacterial restriction endonuclease DNA analysis (BRENDA). *J. Med. Microbiol.* 1982; **15:**331–8.
4. Hookey JV, and Waitkins SA. Numerical analysis of Leptospira DNA-restriction endonuclease patterns. *FEMS Microbiology, Letters* 1985: **29:**185–8.
5. World Health Organisation. Guidelines for the control of Leptospirosis. Edited by S Faine. *WHO publication No. 67* 1982.
6. Fletcher W. Recent work on leptospirosis, tsutsugamushi disease and tropical typhus in the Federated Malay States. *Trans. R. Soc. Trop. Med. Hyg.* 1928; **21:**265–87.
7. Korthoff G. Experimentelles Schlammfieber beim Menschen. *Zentralbl. Bakteriol. Parasit. Abt I Orig.* 1932; **125:**429–34.
8. Stuart RD. The preparation and care of a simple culture medium for leptospires. *J. Pathol. Bacteriol.* 1946; **58:**343–49.
9. Ellinghausen HC and McCullough WG. Nutrition of *Leptospira pomona* and growth of 13 other serotypes: fractionation of oleic albumin complex and a medium of bovine albumin and polysorbate 80. *Am. J. Vet. Res.* 1965; **26:**45–51.

10. Johnson RC and Harris VG. Differentiation of pathogenic and saprophytic leptospires, I Growth at low temperatures. *J. Bacteriol.* 1967; **94**:29–31.
11. Palmer M, Waitkins SA, and Zochowski W. Survival of leptospires in commercial blood culture systems. *Zentral. Bakteriol. Mikrobiol. Hyg. A.* 1984; **257**:480–7.
12. Waitkins SA, Wanyangu S and Palmer M. The coypu as a rodent reservoir of leptospira infection in Great Britain. *J. Hyg. Camb.* 1985; **95**:409–17.
13. Waitkins SA. Update on leptospirosis. *Br. Med. J.* 1985; **290**:1502–3.
14. Ministry of Agriculture, Fisheries and Food; Department of Agriculture and Fisheries for Scotland. *Animal health services in Great Britain. Annual Reports, 1969–72.* London: HMSO 1969–72.
15. Little TWA, Richards MS, Hussaini SM et al. The significance of leptospiral antibodies in calving and aborting cattle in South West England. *Vet. Rec.* 1980; **106**:221–4.
16. Ellis WA, O'Brien JJ, Neill SD et al. Bovine leptospirosis: microbiological and serological findings in aborted foetuses. *Vet. Rec.* 1982, **110**:147–50.
17. Hellstrom JS and Blackmore DK. A cohort analysis study of bovine leptospirosis with a dairy herd. *Proceedings of the Second International Symposium on Veterinary, Epidemiology and Economics*, Canberra, Australia, 1980; 214–19.
18. Ellis WA, O'Brien JJ, Neill SD et al. Bovine leptospirosis: serological findings in aborted cows. *Vet. Rec.* 1982; **110**:178–80.

7. CIRCADIAN RHYTHMS, INTERCONTINENTAL TRAVEL AND SHIFTWORK

J. M. Waterhouse, D. S. Minors and A. R. Scott

BACKGROUND

Man is a diurnal creature, that is, his natural habits and social history indicate that he is used to working in the daytime and sleeping at night. Physically and mentally, his body is more efficient by day, shuts down during the evening and recuperates metabolically in the hours of darkness.

However, for an increasing number of the population, there are occasions when the natural sleep/activity schedule has to be changed. This would be the case for businessmen and holiday-makers after intercontinental flights and for workers starting a night shift. For others, for example aircrew, a combination of these changes takes place. It has been known for some time that changes in sleep/activity schedules are associated with a general malaise, known as 'jet-lag' after time-zone transition. Recent advances[1–3] in our understanding of the body clock and of the mechanisms which control and adjust it enable us to offer advice about how best to deal with these difficulties.

The chapter summarizes the advice which can be given, the reasons for the advice, and the problems of disordered body rhythms resulting from time-zone transition and shift work.

Circadian rhythmicity

Recent technological advances such as the development of microchips and miniaturization of tape recorders have enabled the production of portable devices which can record physiological variables repeatedly throughout the course of a 24-hour period. The development of radio-immunoassay techniques, which require only very small volumes of blood, has similarly enabled the time-course of the concentration in plasma of hormones to be described accurately. Such measurements indicate that most physiological variables are not constant over the 24 hours but rather show a rhythmic change. In most cases, including mental performance, lowest values are found nocturnally.[2]

Intuitively this might be expected on the grounds that we are active, eat, drink and live in a socially active environment in the daytime, in contrast to our inactivity and fasting and the quiet environment at night. If this argument was correct then the rhythms would not be observed in individuals living in an environment from which all rhythmic factors had been removed. Such an environment has been achieved experimentally in a protocol which is described as a constant routine.[4] In this, volunteers are required to stay awake continuously for 24 hours in an environment of constant light, heat, noise etc.; they are also required to eat an identical snack each hour. In such conditions many bodily rhythms persist, but with slightly diminished amplitude. It is argued that the rhythmicity that persists in these circumstances must originate from *inside* the body—it is called the endogenous component—and is attributed to some form of body clock.

The endogenous component and zeitgebers (time-givers)

The properties of this 'internal clock' have been investigated by placing individuals in time-free environments such as the Arctic during the summer, in underground caves and in specially-constructed isolation units.[1-3] In such circumstances the individual's times of retiring, rising and eating will no longer be dictated by the external world but rather by the internal clock. Experiments which have investigated individuals in such conditions have established that the rhythms continue, generally with remarkably little day-by-day variation but with a period closer to 25 than 24 hours. For this reason such rhythms are described as *circadian*, from the latin *circa*—about, and *dies*—a day. In our normal environment such circadian rhythms are synchronized to a period closer to 24 hours by rhythmic changes in environmental factors known as zeitgebers. The nature of these zeitgebers depends upon the living organism under consideration.[5] For example, in many plants it is the alternation of light and dark that adjusts the internal clock each day. Further, due to the changes in times of sunrise and sunset that occur with the seasons, the clock is adjusted to a period of *exactly* 24 hours only at the solstices. In predatory animals, the clock is adjusted by the rhythmic availability or absence of food; and, in the case of certain dwellers in the intertidal zone, it is the regular buffeting from high tides which adjusts rhythms to a tidal or lunar day with a period of about 24·8 hours. For humans living in society and able to make use of artificial lighting and preserved food, none of these factors need be particularly important. Instead it is the structure imposed upon us by society as a whole. Thus there are only certain occasions when it is acceptable or socially convenient to sleep, go shopping, make a noise, expect eating facilities etc. Even if we could be independent with regard to lighting and food, we still have to go to bed at the 'normal' time in order to have enough sleep, before getting up for work, school etc. the next day.

The exogenous component

The external environment not only adjusts the internal clock but also exerts a direct effect upon a rhythm measured under normal circumstances.[1-3] Thus the rhythm of deep body temperature in individuals living in a rhythmic environment demonstrates a higher daytime value than when the individuals were under constant routine conditions; by contrast, night-time values are lower in a rhythmic environment than during a constant routine. These direct effects of the external environment are termed exogenous (in contrast to the endogenous clock) and the normal circadian rhythms are considered to result from the sum of endogenous and exogenous components, the relative sizes of which are a characteristic of each variable. Moreover, these two components are normally in phase with each other, for example, both raising body temperature during the daytime and lowering it in the night.

Advantages and disadvantages of a circadian system

The advantage of possessing a circadian clock is that it enables the individual to adjust better to a rhythmic world.[5] Thus, a circadian system enables the body to prepare for waking—by increasing blood pressure, plasma cortisol, plasma adrenaline and body temperature towards the end of the night—and to prepare for sleep by enabling us to 'tone down' in the evening. It also enables plants and animals—and those societies which eschew artificial lighting and alarm clocks—to adjust their rhythms to the changing hours of sunrise and sunset as determined by the seasons.

If the evolutionary value of such a circadian system is clear, then its disadvantages must be considered also.[6] The internal clock is slow to adjust to changed schedules of sleep and activity. This inertia of the circadian system has the advantage that it is not inappropriately adjusted by capricious events—such as an accidental nocturnal wakening or a particularly early dusk due to poor weather. It is disadvantageous, however, when time-zone transitions and night work are considered. It is during such mismatchings of the internal clock and environmental rhythmicity that the inconveniences of jet-lag and the general malaise of night workers are felt.

TIME-ZONE TRANSITION

Jet-lag

Rapid movement from one time-zone to another results in a temporary malaise known as jet-lag.[7] Individuals feel tired during the daytime of the new time-zone and yet have difficulty in sleeping at the new night time. They lose their appetite, might suffer from gastrointestinal disorders, feel irritable, suffer headaches and generally feel 'below par'. The severity of these symptoms tends to increase with the number of time-zones crossed and is generally perceived to be worse after a flight to the east than after one to the west.

Many of these symptoms might reasonably be attributed to the excitement or stress of a long journey, together with the loss of sleep, fatigue and the change of food and customs that will be encountered. Against this view is the observation that these symptoms are as pronounced when one returns home and so could not be due to 'culture shock'. Also, jet-lag is less troublesome and persistent after a north–south flight of comparable length and inconvenience. In other words, in some way these symptoms are due to the time-zone transition itself.

Adjustment of rhythms

Measurements of the circadian rhythms of a variety of variables show that they do not adjust immediately to the time-zone transition.[7–9] The rate of adjustment depends upon the individual, the number of time-zones crossed and the direction of flight as well as the particular variable being considered. Until adjustment to the new time-zone is complete the following abnormalities will exist:

- mismatching between the internal clock and the new external environment,
- an abnormal timing between different circadian rhythms, and
- a shifting time-structure in the individual.

The relationship between these abnormalities and the severity of the symptoms of jet-lag (including interindividual differences) has not been investigated in any detail.[7] Indeed jet-lag might result from any combination of the above abnormalities or might be due to quite different factors. However, it is important to realize that the distractions of important business, holiday excitement or the work schedule in the new time-zone will considerably affect an individual's perception of the severity and persistence of the symptoms.

Advice to travellers

As might be expected therefore, the advice for travellers must be of a rather general nature.[6] It depends upon whether they plan to stay in the new time-zone for a short or longer period of time.

Short stays

If a short stay is envisaged then adjustment cannot occur. In such cases the traveller is advised:

— to take a nap as soon as possible after the journey in order to recuperate from stress, loss of sleep etc.

— to arrange meetings to coincide with daytime on the old time where possible. This entails meetings later in the day on new local time after an eastward flight and as early as possible in the morning after a flight to the west. Morning meetings on local time after an eastward flight (or in the evening after a westward flight) are to be avoided since they coincide with night on old time.

Long stays

By contrast, if the stay is to be longer, then it is advisable to try to adjust circadian rhythms to the new time-zone as rapidly as possible. Several means have been suggested:

Make full use of the zeitgebers in the new time-zone

Under normal circumstances the process of adjustment of the internal clock to the new time-zone is achieved by zeitgebers in the new time-zone. Therefore the traveller should ensure as full exposure to these as possible. In addition, he should attempt to adopt a strict regimen with regard to sleep and activity in phase with the new local time. In practice this implies the need for strong social contacts in the new time-zone. There is some evidence in favour of this suggestion;[10] one group of volunteers whose social life was restricted to that in a hotel room after a time-zone transition adjusted less quickly than did another group who were also allowed access to the new social life outside the hotel.

The use of drugs

Hypnotics and stimulants might be used at appropriate times in the new time-zone, but this does not seem to have been investigated formally. However, there has been a search for a drug which could promote adjustment of the internal clock—a chronobiotic.[11] A variety of substances have been tried but with no substantial effect. Since we are unclear as to the exact site and nature of the internal clock, as well as the means by which zeitgebers affect it, the discovery of an appropriate drug is likely to be fortuitous.

Changing the type of meals

The regular timing of meals might be part of a highly structured life style which, as described above, can make an individual more fully aware of his new local time.

In addition, it is believed that the suitable *choice* of food can influence the body clock in a further way.[11] Thus, as a result of the hormonal changes associated with eating, with the relative amounts of protein and carbohydrate ingested and with the intake of theophylline derivatives—in coffee, for instance—brain function can be altered via changes in activity of its neurotransmitters. Arousal is mediated by adrenergic mechanisms and so is promoted by ingestion of caffeine, which inhibits breakdown of cyclic-AMP, and by high protein intake, which promotes brain uptake of tyrosine, the precursor of adrenaline and noradrenaline. By contrast, sleep is believed to be mediated by serotoninergic mechanisms and the uptake into the brain of the precursor of this transmitter (tryptophane) is promoted by high-carbohydrate diets. In one study of military personnel the efficacy of these dietary modifications in promoting adjustment to time-zone transitions—essentially a high-protein breakfast and a high-carbohydrate supper—appeared to be slight.[11] However, these effects were measured under field conditions in which external disturbances can be rather marked, and render difficult the interpretation of results. Even so, it is generally believed that food intake in man is not a strong zeitgeber.[12]

The pre-adjustment of sleep-activity schedules

If possible, this expedient is best, since the traveller is fully adjusted to the new time-zone immediately on arrival. One could adjust one's times of rising, meals and retiring before the journey by about 2 hours per day in the appropriate direction, that is 2 hours earlier each day for 4 days if flying eastwards through 8 time-zones. This does not seem to have been tested formally and in some cases other commitments would make it impossible to live an 'abnormal' schedule on home time in the days before the flight. It also raises another problem, namely that it would require individuals to live at variance with the rest of society, and this is the dilemma that is particularly marked for those who work shifts, especially at night.

NIGHT WORK

Shift systems

About 25 per cent of the working population work some form of shift system and this often requires them to work at night.[13] When shift systems are considered in detail, it soon becomes apparent that many different forms exist. Some are peculiar to one factory, having been devised collectively by workers and management. Even so, a general classification can be made.[13] The major types are described below.

The 3-shift system

A fairly common example of this would have 8-hour shifts starting at 0600, 1400 and 2200. These shifts often rotate amongst the workforce. Thus all workers share the shifts equally. Questions then arise as to how rapidly should the shifts rotate (each day, alternate days, weekly, monthly), in which direction (morning–evening–night or night–evening–morning) and when should 'days off' be taken?

The 2-shift system

This tends to consist of the 12-hour shifts, 'day' (0800–2000) and 'night' (2000–0800). As before, these shifts can rotate and so the same problems can be raised. By contrast it is sometimes found that the shifts do not rotate so that two groups, one of 'permanent' night workers and the other of day workers, will exist. There is the additional problem with 12-hour shifts that fatigue might become marked towards the end of the shift.

Irregular sytems

There are some shift systems, often in a military context, in which extreme flexibility is required to provide emergency round-the-clock cover or to compensate for loss of personnel.[14] This necessitates irregular hours of work and sleep from one day to the next. In other cases, for example, US submariners, the sleep-work schedules have an 18-hour periodicity being based on a 6-hour 'watch' system.[15] Finally, in the case of aircrew on long-haul flights undergoing time-zone transitions, there is also an extremely irregular schedule since they have to contend with repeated time-zone transitions. In addition, flight restrictions at airports are likely to increase the irregularity of the crew's duty schedule as judged both on local and home time.

Irregular shift schedules will tend to produce two peculiar abnormalities. These are:

- the length of duty periods will vary and so some might be particularly long.
- there will be a loss of zeitgebers from the external environment and the irregularity of habits. This might cause the internal clock to run with a period greater than 24 hours and closer to its free-running period of about 25 hours.

Problems for the night worker

In the present article, unless stated otherwise, 'night worker' refers to all those who work at night, whether they are working 8- or 12-hour shifts and whether or not their shift system rotates.

The night worker is required to alter his lifestyle by working at night and sleeping during the daytime. In some ways, this change in timing of habits is rather similar to that experienced by the traveller after time-zone transition and similar symptoms arise; for example, fatigue, difficulty in sleeping at 'bedtime' and gastrointestinal disorder are predicted.[13,16,17] Fatigue during night work is not unexpected, being a reflection of the 'troughs' of circadian rhythms in mood, performance, adrenaline level and deep body temperature. However, fatigue during the daytime is also observed and this results from a cumulative loss of sleep. This lack of total sleep is not wholly due to the difficulties of sleeping in a daytime environment. Subjects studied in the sound-proofed conditions of a sleep-laboratory have greater difficulty in getting to sleep and staying asleep and their sleep tends to be broken more in the daytime than during the night-time.[18]

This is because sleep during the day-time, and particularly during the morning after a night shift, is being attempted when body temperature and the sympathetic nervous system are preparing the body for a new spell of activity. Even if sleep is achieved, a full bladder is likely to curtail it as urine production also rises. In other words, many of the night workers' problems arise because of a mis-matching between his lifestyle and his internal clock. However, external noise does play a part in reducing daytime sleep, particularly in inner city areas with more traffic and poorer housing which is smaller and less well insulated against noise.

The gastrointestinal disturbances might result from eating food at 'wrong' times and changes in gastric acid secretion are likely. Other possibilities include: the lack of provision of hot food so that there is a reliance on sandwiches etc.; the tendency to nibble rather than take full meals; the higher intake of carbohydrate, caffeine and alcohol and the higher consumption of tobacco.[16,19]

Differences between the problems for the night worker and traveller

Even though the traveller suffering from 'jet-lag' and the night worker share some symptoms and the causes for them, the plight of the night worker is far worse than that of the traveller. First, the changes can continue throughout the individual's working life. Thus, every time the shift changes, another change of routine and a temporary mis-matching between routine and internal clock will occur. Second, for all night workers, even those who are permanent night workers (e.g. some in the newsprint industry), there will be social pressures to adjust to a 'normal' existence during days of rest.[20] Such pressures are likely to be greater in the case of night workers on rotating schedules. These social pressures will be an influence even during work periods. Thus, unless the individual is a social 'loner' or part of an isolated community in which night work is the norm, there will be conflicting zeitgebers as far as the adjustment of his body rhythms is concerned. Therefore, the observation that circadian rhythms adjust slowly and only partially to night work is, perhaps, not surprising.[16,21] The possibility of eventual adjustment to nightwork is rendered less likely when it is realized that during days of rest, when a normal routine is adopted, *all* zeitgebers constrain to adjust rhythms to a normal pattern of nocturnal sleep and diurnal activity. That is, adjustment to night work is slow but loss of adjustment during days of rest is far more rapid.

Are there workers who should be warned against night work?

There are certain groups who would be advised to think very carefully before performing night work and require careful medical counselling.[6] These groups are:

Epileptics

Epileptics are susceptible to seizures when fatigued and so cumulative sleep loss should be avoided. It is possible, although this has not been tested formally, that a rapid rotation of shifts is desirable, since the cumulative sleep loss will be less in this circumstance.

Asthmatics and those with respiratory disorders

Allergic reactions to house dust, etc. are often worse overnight, probably due in part to low plasma concentrations of the endogenous anti-inflammatory agent, cortisol. Histamine release in response to such challenges is most marked then and the severity of asthmatic symptoms owes something to histamine release. This is one reason why asthmatic symptoms are worse at night though irritants in the workplace will exacerbate the position.[22]

Diabetics and others on chronic medication

A problem for all who regularly take drugs is the interpretation of instructions such as 'three times per day with meals', 'once a day on rising' etc. if their schedules are continually being changed. For example, the diabetic's insulin regimen will be very difficult to judge accurately with irregular mealtimes and for this reason he would be advised against shift work in general. As a further example, arthritic pain is often worst in the morning—due in part to a lack of cortisol—and some medication is often taken at night, when cortisol concentrations are low, to reduce pain on waking. With a changed sleep/activity schedule, should the patient be advised to take the medication always before bedtime or to try and take it instead at a time coincident with the trough of his plasma cortisol rhythms?; and, if the latter advice is given, then by how much will his cortisol rhythm have adjusted to the changed schedule?

There is also evidence that the efficacy of medication may vary according to the time of day that it is administered.[23] For example, the response to a morning dose of oral antihistamine may last three times longer than with evening administration; conversely both digitalis and dexamethasone are more effective when taken at night. Even though this new field of chronopharmacology is based mainly upon work performed with animals, the results are beginning to apply to man also and the occupational physician should be aware of the possibilities.

Differences between night workers

Where a choice exists, an individual's decision whether or not to undertake night work is often influenced far more by the advantages of financial reward, promotion, positions of responsibility than by any physiological considerations.[24] For most workers, there are some stages of their lives when these other factors are seen as outweighing the physiological disadvantages. Such a stage is liable to be reached in the early years of married life, when the home and family are being set up and extra money and status are welcome. There is even a small group of workers—often single and socially independent—who *choose* night work because of the blocks of days off associated with it, and the use that can be made of these in pursuing hobbies such as fishing in less crowded circumstances than are available to the day worker at weekends. At other stages of one's career, such as before marriage and when the family is beginning to leave home, the social disadvantages of night work dominate.

However, there is a group of night workers who, in due course, are forced to leave night work for health reasons.[17,25] It would be useful to identify these individuals early in their careers so that they could be forewarned. At the present time no reliable predictive tests are known. Even so, differences between workers who are tolerant or intolerant of night work have been described. On the one hand there are factors which are part of our 'make-up' while, on the other, there are factors that can be attributed to experience. In practice, there is a considerable overlap between these divisions.

The role of 'nature'

This includes the effects of age, circadian rhythm phase and amplitude, and certain psychological indices.[26]

Age

It has been suggested that ageing is associated with a decreased tolerance to changed schedules. It has also been suggested that rhythms adjust more slowly with age, but the evidence from human studies to substantiate this suggestion is most flimsy. It will be noticed that if both suggestions were true then there would be the implication —at least for the younger night worker—that to encourage adjustment of circadian rhythms to night work would be desirable.

The phase of circadian rhythms

Body temperature and adrenaline rhythms in all individuals are not phased identically. In a small proportion of the population—the 'tails' of a normal distribution—rhythms rise earlier in the morning and fall earlier in the evening. Such individuals—'larks' or 'morning types'—are better able to go to sleep and get up early and so are more suited to the morning shift or to an eastward flight. By contrast 'owls', or 'evening types'—with slightly delayed rhythms when compared with most of us—are better suited to night work, since they can 'lie-in' till the afternoon and so get sufficient sleep after having gone to bed at about 0700. They also adjust more readily to a westward flight.

Obviously neither group is generally advantaged with rotating shift systems.

Amplitude

Individuals with circadian rhythms of an amplitude that is higher than average are more tolerant of shift work. The reason for this association is unknown but it has been suggested that such high amplitudes might result in a day-by-day stability of the phasing of rhythms and that this is desirable (*see* The role of experience, *below*). However, since any rhythm is the sum of endogenous and exogenous components, the high amplitude might result, instead or in addition, from a *stability of habits*. This point will be taken up again later.

Psychological make-up

Some studies have indicated that certain psychological characteristics—such as flexible rather than inflexible sleeping habits and ability rather than inability to overcome drowsiness—are found more frequently in those who are tolerant of night work. In practice, these differences imply that those who can sleep at 'unusual' times are advantaged and that the individual who, having been woken up during a daytime sleep following night work, turns over and tries to go back to sleep, is at an advantage compared with another who loses his temper at having been disturbed!

It will be noted that these characteristics of an individual's circadian rhythms might not only have a genetic component but also be seen as resulting, at least in part, from habits accrued by the individual as a result of prolonged night work. This implies that there is a role for experience.

The role of experience

Presumably the lifestyles of experienced aircrew and night workers can be expected to provide information as to how to tackle successfully the problems involved.[6,16] However, first, there is a methodological problem when considering experienced night workers for instance. The problem is that the experienced group is self-selected, since those who have suffered most adverse effects will have decided to leave night work. Therefore, any better adjustment to night work that might be measured in experienced workers need not be an indication that the process of adjustment becomes easier with practice, but might indicate instead that individuals learn how to deal with night work—or that there has been a loss from the sample of those who, for whatever reason, had greater problems with adjustment. In military and civil aviation circumstances, further severe selection measures exist, so that any individuals with difficulties in coping would soon be 'weeded out'.

A better protocol would be that of a prospective study with an age-matched group of day workers. Both groups would have to be studied for a number of years and would be required to remain on the same shift schedule throughout. Measurement of adjustment of rhythms, of perceived difficulties and of commitment would all have to be made regularly. This is obviously a Herculean task. Those results which are available at the present time appear to stress the role of self-selection.[27]

When the logs of aircrew are studied, they often show irregular times and lengths of sleep during long-haul flights, that is, round-the-world trips.[2,3] However, this is not always the case. Sometimes, at least some of the sleep appears to be taken progressively later each day, as would be the case if the individual were following an internal clock running with a period in excess of 24 hours or if the individual was abiding by local time during a westward circumglobal navigation. Another pattern is one in which a split sleep is adopted, with a portion of the sleep being taken during the night on home-time, whenever that might be by local time. There is independent evidence based upon laboratory studies to indicate that a portion of sleep, taken at the same time each day, can stabilize rhythms to a 24-hour period in spite of an otherwise irregular schedule.[28]

In other words, on irregular schedules of sleep and waking, some advantage might be gained by attempting to stabilize circadian rhythms to a 24-hour periodicity. Alternatively, if the schedule permits, to allow rhythms to show a period of about 25 hours seems an acceptable solution. In both cases, the requirements will conflict with a conventional social life in the new time-zone because the individual gives priority to a consideration of his rhythms rather than to his social lifestyle.

When experienced night workers are considered, then similar changes—called commitment or motivation—have been observed.[16] In one study, those nurses who coped best with night work put need for sleep above social considerations. In another, nurses who worked 3 or more nights in succession adopted a regular routine of meals and took naps in the afternoon before their first night shift. By contrast, those who worked only 1 or 2 nights consecutively showed far less regular routines and appeared to skimp on nocturnal meals and diurnal sleep.

Advice on shift schedules with night work

We have considered and explained where possible the problems that are associated with night work. We have also considered the factors which appear to affect the ease with which individuals can adjust to night work. On the basis of this we can offer advice with respect to the least troublesome types of shift schedule involving night work and the means by which a worker can best deal with his particular schedule of night work.

Permanent night work

If night work is permanent then as much adjustment to it as possible is advised. This requires the individual to be highly motivated or committed to the work pattern. Such motivation will be manifest in a variety of ways, but in general it will require a regular and 'abnormal' time to be set aside for sleep each day which will intrude upon conventional social hours. Facilities for uninterrupted daytime sleep—a quiet and darkened room—should exist. If sleep immediately after night work is found to be difficult, then attempting sleep at about 1400—at a time when most people feel tired temporarily—might be useful. The chronic use of hypnotics is not recommended nor is the use of alcohol as its diuretic action will cause sleep to be interrupted by a full bladder. It is also desirable to adjust work hours as much as possible so that one might have a cooked meal in the middle of the night (i.e. 'lunchtime').

Such a regimen is easier to adhere to if it is practised also by colleagues and acquaintances—or at least if they accept it as necessary. It is also beneficial if there are no strong time cues—to indicate the abnormal timing of the behaviour of the individual—as in the cases of work in motorway cafeterias, oil-rigs, submarines etc. However, if conditions are such that daytime sleep is poor so that a substantial cumulative sleep loss is likely, then this should be minimized by working only a limited number of nights (3–5) rather than longer stretches before days off. Where possible, during days off the worker should not revert fully to a 'normal' diurnal life-style and there is a role here for daytime naps (*see* The role of experience, *above*).

Rotation of shifts

If shifts are rotated, then the direction of rotation should be delayed rather than advanced. That is, the sequence 'morning shift, evening shift, night shift' is preferable to that of 'evening shift, morning shift, night shift'. The reason is that the internal clock, with an inherent period greater than 24 hours will better adjust to delaying than advancing shifts, for the same reason that adjustment to westward time-zone transition is more rapid than that to eastward journeys. There is also the general rule that days off should be taken after night work so that any cumulative sleep loss can be made up.

When rotating shifts are considered, the problems of cumulative sleep loss and social disruption are potentially most marked when a slow rotation of shifts is involved. A worker in this case will benefit greatly from colleagues, friends and a family who understand and accept the difficulties. On the other hand, a slow rotation of shifts does give the greatest opportunity for adaptation though, as has already been described, some studies indicate that this is still incomplete even after 3 weeks of night duty.

Where a weekly rotation of shifts is practised, this does not give enough time for rhythms to adjust much but it does raise the issue of social acceptability. For many people, long stretches of night work are undesirable but weekly stretches less so. Further, the unit of social planning is often the week, with special importance being attached to the regular occurrence of weekends, 'wash days' etc. Many prefer a weekly rotation of shifts for these reasons.

At the other extreme to permanent night work are the 'Continental' or 'Metropolitan' systems by which shifts rotate every one or two days. In such circumstances, adjustment is not possible. Accordingly, the advice would be to stabilize one's rhythms to a 24-hour period (and remain *unadjusted* to night work) and to achieve this by observing a regular diurnally-orientated routine whenever possible. Such a regular routine would also reduce the tendency for an individual to become disorientated in time, a malady sometimes associated with rapid rotations of shifts. It will be realized that, with a rapid rotation of shifts, on every occasion that night work is being performed, circadian rhythms will be phased appropriately for sleep. This might have a deleterious effect upon performance on the night shift especially if tasks are repetitive or require prolonged concentration.

There has been a trend recently towards 12-hour shifts. These are normally divided between night and day shifts. The 12-hour night shift raises another problem that occurs in irregular shifts also (*see* Shift systems, *above*); namely, what are the effects of prolonged work hours? The area is poorly researched, but the following general comments are likely to apply:[6,16,25]

— for 'interesting' and varied tasks, performance does not fall off appreciably with longer hours of work
— for repetitive, boring and vigilance-type tasks performance might deteriorate, or at least be harder to sustain
— as with shifts of any length, performance, especially in boring and vigilance tasks, tends to worsen with sleep loss and when circadian rhythms are at their trough.

These decrements in performance due to different factors tend to be additive. This will have some obvious implications for the night worker who is simultaneously working a long shift, suffering from sleep loss and has an uninteresting job. The comments imply equally obvious ways by which attempts to minimize performance decrement could be made.

The 'ideal' night worker

Perhaps many of these factors can be assimilated into a model that describes a 'committed' or 'motivated' night worker.[16] He or she will accept the changes in lifestyle that are involved and attempt to make use of the advantages it offers rather than be irked by the disadvantages. This will require a dedication to work rather than conventional social life, or at least require social activities to be shared with others of a similar position. This dedication will manifest itself as a regular lifestyle with regard to times of sleep, mealtimes and times for chores such as shopping and appointments which will require to be performed at 'normal' day times. These regular influences have been called 'personalized zeitgebers'. Such regularity will act as a strong exogenous component, increasing the amplitude of the circadian rhythm and so helping to stabilize circadian rhythms to a 24-hour period. The details of such a lifestyle will require experiment, of course, but will come with experience.

Such a model would be of use to those who work for long spells on night work since they could phase personalized zeitgebers in such a way that they could promote adjustment to night work. The model could also help to stabilize the rhythms of those on rapidly rotating shift systems or on irregular schedules to a 'normal', 24-hour day. A similar lifestyle could be practised by aircrew during a circumglobal flight if they wished to remain adjusted to 'home' time.

Such advice cannot remove the difficulties, but can only attempt to minimize them. Some will find ultimately that the problems are greater than any advantages and will choose to stop night work; others will have to do so due to ill-health. It is a consolation to know that those who leave night work through ill-health find a regression of some of their symptoms and no worsening of others.[27]

THE FUTURE

There is pessimism associated with the view that time-zone transitions and night work are biologically unnatural and undesirable. Even if, in future, the biological clock can be manipulated by some means or another, the advisability of doing so may be questioned. Instead, we must continue to try to improve the types of shift systems that are worked by reference to 'chronobiological principles' as described in this article. We must continue to give, and improve upon, advice to the individuals concerned. Perhaps future work will enable better predictive tests to be devised so that those likely to experience most difficulties can be forewarned.

In the meanwhile, we must accept that some will find the circumstances too 'stressful'. These can then be identified and advised accordingly. It will be of interest to see if the increasing number of female night workers will result in an increase in menstrual disorders. Sporadic, but unconvincing, reports of such disorders after time-zone transition exist but a population of female night workers would provide material for a fuller study.

SUMMARY OF RECOMMENDATIONS

A *For the traveller crossing time-zones:*

1. If the stay is a short one, then *do not adjust* to the new time-zone but arrange appointments to coincide with daytime on new and old local times.
2. If the stay is longer, then *adjust* to the new time-zone by gaining as much access as possible to zeitgebers in the new environment.
3. If multiple journeys are involved, then either adopt a lifestyle which becomes progressively delayed or stabilize it to remain in phase with home time.

B *For those devising shift systems involving night work:*

1. Ensure that personnel have adequate facility for daytime sleep and leisure. The importance of this increases with the number of successive nights that are worked.
2. If shifts are to be rotated, then rotate them in the direction—morning, afternoon, night. Adjustment to this direction of rotation

requires a lifestyle with a period of more than 24 hours and this is easier to achieve than one which becomes progressively earlier.

3. After night work, allow days off for recuperation from sleep loss.

C *For workers involved in night work:*

1. If the night shift is permanent, semi-permanent or slowly rotating, then attempt to adjust to it. This requires an adjustment of sleeping habits, mealtimes and leisure activities—that is, a radical reorganization of lifestyle. Colleagues, family and friends must be encouraged to help in this.

2. If the night shift rotates rapidly or the work schedule is irregular then attempt to stabilize your own lifestyle to a 'normal', diurnally-orientated day. Again, this involves an appropriate timing of 'personal zeitgebers'.

D *For physicians concerned with the effects of night work on health:*

1. Discourage sufferers from asthma and diabetes from working at night. Forbid people with epilepsy from working at night except those whose convulsions are associated with sleep.

2. Look out for symptoms of, particularly, chronic fatigue and gastrointestinal disorder.

3. Be prepared to offer advice to those on chronic medication with regard to the time at which drugs should be taken during abnormal hours of work.

REFERENCES

1. Wever RA. *The Circadian System of Man: Results of Experiments under Temporal Isolation*. New York: Springer-Verlag, 1979.
2. Minors DS and Waterhouse JM. *Circadian Rhythms and the Human*. Bristol: Wright PSG, 1981.
3. Moore-Ede MC, Sulzman FM and Fuller CA. *The Clocks that Time Us*. Cambridge, Massachusetts: Harvard University Press, 1982.
4. Minors DS and Waterhouse JM. The use of constant routines in unmasking the endogenous component of human circadian rhythms. *Chronobiol. International* 1984; **1**:205–16.
5. Cloudsley-Thompson JL. *Biological Clocks. Their functions in Nature*. London: Wiedenfeld and Nicolson, 1980.
6. Minors DS, Scott AR and Waterhouse JM. Circadian arrhythmia: shiftwork, travel and health. *J. Soc. Occup. Med.* 1986; **36**:39–44.
7. Winget CM, De Roshia CW, Markely CL et al. A review of human physiological and performance changes associated with desynchronosis of biological rhythms. *Aviat. Space Environ. Med.* 1984; **55**:1085–96.
8. Aschoff J, Hoffman K, Pohl H et al. Re-entrainment of circadian rhythms after phase shifts of the zeitgeber. *Chronobiologia* 1975; **2**:23–78.
9. Wegmann HM and Klein KE. Jet-lag and aircrew scheduling. In: Folkard S and Monk TH (ed.) *Hours of Work: Temporal Factors in Work Scheduling*. Chichester: John Wiley, 1986, pp. 263–76.

10. Klein KE and Wegmann H-M. The resynchronization of human circadian rhythms after transmeridian flights as a result of flight direction and mode of activity. In: Scheving LE, Halberg F and Pauly JE (ed.) *Chronobiology*. Tokyo: Igaku Shoin, 1974, pp. 564–70.
11. Graeber RC, Sing HC and Cuthbert BN. The impact of transmeridian flight on deploying soldiers. In: Johnson LC, Tepas DI, Colquhoun WP et al. (ed.) *Biological Rhythms, Sleep and Shift Work*. Lancaster: MTP Press, 1981, pp. 513–37.
12. Reinberg A. Chronobiology and nutrition. In: Reinberg A and Smolensky MH (ed.). *Biological Rhythms and Medicine. Cellular, Metabolic, Physiopathologic, and Pharmacologic Aspects*. Berlin: Springer-Verlag, 1983, pp. 265–300.
13. Kogi K. Introduction to the problems of shiftwork. In: Folkard S and Monk TH (ed.) *Hours of Work. Temporal Factors in Work Scheduling*. Chichester: John Wiley, 1985, pp. 165–84.
14. Nicholson AN, Stone BM, Borland RG et al. Adaptation to irregularity of rest and activity. *Aviat. Space Environ. Med.* 1984; **55**:102–12.
15. Naitoh P, Beare AN, Bearsner RJ et al. Altered circadian periodicities in oral temperature and mood in men on an 18-hour work/rest cycle during a nuclear submarine patrol. *Int. J. Chronobiol.* 1983; **8**:149–73.
16. Folkard S, Minors DS and Waterhouse JM. Chronobiology and shift work: current issues and trends. *Chronobiologia* 1985; **12**:31–54.
17. Harrington JM. *Shiftwork and Health. A Critical Review of the Literature*. London: HMSO, 1978.
18. Akerstedt T. Work schedule and sleep. *Experientia* 1984; **40**:417–32.
19. Reinberg A, Migraine C, Apfelbaum M et al. Circadian and ultradian rhythms in the eating behaviour and nutrient intake of oil refinery operators (Study 2). *Chronobiologia* 1979; **Suppl. 1**:89–102.
20. Walker J. Social problems of shiftwork. In: Folkard S and Monk TH (ed.) *Hours of Work. Temporal Factors in Work-Scheduling*. Chichester: John Wiley, 1985, pp. 211–25.
21. Akerstedt T. Adjustment of physiological circadian rhythms and the sleep–awake cycle to shiftwork. In: Folkard S and Monk TH (ed.) *Hours of Work. Temporal Factors in Work-Scheduling*. Chichester: John Wiley, 1985, pp. 185–97.
22. Clark TJH. The Philip Ellman lecture. The circadian rhythm of asthma. *Br. J. Dis. Chest* 1985; **79**:115–24.
23. Reinberg A and Smolensky MH. Circadian changes of drug disposition in man. *Clin. Pharmocokin.* 1982; **7**:401–420.
24. Wedderburn AAI. How important are the social effects of shiftwork? In: Johnson LC, Tepas DI, Colquhoun WP et al. (ed.) *Biological Rhythms, Sleep and Shift Work*. Lancaster: MTP Press, 1981, pp. 257–69.
25. Rutenfranz J. Occupational health measures for night- and shiftworkers. *J. Human Ergol* 1982; **11 Suppl.** 67–86.
26. Monk TH and Folkard S. Individual differences in shiftwork adjustment. In: Folkard S and Monk TH (ed.) *Hours of Work. Temporal Factors in Work-Scheduling*. Chichester: John Wiley, 1985, pp. 227–37.
27. Verhaegen P, Maason A, Meers A. Health problems in shiftworkers. In: Johnson LC, Tepas DI, Colquhoun WP et al (ed.) *Biological Rhythms, Sleep and Shift Work*. Lancaster: MTP Press, 1981, pp. 271–82.
28. Minors DS and Waterhouse JM. Anchor sleep as a synchronizer of rhythms on abnormal schedules. *Int. J. Chronobiol.* 1981; **7**:165–88.

8. PSYCHIATRIC ASPECTS OF FITNESS FOR WORK

B. S. Ricketts

THE EFFECT OF MENTAL DISORDER ON WORK

It has been estimated recently that in the United Kingdom 37 million working days are lost annually through mental disorder while the lower number of 23 million are lost through injuries at work. Minor mental disorders contribute substantially to sickness absence and reduced work output. The world literature provides examples of this in groups ranging from officers in the British Civil Service to Romanian cotton workers. The prevalence of minor mental disorders in the general adult population had been shown to be high, between 10 per cent and 30 per cent.

A proportion of sickness absence is due to relapse. It is therefore useful to consider how many people at work have already had at least one period of absence through mental disorder. Wansbrough and Cooper[1] established the proportion of the working population in England and Wales in 1971 who had previously required hospitalization for a mental disorder. Out of a total of 13·9 million men in employment up to 400 000 (2·88 per cent) had been hospitalized thus; out of a total of 7·6 million women in employment up to 300 000 (3·95 per cent) had been hospitalized. The majority of those hospitalized will have suffered from episodes of psychotic illness, whether depressive, manic or schizophrenic.

THE EFFECT OF WORK ON MENTAL HEALTH

Worldwide financial recession has led to new interest in the psychological consequences of unemployment.[2] We should not allow this to overshadow the corollary that work may enhance mental health. Wansbrough and Cooper have written eloquently about the impact that work may have after mental illness. It will provide not only earnings and greater purchasing power but also evidence of recovery and restored status. It has been shown to have a direct effect of reducing symptoms in some circumstances.

It follows that it may be in the interest of someone with residual mental illness to return to work. At the same time it may be in the interests of the employer that such a person should not, or at least should not do so yet. The occupational physician may have a difficult conflict of interests to balance. There is the additional question as to whether certain forms of work may be stressful to the detriment of mental health. This is touched upon in the paragraph below concerning the outlook after schizophrenia and in the final section of the chapter.

NEUROSIS AND WORK

Neuroses are common in working populations as they are in the general adult population. For example, Jenkins[3] has shown in a recent epidemiological study of 321 unselected young executive officers in one department of the British Civil Service that 33 per cent were suffering from minor mental disorders, 30 per cent doing so to the point of qualification as clinical cases. She found depression to be the most common diagnosis, present in 30·07 per cent and anxiety to be the next most common, present in 7·06 per cent.

The impact of neurosis on working capacity

The purpose of Jenkins' study was to investigate the contribution of minor psychiatric morbidity to sickness absence in young men and women. The contribution was found to be important. And we may turn to Aubrey Lewis,[4] one of the fathers of modern psychiatry, for a balanced judgement about the bearing that neurosis may have on work capacity. Writing in 1944 about the vocational aspects of neurosis in soldiers he commented:

> 'Neurotic symptoms and the ability to do good work sometimes go together; but the neurotic illnesses which occur in some uprooted men and women in the armed forces are seldom compatible with good military work.'

It is likely that we can all bring to mind an example of a person who suffers from a neurotic illness and yet at the same time is most capable in work. It is equally likely that we can each see from experience in our own fields how certain kinds of exacting work demand freedom from neurosis.

Jenkins has argued soundly that recognition of minor psychiatric disorder by a doctor increases the prospect of improvement. And so the recognition has purpose. We must then ask how far the severity of the disorder relates to the degree of incapacity and amount of absence. The relationship may not be close.

A single man of 45 was treated successfully as a day-patient for a severe depressive neurosis—his first episode of mental illness—and for longstanding lack of social confidence. On discharge he considered his social confidence to be greater than at any point in the past. However, he was referred back after an interval of four months having still not returned to his previously long held employment as a machine tool operator. His failure to return proved unrelated to the conditions for which he had been treated. Support of elderly parents with whom he lived had occupied him increasingly. It was only after measures appropriate to this had been taken and after expiry of sick pay that he resumed work.

A danger of misdiagnosis

Anxiety neurosis presenting for the first time after the age of 40 is not common. Anxiety phenomena presenting freshly in the older half of the working population may be the first manifestions of depressive psychosis or dementia. Such depressive illness is potentially responsive to any antidepressant drug of the tricyclic group (amitriptyline, clomipramine, etc.) and if not so in practice is likely to respond to other treatments in specialist hands.

MANIC-DEPRESSIVE PSYCHOSIS AND WORK

The illnesses subsumed under this heading are bipolar manic-depressive illnesses, unipolar manic illnesses, unipolar depressive psychoses and mixed affective states. The clinical features of these illnesses are well known. Yet in practice many cases of depression are found with a mixture of neurotic and psychotic features, defying the classification. A further semantic confusion arises through the varied use of the term 'endogenous'. The term has often been used synonymously with 'psychotic'. However, it is clear that some episodes of depressive psychosis are not truly endogenous for they may be precipitated by stress. It is necessary in each case of depression to determine whether the features of the illness lie predominantly in the neurotic or psychotic sphere before selecting treatments. The more psychotic the illness the more appropriate is the use of a tricyclic antidepressant drug.

Readers unsure of physical disorders that may present or coexist with depression should remind themselves of the wide range including cerebral tumour, Parkinson's disease, neurosyphilis, hypothyroidism, Cushing's disease, hypocalcaemia, hypercalcaemia and anaemia. The centrally acting antihypertensive drug methyldopa has been notorious for causing depression as a side effect. Although alcohol-related problems are discussed in Chapter 9 it should not be forgotten here that regular alcohol excess aggravates depressed mood and reduces the effectiveness of antidepressant drug treatment.

Kraepelin at the end of the last century drew a fundamental distinction between manic-depressive psychosis and an illness he described as 'dementia praecox'. This latter illness was yet to be given its twentieth century name of 'schizophrenia'. Kraepelin's distinction between these two major illness groups was that episodes of manic-depressive illness were followed by full recovery, however long or severe they may have been, while episodes of dementia praecox led on to some permanent mental impairment. This distinction still carries an important message for the occupational physician who has a patient on sick leave on account of a manic-depressive illness: full recovery and return to work is generally to be expected, however long it takes.

A few suffer from only one manic-depressive episode in a life-time and others suffer from episodes spaced some years apart. Sadly, some suffer so frequently that the frequency itself impedes employment. Not all manic-depressive illnesses prove responsive to lithium treatment.

The reasons for incapacity at work are sometimes complex.

The experienced warden of a children's home suffered from one depressive and one manic episode in a period of two years. It was therefore considered that she was unfit for the shift work involved. The alternative work that she was given in administration offered standard working hours. However, it failed to provide her with the responsibility and the continuity of child care that she had relished. She developed a mild but persistent reactive depressive neurosis and it was this rather than her manic-depressive illness that led to her medical retirement.

The effect of shift work on manic-depressive illness has been the subject of little study. Yet there are peripheral reasons for thinking that the effect might be disturbance of mood equilibrium. Disturbance of circadian rhythm in depressive psychosis is exemplified by alteration of the normal circadian pattern of cortisol release. And sleep deprivation has been shown to elevate mood in depressive psychotic episodes.

There are rare situations in which mild manic illness proves productive.

A college lecturer in natural sciences suffered from occasional frank manic illnesses necessitating brief absence from work. However, his mood at other times was as often mildly manic as it was normal. As a result his lectures were usually most stimulating and he had an enthused following of students.

Treatment and its effect on psychomotor performance

Tricyclic antidepressant drugs remain the first line of treatment in all but the most severe of depressive psychoses. There is evidence that continuation of a tricyclic drug after initial recovery reduces relapse rate.[5] Therefore it is common for people to be fully recovered and back at work while still taking such a drug. It is therefore useful to know that psychomotor performance is impaired more by some of the tricyclic group than by others. Greater impairment of psychomotor performance and greater sedation go hand in hand. Those in common use that are more impairing in these ways include amitriptyline and doxepin while those that are less so include clomipramine, nortriptyline and protriptyline. The tetracyclic drug mianserin is relatively sedative. However, these side effects all diminish as treatment is continued. An informative review of the effects of antidepressants on skilled performance is given by Seppala and Linnoila.[6] Adjacent to this is a paper by Hindmarch et al.[7] reporting the effect of amitriptyline in single doses of 50 mg on the car driving of volunteers, by comparison with placebo and another drug on brake reaction time. Amitriptyline is shown to increase the brake reaction time significantly.

Lithium carbonate as used in prophylaxis has the advantage of causing no sedation whatever in normal therapeutic dosage. A fine and mild tremor is a dose-related side effect and is not uncommon.

SCHIZOPHRENIA AND WORK

Kraepelin made the point as mentioned above that schizophrenic illnesses are followed by lasting impairment. However, this rule does not hold good in all cases.

A man of 22 with no history of previous psychiatric illness was hospitalized for an acute schizophrenic episode of abrupt onset. This followed shortly after his marriage, a move with his wife into their first mortgaged house and some situational difficulties with the house. These life events were assumed to have been precipitants for the illness. He was symptom-free within two months, and able to resume his work as a skilled machine operator within three months. Long-term prophylactic drug therapy was not considered necessary. When he presented again six years later it was learned that he had been entirely well in the interim, had been promoted by his firm and had become a contented father. His new presentation was with a similar schizophrenic episode occurring in stressful circumstances: he was involved in running a course on behalf of his firm and for this he had made his first ever journey from his temperate home country to a tropical climate. He made a similar swift and full recovery.

There is no absolute definition of schizophrenia and various diagnostic criteria have been proposed. The case just described was classical in its symptomatology yet it would not meet a criterion used by some that symptoms must have been present for at least six months.

In those cases meeting all accepted diagnostic criteria and thus displaying an illness of at least six months' duration there is usually some lasting residual impairment. With the possibility of return to work in mind it is important to consider what impairment this might be. It may be either a persistence of mild so-called 'positive' symptoms such as hallucinations and delusions, or a collection of so-called 'negative' symptoms such as social withdrawal, flattening of emotional responsiveness and reduced drive, or a mixture of both. The impairment is not necessarily incompatible with work. However, the school teacher left with negative symptoms after a schizophrenic illness may then be unsuited to class room work and may have to consider alternative work as, for example, a library assistant.

Prognosis in schizophrenia

There is known to be a statistical correlation between sudden onset and good prognosis. However, it is not possible to set down clear rules for making a prognosis. The most comprehensive study of schizophrenia, its diagnosis and course has been the International Pilot Study of Schizophrenia conducted by the World Health Organization[8] involving nine countries and 1202 patients. Although five factors were identified as best predictors of poor outcome in a two-year follow-up these accounted for only 19 per cent of the variance. The factors were: long duration of episode before examination, social isolation, history of past psychiatric treatment, being widowed, divorced or separated and abnormal behaviour before admission.

Improving the outlook after schizophrenia

The value of long-term neuroleptic medication in preventing relapse and reducing residual symptoms is well known and needs no further mention here. However, there has been a recent development in the aftercare of schizophrenia not yet sufficiently known outside psychiatric circles. It has been shown that relapse is more likely when any relatives with whom the recovered person lives demonstrate hostility, make critical comments and are over-involved with the person. These relapse-provoking behaviours in relatives are spoken of as comprising *high levels of expressed emotion*. Furthermore, it has been shown that these behaviours can be modified and the relapse rate reduced by a package of measures including education about schizophrenia for relatives, supportive group meetings for relatives and family interviews to modify interaction between the recovered person and the relatives. A successful controlled trial of social intervention on these lines is reported by Leff et al.[9] and a review of the whole field of expressed emotions in schizophrenia is provided by Leff and Vaughn.[10] It does not require too great a conceptual leap to see how these findings might be applied to the relationships of recovered people in the workplace.

Who succeeds in work after schizophrenia?

Wansbrough and Cooper[1] conducted an extensive survey of open employment after hospitalization for mental illness. Their findings are especially applicable to those who have suffered from schizophrenia. They found that those returning to work immediately after discharge from hospital failed more often than those returning after an interval. Another principal finding of theirs follows from the point already made about the difficulty of prognosis in schizophrenia: the only effective and realistic way available of assessing employability in most cases is by practical trial. There are few rehabilitation units and sheltered facilities equipped for such trials. Therefore, in practice, the trial usually has to take place in open employment. Sergeant[11] has recently drawn attention to the need for a long probationary period in work after schizophrenia. From his position as a psychiatric referee carrying out pre-employment psychiatric examinations he has observed that the Employment Protection Act passed in the United Kingdom in 1975 has appeared, by reducing the maximum health probationary period to one year, to reduce the number of those with a history of schizophrenia who have been accepted for employment.

Successful return to work after schizophrenia among the semi-skilled and unskilled has been shown by Morgan and Gopalaswamy[12] in a study from a psychiatric rehabilitation hospital to correlate strongly with low social withdrawal, the absence of socially embarrassing behaviour, better work performance and better relationships with other people. This correlation supports a commonsense conclusion that the degree of success in return to lower grades of employment is related to the degree of recovery. Our present knowledge suggests that successful return into higher grades requires not only a more complete recovery but a position entailing little interpersonal stress.

DEMENTIA AND WORK

Suspicion of dementia calls for specialist psychiatric assessment. This is so not only because the employment implications of the diagnosis are ultimately so grave but also because much skill is needed in some instances to distinguish between eminently treatable depressive illness and untreatable dementia. Psychometric tests can sometimes aid diagnosis but their administration and interpretation are not tasks for the enthusiastic general practitioner.

THE IDENTIFICATION OF CASES BY QUESTIONNAIRE

Questionnaires and self-rating scales have been developed in the psychiatric field both for case identification and for the monitoring of clinical change. Instruments designed for accurate distinction between cases and non-cases are not well suited to measurement of severity within a particular illness and vice versa. Case identification is of most interest in the employment field. And it is identification of minor disorders that calls for aids of this kind, because the major disorders are usually self-evident as illness when they develop. There are two instruments that qualify well for mention here, their validity and reliability having been established carefully.

1 The General Health Questionnaire (GHQ)

This is a self-administered screening test developed by Goldberg[13,14] for the identification of minor psychiatric disorders. None the less it has been shown in practice to identify schizophrenia and depressive psychosis satisfactorily. The full version consists of 60 items each to be scored as 'less than usual', 'no more than usual', 'rather more than usual' or 'much more than usual'. Shorter versions with 30, 28 and 20 items exist but as the full version takes only 6–8 minutes to complete it is generally preferable. The result is case identification without allocation of diagnosis. The GHQ is intended for completion as an immediate prelude to medical consultation and it is in this setting that it has been validated.

There are some noteworthy limitations to the GHQ. It was the instrument used by Jenkins[3] in conjunction with a psychiatric interview formalized by schedule in the study already mentioned. She found that 1 in 6 of those who were GHQ negative proved to be cases in interview whereas 1 in 3 of those GHQ positive proved not to be cases. This discrepancy serves principally to demonstrate the inherent difficulty of setting a finite limit around categories of illness such as neuroses the margins of which are marked with no natural line. The GHQ has been ineffective in situations where the subject has set out to 'fake healthy'. There has been low case yield in a survey intended to identify stressed executives. This serves to emphasize the value of using the GHQ in tandem with clinical interview.

2 The Leeds scales for self-assessment of anxiety and depression

These include specific and general scales. It is the *Leeds Self-assessment of Anxiety General Scale* and the *Leeds Self-assessment of Depression General Scale* that may be applied to case identification.[15] They provide good cut-offs between healthy and sick populations. The anxiety scale scores require correction for age as there is significant negative correlation between anxiety score and age. The form of depression identified by the depression scale is 'endogenous' indicating that it is of psychotic rather than neurotic type.

3 Other instruments of possible interest

The State-Trait Anxiety Inventory described by Spielberger et al.[16] may be applied to both the identifying and monitoring of anxiety states. It consists of two self-rated scales each consisting of 20 statements. The 'state' scale records state of feeling at that precise moment while the 'trait' scale records the general underlying feeling. It is therefore the 'state' scale that will monitor change, this taking only 6–10 minutes to complete.

Both the Beck Depression Inventory[17] and the Carroll Rating Scale for Depression[18] have been developed for measuring the severity of depressive illness. Their principal uses are in trials of treatment, measuring the change that may occur.

4 The use of questionnaires in occupational health settings

It is important to appreciate the limitations of the questionnaires described above. Some doctors whose practice is rooted in the traditions of organic medicine are ill at ease with the relatively subjective nature of the diagnostic process in psychiatry. For them such questionnaires may have particular appeal. However, there lies a fundamental difference between these questionnaires and the mainly objective confirmatory tests provided by pathological and radiological services: the questionnaires provide quantitative measurement of mental phenomena that still depend on subjective report. The scores from such questionnaires cannot be taken safely as complete diagnoses. They are to be integrated with the results of clinical interview. Thus a questionnaire is an adjunct to interview rather than an alternative to it. Once a questionnaire has helped in the identification of 'a case' there remains the point that such identification does not necessarily establish unfitness for work.

The administration of the questionnaires described above is a task that may be taken on by a nurse in an occupational health setting, assuming that the nurse has read the manual for the questionnaire concerned.

The questionnaire of choice for general pre-employment screening is the General Health Questionnaire. The investigations of selected working groups believed to be specially at risk for anxiety or depression are tasks for which the Leeds Scales are well suited, whichever scale is appropriate to the suspected disorder. The monitoring of anxiety in the work setting, once recognized, could be undertaken with the 'state' scale of the State-Trait Anxiety Inventory. As for the monitoring of depression in a person who remains at work and the re-assessment of depression at the end of sick leave with a return to work in mind, it must be said that few practising psychiatrists would use questionnaires in clinical situations of this kind. Should there be a particular reason for using one, such as a need to ensure high performance capability, then either the Beck Depression Inventory or the Carroll Rating Scale for Depression might prove helpful.

SOURCES OF STRESS IN WORK

It might be imagined that questionnaires would serve to illuminate sources of stress in the work-place or in the context of work by revealing subgroups of workers with above-average incidence of stress-related mental disorders.

However, Kasl[19] found when reviewing the literature on this topic in 1973 that the association between job level, job satisfaction and mental health was "not particularly impressive". A more recent review by Cooper[20] highlights findings such as the following: greater dissatisfaction among older as opposed to younger operational police sergeants; negative correlation between status congruency and the incidence of psychiatric disorders among Navy employees; and lower incidence of psychiatric disorders when there is participation in the decision-making process. However, no very strong overall principles emerge. We might attribute this to inadequate research. Yet would it not be equally reasonable to postulate that it is due to the very individual nature of susceptibility to stress? One man's meat may be another man's poison. This adage is a reminder that stress may reveal for one person the painful limitation of his appropriate coping skill while bringing a satisfying knowledge to another of an ability he had not previously known himself to possess.

It is fitting to close this chapter with a reminder that mental disorder rarely leads to dysfunction in every sphere of a person's life. While in one sphere there is dysfunction in another there may be valuable strength.

REFERENCES

1. Wansbrough N and Cooper P. *Open Employment after Mental Illness*. London: Tavistock Publications, 1980.
2. Smith R. 'What's the point. I'm no use to anybody': the psychological consequences of unemployment. *Br. Med. J.* 1985; **291:**1338–41.
3. Jenkins R. Minor psychiatric morbidity in employed young men and women and its contribution to sickness absence. *Br. J. Ind. Med.* 1985; **42:**147–54.
4. Lewis A and Goodyear K. Vocational aspects of neurosis in soldiers. *Lancet* 1944; **ii:**105–9.
5. Coppen A and Peet M. The long term management of patients with affective disorders. In: Paykel ES and Coppen A (ed.) *Psychopharmacology of Affective Disorders*, 248–56. Oxford: Oxford University Press, 1979.
6. Seppala T and Linnoila M. Effects of zimeldine and other antidepressants on skilled performance: a comprehensive review. *Acta Psychiatr. Scand.* 1983; **68:** Suppl. 308, 135–40.

7. Hindmarch I, Subhan Z and Stoker MJ. The effects of zimeldine and amitriptyline on car driving and psychomotor performance. *Acta Psychiatr. Scand.* 1983; **68:** Suppl. 308, 141–6.
8. World Health Organization. *Schizophrenia: an International Follow-up Study*. New York: Wiley, 1979.
9. Leff J, Kuipers L, Berkowitz R et al. A controlled trial of social intervention in the families of schizophrenic patients. *Br. J. Psychiatry* 1982; **141:**121–34.
10. Leff J and Vaughn C. *Expressed Emotion in Families: its Significance for Mental Illness*. London: Guildford Press, 1985.
11. Sergeant H. Pre-employment psychiatric examinations. *Lancet* 1984; **ii:**212–4.
12. Morgan R and Gopalaswamy AK. Employability of the mentally disabled in the 1970s. *Br. J. Psychiatry* 1983; **142:**572–4.
13. Goldberg D. *The detection of psychiatric illness by questionnaire* (Maudsley Monograph 21). London: Oxford University Press, 1972.
14. Goldberg D. *Manual of the General Health Questionnaire*. Windsor: NFER Publishing company, 1978.
15. Snaith RP, Bridge GWK and Hamilton M. The Leeds scales for the self-assessment of anxiety and depression. *Br. J. Psychiatry* 1976; **128:**156–65.
16. Spielberger CD, Gorsuch RL and Lushene RE. *Manual for the State-Trait Anxiety Inventory* (Self-evaluation Questionnaire). Palo Alto, California: Consulting Psychologists Press, 1970.
17. Beck AT, Ward CH, Mendelson M et al. An inventory for measuring depression. *Arch. Gen. Psychiat.* 1961; **4:**561–71.
18. Carroll BJ, Feinberg M, Smouse PE et al. The Carroll Rating Scale for Depression: I, Development, reliability and validation. *Br. J. Psychiatry* 1981; **138:**194–200.
19. Kasl SV. Mental health and work environment: an examination of the evidence. *J. Occup. Med.* 1973; **15:**509–18.
20. Cooper CL. Identifying stressors at work: recent research developments. *J. Psychosom. Res.* 1983; **27:**369–76.

9. IDENTIFYING AND HELPING PROBLEM DRINKERS AT WORK

A. Ward Gardner

INTRODUCTION

World consumption of alcohol, measured in litres of 100 per cent alcohol per head, rose from 2·2 in 1965 to 2·5 litres in 1980, the latest date for which figures are available (*Table 9.1*).[1]

World total known alcohol production has also risen from 73·8 to 110·1 million hectolitres of 100 per cent alcohol over the same period (*Table 9.2*).[1] On this basis the present problems associated with alcohol will not go away. There is little reason to believe that there will be any change in the continuing upward trend, although a continuance cannot be assumed.

Table 9.3 shows the growth in alcohol consumption in European and non-European countries. Note the rapid increase in alcohol consumption in the non-European countries.

In relation to health effects it is not usually important whether alcohol is consumed as beer, wine or spirits—or other forms, or whether the ingestion is intermittent or continual: what matters is the total amount of alcohol consumed, measured as 100 per cent alcohol. This total amount of alcohol consumed per head is what correlates best with the number of people who become dependent on alcohol or who are harmed by it. Alcoholics Anonymous (AA) say that alcoholism is an illness. There is no doubt that ingestion of alcohol in sufficient quantities can lead to dependence or harm—including the production of disease(s). However, the illness model now appears outmoded. ACCEPT (Addictions Community Centres for Prevention Education and Treatment) believe that alcohol addiction is learned behaviour and that predisposing factors may be extraversion and impulsiveness. Whatever the reasons, rationalizations or whys of the matter, one feature distinguishes alcohol addicts, problem drinkers, alcoholics or whatever they are called from other people: they drink more in quantity (and they usually drink more often) than most people. Why some people drink more than others is another question. Alcohol abuse is often blamed on stress—but who cannot

Table 9.1. World alcohol consumption (litres of 100 per cent alcohol per head) (reproduced by permission of WHO)

Region or country	1965				1980			
	Beer	Wine	Spirits	Total	Beer	Wine	Spirits	Total
Africa	0·1	0·8	0·1	1·0	0·4	0·3	0·1	0·7
USA and Canada	2·7	0·5	1·5	4·7	4·5	0·9	2·7	8·1
Latin America and Caribbean	0·4	1·1	0·9	2·4	0·9	1·0	0·5	2·4
Asia, excluding Japan	0·0	0·0	0·0	0·1	0·0	0·0	0·1	0·2
Japan	0·9	0·0	1·2	2·1	1·7	0·0	2·2	4·0
Europe, excluding USSR	2·4	5·1	1·1	8·7	3·3	5·4	2·1	10·9
USSR	0·6	0·6	3·2	4·4	1·0	1·3	3·3	5·6
Oceania, excluding Australia and New Zealand	0·0	0·0	0·0	0·0	2·0	0·0	0·0	2·0
Australia and New Zealand	5·0	0·7	0·7	6·4	5·6	2·8	2·2	10·6
WORLD	0·7	0·9	0·6	2·2	0·9	0·8	0·8	2·5

Based on an assumed alcohol content as follows: beer, 4·4%; wine, 11·0%; spirits, 40%. Totals may not be exactly the sum of the individual entries since these have been rounded off.

Table 9.2. Total alcohol production (beer, wine and spirits). Volume (million hl of 100 per cent alcohol) (reproduced by permission of WHO)

Region or country	1965				1980			
	Beer	Wine	Spirits	Total	Beer	Wine	Spirits	Total
Africa	0·35	2·5	0·2	3·1	1·9	1·2	0·2	3·3
USA and Canada	5·4	1·0	3·2	9·6	11·0	2·2	6·5	19·7
Latin America and Caribbean	1·1	2·75	2·3	6·1	3·3	3·5	2·0	6·8
Asia, excluding Japan	0·3	0·1	0·7	1·1	1·1	0·2	3·0	4·3
Japan	0·9	0·0	1·2	2·1	2·0	0·0	2·6	4·6
Europe, excluding USSR	10·9	22·8	5·1	38·8	16·2	26·3	10·2	52·7
USSR	1·4	1·5	7·4	10·3	2·7	3·5	8·8	15·0
Oceania, excluding Australia and New Zealand	0·0	0·0	0·0	0·0	0·1	0·0	0·0	0·1
Australia and New Zealand	0·7	0·1	0·1	0·9	1·0	0·5	0·4	1·9
WORLD	22·4	30·9	20·5	73·8	39·3	37·5	33·3	110·1

Table 9.3. Growth in alcohol consumption (in litres of 100 per cent alcohol per head) in countries in which it has increased rapidly (reproduced by permission of WHO)

European countries

Country	1960				1981				Increase 1960–1981 (%)
	Beer	Wine	Spirits	Total	Beer	Wine	Spirits	Total	
Netherlands	1·0	0·2	1·1	2·3	3·9	1·4	2·6	7·9	243
Denmark	4·2	1·1	1·9	7·2	6·5	2·2	2·9	11·9	61
Finland	1·1	0·1	1·3	2·5	2·5	0·5	2·8	5·8	132
German Democratic Republic	3·1	0·3	0·6	4·0	5·7	1·8	1·5	9·0	125
Canada	3·4	0·4	1·4	6·2	6·2	1·1	4·8	12·1	95
Hungary	1·6	3·2	1·4	7·2	3·9	3·5	4·8	12·2	69
Poland	1·0	0·4	2·4	3·8	1·3	0·8	4·3	6·4	68
Germany, Federal Republic of	2·5	0·2	1·5	4·2	3·8	1·0	3·3	8·1	93
Austria	3·2	2·3	2·4	7·9	4·6	3·8	1·5	9·9	25

Non-European countries

Country	1960				1981				Increase 1960–1981 (%)
	Beer	Wine	Spirits	Total	Beer	Wine	Spirits	Total	
Japan	1·1	1·3*	1·2	3·6	4·3	2·2*	3·2	9·7	169
Mexico	1·0	—	0·5	1·5	1·8	—	0·9	2·7	80
Repubic of Korea	0·1	—	0·7	0·8	1·5	—	5·4	6·9	762

* Including sake.

plead stress at some time? A chicken and egg relationship may well be involved but the causal relationship remains doubtful. Alcohol has also been said to be a good solvent for anxiety: unfortunately, like other addictive drugs it also appears to be an excellent solvent for both conscience and reason.

The problems associated with alcohol use and abuse are international, national, local and individual. The family aspect is often forgotten in looking at people who have alcohol-related problems: it is important to remember that families, friends, work colleagues and others are always involved as part of the problem of any individual. Families often suffer greatly.

Alcohol-related problems can be described in two main categories: acute alcohol intoxication (drunkenness) and chronic alcohol intoxication/addiction/problem drinking/alcoholism. This classification is inexact and overlaps exist, but it is a useful way to think about the problems in a work setting—and perhaps in other settings too. The problems of *acute alcohol intoxication* ('drunkenness') can be classified as medical, social and legal and are listed below.

Medical problems of alcohol intoxication

Acute alcohol poisoning or overdose; amnesic episodes; drug overdose; suicidal behaviour; acute gastritis; pancreatitis; trauma; head injury; accidents and injuries; epilepsy; hangover; fetal alcohol syndrome; congenital abnormalities in the offspring of women who take alcohol during pregnancy.[2,3]

Social problems of acute alcohol intoxication

Social isolation; aggressive behaviour; passive behaviour; domestic violence; child abuse; child neglect; sexual problems; domestic accidents and injuries; industrial accidents; and injuries; lying, cheating and deception; absenteeism; poor time-keeping and lateness.

Legal problems of acute alcohol intoxication

Driving offences; drunkenness offences; theft; shoplifting; taking and driving a vehicle; criminal damage to property; hooliganism and vandalism; fraud; deception; assault; homicide.

The problems of *chronic alcohol intoxication/addiction/problem drinking/alcoholism* (the alcohol-dependence syndrome) can be classified as medical, social and legal and are listed below.

Medical problems of chronic alcohol intoxication addiction/problem drinking/alcoholism

These are given in detail in *Table 9.9*.

Social problems of addiction/problem drinking/alcoholism

Debt; homelessness; family problems; marital problems; lying, cheating and deception; sexual problems; absenteeism; employment problems; stigma.

Legal problems of addiction/problem drinking/alcoholism

Theft; fraud; deception; vagrancy.

The rest of this chapter discusses alcohol abuse in relation to occupation, alcohol problems at work and what may or could be done about the problems in an occupational setting.

OCCUPATIONS WITH A HIGH INCIDENCE OF PROBLEM DRINKING

Problem drinkers are found in greater numbers at both ends of the social scale—in social classes I and V. In social class I, which includes medical practitioners, both freedom from close supervision at work and being able to afford to drink in quantity contribute to the raised incidence. The notion that most problem drinkers are 'winos' or 'skid row bums' is far from the truth: most people who have drinking problems are still in employment, the usual estimates are 75–80 per cent. Their drinking behaviour is usually by deliberate design as low-profile as possible, quite unlike the 'don't care' attitudes of down-and-out drinkers. However, some problem drinkers do have a high profile of drinking behaviour. Examples of this are senior executives in some organizations, younger problem drinkers who make very little attempt to disguise their drinking, usually because they see no need, and people who are in the early stages of the dependence continuum.

Anyone in an occupation which has access to alcohol is at high risk—for example publicans, barmaids and those manufacturing and selling alcoholic beverages. Sales people and those who have to entertain others as part of their jobs also have an above-average risk of becoming problem drinkers. Seafarers, including officers, pilots and ratings, are notoriously high on the scale on account of their access to cheap alcohol and their freedom from normal shore-side social restraints. Custom, practice and social and culture norms can also make a contribution to the amount of alcohol which any individual will take. Peer group influences play an important part in determining drinking styles. Thus, if a person joins a heavy-drinking group there is then a greater chance of drinking heavily—for example medical students in the UK. Women may form the bulk of some workforces and, in view of present indications of increased excessive drinking by women, it would be a mistake to assume that few or no problems will be present in predominantly female groups.

Cirrhosis of the liver can be used as a marker because it is caused by *excessive* alcohol ingestion in about 80 per cent or more of cases.

DEALING WITH ALCOHOLISM AND PROBLEM DRINKING AT WORK

Two useful reviews—one of alcohol policies in national health and development planning[4] and the other of guidelines for investigating alcohol problems and developing appropriate responses[5]—have recently been published by the World Health Organization. These documents are probably too wide in outlook for those with only a local or organizational interest in alcohol problems but are recommended as useful background reading.

What steps can or could be taken in any organization to deal with alcoholism and problem drinking? One manoeuvre which is frequently employed is to deny that a problem exists. This belief may be genuinely held, and in a very small workforce may be true. But the statistical probability is that there will be 1 or 2 people with problem drinking per 100 people in a workforce, and it is likely that these people will be the cause of much social disruption in their work groups. A suggested list of steps for tackling alcohol-associated problems in any organization could be as follows:

— inform the workforce of the problems
— alter the social climate with regard to alcohol and drinking and identify problem people
— try to remove 'cover-up'
— have a clearly stated system for help.

Inform the workforce of the problem

This step will include giving factual information and producing increasing awareness of alcohol-related problems by discussion with management, unions and joint consultative bodies. Seminars about alcohol may be a useful method of raising awareness and of allowing discussions and questions. The usual techniques of health education—for example posters and booklets—can also be employed but are often too impersonal to change attitudes in the group.

There is, of course, a case to be made for trying to give information about alcohol, its effects and its problems as a part of general health education. There may also be good reasons for developing an educational programme specific to a given organization which is concerned about alcohol and problem drinking. Such a programme can be of special value to the person who is on the borderline of drinking too much and who is able to cut back on consumption before damage or dependence occur.

Two important steps should follow when awareness and interest have increased: *first* agree a policy about alcohol and *second* spell out a system for help. A policy statement which is communicated positively to the workforce can do much to allay the very real fears which every problem drinker has about his or her job security. However, it will not remove that fear. The policy should also include a statement about help. Any policy statement must have the full agreement and support of top management or it is bound to fail. It would be nice to add 'and trade unions' to the preceding sentence, but the attitude of many unions is still unfortunately rather negative. An example of a policy statement is given in *Table 9.4*—not as a model

Table 9.4. A policy on problem drinkers and alcoholics

The use of any drug including alcohol which interferes with safe and efficient job function is a matter of company concern and will be dealt with in an appropriate manner.

The company recognizes that people who have a drinking problem should be offered help, which may be medical or may be from other sources.

When a person is noticed as having a drinking problem that individual should be referred at once to a counsellor who may be either a doctor or a person specially trained for this role so that he or she can be given help and treatment on the clear understanding that:

- The company will offer help towards the treatment of a drinking problem if the individual wants this help for its own sake.
- Review of the individual's future with the company will be undertaken at an appropriate time, depending for example on the severity of the condition, on the understanding that successful treatment does not automatically lead to reinstatement or continued employment. The whole circumstances will need to be taken into consideration, including the individual's particular duties.
- If the individual refuses help or if after having had suitable help and treatment given he again places his problem drinking before his job, then it will be necessary to deal with these situations as disciplinary problems.

policy, but as an example of the kind of statements which could usefully be made.

People who have a drinking problem will inevitably become increasingly poor performers. This lowering of standards will always present problems both for the person and for his manager. If steps are not taken by the individual with or without help to deal effectively with problem drinking, the likely outcome can only be painful and may end in dismissal. So, the sooner effective referral and other steps are taken, the sooner will the problem(s) be resolved. This is clearly in the interest of all concerned.

Alter the social climate with regard to alcohol and drinking

Management can take decisions about whether alcohol will or will not be available for consumption on company premises in any dining room, cafeteria, visitors' hospitality room or canteen. Concern about alcohol can often be communicated clearly to every member of the workforce by not allowing alcohol at work. In organizations which adopt such a policy, the statement is an everyday one. In organizations which permit drinking in any of their dining or cafeteria facilities a strong impact on social climate regarding alcohol consumption can be made by banning alcohol. The step may be unpopular initially but will usually be seen to be a wise one. The ban should affect everyone, including directors, managers and visitors. Another strategy may be to limit the amount of alcohol available. This has been used on board ships where the quantity of alcohol put on board can be related to crew size, and issued on a daily basis.

Late return to work after a lunch break is often a significant marker of over-indulgence in alcohol, so a hard-nosed attitude to being back on the job on time after lunch can make a strong statement about the organization's attitude. Challenging all latecomers, especially persistent ones, can help to identify problem drinkers.

A high rate of short-term absences, especially on Mondays and Fridays, may be identified in those who have problems, provided that suitable records are kept. Again, the individual should be challenged and absences should not be tolerated.

Identify problem people

A few of the identifying features have been mentioned above—lunchtime lateness and high short-term absences especially on Mondays and Fridays. Other indicators are statements like 'he's a good chap but'—the 'but' often refers to his known drinking. This statement should be one which both alerts managers and supervisors, and leads to action. Too often the action phase does not follow. Performance deterioration without any obvious explanation is an especially important early identifier. Work performance deteriorates even more markedly as time passes and there may also be a deterioration in appearance and a noticeable smell of drink. Around this stage excuses are the order of the day. Later still the hangovers and the pale sweaty and trembling appearance of 'the shakes' may be observed.

Injuries are more common in problem drinkers, especially those sustained from road traffic accidents. Work injuries have been shown to be about three times more common in people who have been drinking, especially on the afternoon shift.

The removal of cover-up at work may expose the dilemma of 'incomplete intervention' which has to be faced by occupational health services: a dependent individual is alerted to the effects of dependence at work but not elsewhere.

Remove cover-up

A natural reaction in most people is to help others. People with alcohol-related problems are very good at playing on the sympathies of their colleagues and supervisors and use this so that their lateness, hangovers and decreasing effectiveness are 'covered' by others. In the long term, other people become unable to cover up and/or lose patience in trying to do so. During this time, the problem drinker is allowed to get worse and becomes more of a problem. The lesson is simple: cover-up only allows everything to take longer and to get worse. So it should be removed at once. Every effort should be made to focus on the real problem. The job of managers and supervisors is not to talk to problem drinkers about alcohol but to discuss lateness, attendance, effectiveness and performance. If alcohol abuse is suspected, however slightly or diffidently, as a reason for these adverse behavioural effects, the person should be referred *at once* for diagnosis and counselling. Referral should be a formal procedure—so a record is made of the person being referred on a given date on account of stated problems. This move prevents occupational health groups being played off against the management by problem drinkers, and others. It also allows objectives and time scale to be recorded so that the employee, the manager and the doctor or counsellor have common recorded understanding. Failure of the employee/problem drinker to cooperate or comply should lead to the disciplinary route for dealing with the problem(s). Accurate records of lateness, absences, effectiveness and performance can be used to confront the person with facts. Lastly, follow-up must be insisted on. Both the supervisor and the doctor or counsellor should arrange regular and recorded follow-up until everyone is sure that a problem no longer exists. This usually takes a minimum of 3–5 years.

The role of the occupational health team

The occupational health team should act as a catalyst in their organization to try to produce enlightened attitudes about problem drinking.

The doctor, the nurse and the counsellor all have responsibilities in dealing with individuals who have or are suspected to have alcohol-related problems. The occupational health team should be well informed about how to search for problem drinking in the history, past medical history, examination and use of tests in relation to their patients or clients.

History

Tables 9.5–9.8 list the points which should be sought. Items marked * are dealt with in more detail later in the chapter. Direct questions should be asked—for example about relatives with alcohol problems, about work performance, about injuries and about criminal offences. It is not good enough to hope that the patient will volunteer such information. The order of appearance of symptoms may also indicate the extent of the problem (*Tables 9.5, 9.6*).

Table 9.5. History

Area of person's history	Aspects of history that should be questioned
Present	diseases*
	injuries, especially road traffic accidents
	alcoholic symptoms*
	detailed drinking history*
Past	diseases*
Family	other relatives with alcohol problem
	very strict about alcohol or teetotal
	financial problems, debt
	neglect of children
	wife beating, non-accidental injuries in children
Occupation	some occupations carry high risk*
Criminal offences	drunk and disorderly
	drunken driving
	other offences under influence of alcohol
Work	becomes less effective*

* These items are dealt with later in the chapter.

Drinking history

In taking a detailed drinking history, it is more informative to ask 'what did you have to drink yesterday' than to ask questions like 'how much beer do you normally drink'. Begin on waking (an early morning drink may be taken to 'cure' the shakes) and continue till bedtime (does he remember? did he have a nightcap?). If the reply to 'what did you have to drink yesterday' is 'none' go back a day

Table 9.6. 'Typical' order of appearance of symptoms in men admitted to an alcohol problem clinic

completely unable to keep to a limit
needing more drink than companion
difficulty in preventing getting drunk
spending more time drinking
missing meals
memory lapses and 'blackouts'
difficulty in cutting down
giving up hobbies and interests
restless without regular drinking
change to drinking the same amounts on work days as on weekends
organizing to ensure a supply of alcohol
tense on waking
passing out while drinking
can't think of anything else
retching (often provoked by brushing teeth)
pale and sweating
decreased tolerance to alcohol (=liver failure)
working feeling frightened
hallucinations (DTs)

or two. If a person is a teetotal, ask why. He or she may be an alcohol abuser who is recovering or who has been off drink for a longer period. Remember, too, that answers given to questions about drinking will often be untruthful—and in the case of those with established problem drinking, will almost certainly underestimate the total. 'Admitted consumption' is a useful phrase in the notes. Try to get answers in numbers for each category of drink when taking the history, but wait until you have explored the consumption of beer, spirits and wine—and any other beverage—before adding it all up (1 drink = $\frac{1}{2}$ pint of beer = 1 glass wine = 1 tot of spirits). Heavy smoking often goes with heavy drinking, so always ask about smoking. It is

Table 9.7. Detailed drinking history aspects that should be questioned

What did you have to drink yesterday?
— begin on waking and continue till bedtime
Bout drinking?
What is taken?
— question intake from beer, spirits and wine
How much of each?
— answers here often evasive and low!
Duration of drinking and build-up to present levels
Smoking history
— heavy smoking often goes with heavy drinking

useful also to enquire about any medicines or drugs which are taken regularly so that confusion with side effects of these can be avoided.

Work history

A work history should be taken from the patient but in all cases it should be compared with an account of his or her work from supervisory sources. The points to note are shown in *Table 9.8*.

Table 9.8. Work history (best from a third party)

Performance suffers, less effective
Lateness, especially after lunch
Absences rise
— measure both the number of *spells* and the duration of spell; and the number of days lost from work in a given period
— especially short-term absences
— usually worse on Mondays and Friday and around holidays
Unexplained lapses of all kinds

Alcoholic symptoms

It is always wise in taking a history from anyone suspected of having an alcohol-abuse problem to ask if they ever had 'the shakes', 'blackouts', a fit or 'the DTs', to enquire also about peripheral neuropathy and to seek for items in the alcohol-abuse symptoms, diseases and injuries list in *Table 9.9*.

Symptoms, diseases and injuries related to alcohol abuse

The alcohol-related significance of many diseases and conditions from which problem drinkers suffer may not be realized by those looking after them.[6] If the item is recognized as alcohol-related by the patient, it may be concealed deliberately. Doctors will sometimes write 'gastritis' on a medical certificate when they mean hangover or alcoholism; 'neurasthenia' when they see the shakes; or 'head injury' when a road traffic accident has occurred involving a car which has been driven off the road by its sole drinker occupant at 1 a.m. on a Sunday morning —the so-called 'one-person accident'. Each of these examples disguises the real problem. However, if good records are kept and the person who looks at the records is alert, the possibility of alcohol abuse as a root cause should not be overlooked. A combination of several of the diseases or injuries in *Table 9.9* should suggest that questions about alcohol abuse be asked and that further examinations and tests be carried out. I would go so far as to suggest that if any three of the diseases listed appear on the records of any individual, then this person should be considered to have a drinking problem until careful investigations have shown this to be untrue.

A few words about some of the conditions may be in order. Bleeding from the gastro-intestinal tract should always raise *suspicion* about alcohol intake in spite of whatever other label may be attached. If other diseases in the disguise list are found, the index of suspicion can rise. Alcoholic cardiomyopathy—which presents as congestive failure of 'unknown' origin—is often diagnosed by exclusion, so seek other alcohol markers. Arrhythmia may also be alcohol-related.

A 'brewer's chest', the beer drinker's paunch, is usually identified easily as alcohol related. However, 'flabby jelly' fatness—the appearance of a man who shakes like a jelly and often has gynaecomastia—is not so often recognized to be connected with alcohol intake.

Epilepsy which occurs for the first time in a patient over the age of 25 years with no apparent cause may be, and often is due to *withdrawal* from alcohol. The characteristic feature of an alcohol withdrawal fit is that it is the first fit in a person over the age of 25. The interval between the fit and the last drink is usually 6–48 h.

It is a great time saver to take a history with alcohol abuse in mind and to do the alcohol investigations before the full neurological investigations because many such presentations of 'illness' will be due to alcohol abuse. A blackout without obvious cause may also be part of alcohol abuse—so again it is wise to remember this early in the differential diagnosis so that appropriate examination and tests can be done.

Many years ago it was taught that syphilis was the great mimic—and that it should always be thought about. Today, late syphilis is a comparatively uncommon disease, but alcohol-induced and alcohol-related illnesses are common and should always be considered in the differential diagnosis.

Examination

The appearances of some alcohol abusers have been mentioned above —the beer drinker's paunch, obesity, and 'flabby-jelly' fatness in men. Wasted legs, due to alcoholic myopathy, may also be noted. Some alcohol abusers have high colour with red eyes, spider naevi on their cheeks and a red nose. Others, particularly if withdrawing or suffering from a hangover, will look pale, sweaty and trembling, so appearances can vary. The smell of drink on the breath should always be sought and should be related to the time of day when the patient is seen. The more inappropriate the timing, the higher is the possibility of

Table 9.9. List of symptoms, diseases and injuries which may be used to disguise alcohol abuse or which may be used to indicate alcohol abuse

1. *Gastrointestinal*
   — cancer of mouth, pharynx and oesophagus
   — haematemesis
   — oesophagitis
   — oesophageal varices
   — Mallory–Weiss syndrome
   — gastritis
   — pancreatitis
   — hepatitis
   — cirrhosis of liver
   — Zieve's syndrome (haemolytic anaemia associated with cirrhosis, hyperlipidaemia and jaundice)
   — hepatocellular cancer
   — melaena
   — cancer of the rectum[7]

2. *Circulatory*
   — hypertension (alcohol may be related to about 25% of cases of 'essential' hypertension)[8]
   — congestive heart failure of 'unknown origin'
   — cardiomyopathy (alcoholic)
   — cardiac beri-beri
   — atrial fibrillation[9] after binge drinking in apparently healthy people
   — cerebral infarction and spontaneous subarachnoid haemorrhage leading to strokes,[10] especially strokes which occur at or 24 h after a weekend and strokes in young men after binge drinking[11,12,13]

3. *Nutritional*
   — obesity, especially 'flabby jelly' fatness
   — vitamin deficiencies: peripheral neuropathy, cardiac beri-beri, thiamine deficiency (Wernicke's encephalopathy)

4. *Infective*
   — tuberculosis

5. *Nervous*
   — epilepsy—a fit for the first time after the age of 25 with no obvious or apparent cause is most often an alcohol *withdrawal* fit[14]
   — unexplained 'blackouts' (6–48 h after last drink)
     — these may be withdrawal fits
   — peripheral neuropathy
   — anxiety state, phobic anxiety
   — depression
   — attempted suicide
   — cerebral degeneration, Korsakoff psychosis
   — Wernicke's encephalopathy
   — paranoid states
   — psychoses
   — hallucinations
   — lying (often gross), with confabulation—often unrecognized until whole fleeces of wool are pulled over the eyes of the trusting and the believing before suspicions are aroused
   — fugue states (alcoholic 'blackouts')
   — memory impairment in younger people

— 'neurasthenia'
— delirium tremens ('the DTs') (which usually occurs about 40–96 h after the last drink)
— 'the shakes' (withdrawal tremor)
— tremor (gross—may be 'the shakes')
— toxic amblyopia, diplopia, ophthalmoplegia

6. *Endocrine*
— diabetes
— testicular atrophy
— impotence and sexual failure(s)
— feminization

7. *Musculoskeletal*
— acute myopathy (alcoholic myopathy in young men after binge drinking)
— chronic myopathy (alcoholic myopathy)
— rib (and clavicle) fractures seen on chest radiography
— gout/hyperuricaemia

8. *Blood*
— anaemia
— Zieve's syndrome (haemolytic anaemia with cirrhosis, hyperlipidaemia and jaundice)
— altered morphology of red blood cells: raised mean corpuscular volume (MCV), stomatocytes, cells with two (enzocytes) or three (triangulocytes), large indentations[15]

9. *Poisoning*
— intoxication by alcohol
— tolerance of alcohol in amount which would result in (gross) intoxication in normal people

10. *Injuries*
— road traffic accidents, especially one-car accidents late at night at weekends; pedestrian injuries
— work injuries (of all kinds)
— home injuries (of all kinds)

alcohol abuse. An aggressive, irascible, irritable and restless demeanour is common. Sometimes when such people stop drinking they revert to quiet, calm, pleasant individuals. This Jekyll and Hyde behaviour is often seen in problem drinkers. Spider naevi should be sought and counted below the root of the neck, more than 5 below this level is said to be a significant finding. The liver edge should always be sought and the size and type of any enlargement noted.

Use of tests

The blood-alcohol level is the most usual of the direct measurements. Breath-alcohol and urine-alcohol may also be used. These tests give specific answers at or around the time when the test was taken. Blood alcohol measurement is greatly under-used and can often serve to verify or deny the veracity of the history of today's intake.

However, alcohol abuse is a long-term problem and so other tests can be used which show the secondary effects of alcohol abuse—for example on liver enzymes.[16] The most generally useful test is the serum gamma-glutamyl transferase (γ-GT) level. This is said to be raised by recent alcohol abuse. However, in some people who have been drinking heavily for some years the γ-GT may not be raised and so the test is not sensitive.

The serum level of γ-GT is increased in a number of clinical conditions including other liver disease, myocardial infarction, diabetes mellitus, and by many drugs including anticonvulsants and antidepressives.

When the liver is further damaged by alcohol the aspartate aminotransferase (AST) is usually the next enzyme to show a raised level. This enzyme is a sensitive indicator of hepatocellular necrosis but is not specific for liver disease—raised levels are found for example after myocardial infarction and in myositis.[16] It must, of course, be assumed that there is no other obvious cause for altered liver function before presuming that any abnormality is alcohol related. Glutamate dehydrogenase (GDH) is a reliable reflector of liver cell necrosis in chronic alcoholics.[17] The predominantly centrilobular intrahepatic localization of GDH may help to explain the sensitivity of this enzyme in alcoholic liver injury. The test is a good deal less invasive than liver biopsy and identifies patients with active alcohol liver disease, including those who are asymptomatic. This information can and should be communicated to the patient to assist in gaining cooperation and to concentrate and increase therapeutic effort on those whose problem drinking is causing active liver damage.

The mean cell volume (MCV) increases by about 1·7 fl for every 10 g of alcohol taken daily[18] although there is very wide individual variation. The increase in MCV corresponds to 1·4 fl per drink since the average drink contains 8 g of alcohol. It follows therefore that in people who take 10 or more drinks a day (80 g or more of alcohol) that the MCV will be outside the normal range.

In interpreting raised MCV levels the possibility of reasons other than alcohol should be borne in mind—these include, for example, B_{12} and folate deficiency, radiotherapy and recent surgery.

The MCV rises slowly in alcohol abusers and takes a similar time to settle. In the absence of other obvious explanations, the combination of a raised γ-GT level with a raised MCV can be presumed to be due to alcohol abuse. The γ-GT and MCV are unlikely to be raised in early dependence and are sometimes normal in quite severe dependence. So, these tests are useful only if positive. Triglyceride and serum urate levels are also said to be raised but they are not of much value as identifying markers. Folate deficiency has also been shown to be present in about two-thirds of a series of 30 patients investigated.[19] Alcohol suppresses the reticulocyte response to folic acid and twice the physiological daily dose (150 μg) by injection was required to overcome this effect.

A comparison of the use of questionnaires with laboratory tests in the detection of excessive drinking and alcoholism was reported.[20] The screened population was high-risk groups such as hospital patients. In these groups, the questionnaires performed better as predictors than did laboratory tests. Information is not at present available on the performance of these questionnaires in a general population, so it would be prudent to reserve judgement. The questionnaires used were the Michigan Alcoholism Screening Test (MAST), the brief MAST[21] (selected questions from the full MAST) and the Reich test. All of these questionnaires were validated using high-risk groups—again, not the situation in which a screening test would be most useful in an occupational setting.

Treatment

Treatment is not normally a responsibility of occupational health services, but a few words on this subject may be appropriate. There are two main steps in treating any patient who has an alcohol-related problem:

— the *first* step, drying out (detoxification) is not always required, but when it is, it must be accomplished, usually in hospital, before any other therapy can be attempted. It should be remembered always that DTs may be life-threatening and so requires efficient management in hospital.

— the *second* step is to try to get the full cooperation of the patient; to allow him to take a hard, cold and long look at his or her condition; at the damage which drinking is causing to him or herself and to the family and at work; and to have a plan to stop drinking and carry it out.[22]

A study compared the results, of a 3-hour assessment of the patient and of his wife, with inpatient treatment of 6 weeks on the one hand, with a 1-hour counselling session in which the patient was confronted with the problems outlined in step two above.[23] In the 1-hour counselling session the couples were told in sympathetic and constructive terms that they were responsible for the stated goals. No further appointments were given. They were told that someone would call monthly to monitor progress and that if there were withdrawal symptoms they should get in touch with their family doctor. The 6-week treatment group were given an introduction to Alcoholics Anonymous (AA), drugs to cover withdrawal, psychiatric appointment(s) and a 6-week stay in an alcohol unit. The wife was offered help with the problems of her husband's dependence on alcohol, with her own problems and with the marital situation. One year later there were no apparent differences in outcome between the two groups. The costs, however, were vastly different.

The daily *supervised* administration of disulfiram (*antabuse*) is a useful treatment in some cases. The patient must be seen by a reliable observer *to swallow* the daily dose.

Disulfiram produces unpleasant symptoms if the patient takes alcohol. This treatment is reliable only as far as the disulfiram is taken. Experience shows that the observer has to be very reliable indeed and that giving the disulfiram to the patient to take leads to high relapse rates.

The level of intervention required to effect change will vary with time and with the level of dependence (*Fig. 9.1*).

spontaneous recovery as a result of chance environmental influences → recovery after brief counselling from a professional or friend → recovery after formal treatment → intervention unlikely to effect recovery

Fig. 9.1. The level of duration of dependence will affect the intervention necessary to produce change.

A very few cases of problem drinking recur secondary to depressive illness. This group does well in contrast to the ordinary group of problem drinkers where about a third will do well, another third will go on drinking with ups and downs, and the other third will head steadily down the drinking and problem scales. With earlier identification, more people will be seen who are damaging their health but who are not addicted and who are able to remain at work. This group can be helped before serious social, employment or health problems arise. So there may be a case for some routine screening procedures such as questionnaires, γ-GT and MCV tests in order to identify these people as soon as possible.

Further help

Many courses are now available both about the diagnosis and about the treatment of problem drinking. Health professionals can also improve their skills by visiting good alcohol units and working with their colleagues there. It may be particularly useful for doctors and nurses to attend assessment-type interviews when people with suspected or actual drinking problems are interviewed by experienced colleagues. Experienced colleagues and alcohol counsellors can also give advice both about further help, about clinical problems and about producing alcohol programmes and policies in an occupational setting.

Acknowledgements

I am grateful to: Norman Stott, Honorary Editor of the *Journal of the Society of Occupational Medicine* for permission to use material from an article which I wrote (*J. Soc. Occup. Med.* 1982; **32:**171–79); Marcus Grant, Brendan Walsh and the World Health Organization for allowing me to quote the figures for world production and consumption per head of alcohol which appear in *Public Health Implications of Alcohol Production and Trade*, Geneva: World Health Organization, 1985; David Coggon and Chris Martyn for useful discussions and help while preparing this chapter; and Marian Elliott for typing this chapter and for her help with the book.

Appendix

Comparative equivalent blood, urine and breath alcohol levels

Blood mg/100 ml	Urine mg/100 ml	Breath mg/100 ml
80	107	35
83	110	36
85	113	37
87	116	38
90	119	39
92	122	40
94	125	41
97	129	42
99	132	43
101	135	44
103	138	45
106	141	46
108	144	47
110	147	48
113	150	49
115	153	50
117	156	51
120	159	52
122	162	53
124	165	54
126	168	55
129	171	56
131	174	57
133	177	58
136	181	59
138	184	60
140	187	61
143	190	62
145	193	63
147	196	64
149	199	65
152	202	66
154	205	67
156	208	68
159	211	69
161	214	70
163	217	71
166	220	72
168	223	73
170	226	74
172	230	75
175	233	76
178	236	77
180	239	78

REFERENCES

1. Walsh B and Grant M. *Public Health Implications of Alcohol Production and Trade*. Geneva: WHO, 1985.
2. Davis PMJ, Partridge JW and Storrs CN. *Arch. Dis. Child.* 1982; **57**:940–43.
3. Woolan DMH. *J. R. Coll. Physicians Lond.* 1980; **14**:213–17.
4. Moser J (ed.). *Alcohol Policies in National Health and Development Planning*. Geneva: WHO, 1985.
5. Rootman I and Moser J. *Guidelines for Investigating Alcohol Problems and Developing Appropriate Responses*. Geneva: WHO, 1984.
6. Wilkins RH. *The Hidden Alcoholic in General Practice*. London: Elek Science, 1974.
7. *Lancet* 1981; **i**:762–63.
8. Saunders JB, Beevers DG and Paton A. *Lancet* 1981; **ii**:653–56.
9. Thornton JR. Atrial fibrillation in healthy non-alcoholic people after binge drinking. *Lancet* 1984; **ii**:1013–14.
10. *Lancet* 1983; **ii**:660–61.
11. Gill JS, Zezulka AV and Beevers DG. *Br. Med. J.* 1985; **291**:1645.
12. Stott DJ et al. *Br. Med. J.* 1985; **291**:1645.
13. Wilkins MR and Kendall MJ. *Br. Med. J.* 1985; **291**:1342.
14. De Keyser J, Michotte A and Ebinger G. Television induced seizures in alcoholics. *Br. Med. J.* 1984; **289**:1191–92.
15. Homaidan FR, Kricka LJ and Whitehead TP. *Lancet* 1984; **i**:913–14.
16. Hargreaves T. *Health Trends* 1985; **17**:78–80.
17. Van Waes L and Lieber CS. *Br. Med. J.* 1977; **ii**:1508–10.
18. *Lancet* 1983; **ii**:397.
19. *Br. Med. J.* 1978; **i**:1504–5.
20. Bernadt MW, Mumford J and Taylor C. *Lancet* 1982; **i**:325.
21. Pokorny AD, Miller BA and Kaglan HB. *Am. J. Psychiat.* 1972; **129**:118–21.
22. Bissel D, Paton A and Ritson B. *Br. Med. J.* 1982; **284**:495–97.
23. Edwards G, Oxford J and Egert S. *J. Stud. Alcohol* 1977; **38**:1004.

USEFUL FURTHER READING

Scottish Health Education Coordinating Committee. *Health Education in the Prevention of Alcohol-related Problems*. Edinburgh: Scottish Home and Health Department, 1985.

Report of a Special Committee of the Royal College of Psychiatrists. *Alcohol and Alcoholism*. London: Tavistock Publications, 1979.

Edwards G. *The Treatment of Drinking Problems*. London: Grant McIntyre, 1982.

Sherlock S. *Alcohol and Disease*. London: Churchill Livingstone, 1982.

Kessel N and Walton H. *Alcoholism*. Harmondsworth: Penguin, 1969.

Gardner AW. *How to Prevent Absenteeism*. Southampton: Applecross Books, 1983.

Paton A and Smith R. *Alcohol Problems*. London: *British Medical Journal* 1982.

Heather N and Robertson I. *Problem Drinking—the New Approach*. Harmondsworth: Penguin, 1985.

10. ACIDIC DEPOSITION

D. W. Gardner

INTRODUCTION

Acidic deposition, or acid rain, has been implicated as the cause of environmental damage in both Europe and North America and as such has generated considerable public concern. Adverse effects attributed to acid rain are losses of fish from lakes and rivers in Norway, Sweden and parts of the north-eastern United States, the decline of forests in central and southern Germany, and acid damage to soils and buildings. Classically, these adverse effects were attributed to man-made emissions of sulphur dioxide (SO_2) and nitrogen oxides (NO_x) which were converted to sulphuric and nitric acids in the atmosphere and eventually deposited into acid-sensitive ecosystems by rainfall. However, none of the effects ascribed to acid rain has been proved unequivocally to be due solely to acid rain, but some of these effects may be caused by a complex of factors which can collectively be called 'acidic deposition.'

Acidic or atmospheric deposition therefore, includes the dry deposition of the gases SO_2 and NO_x themselves or of their reaction products, principally ozone (O_3) and aerosols, and the wet deposition in rain, snow and fog of sulphates and nitrates which may or may not be associated with acidity. The exact chemical form of acid deposition is controlled by a variety of atmospheric and environmental factors, and the susceptibility of the receptors, plants, soils, etc., is similarly controlled by a multitude of environmental and man-made factors. Many of these processes are still poorly understood because of their extreme complexity and much of the current controversy over acid rain has arisen from a number of erroneous assumptions about the chemistry of rain itself.

The principal error in assessing the potential effects of acid rain on the ecosystem has been to equate the presence of sulphate and nitrate ions in rainfall as being exactly equivalent to the presence of sulphuric and nitric acids. This assumption is untenable, and in most studies concerning the effects of acid rain, many researchers have used 'worst-case scenarios' for predicting the effects of acid rain. Plants, soils, fish and aquatic ecosystems have been subjected to solutions of sulphuric and nitric acids with pHs as low as 2·0—with understandably deleterious effects. These results in highly artificial laboratory experiments have, unfortunately, often been used to make predictions about the effects of acid rain on the natural ecosystem. Such experiments make unwarranted assumptions about the applicability of simple laboratory experiments as models for the real world.

Having stated what acid rain is not, that is, a mixture of sulphuric and nitric acids, the question arises, what are the chemical constituents of rainfall? This question is easier to answer by considering how pollutants such as SO_2 and NO_x move through the atmosphere, how they are chemically transformed and how they are deposited to the surface of the earth as acidic deposition, of which acidic precipitation, or 'acid rain' is but one type.

SOURCES OF POLLUTION

The principal source of man-derived SO_2 in the atmosphere is power plants which burn sulphur-containing coal, oil or lignite. Power plants account for 80 per cent of the man-made total, and in rural areas approximately 90 per cent of the SO_2 at ground level can be attributed to man's activity.[1] Natural processes such as decay in oceans, marshes and soils, and volcanic activity account for the remainder. By contrast, in rural areas, only 50 per cent of the NO_x at ground level is man-derived. Natural processes, such as the denitrification of soils by bacterial action, account for the other half. Of the man-made total, half of the NO_x comes from power plants and half from automobile exhausts. In urban areas, however, the man-derived portion of the total NO_x in the atmosphere can be in excess of 90 per cent, and on a world-wide basis the ratio of man's emissions of SO_2 to NO_x is approximately 3 : 2.

Power plants and automobile exhaust therefore, are the principal sources for man-made SO_2 and NO_x. Densely populated industrial regions such as the Ruhr Valley in Germany and the Ohio River Valley in the United States are major sources of SO_2 and NO_x. Pollutants from these regions can be carried hundreds of kilometres before being deposited, and in Norway for example, over 90 per cent of the sulphate deposited within its borders originated in some other European country.[2] SO_2 and NO_x emitted in one region or country therefore, can spend from 1 to 5 days moving through the atmosphere, being chemically transformed or forming secondary reaction products before reaching the ground as acidic deposition at sites remote from their point of origin.

POLLUTANT DISPERSAL

When SO_2 and NO_x are released, the physical and chemical conditions of the atmosphere determine where and in what form these pollutants are dispersed and deposited.

Major wind patterns associated with the movement of high and low pressure systems are a key factor in the dispersal of pollutants because these systems are capable of carrying SO_2 and NO_x over long distances. Wind speed and direction at the source of pollution therefore, do not usually indicate where the eventual receptors will lie because in the northern hemisphere, winds move counter-clockwise around low and clockwise around high pressure systems. Thus, in many instances, pollutants are carried on curved trajectories from their source to their receptors.

Another factor which affects how effectively pollutants are dispersed is the variation in air temperature with altitude. During the day, the ground warms the adjacent air and thus the temperature of the air decreases with height. This condition allows pollutants such as SO_2 and NO_x to disperse both vertically and horizontally. In addition, when air near the ground is warmed during the day, the resulting turbulence enhances pollutant mixing. Thus, when stable high pressure systems move slowly over populous industrial regions, SO_2 and NO_x are entrained, mix and can be carried by such weather systems over long distances. In many instances, these high pressure systems encounter a low, a frontal system forms and rain results.

Meteorological conditions, however, are extremely variable—the predominance of high and low pressure systems varies seasonally, air temperature varies diurnally and wind speed and direction can vary hourly. Thus, many physical factors in the atmosphere affect the path of pollutants from the source to the receptors. This variability is reflected in the extreme variability of the chemistry of the rainfall at a specific receptor site.

CHEMICAL TRANSFORMATIONS

The chemical process of transformation of SO_2 and NO_x can be divided into two basic types of reaction:

— gas-to-gas reactions
— reactions in which gases combine with liquid droplets or solid particles to initiate rain droplet formation.

Gas-to-gas reactions

The gas phase reactions of SO_2 and NO_x are primarily oxidative and involve free radicals (denoted by the symbol ·) which contain an unpaired electron. These are extremely reactive. The principal type of free radical in the atmosphere is the hydroxyl radical, OH· which reacts with SO_2 and NO_x to form an acidic aerosol. These acid aerosols are initially a mixture of sulphuric and nitric acids and exist as a small solid or liquid particle which ranges in size from a few times larger than a gas molecule to about 0·1 μm in diameter. The production of acid aerosols is favoured by the polluted atmospheres of populous industrial regions which are enriched in NO_x and in hydrocarbons, both of which promote the production of oxidizing radicals. The major source for hydrocarbons in the urban atmosphere is unburned gasoline released by automobile exhausts and by evaporation from petrol-tanks and carburettors. Emissions from oil refineries are secondary. Hydroxyl radicals react with SO_2 as follows:

$$OH\cdot + SO_2 \rightarrow HSO_3\cdot \quad (1)$$
$$HSO_3\cdot + O_2 \rightarrow HSO_5\cdot \quad (2)$$
$$HSO_5\cdot + NO \rightarrow HSO_4\cdot + NO_2 \quad (3)$$
$$HSO_4\cdot + NO_2 + H_2O \rightarrow H_2SO_4 + HNO_3 \quad (4)$$

The production of a strongly acid aerosol in the urban atmosphere is clearly dependent upon the supply of SO_2 and NO_x (that is, NO and NO_2), but the rate-limiting step is the concentration of hydroxyl (and other) radicals in the atmosphere. Hydroxyl radical production depends primarily on the presence of a complex mix of precursors both natural and man-made and upon the presence of sunlight which drives many of these reactions. Thus, conditions which favour the more rapid oxidation of SO_2 to an acidic aerosol, are an urban atmosphere and high solar radiation. These conditions are found during the daytime, in summer and near heavily industrialized cities. Under such conditions SO_2 can be converted to an acid aerosol at a rate of approximately 0·5 to 5 per cent per hour.[3] This rate is significantly lower at night, in the wintertime, and in unpolluted rural areas.

Nitrogen oxides, NO_x, are emitted primarily as nitric oxide (NO) with some nitrous oxide (N_2O). Nitric oxide is oxidized in the atmosphere to nitrogen dioxide (NO_2). Both nitric oxide and nitrogen dioxide are converted to acidic aerosols via reactions 3 and 4. Nitrogen dioxide also participates in an important series of reactions which lead to the production of ozone (O_3). Conditions which favour the production of acidic aerosols from SO_2, also favour the production of O_3 from NO_2, that is, an urban atmosphere, during the day in the summer. The production of O_3, however, is more directly dependent upon solar radiation which causes the photoconversion of NO_2 to O_3.

$$NO_2 + \text{light energy} \rightarrow NO + O\cdot \quad (5)$$
$$O\cdot + O_2 \rightarrow O_3 \quad (6)$$

Ozone, however, also reacts with nitric oxide in a back-reaction to form nitrogen dioxide.

$$O_3 + NO \rightarrow NO_2 + O_2 \qquad (7)$$

Under constant conditions, the reaction would reach the following equilibrium:

$$NO_2 + O_2 \rightleftharpoons NO + O_3 \qquad (8)$$

However, in polluted urban air, hydrocarbons react continuously with NO, and because NO is removed from the atmosphere, the equilibrium of the reaction shifts to the right and ozone accumulates. The conditions which favour the production of ozone, that is high light intensities, also favour its destruction.

$$O_3 + \text{light energy} \rightarrow O_2 + O\cdot \qquad (9)$$

This is a very important reaction because the excited oxygen atom O· reacts with water vapour in the atmosphere to produce excited hydroxyl radicals.

$$O\cdot + H_2O \rightarrow 2\,OH\cdot \qquad (10)$$

The hydroxyl radicals produced are now available to oxidize two molecules of SO_2 (via reaction 1) and so initiate another cycle of reactions.

In summary, these reactions convert the gases SO_2 and NO_x to small solid or liquid particles, that is, to aerosols. Initially these aerosols are only slightly larger than the gas molcules SO_2 and NO_x, but they grow rapidly in size as water vapour condenses on their surface. This leads to a second important series of reactions which occur as gases and other types of aerosol react together in the liquid layer of the particles, leading to a further increase in size.

Gas-to-particle reactions

Once a liquid aerosol is formed, SO_2 is able to dissolve in the water layer where a further conversion to sulphate can occur. Many factors govern the rate of conversion and the final chemical form of the aerosol, but again ozone and hydrogen peroxide, produced by oxidative reactions in the atmosphere, convert dissolved SO_2 to sulphuric acid (H_2SO_4). This process may be catalysed by trace metals such as manganese and perhaps iron which are present in fossil fuels and which are emitted with SO_2 and NO_x upon combustion.

At this stage, in many instances, the air mass into which the pollutants were originally emitted has moved away from the industrial sources into the rural environment where it encounters elevated concentrations of ammonia (NH_3), which is produced by decay processes in the soil and as a result of man's use of ammonia-containing fertilizers. NH_3 is absorbed by the growing aerosol and a partial or almost complete neutralization of the sulphuric and nitric acids present occurs with the formation of ammonium sulphate and ammonium nitrate. At this stage further neutralization may occur because the aerosols produced from the conversion of SO_2 and NO_x are large enough to interact physically and chemically with another class of aerosol present in the atmosphere, namely those in the size range of 1·0 μm and larger in diameter which are produced by a variety of natural processes. When wind passes over the sea or land these aerosols are entrained and carried aloft, and in areas where intense agricultural activity disturbs the soil and/or leaves large areas of bare soil exposed, a significant portion of the total atmospheric burden of aerosol can be produced naturally. This type of aerosol is rich in alkaline components which reflects its terrestrial or marine origin. For example, winds passing over the sea entrain sodium chloride and magnesium sulphate aerosols in particular, and winds passing over land entrain aerosols rich in magnesium and calcium carbonates and sulphates, in addition to smaller quantities of salts of aluminium and silicon. The presence in the atmosphere of these alkaline aerosols can lead to a further neutralization of sulphuric and nitric acids formed from SO_2 and NO_x.

With the formation of an aerosol which is composed largely of ammonium sulphate and ammonium nitrate with smaller amounts of other ions such as magnesium, calcium and aluminium, and in coastal areas, sodium and chloride ions, the SO_2 and NO_x have reached a chemical end point and may be removed by rainfall. These aerosols, however, can be removed from the atmosphere by rainfall at any stage of their development and this accounts for some of the chemical variability found in rain. However, in many parts of the northern hemisphere the weather is dry more often than it rains and hence the dry deposition of gases and aerosols may be a more important process than their removal in rainfall.

Thus, close to sources of SO_2 and NO_x, some of the SO_2 will be deposited directly onto the ground as gaseous SO_2. NO_x is rapidly converted to NO_2 and this again is directly deposited to the ground as gaseous NO_2, but at a higher rate than SO_2. In addition, NO_2 is rapidly converted to nitric acid aerosols and these can be deposited very rapidly onto the ground. At this early stage, very light rains may produce a more acid rainfall in which nitrates predominate over sulphates. This is, however, a fairly rare occurence in most parts of the northern hemisphere—less than 5 per cent of all rains. Higher rain volumes will cause dilution and a rainfall having a higher pH. Further from pollution sources, gaseous SO_2 and NO_2 are increasingly converted to sulphates and nitrates, and interaction with ammonium and other alkaline aerosols produces an aerosol which is predominantly composed of ammonium sulphate and ammonium nitrate. These aerosols are themselves deposited directly to the ground during dry periods or may be removed by rainfall, and again, higher rain volumes result in rains having a higher pH.

RAIN CHEMISTRY

Acid rain, classically, has been defined as any rainfall having a pH lower than 5·65 which is the pH of carbonic acid formed by the reaction of water and carbon dioxide at normal atmospheric concentration. This definition is however artificial because of the assumption that in an atmosphere unpolluted by man, rainfall would consist only of carbonic acid. As indicated previously, there are natural sources for the majority of ions in rainfall, and a number of natural sources for the acidity found in rain. Recent studies[4,5] suggest that the pH of rain in an unpolluted atmosphere could be as low as 4.5 and that about half of this acidity is produced by weak acids, and not only by strong acids such as sulphuric and nitric acids.

In the past, nearly all of the data concerning the chemistry and the acidity of rain have been collected using rain samples which have undergone considerable chemical change prior to analysis. Unfortunately, one of the major changes is a signficant increase in the acidity of many of the samples. These collection problems have led to an over-estimation of the actual amount of acidity present in rain and a significant under-estimation of the role which other ions, such as NH_4 and volatile weak acids, play in the chemistry of rain. Studies using rain samples where chemical changes prior to analysis are minimized, have shown that SO_4 and NO_3 in rainfall is not usually present as H_2SO_4 and HNO_3. Much of the SO_4 and NO_3 in rainfall is accounted for by sulphate- and nitrate-containing salts of ammonia, and to a lesser degree, of magnesium, sodium and calcium.

EFFECTS

Soils

Many of the studies which predict adverse effects on soils have used mixtures of sulphuric and nitric acids to simulate rainfall and have used rainfall rates that are often several hundred-fold higher than those which occur naturally. Potential effects predicted from this type of experiment are a depletion of the alkaline components in soil which may lead to further acidic inputs to acid-sensitive aquatic systems, and the mobilization of potentially toxic metal ions, such as aluminium, which are released when strong acids are added to soils. Such experiments often conclude using the most pessimistic assumptions, that damage of this type might take from 100 to 500 years to develop. Thus, the majority of soils contain sufficient quantities of calcium, magnesium, potassium salts and clay to neutralize the minimal quantities of acid added by acidic deposition from the atmosphere. Under natural conditions, therefore, the majority of soils are very well buffered.

Acidic inputs by rainfall to soils in temperate regions, however, are negligible because the majority of soils are already acidic. Soils in many parts of Europe and in the eastern half of the United States are acidic and are becoming more acidic due predominantly to the natural processes which produce, modify and alter the soil environment. In natural soils the pH of the soil depends upon the underlying bedrock, the chemistry of the soil and upon the type of vegetation growing in the soil. However, in most soils the predominant factors which regulate soil acidity are the production of carbon dioxide and hence carbonic acid from both microbial activity and root respiration, and from the production of humic and fulvic acids by micro-organisms which degrade plant litter. Decay processes are particularly important in heathland vegetation and in pine and spruce forests where the degradation of fallen needles can produce large quantities of natural organic acids. Other natural processes which lead to soil acidification are the oxidation of iron sulphides to sulphuric acid, and the release of hydrogen ions by micro-organisms in the soil which utilize ammonium ions and oxygen as a source of energy. The reaction is as follows:

$$NH_4^+ + 2O_2 \rightarrow 2H^+ + NO_3 + H_2O \qquad (11)$$

When ammonium-containing fertilizers are added to soils, the majority of the ammonium ions are utilized by micro-organisms which take-up one ammonium ion from the soil and release two hydrogen ions in exchange. The remainder of the ammonium ions are absorbed by plant roots which in exchange release one hydrogen ion to the soil. Agricultural soils, therefore, are gradually acidified as plants and micro-organisms release hydrogen ions to the soil in exchange for ammonium ions added in fertilizers. This process is the predominant source of acidity in agricultural soils. In addition, soils are made alkaline by the addition of lime and other agricultural chemicals. Thus, in agricultural soils, acidic imputs from the atmosphere are inconsequential.

As a final note, in many soils the availability of nitrogen is the factor which most limits plant growth, and in marginal soils there is some evidence that inputs of nitrogen from rainfall may in fact have beneficial consequences.

Aquatic ecosystems

Reductions in fish populations in a small percentage of lakes, streams and rivers in southern Sweden and Norway, and in the Adirondack Mountains in the north-eastern United States have been attributed to acidic inputs derived from acid rain. Classically, sulphuric and nitric acids present in snow were thought to accumulate during the winter months and to melt in the spring releasing a sudden flood of strong acid to aquatic ecosystems. These acids were thought to mobilize aluminium ions from the soils, and the combined acidity and elevated aluminium ion concentrations were thought to cause extensive mortality of fish fry which hatch in the spring.

There is good evidence that in these areas in the spring, the pH of surface water entering these lakes and streams is more acid than at other times of the year, and that this is accompanied by elevated concentrations of aluminium ions. Also, there is good evidence that fish fry are particularly sensitive to increased concentrations of hydrogen and aluminium ions. However, these ecosystems are all in regions where temperate coniferous forests are the dominant type of vegetation, and as outlined previously, the soils in these areas are naturally acidic. Recent evidence,[6] therefore, suggests that much of the observed acidity and the resultant increase in aluminium ion concentrations, is caused primarily by extensive leaching of the natural organic acids present in soil. In the spring, snow-melt and in many instances a seasonal increase in rainfall volume, cause the water to run-off through superficial, organic acid-containing layers of the soil, rather than entering lakes and streams by percolation through the bedrock which occurs during the rest of the year when the volume of water entering the soil is much reduced. It seems, therefore, that much of the acidity observed in spring in lakes and streams arises as a result of natural soil and environmental processes, and that atmospheric inputs of acidity due to man's activities are minimal.

Forests

In the early 1970s in central and southern Germany, a number of tree species were noted to be in decline. Symptoms such as loss of needles or leaves in silver fir, Norway spruce and beech were observed. Other symptoms in conifers were a yellowing of the needles, particularly on the upper side, a reduction in growth of the crown of the tree and a decrease in root growth. In addition, the most severe symptoms were observed both at higher elevations—that is at altitudes greater than 500 m—and at rural sites remote from heavily industrialized regions. Initially, acid rain was thought to be solely responsible for the forest decline, but more recent evidence[7] suggests that ozone is probably the primary cause. Several lines of evidence point to this conclusion. Peak ozone concentrations in excess of 500 $\mu g\ m^{-3}$ have been measured in areas where the forest decline is most severe. Ozone concentrations in this range have caused well-documented reductions in the productivity of coniferous species in the San Bernardino Mountains of California. In addition, species known to be sensitive to ozone, such as tobacco, have yielded the classic symptoms of ozone injury when planted in areas where forest decline is occuring. Additional evidence comes both from studies which show that ozone concentrations generally increase with altitude and from the known increase of 50 per cent in NO_x emissions which have occurred in Germany during the last 15 years. NO_x is an essential precursor for the formation of ozone. Although ozone appears to be the primary cause of forest decline, acidic deposition either in the wet or dry form, appears to play a critical role in symptom development, particularly in the yellowing of needles. Acidic fogs or mists, or non-acid fogs and mists which dissolve acids previously deposited on the needles, cause the loss of magnesium ions from needle tissues. This loss appears to be promoted by sunlight, hence symptoms are most severe on the upper surfaces. Magnesium is a key component of chlorophyll and when chlorophyll is inactivated by the loss of magnesium, the metabolic fitness of the plant is reduced. This renders it more susceptible to injury by ozone.

Building materials

Evidence that acid rain alone causes damage to building materials is lacking, but numerous studies have shown that the dry deposition of SO_2 causes corrosive degradation of zinc, iron and stone—particularly limestone, sandstone and marble. Such studies have demonstrated a linear dose-response relationship between SO_2 and weight losses in metals and stone. However, the presence of moisture and of catalysts such as soot, iron, manganese or sulphate ions, is important in converting SO_2 to the sulphuric acid which actually causes the corrosion.

Human health

The effects of acid rain on human health lie in the region of potential concerns rather than proven facts. The primary health concerns centre on the possibility that humans may be exposed to increased levels of heavy metals both in drinking water and by consuming fish. In areas such as Sweden, the Adirondack Mountains and other areas where the groundwater is acidic, these concerns are clearly valid. The relationship between acidic groundwater and acidic deposition has not, however, been established. Humans, therefore, may be exposed to increased levels of lead and copper which are leached from lead and copper plumbing by acidic groundwaters, and to increased levels of methyl mercury by consuming freshwater fish such as pike, perch and trout. In acidic waters methyl mercury is concentrated in the food chain and reach the highest concentration in fish which lie at the end of the chain.[8]

FUTURE TRENDS

I believe that in the future, public concern over acid rain will decline because its acidity and its possible effects upon the environment have been over-estimated. The phenomenon of acidic deposition, however, warrants continued public interest because the exact implications of the continuing deposition of man-derived SO_2 and NO_x have not been fully determined. More attention is likely to be focused in the future on the role of NO_x in the environment because the current increasing trend in its emissions is likely to continue. In addition, NO_x is an important precursor for the formation of ozone which at relatively low concentrations can induce economically significant losses in crop and forest productivity. The importance of acid rain as a topic of public concern in the future is therefore, likely to decrease.

REFERENCES

1. Stern AC. *Air Pollution*. New York: Academic Press, 1976.
2. Environmental Resources Limited. *Acid Rain: a Review of the Phenomenon in the EEC and Europe*. London: Graham and Trotman, 1983.
3. National Research Council. *Acidic Deposition: Atmospheric Processes in Eastern North America*. Washington, DC: National Academy Press, 1983.
4. Galloway JN and Gaudry A. The composition of precipitation on Amsterdam Island, Indian Ocean. *Atmos. Environ.* 1984; **18**:2649–2656.
5. Keene WC, Galloway JN and Holden JD Jr. Measurement of weak organic acidity in precipitation from remote areas of the world. *J. Geophys. Res.* 1983; **87**:8771–8786.
6. Lefohn AS and Klock GO. The possible importance of forest soil processes in defining surface water pH depressions. *J. Air Pollution Contr. Assoc.* 1985; **35**:632–637.
7. Prinz B, Krause GHM and Jung KD. In *Nato Advanced Workshop. Effects of Acidic Deposition and Air Pollution on Forests, Wetlands and Agricultural Ecosystems*. May 13–17 1985, Toronto: (in press).
8. United States Environmental Protection Agency. *The Acidic Deposition Phenomenon and its Effects: Critical Review Papers*. Washington, DC: United States Environmental Protection Agency, 1983.

11. TOXICOLOGY, ITS APPLICATION AND INTERNATIONAL TOXICOLOGY PROGRAMMES

Virginia Murray

Society's widespread anxiety about potential chemical disasters and its awareness of the ever increasing use of chemicals throughout the world has led to the recent rapid development of the science of toxicology in reponse to the need for greater understanding and control. Toxicology is the study of the adverse and potentially adverse effects of chemicals which have, or may have, the capacity to cause injury to living organisms. An examination of the roles and methods employed within the science may clarify the functions leading to greater understanding, and enable the information generated to be interpreted more usefully.

When considering the applications of toxicology, it is necessary to have an appreciation of the many and varied roles of the professionals in the field and their differing methods, activities and contributions. Toxicologists come from many different areas of fundamental science including biology, chemistry, pathology, medicine, statistics and epidemiology, thus providing many useful and varied skills and perspectives. In their careers,[1] these scientists pursue the following main areas of specialization:

The *experimental toxicologist* is particularly concerned with chemical exposure of animals in controlled laboratory conditions. This yields, amongst other information, valuable data for the decision makers and, in circumstances where little human toxicology data is available, their research can provide the only basis for assessment of a chemical. They carry out various well documented routine toxicity tests[2,3,4] which are designed to show the overall degree of effect from exposure to a substance. The most commonly quoted acute toxicity tests is the somewhat unsatisfactory Lethal Dose 50 (LD50), an acute toxicity measurement by which a known dose of a chemical results in death of 50 per cent of the dosed group within 7–10 days. Other routine tests include chronic long-term exposures which provide information on carcinogenicity and mutagenicity, and may involve special tests for skin irritation and sensitization, eye contact and reproduction studies. Animals, even inter-species, have varied metabolic pathways and different susceptibility to the effects of a substance and this leads to a noticeable inconsistency of many of the results obtained from these tests. The antivivisection view is widely held, and given the unreliability and cost of many of the tests, it is understandably held by much of the non-scientific group of the population. However, the exposure of experimental animals to toxic agents is justified as a means of gaining information on possible hazards, especially where there is no opportunity to gather human toxicological data or where a new chemical needs to be examined prior to introduction into the community.

The diagnosis and treatment of the poisoned patient are carried out by the *clinical toxicologist* who, working within the obvious limitations of human ethics, endeavours to gather and evaluate a similar depth of information as his experimental colleague. The advantage is that his researches are of more immediate relevance to the human condition. Acute poisoning occurs relatively rarely in developed countries and chronic poisoning may frequently mimic other human pathology; the collection and interpretation of these data are particularly challenging. A high level of training is therefore necessary to help the clinical toxicologist to draw more valid conclusions from the limited data at his disposal.

The *analytical toxicologist* carries out qualitative and quantitative analysis on human and, more rarely, on animal samples. These results can provide the means to evaluate the degree of exposure to a chemical and, where possible, supply information on toxicokinetics and toxicodynamics to aid the understanding of the pattern of distribution, metabolism and excretion of a substance. It is not appropriate here to discuss in detail the methods used in analytical toxicology but some of the equipment used includes ultraviolet spectrophotometry, spectrophotofluorimetry, immunoassay, high performance liquid chromatography, atomic absorption spectrometry and mass spectrophotometry.

The practical application by the *occupational toxicologist* of toxicological data enables him to recognize better the risk and the potential hazards of chemicals, and to seek methods of risk reduction and containment in their industrial and social context. During the continuous process of monitoring the industrial environment and its workers, the occupational toxicologist gathers thorough, detailed and principally chronic human toxicological data, which in its turn, should provide a valuable contribution to the overview of a chemical's toxicity.[5]

The *forensic toxicologist* is concerned with the medicolegal aspects of the harmful effects of chemicals on humans and animals. He usually works as part of a team in homicide investigations and depends, to a great extent, on data collected from other toxicological research and from analytical techniques, in order to understand post-mortem findings more clearly.

It is important to be aware of the role of the *environmental toxicologists* who are concerned with the occurrence of chemicals and pollutants in air, soil and water and also in food additives and contaminants. Their continuous monitoring and reassessment of chemicals in the more botanical context mirrors the endeavours of the whole toxicological field, enabling the regulatory toxicologists to reach their decisions with the benefit of as much of the available data as possible.

The decisions reached by the *regulatory toxicologists* working with governments and legislators are a crucial factor in seeking to reduce risk and limit hazards. It is their role to provide an achievable balance of risk against benefit, and it is therefore valuable that they should be supplied with the best possible information on which to base their judgements.

Acceptable risk decisions need to be considered in the light of the following factors:[6]

—the need met by the substance,
—the adequacy and availability of alternative substances to meet the need identified,
—the anticipated extent of public use,
—employment considerations,
—economic considerations,
—effects on environmental resources
—conservation of natural resources.

THE APPLICATION OF TOXICOLOGY

One of the methods used in the application of toxicology is the consideration of various factors which influence the toxicity of a chemical and its effects on humans. These factors are summarized in *Table 11.1* and relate to the chemical characteristics, the exposure, the occupational environment and the individual's own susceptibility.[7] The hazardous effect from exposure to the chemical is considered and how these effects can be mitigated or prevented.

Whilst considering a *chemical's characteristics*, obviously its chemical structure is of prime importance but its activity may depend on such things as its pH environment. The stability of a chemical and its conditions of storage must be looked at as deterioration may have occurred leading to possible increased toxicity consequent to continuing chemical reactions. The medium or vehicle in which a chemical

Table 11.1. Summary of some of the factors influencing toxicity of chemicals and their effect on humans

Chemical characteristics	
Chemical composition (pH, choice of anion, etc.)	
Physical characteristics (particle size, method of formulation, etc.)	
Presence of impurities or contaminants	
Stability and storage characteristics of the toxic agent	
Solubility of the toxic agent in biological fluids	
Choice of the vehicle	
Presence of excipients: adjuvants, emulsifiers, surfactants, binding agents, coating agents, colouring agents, flavouring agents, preservatives, antioxidants, and other intentional and non-intentional additives	
Exposure	
Dose, concentration and volume of administration	
Route, rate and site of administration	
Duration and frequency of exposure	
Time of administration (time of day, season of the year, etc.)	
Occupational environment	
Temperature and humidity	
Barometric pressure (hyper- and hypobaric effects)	
Ambient atmospheric composition	
Light, noise and other forms of radiation	
Occupational hygiene controls	
Individual susceptibility	
Genetic status (siblings, etc.)	
Immunological status	
Nutritional status (diet factors, state of hydration, etc.)	
Hormonal status (pregnancy, etc.)	
Age, sex, body weight and maturity	
Central nervous system status (activity, etc.)	
Presence of disease or specific organ pathology	
Effects	
Acute	Local
Chronic	Systemic (involving the whole body)
Acute-on-chronic	

is formulated affects its toxicity. For example, a gas is less easy to contain and is therefore more likely to be dangerous than a solid. Contaminants, additives or impurities may also be present and these may be considerably more, or less, toxic than the chemical itself. For instance, adjuvants can enhance the pharmacological or toxic effects of the active ingredient by acting on the same biological receptor and producing a synergistic, additive, effect.

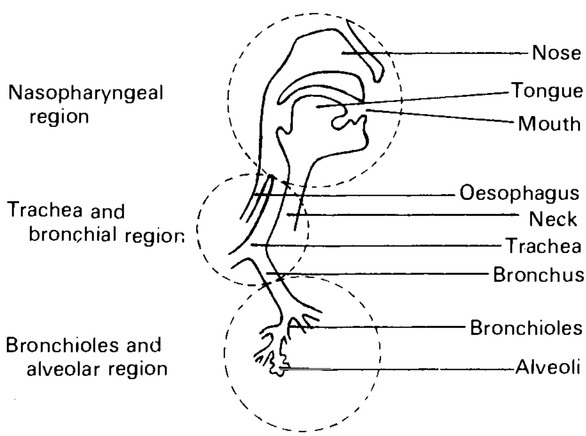
A chemical's physical characteristics include particle size which can both influence the aerodynamic density and the drag, thus resulting in varying degrees of penetration and areas of deposition in the respiratory tree. Suspended aerosols which can be inhaled can be defined as variable-sized particles capable of remaining airborne, and these include some of the following descriptions:

Dusts: caused by mechanical abrasion or fragmentation of solids, e.g. asbestos, silica; size of particles 0·1–100 μm.
Fumes: produced by combustion, sublimation or condensation of volatile solids, metal oxides, e.g. cadmium; size of particle usually less than 0·1 μm, but fumes have a tendency to flocculate and produce larger particles as aerosol ages.
Smokes: suspension of solid particles produced by incomplete combustion of organic materials; size of particle usually less than 0·5 μm and do not settle readily.
Mists and fogs: condensation or uptake of liquid via hydroscopic particles to produce liquid aerosols.

The inhalation of a particle into the respiratory tract presents a changing pattern of risk because of the varying anatomy and physiology of the tree (*Table 11.2*). For instance, the change due to the reduction

Table 11.2. The respiratory tree and the factors influencing the deposition of particles

Anatomical site	Air velocity from breathing	Size of particle deposited	Example of particle deposited
Nasopharyngeal	High	5–30 μm	Smoke
Trachea and bronchial region	Slowing	1–5 μm	Fumes (metal oxides, cadmium)
Alveolar region	Minimal (very slow)	1 μm	Dusts (asbestos, silica)



in size from the wide nasopharyngeal area to the tiny bronchioles and alveoli restrict the deposition of all but the very small particles. The dose received will relate to the amount of substance in the aerosol and the rate of inhalation so that faster, deeper breathing, associated with heavy physical work, will increase the dose received.

The exposure to a chemical and the number, duration and concentration of doses influences the effect of a chemical. The route of exposure affects the absorption of a chemical and these routes include inhalation, ingestion, injection, skin and eye contact. *Fig. 11.1* summarizes the factors influencing exposure effects and shows how the response to exposure may produce an effect after a time interval of minutes or years. A chemical that is quickly absorbed systemically is likely to have an immediate effect while one which has a local site of action and is not systemically absorbed will be slower to show a dose response effect.

The *occupational environment* has a major influence on the dose that an individual can receive. Good ventilation will reduce the risks of exposure to gases, vapours and aerosols although some chemicals, such as methylisocyanate, are so toxic that they should always be contained. A coal mine demonstrates the problems of temperature and humidity and the use of environmental controls and occupational hygiene equipment to reduce dust in machine mining, shows how good mining practices can reduce the incidence of coalworker's pneumoconiosis.[8]

Individual susceptibility shows marked differences of response under conditions of similar exposure.[9] This area is not fully understood but some factors have been shown to have definite roles:

- —the age and sex of an individual have varying effects—for instance, elderly people have slowed metabolic rates making them more susceptible to some exposures such as solvents;
- —pregnancy causes possible risks to the fetus, which may be more susceptible to, for instance, mercury as this crosses the placenta readily;
- —previous illnesses can be significant, such as previous respiratory infections, particularly tuberculosis, and are known to enhance susceptibility to silicosis;
- —obesity is thought to be an important predisposing factor for individuals working with organic solvents and related compounds;
- —liver enzyme activity can alter the speed of some metabolic pathways, and can reduce the risk to young children in some instances where their metabolism produces toxic metabolites at slower rate;
- —alcoholic beverages are known to increase the risk of occupational poisoning by some substances such as organic solvents;

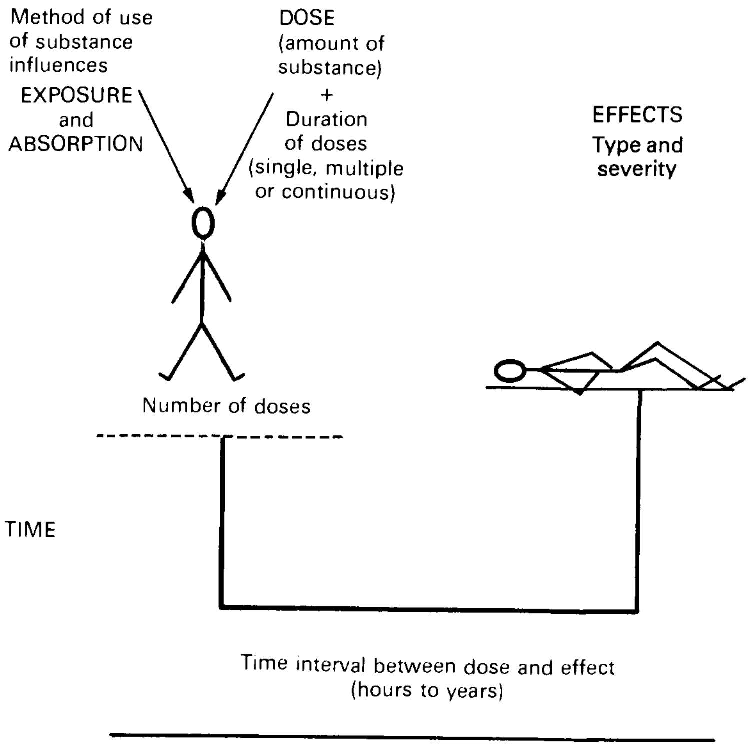


Fig. 11.1. Exposure effects.

—differences in the anatomical structure of noses may produce different degrees of efficacy in filtering out harmful dusts; and
—a high working speed with increased, deep respirations may result in the individual being exposed to a greater dose of aerosols.

The principle effects of chemical exposure are acute illness, continuing chronic illness and cancer. These effects can be summarized as acute, chronic and acute-on-chronic, where an acute illness is precipitated by exposure to a larger dose or to a second chemical insult; for instance, exposure to carbon monoxide on a non-elastic fibrotic lung. These effects reflect the route of absorption and may be local or systemic. The respiratory tract provides examples of the pathological responses to chemical exposure and these include local irritation, cellular oedema and damage, allergic responses, development of fibrosis and pulmonary carcinoma:

—Local irritation to respiratory tree: oedema (swelling) often occurs and a secondary infection or bronchitis can compound the damage; usually an immediate response occurs to exposure; little systemic absorption usually occurs. Example: ammonia (NH_3), used in manufacture of fertilizers and explosive, causes immediate effect on the upper airway resulting in upper and lower respiratory tract irritation and oedema.

—Cellular oedema and damage with increased permeability by the chemical may lead to systemic absorption. Onset of effect may be delayed and the individual may be unaware, at the time of exposure, of the potential danger and the possibility that death might occur if suitable treatment is not given rapidly. Example: phosgene, found in the production of plastics and fertilizers, hydrolyses in the moisture of the respiratory tree to hydrochloric acid and carbon dioxide and can destroy the alveolar cell membrane permeability; the full blown syndrome of severe oedema flooding the lungs can take up to 24–48 hours to develop.

—Allergic response with constriction of airways occurs as a widespread specific antigenic response to the inhalation of usually simple materials. Circulating antibodies or positive skin tests can demonstrate an individual's sensitivity to a chemical. Example: toluene di-isocyanate (TDI) occurs in plastics manufacture and is a very reactive material that can conjugate with proteins. TDI causes a classic decrease in pulmonary function during a working week, where the deterioration is measured by a simple test of breathing: forced expiratory volume in 1 second (FEV_1); reduction in pulmonary function appears to be cumulative over time with repeated exposure causing multiple doses to sensitive individuals. This asthma-like response to TDI can seriously limit effective working. Although it can occur in any worker, it causes most problems in individuals who are already asthmatic and preventive occupational health screening at pre-employment limits the hazard.

—Production of fibrosis can become massive and lead to obliteration of the respiratory capacity of the lung; occasionally local fibrosis can occur limiting the elasticity of the lung and possibly leading to pain from the irritation of the pleural surface. This can result in severely debilitating disease and was recognized as one of the earliest types of occupationally related disease.[10] Example: silica (SiO_2), produced by mining, quarrying and stone cutting, occurs in several different crystalline forms of which tridymite had possibly the greatest fibrogenic potency. Numerous doses without adequate protection from exposure leads, after a time interval of something more than ten years, to the classic debilitating disease of silicosis. Systemic toxicity does not occur and few medical treatments have much influence on the disease process.

—Pulmonary carcinoma is clearly associated with cigarette smoking, and also with exposure to asbestos. Smoking significantly multiplies the risk of development of lung cancer. Example: nickel (Ni) from nickel ore extraction, nickel smelting and electronic electroplating can cause both nasal and lung cancer; nickel exposure may result in changes in DNA (deoxyribonucleic acid) by selective alteration in the DNA replication system; this affects principally the squamous cells which line the nose and lung and thus may produce squamous-cell carcinoma.

When the hazardous effects of a chemical are observed or are expected, a thorough risk assessment* must be carried out with analytical toxicology data where available. A review of the factors which influence a chemical's toxicity and its effects on humans must be undertaken. Reduction of risk may be achieved by altering, perhaps only slightly, one of these factors such as the chemical or its characteristics, the exposure, the occupational environment or the individual susceptibility.

The effect of a *chemical*, such as carbon tetrachloride which causes liver damage, may be so severe that substitution by a safer solvent may be the most effective means of reducing the hazard.

Where the use of a toxic chemical is absolutely essential to a manufacturing process and no alternative chemicals are available for substitution, at a reasonable cost, personal protection may be the only possible method for limiting *exposure*. Health education is an invaluable means of hazard reduction; for instance, the use of protective gloves when handling hydrofluoric acid and its relevant first aid treatment.

Control of the occupational environment, to reduce an individual's exposure to aerosols, can be the most effective means of reducing the incidence and severity of fibrosis but this is usually a very costly means of hazard reduction. Frequently, workers do not cooperate in the use of personal protection so that environmental control is much the most effective method of limiting exposure; air-moving devices can be used to limit exposure to the inhalable fraction of cotton dust and reduce the incidence of byssinosis.

Cigarette smoking has been clearly shown to increase the incidence of an individual's susceptibility to developing lung cancer, especially whilst occupationally exposed to asbestos fibres. Again, active health education against smoking can help to reduce this hazard.

INTERNATIONAL TOXICOLOGY PROGRAMMES

Much of the information needed to understand the toxicity of a chemical and to institute reasonable methods of control is often incomplete.

* Here risk is used as a statistical concept defined as the expected frequency of undesirable effects arising from exposure to pollutant.[11] Estimates of risks may be expressed in relative or absolute terms. The relative risk is the ratio between the risk in the exposed population and the risk in the unexposed population. The absolute risk is the excess risk due to exposure.

International programmes to prevent and reduce hazards from exposure to chemicals present an excellent opportunity to improve chemical information by better data collection and exchange, and evaluation techniques.

Following chemical disasters such as Minamata disease (mercury poisoning in Japan in 1959), polychlorinated biphenyls in animal fodder (USA, 1968), vinyl chloride monomer carcinogenicity (1974) and the Seveso incident (release of dioxin in Italy, 1976), the need to exert some control over chemicals at an international level was recognized. In 1967, the initial step was undertaken by the European Economic Community (EEC) with the issuing of a Directive relating to the classification, packaging and labelling of dangerous substances. With the development of further data on the potential dangers of chemicals to man, this Directive was amended a number of times until, in 1979, a further Directive called for at least some testing of all new chemicals marketed from September 1981. Testing requirements related to the quantities introduced each year. At the same time the Organization for Economic Cooperation and Development and the Council for Mutual Economic Assistance embarked on an examination of test procedures that could be undertaken by different countries.

Other international groups in the United Nations system who have carried out work on the control of chemicals include the United Nations Environment Programme (UNEP), the International Labour Organization (ILO), the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO). Special groups have been developed from these according to their particular responsibility and include the Joint FAO/WHO Expert Committee on Food Additives (1956), the Joint FAO/WHO Meetings on Pesticide Residues (1961), the Joint ILO/WHO Committee on Occupational Health (1969), the International Register of Potentially Toxic Chemicals (IRPTC) and, within this framework, the Global Environmental Monitoring System (1975) and the United Nations Scientific Committee on the effects of Atomic Radiation.

Further increasing concern of member states of WHO, UNEP and ILO resulted in the establishment of the International Programme on Chemical Safety (IPCS) in 1980, in order to provide support to countries in the development and establishment of chemical safety programmes through:[12]

— the evaluation of the risk to human health and the environment from exposure to chemicals;
— the development of methods for assessing health and environmental risks, and exposure hazards from chemicals;
— the promotion of effective international cooperation with respect to emergencies and accidents involving chemicals;

—the encouragement of technical cooperation among countries as regards to the assessment and control of environmental and health hazards of chemicals;
—the promotion of manpower training needed for testing and evaluation of the health and environmental effects of chemicals and the regulatory and other control of chemical hazards;
—the coordination of laboratory testing and of epidemiological studies, and the promotion of research to improve the scientific basis for health and environmental risk assessment and for control of chemical hazards.

IPCS has initiated several series of publications, including the Environmental Health Criteria, which summarize, review and evaluate the available information on the effects of a specific chemical (including, where relevant, impurities, metabolites and degradation products) or a group of chemicals (of the same class or otherwise related), that may influence human health and the environment, and to provide a scientific basis for decision makers. These Criteria documents and their Supplements provide differing levels of information on the diagnosis and treatment for three main categories of information users:

—non-medical professions dealing with accidents involving chemicals, for instance firemen, police and occupational first aid workers;
—medical practitioners, nurses, pharmacists and other paramedical staff not directly associated with a poison centre or clinical toxicological service;
—poison centres and clinical toxicology services responsible for the diagnosis and treatment of poisoning.

IPCS is collaborating with the EEC and WHO Regional Office for Europe (EURO) to spearhead a public health response to acute poisoning: the poison control programme. Poison control programmes provide the framework for both response to and prevention of poisonings by chemicals and are important elements in national chemical safety programmes.[13]

Response to poisoning, on an individual level, requires clinical and analytical diagnosis and treatment and, in the case of a major incident involving chemicals, where many people have been poisoned at the same time by the same agent, coordinated measures should be taken to minimize the overall impact of the accident.

A poison control programme seeks to prevent poisoning by:

—safety measures such as those concerning packaging, design, labelling and handling of hazardous products;
—limited availability or withdrawal of selected toxins;
—information, for instance, on specific populations at risk, occupations at risk, various institutions and organizations;

—health education, for instance, to specific populations at risk including specific occupations.

As a result of these tasks, the elements of a programme are diverse and should include diagnosis of acute and chronic poisoning; care, treatment and rehabilitation; information and documentation; prevention, epidemiology and toxicovigilance (the active observation and evaluation of toxic risks and phenomena in a community with a view to developing and implementing preventive measures); relevant education and training, including clinical toxicology; expert counselling and qualified advice; measures for the response to major incidents involving chemicals; experimental toxicology and research; appraisal and testing of chemicals; and legislation and regulations.

The fulfilment of such a programme requires a number of partners including clinical toxicological centres; analytical toxicology laboratories; medical and paramedical professionals (such as hospital doctors, general practitioners, occupational physicians, coroners and medicolegal experts, psychiatrists and pharmacists); industries, including manufacturers and users of chemicals; experimental toxicologists and research workers; governments and local authorities; the public; and poison centres.

The poison centre must be able to provide, with well-trained staff, information and tailor-made advice about any chemical involved in a poisoning incident on a 24-hour basis; these centres are able to respond to an immediate emergency.[14] They are the pivot of a poison control programme and are one of the primary sources for generation of acute clinical toxicology information. They are also involved in toxicovigilance, antidote evaluation, planning for and providing toxicological elements for response to major incidents involving chemicals, participation in prevention actions, training and dissemination of information.

Many poison centres have been active for twenty-five years but frequently have a poor understanding of occupational toxicology.[15] This reflects the fact that acute occupational poisoning is relatively rare in the developed countries. Poison centres are now receiving enquiries about chronic occupational exposure. Wider availability of occupational toxicologists within poison centres will improve their understanding with occupational health services. It is vital that poison centres, such as the National Poisons Unit in London, be made greater use of by all doctors, but particularly by occupational physicians, as this will improve the quantity and quality of data held by these centres and will thus increase the value of the poison control programme's response to, and prevention of, poisoning.

It is sincerely hoped that, with heightened public awareness and concern, governments and industries, recognizing the public relations advantages, will now become more closely and enthusiastically involved with supporting poison control programmes, which are, after all, to the mutual benefit of the community, its leaders and industry alike. The commitment and support of international organizations must, by now, lend credence to the fact that poison centres are ethically independent and occupy neutral ground between industry and the community; without this independence, the work of the poison centres and their critical evaluation of toxicological data could be brought into question.

REFERENCES

1. Commission of the European Communities. *Manpower Development and Training in Toxicology and Chemical Safety*. Copenhagen: CEC/WHO, 1984 (WHO/EURO Interim Document 18, IPCS Joint Seminar 9, CEC/EUR 9619).
2. Cooke MA. Toxicity testing. In: Ward Gardner A (ed.) *Current Approaches to Occupational Medicine*. Bristol: Wright, 1979; 137–150.
3. Bridges JW and Hubbard SA. Principles, practice, problems and priorities in toxicology. In: Ward Gardner A (ed.) *Current Approaches to Occupational Medicine*. Bristol: Wright, 1982; 251–282.
4. World Health Organization. Environmental Health Criteria. *6. Principles and Methods for evaluating the Toxicity of Chemicals, Part 1*. Geneva: WHO, 1978.
5. Lawreys RR. Occupational toxicology. In: Doull JL, Klaasen CM and Amdur MO (ed.) *Casarett and Doull's Toxicology, the Basic Science of Poisons*, 2nd ed. New York: Macmillan, 1980; 669–709.
6. Klaassen CN and Doull JL. Evaluation of safety: toxicologic evaluation. In: Doull JL, Klaasen CM and Amdur MO (ed.) *Casarett and Doull's Toxicology, the Basic Science of Poisons*, 2nd ed. New York: Macmillan, 1980; 11–27.
7. Doull JL. Factors influencing toxicology. In: Doull JL, Klaasen CN and Amdur MO (ed.) *Casarett and Doull's Toxicology, the Basic Science of Poisons*. 2nd ed. New York: Macmillan, 1980; 56–69.
8. Dick J. Occupational health in coal mining, Part 2: respiratory disease. In: Ward Gardner A (ed.) *Current Approaches to Occupational Health—2*. Bristol: Wright, 1982; 43–61.
9. Durham WF. Toxicology. In: Sax NI (ed.) *Dangerous Properties of Industrial Materials*, 5th ed. New York: van Nostrand Reinhold, 1979; 271–279.
10. Hunter D. *The Diseases of Occupations*, 4th ed. Boston: Little Brown, 1969.
11. United Nations. Preparatory Committee of the United Nations Conference on the Human Environment, third session. New York: United Nations, 1971; 13–24. (A/Conf. 4818 pp. 45–46).
12. World Health Organization/International Programme on Chemical Safety: The International Programme on Chemical Safety, Geneva: WHO.
13. WHO Regional Office for Europe/IPCS/CEC: A working group on public health response to acute poisonings—poison control programmes, Munich, 16–20 December, 1985, in press.
14. World Federation of Associations of Clinical Toxicology Centres and Poison Control Centres/CEC/IPCS: Report of the meeting on the role and responsibilities of Poison Control Centres within a Poison Control Programme, Geneva, 7–9 October 1985, in press.
15. Murray VSG. Report on the ESSO Travelling Fellowship 1984, Faculty of Occupational Medicine Newsletter, 1985.

12. THE OIL INDUSTRY

A. Downie

To the outsider there is a certain mystique about this highly competitive, capital intensive industry which traverses international boundaries and can influence the political decisions of governments. Its history over the past hundred years has been one of continued change to meet market demands and is characterized by the saying 'you never stop building a refinery'. The industry has its own language—a combination of Americanisms and neologisms—and a short glossary of terms in common use is included. A basic introduction to the chemistry of hydrocarbons is followed by an outline of the major processes with their potential exposures. Space precludes a discussion of the associated gas and petrochemical industries.

BASIC CHEMISTRY

Hydrocarbons are organic compounds composed of carbon and hydrogen atoms. Estimations of the number of different hydrocarbons in crude oil vary between 50 000 and over a million. They are classified by carbon number and by molecular structure.

Carbon number

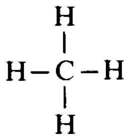
This is based on the number of carbon atoms found in a given hydrocarbon molecule. For example, methane (CH_4) is a C_1, ethane (C_2H_6) and ethylene (C_2H_4) though different compounds with different properties are both C_2. The higher the carbon number, the higher the boiling point, the greater the viscosity and the higher the density. At atmospheric temperature and pressure C_1–C_4 are gases, C_5–C_{17} liquids and C_{17}–C_{40} solids.

Molecular structure

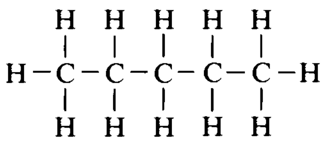
Aliphatic hydrocarbons

1. Paraffins (Alkanes). These contain single bonds (saturated) and have a general formula of C_nH_{2n+2}. They are relatively stable, insoluble in water and the vapour burns in air.

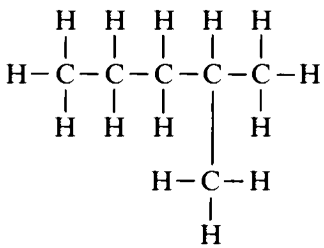
The simplest example is methane CH_4.



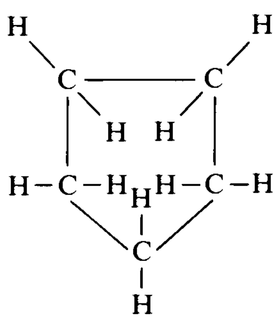
Normal paraffins are straight-chained, for example, n pentane C_5H_{12}.



Paraffins may also be branch-chained, for example, isohexane C_6H_{14}.

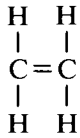


Cycloparaffins have the general formula C_nH_{2n}. An example is cyclopentane C_5H_{10}.



2. Olefins (Alkenes). These contain at least one double bond (unsaturated) and have a general formula of C_nH_{2n}. They are produced

in refining by cracking alkanes and burn to form similar products but with a slightly smoky flame because of the greater proportion of carbon. An example is ethylene C_2H_4.

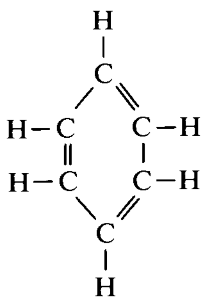


3. Acetylenes (alkynes). These contain a triple bond and have the general formula C_nH_{2n-2}. An example is acetylene C_2H_2.

   $H{-}C{\equiv}C{-}H$

Aromatic hydrocarbons

These are based on the benzene ring with alternate single and double bonds (unsaturated). Aromatic hydrocarbons occur naturally in crude oil and are also produced during refining, especially in catalytic cracking. This group includes the carcinogenic polycyclic aromatic hydrocarbons. The simplest example of an aromatic hydrocarbon is benzene C_6H_6.



Polymerization

This process is the linking together of monomers to form polymers which have a large molecular structure. It takes place only in organic molecules containing double or triple bonds (unsaturated) and is usually dependent on temperature, pressure and the presence of a suitable catalyst. Simple addition polymerization occurs by the removal of double bonds. Condensation polymerization occurs by the loss of a small molecule, often water.

Octane rating

This is a classification of gasoline according to its 'antiknock' qualities. The higher the octane number the greater the antiknock qualities of the gasoline. Knock occurs when the gasoline-air mixture in the cylinder explodes on compression by the piston rather than burns evenly. This causes increased engine wear and inefficient use of fuel. The octane rating is directly related to the molecular structure of the hydrocarbon. Iso-octane is given the standard octane number of 100 and normal heptane the octane number 0. Fuels are rated against these two. A gasoline with an octane rating of 95 has the same knocking characteristics as a mixture of 95 per cent iso-octane and 5 per cent n-heptane. A number of refining processes increase the octane rating of the hydrocarbon by changing its molecular structure. Long straight-chain paraffins have low octane numbers and aromatics have high numbers.

Crude oil

Crude oil and natural gas are formed by the decomposition of organic matter of marine origin and consist of a mixture of thousands of different hydrocarbons with a wide range of boiling points. Crude oil is composed of the heavier constituents which normally occur in liquid form while natural gas comprises the lighter constituents which normally occur in gaseous form.

A commercial reservoir results from a migration of oil and gas through a porous bed of rock due to increases in pressure from a variety of causes. The reservoirs are found mainly in sedimentary rocks (sandstones and carbonates) and are trapped, usually by being overlain by a dense impermeable caprock (shale or similar rock). As the oil and gas permeate the porous water-filled reservoir rock, there is a gradual separation into layers of gas, oil and water and a build-up of pressure due to the trap.

Crude oil may vary from an almost clear liquid to a thick tar-like liquid which can readily solidify in pipelines and contains varying amounts of sulphur, nitrogen and trace metals such as nickel and vanadium. Depending on its constituents, crude oil is subdivided into light, medium or heavy, a classification which affects both the price per barrel and the ultimate end product. A further subdivision into sweet or sour crude (or gas) is made depending on the sulphur content.

EXPLORATION

Oil was first located due to surface seepage through geological faults in the impervious layers of rock. As bitumen it was used as a building material and caulking material in Babylon and ancient Egypt. By the 1800s kerosene found floating on the surface of ponds was being collected and distilled for use in lamps in USA and Russia. In 1859

Edwin Drake, a man experienced in drilling for brine, located oil by drilling to a depth of 21 metres in Titusville, Pennsylvania. Today, exploration relies on geologists who are concerned with rock formation and type and on geophysicists who study the physical properties of the earth's crust such as gravitational, electrical, mechanical and magnetic properties. Satellite photography, seismic, magnetic, gravimetric and stratigraphic analysis all play a part in locating possible oil fields.

The only positive proof of the presence of oil in commercial quantities comes with the drilling of an exploration borehole (wildcat). During drilling, samples of the rock coming from the hole are collected and constantly examined. Speed of drilling, pressure and temperature are also important. From time to time core samples of rock are taken and transported back to the laboratory in stainless-steel vessels under mercury to keep gas in solution.

The sites of location of crude oil vary from arctic to hot desert climates onshore and often beneath deep hostile seas.

HAZARDS OF EXPLORATION

Cold stress

Proper insulated protective clothing is essential to maintain deep body core temperature from falling below 36 °C. It is vital to educate the employee on the safety precautions related to cold, especially wind chill factor.[1] A 10 mph wind at −40 °C is equivalent to a temperature of −70 °C.

Heat stress

This may prove to be a problem in some jobs where impervious clothing has to be worn. The general experience in Saudi Arabia, however, is that it is a very rare condition in oil workers, so long as adequate hydration is maintained and time has been given for acclimatization. Salt supplements are not in general use. The ACGIH gives recommended work/rest schedules for exposure to heat and cold.[2] From personal experience in hot climates, if one were to apply the TLV for heat there are many days in the year when little or no work would be performed outdoors.

A recent unusual hazard encountered by an exploration crew in North Yemen was to be fired upon by hostile tribesmen. Fortunately the intention was to frighten rather than to injure the team.

Drilling rigs both onshore and offshore are noisy, dangerous working areas and even today a driller's length of service may sometimes be estimated by his number of surviving phalanges and how loud one has to speak to converse with him!!

The drilling derrick houses the motor which drives the drill and a crane for lifting the drill pipe in and out of the hole. The cutting bit is weighed by 9-m long steel collars and extended by joining 9-m lengths of steel pipe. As the hole is drilled, it is lined with a steel casing. Cement is forced into the space between the casing and the walls of the hole to prevent leakage.

Drilling mud is pumped down the hole to cool the drilling bit, to remove fragments of rock and to maintain pressure. Mud is a complex mixture of barium sulphate, clay and liquids (fresh water, diesel or crude oil) to which a variety of conditioners are added. The composition varies depending on drilling conditions but it often contains talc, biocides and sometimes formaldehyde. Until 1975 asbestos was a standard hole-cleaning solid suspension additive and there is a suggestion that in some countries it might be re-introduced in a form which the manufacturer claims is 'improved to reduce airborne fibres'. Hydrochloric or hydrofluoric acid, caustic and vegetable fibre may from time to time be added. The platform employee has the potential for exposure to any of the above substances.

While drilling, the safety measures to prevent a blow out include an elaborate system of one-way valves and plugs which expand under pressure. The drilling mud may be thickened but most important is the constant supply of information on drilling rates, depths and the nature of the rock fragments brought up by the mud.

Once a well is in production a permanent well head, fitted with safety valves is cemented in position. This is given the quaint name of a Christmas tree.

Work on a drilling platform has a high physical content. There is a danger on wet surfaces of slipping and the potential of injury from heavy pipes is high. The crew work 12 hour shifts and often experience boredom in their off time. The limited area of a platform contains noisy equipment in the form of pumps, compressors and motors so noise is at a nuisance level and intrudes into the sleeping accommodation. Personal protection in the form of hard hats, gloves, safety shoes and ear defenders require to be worn at all times. The safety of a platform relies heavily on the commitment of the superintendent and when it comes to hearing conservation this is not always high.

The danger of a blow out is not simply the escape of flammable hydrocarbon but the release of hydrogen sulphide. Two factors regarding this highly toxic gas should be remembered, first it is flammable and an explosive hazard and second, it is denser than air. The latter property was forgotten to their cost some years ago by the crew of a drilling platform in the Arabian Gulf who jumped into the sea to escape a blow out and perished in a cloud of hydrogen sulphide.

Before leaving exploration, two hazards of the laboratory are worth a mention. Hydrofluoric acid is used in the palaeontology laboratory to dissolve silica. It is therefore advisable to keep a supply of calcium gluconate gel available for first aid purposes. Recent trials[3] reinforce the value of this topical treatment for skin burns. Eye burns require immediate irrigation with water, normal saline or isotonic magnesium chloride.[4]

Core samples transported under mercury are not in general a problem while being collected. In the laboratory, however, there is a potential for spillage of mercury which can lodge in crevices or in sink traps. Over a period of time this can give rise to the vaporization of significant quantities of mercury into the atmosphere. Precautions include laboratory design which as far as possible excludes crevices where mercury might lodge, regular air monitoring for mercury levels and an adequate clean up of mercury spills. A specific heavy-duty vacuum cleaner is available for this purpose. A decision on whether to perform biological monitoring on the laboratory staff must depend on the potential for exposure.

PRODUCTION AND ITS HAZARDS

Crude oil coming from a well is a high velocity, turbulent, constantly expanding mixture of gases and hydrocarbon liquids, water, solids and other contaminants. Flowing from the hot, high pressure reservoir the stream undergoes continuous pressure and temperature reduction and as it emerges the high velocity gas is carrying liquid droplets while the liquid is carrying gas bubbles. The first plant this stream reaches is the gas oil separation plant (GOSP). Here the stream passes from a narrow diameter collecting pipe to a wide diameter vessel giving a further drop in pressure and a slowing of the flow rate. Gravity separates the vapours and solids from the liquids.

Natural gas is a mixture of hydrocarbon gases and impurities. The hydrocarbons are mainly methane (C1), ethane (C2), propane (C3), butanes (C4), and pentanes (C5), with small amounts of heavier gases. Usually the C3 and above fractions are removed for additional processing because of their commercial value as gasoline blending stock and chemical plant raw feedstock. The C1 and C2 fraction may be used as fuel or as a feedstock for the manufacture of chemicals. The levels of H_2S are high at this stage. Noise levels often exceed 100 dB. Because of high acidity, corrosion inhibitors are used in large amounts and some of these may include chromates. Offshire production platforms are somewhat less dangerous than drilling platforms though they do share many of the drawbacks. Travel to and from the workplace is usually by helicopter giving noise exposure in excess of 100 dB. In addition, there is the potential danger of the helicopter ditching in the sea and, while flotation devices are fitted, helicopter escape training is essential for offshore workers.

Diving teams provide support in construction, maintenance and inspection of offshore facilities. Much of this work is done by saturation teams since this is both economical to the company in number of hours work obtained per diver and safer in respect to the dangers of decompression sickness and barotrauma. In shallow waters to depths of 50–55 m surface-supplied air diving and diving with self-contained breathing apparatus (scuba) are still carried out. Scuba diving even using the buddy system of 2 divers is generally thought to be unsatisfactory in commercial operations. Frequent air diving at depths of 50 m requires constant vigilance on the part of the supervisor to ensure a satisfactory safety performance. As work proceeds to depths of 200 m and below, it seems likely that submersibles and remote control units will gradually replace divers. For details on diving hazards the reader is referred to a standard text such as Edmonds, Lowry and Pennyfather.[5]

TRANSPORTATION AND ITS HAZARDS

Crude oil after having gas and water removed, is transferred by pipeline to storage tanks either at a refinery or at a shipping terminal. Pipelines may be of varying diameter up to about 1 m and are laid on sea beds, below and above ground and many run for thousands of kilometres. There are pumping stations located at intervals of between 100 and 300 km to boost the flow. The constant internal flow of crude keeps the pipe in good condition but externally it must be protected from corrosion. Regular inspections are made using ultrasound and occasionally x-rays.

Health hazards include noise at pump stations and hydrogen sulphide during maintenance. Pipelines are often in trenches about 1 m deep and contain crude *before* desulphurization. This is one of the more common sources of H_2S exposures.

Oil tankships remain a major transporter of crude oil to refineries throughout the world. The rapid growth in the 1970s which saw the supertankers of 30 000 deadweight tonnes (dwt) being replaced by very large crude carriers (VLCCs) of over 250 000 dwt has now almost stopped, but oil tankers still account for a large proportion of the world's merchant shipping.

The hazards of oil tankships are mainly those of the volatile hydrocarbons and fires or explosions due to their ignition from sources such as direct heat (cigarettes and welding), mechanical sparks, electrical equipment and static electricity within the storage tanks. Collisions due to navigational errors are regrettably an all-too-frequent cause of such fires.

Autoignition may occur due to oil being in contact with an overheated pump bearing. In this instance the oil reaches a temperature at which it will burn without an external source of ignition.

Iron scale on the inside of tanks may become impregnated with sulphur from sour crudes forming pyrophoric iron sulphide. Under dry conditions and in high temperatures this scale may glow and provide a source of ignition for flammable vapours.

Inert gas produced from the engine room boilers may be used to reduce the oxygen content of cargo tanks. This usually means replacing most of the oxygen by carbon dioxide. In these oxygen-deficient conditions anyone who enters the area must be protected by breathing apparatus.

Once again hydrogen sulphide may prove to be a hazard in cargoes of sour crude.

REFINING

The refinery is probably the section of the oil industry most familiar to the general public. From the outside it appears to be a collection of tanks, and towers, a maze of piping and a number of chimney stacks! The flares previously burned unwanted hydrocarbon as well as acting to relieve pressure build up in an emergency situation. The rapid rise in crude oil prices in the 1970s has led to better utilization of hydrocarbons and flares now only rarely light up the night sky.

Refining separates crude oil into its different parts or fractions, converts certain fractions into more desirable portions and cleans up the impurities. A refinery may vary from a simple atmospheric and vacuum distillation unit producing a few products to a highly complex multiprocess facility turning out not only fuels and oils but petrochemical feedstocks.

Fractionation

Following desalting the crude undergoes atmospheric distillation (topping). The hydrocarbons boil off at temperatures related to their molecular size, are drawn off and recondensed. The hot bottoms undergo a vacuum distillation where the heavier molecules are separated without being broken down.

Conversion

These processes produce a change in the size and structure of the hydrocarbon molecules in order to obtain more commercially desirable products.

Cracking

This breaks large molecules into smaller lower boiling molecules. Cracking may be thermal, catalytic (aluminium/silicon oxides) coking (using the heavy bottoms of fractionation) viscosity breaking or hydrocracking (can use high sulphur feedstocks).

Combining

This process joins together small hydrogen-deficient molecules (olefins) to produce more valuable gasoline blending stock. It may be alkylation (catalyst—hydrofluoric or sulphuric acid) or polymerization (catalyst—phosphoric acid).

Rearranging

Here the molecule is changed to produce a product of different characteristics. Catalytic reforming strips hydrogen off giving rise to aromatics. The catalyst here is usually platinum based. Isomerization converts straight-chain normal paraffins to branched isomers.

Treating

In the cleaning-up process of hydrocarbons sulphur is one of the major contaminants requiring removal. Hydrogen sulphide may poison catalysts as well as people and is removed at varying stages of the processing of crude. The end product of elemental sulphur has a commercial value. Other impurities requiring removal are mercaptans (which have a strong unpleasant odour), nitrogen, arsenic and various corrosive compounds from processing systems. The catalysts used in these processes include cobalt, aluminium and molybdenum oxides. Large amounts of sodium hydroxide (caustic) are also used in sweetening processes.

Blending

Most products sold by oil companies are blends of the carefully formed pure fractions with certain additives.

Gasolines are a blend of the C4–12 fractions. In order to raise the octane number tetraethyl (or tetramethyl) lead may be added to the blend. Organic lead compounds are highly toxic and may be rapidly absorbed in large amounts by inhalation giving rise to an acute poisoning which affects the CNS. An acute toxic psychosis may progress to coma and death. The clinical features are totally different from the chronic poisoning by inorganic lead. The organic lead compounds may also be absorbed in lesser amounts through intact skin.

The hazards are minimal in blended gasoline but are high in the 'neat' tetraethyl lead (TEL) at the blending stage and more particularly in the sludge during the cleaning of tanks known to have contained leaded gasoline. Monitoring of air concentrations in tanks before and during the removal of sludge, the wearing of protective clothing and respirators and biological monitoring (urinary lead estimation) should all help to ensure that poisoning is prevented.

The environmental move towards lead-free gasoline is altering the gasoline blend formula with the likely introduction of more aromatics to improve the antiknock formerly given by the organic lead compounds.

Lubricating oils are blends of paraffins, isoparaffins, naphthenes and aromatics in the range >C20 and contain additives to act as antioxidants, detergents and corrosion inhibitors.

Kerosines and diesel oils are in the C9–C17 range.

Auxiliary operating facilities

These are plants required to support the major processes of refining and include power plants, steam-generating facilities, cooling water systems, hydrogen production, waste water treatment, light ends recovery, flare and blowdown systems and storage tank farms. The utilities plants tend to be among those with the highest noise levels.

HEALTH HAZARDS IN REFINERIES

At this stage it might be advantageous to draw a distinction between the terms 'toxicity' and 'hazard' in this context. 'Toxicity' is the property of a chemical to harm a living organism once it has been absorbed into that organism. 'Hazard' is the risk or chance that the individual will be exposed to significant quantities of that chemical.

Four major epidemiological studies of refinery workers in UK,[6] US[7,8] and Australia[9] revealed standardized mortality ratios of less than 100. This 'healthy worker' effect is not uncommon in such population studies. There was no significant increase in cancer or leukaemia in these studies with the exception of lymphoma in one of the studies.[8] There was no correlation between the lymphoma and any specific exposure.

Benzene

Rushton and Alderson[10] in a further evaluation of UK refinery employees who had developed leukaemia found no excess compared with national rates and no increase in the cytological types associated with benzene. A rough breakdown into jobs with low, medium and high exposures to benzene did show that the relative risk in the medium and high categories compared to the low approached significance ($P = 0{\cdot}05$) when length of service was taken into account. Their conclusion was that if a risk was present, it applied to a very small proportion of refinery workers.

A recent brief communication[11] from USA did show an unexplained increase in leukaemia in one refinery (not included in the above US study). This excess was of acute myeloid leukaemia and it did not occur in those jobs having the highest benzene exposure.

Benzene occurs naturally in crude oil and is produced in a number of refinery processes, particularly cracking and reforming. Exposures may occur during stream sampling, maintenance work, loading of naphthas, in laboratories and at separators and effluent tanks. With adequate plant maintenance and good work practice exposures should be well below the TLV. Biological monitoring for potential exposures by sampling of exhaled breath can be used to evaluate control measures.[12]

Skin cancer from exposure to mineral oils was described over 100 years ago[13] and is still recognized today. A recent CONCAWE investigation[14] into the relationship of chemical composition of 76 mineral oils from different companies and their ability to cause skin cancer in mice showed no single reliable predictor of activity.

Noise

The ubiquitous health-related problems from the early stages of exploration throughout the refinery processing of crude oil are hydrogen sulphide and noise. While the acute toxic effects of an exposure to the H_2S are dramatic and life-threatening but rare, the insidious onset of noise-induced sensorineural hearing loss is an affliction which is by no means uncommon within the industry. Regulatory noise exposure levels set at 90 dBA in most countries are likely to reduce to 85 dBA. Unfortunately personal hearing protection in the form of ear muffs, glass down or foam ear plugs must continue to be the mainstay of most hearing conservation programmes because noise prevention and suppression are costly and of limited efficiency unless they are introduced in the design stage of projects. As additional countries accept noise-induced hearing loss as a compensable industrial disease routine audiometric testing has become essential within the industry. Fortunately, this screening procedure lends itself well to computerization.

Ideally one thinks of noise as an engineering problem and of course constant attention must be paid to this. Good occupational hygiene input at the design stage of new plants and thorough evaluation of noise levels in existing plants are essential in handling noise.

Most refinery processes are enclosed systems which do not give rise to exposure of the operators except at sampling points or due to leakage. Catalysts have previously been mentioned and may give rise to exposure during replacement. Caustic, acids and ammonia are in common use. Phenol extraction was at one time regularly used in the production of high quality lubricating oil. This is gradually being replaced by more efficient and safer processes. Phenol is highly toxic as well as being corrosive. Not only can it cause severe skin burns but it is rapidly absorbed through the skin giving renal, hepatic and lung damage. Showering with water in such cases is not enough and the exposed skin requires to be swabbed down at once with polyethylene glycol.

The maintenance worker is usually more vulnerable to exposure than his process operator colleague. Plant maintenance is either scheduled or on an emergency basis. The scheduled maintenance may involve plant shutdown which, with good planning, adequate monitoring and safe work practice should not give many problems. The emergency situation is the one which requires assiduous attention to procedure both by the employee and by his supervisor.

The following list, by no means exhaustive, gives some idea of potential exposures:

Asbestos

Refinery insulation which has been in use for years contains asbestos. While the tendency over a number of years has been to renew insulation with a non-asbestos material, care must be taken on the removal of old insulation. Full precautions include wetting of the insulation before removal, the wearing of protective clothing and the use of the appropriate respiratory protection. The old insulation must immediately be bagged and disposed of according to statutory regulations.

Sandblasting

In some countries (particularly where there is a strong US influence) this procedure is still carried out. Substitution by carborundum or other non-silica-containing materials is standard in UK and a positive pressure air supply can cut down dust exposure. Noise exposure during sandblasting or grit blasting can be high.

Vanadium pentoxide

This may build up in the soots of furnaces fired by certain vanadium-containing crudes and during cleaning may reach toxic levels. It is not found in gas-fired plants. While the classic green tongue of vanadium poisoning is a rarity, it is possible to find chest problems as well as increases in urinary vanadium levels during furnace maintenance. Proper education and the wearing of respiratory protective devices are usually sufficient to overcome the problem.

High pressure water jets

These are often used in the cleaning of boiler tubes and the force of this jet is sufficient to sever the muscles of a man's thigh.

Solvents

These are widely used in metal cleaning. Where used to clean hands there is the risk of development of an irritant dermatitis. The neurobehavioural aspects of chronic low level solvent exposure is a controversial area which is worthy of continuing evaluation within the oil industry.

Confined space working

The problems in this situation are the build up or existence of toxic chemicals, explosive atmospheres or lack of oxygen. To these may be added other physical agents such as heat or noise. There must be adequate planning of the work to be done with a definition of all possible hazards and an emergency rescue plan. Regular measurements of chemical and physical agents and of oxygen levels should be carried out before and during entry. Ventilation must include adequate extraction of toxic substances and adequate make up air. Appropriate personal protective equipment must be worn when the confined space atmosphere cannot be assured to be hazard free. To the above requirements must be added education of the worker in safe work practice and the use of respiratory protective equipment. A permit system which ensures all the above conditions is essential as a part of allowing entry to confined spaces: no entry should be allowed until a manager has approved *all* the necessary precautions.

DISTRIBUTION AND MARKETING

A significant proportion of oil industry products go as feedstock to the petrochemical industry. The finished refinery product may be transported in bulk by pipeline, ship, road or rail tanker. Safety considerations are paramount with most countries having a system of identification and hazard rating of the product carried. Bulk loading and unloading are the major sources of exposure to employees.

Marketing of petroleum products is a highly competitive business with rich rewards for successful companies. The health risks at this end of the business are mainly those related to stress and life-style and include for example problem drinking.

OVERALL SAFETY IN THE OIL INDUSTRY

In an industry dealing with highly flammable hydrocarbons, safety must always be a prime consideration of management. This is especially the case in refineries and loading terminals which are located in urban communities. There can be few more horrifying sights than a boiling liquid expanding volume explosion (BLEVE) which suddenly devastates all around its source.

As in all other industry, safety must be considered a line management responsibility and safety performance should be regarded as part of the employee's overall job performance. In the provision of safe operating facilities, personal protective equipment and regular plant maintenance the oil industry normally adopts a higher standard than the legislation required by the particular country in which it is operating. Variations between companies tend to be in areas such as setting realistic practical safety goals, motivation of employees, regular training, monitoring of and acceptance of responsibility for safety performance by management. These topics are well covered in two HSE Occasional Paper Series, *Managing Safety*[15] and *Monitoring Safety*.[16]

Injury and accident statistics provided by individual oil companies frequently show much lower than national average rates for serious injuries and fatalities. Cynics might remark that this is due to the dangerous jobs being given to contractors—exploration drilling, diving and tank cleaning being but three examples. This does raise the important point that the overall responsibility of a manager for safety does extend to all contractor operations within his work site.

Injury and accident statistics must be regarded as one index only of total safety performance and then as an index of failure. They should be interpreted with caution since they take no account of inherent risk. Comparison between different countries is fraught with error. Nevertheless if one looks at death and serious injury rates compiled for USA[17] and UK[18] the oil industry comes out around or only slightly higher than the average rates for industry as a whole.

THE ROLE OF THE PHYSICIAN IN THE OIL INDUSTRY

Oil companies employ physicians for a number of purposes. A number provide primary medical care to employees at remote work locations, some provide a general practice to expatriate employees and dependants in countries where health services are less well developed than in their home country and a few perform regular examinations given to executives as a perquisite of their employment.

The occupational health physician can most usefully function as the leader of a team including the occupational hygienist, the occupational health nurse and the safety officer. Unfortunately, there has been a tendency in some companies for safety to be divorced from medicine. The only difference between an occupational accident and an occupational disease is in the time factor—it is the difference between acute and chronic exposure.

The physician may expect to have a limited role in the medical treatment of some employees. Statutory medical examinations may require to be performed—for example on lead workers and divers. A variety of screening programmes such as audiometry may be performed on vulnerable groups.

Today the oil industry along with other leading industries is moving towards the development of a computerized health information system.[19] This system links work history, potential exposure levels, toxicology information and health records. It has been developed for the following reasons:

—to improve the effectiveness and efficiency of the medical department
—to improve the capacity for undertaking population studies, and
—to improve management reporting and regulatory compliance.

By his relationship with management, the workforce, other professional groups inside and outwith the industry and by a basic understanding of operating processes the physician is ideally placed to be the watchdog of health and safety within the oil industry.

GLOSSARY

Acid gas corrosive gas which forms an acid when mixed with water. Usually hydrogen sulphide or carbon monoxide.

Antiknock compound a substance added to a fuel to prevent its detonation (usually tetraethyl lead).

Asphalt a mixture of heavy hydrocarbons and non-metallic derivatives usually the residue of petroleum refining. May be solid, semisolid or liquid.

Associated gas natural gas that overlies and contacts crude oil in a reservoir.

Barite barium sulphate, a constituent of drilling mud.

Barrel measure of volume for petroleum products
1 barrel = 42 US gallons
33·5 UK gallons

Bentonite a colloidal clay which swells when wet and is a major constituent of drilling mud.

Bitumen see asphalt. Bitumens may occur naturally but in the oil industry the term is synonymous with asphalt.

BLEVE boiling liquid expanding volume explosion.

Blowout preventer valve at wellhead to prevent the escape of pressure.

Bottoms liquids and residue that collect in the bottom of a vessel or the residual fractions at the bottom of a fractionating tower after the lighter components have been distilled off as vapours.

Cap a well to control a blowout by placing a very strong valve at the wellhead.

Cap rock impermeable rock overlying an oil or gas reservoir that tends to prevent migration of the oil or gas out of the reservoir.

Cetane number measure of the ignition quality of fuel oil (similar to octane number in gasoline).

Christmas tree the control valves, pressure gauges and chokes assembled at the top of a well to control the flow of oil and gas after the well has been drilled and completed.

Condensate a light hydrocarbon liquid obtained by condensation of hydrocarbon vapours. It contains little or no C1 or C2 factors and varying quantities mainly of C3, C4 and C5 fractions.

Cracking the breaking down of large hydrocarbon molecules into smaller ones by thermal or catalytic means.

Cryogenic plant a gas processing plant that is capable of producing natural gas liquid products at very low temperatures.

Deadweight ton unit of capacity of tank ships equal to the difference in tons between the ship's displacement when unloaded and loaded.

Derrickman crew member who handles the upper end of the drill string as it is being taken out or lowered into the hole and also responsible for the circulating machinery and the conditioning of the drilling fluid.

Distillate see condensate, or may be heavy gasoline or light kerosine used as fuels.

Drill collar heavy steel tube used to add weight to the drill bit.

Driller employee directly in charge of a drilling or workover rig or crew.

Drill string the column of drill pipe above the drill collar.

Dutchman portion of a stud or screw that remains in place after head has broken off.

Gauger employee who samples crude oil to determine its quality and quantity in a tank.

GOSP Gas oil separation plant.

Heavy ends parts of the hydrocarbon mixture that have the highest boiling point and highest viscosity.

Hydrocracking cracking in the presence of low pressure hydrogen.

Hydroforming passing of naphtha over a catalyst at elevated temperatures in the presence of added hydrogen. This improves the octane rating of the motor or aviation fuel.

Injected gas a high pressure gas injected into a formation to restore reservoir pressure.

Kelly the heavy steel member, 3-, 4-, 6- or 8-sided, suspended from the swivel through the rotary table and connected to the topmost joint of drill pipe to turn the drill stem as the rotary table turns.

Liquified natural gas (LNG) mostly methane and to get it in the liquid state it must be chilled to very low temperatures.

Liquified petroleum gas (LPG) a mixture of the heavier gaseous paraffinic hydrocarbons, mainly butane and propane. A portable source of thermal energy, easily liquified at moderate pressure.

Motorman crew member of rig who is responsible for the care and operation of drilling engines.

Mud liquid circulated through the wellbore during drilling, brings cuttings to the surface, cools and lubricates the bit and increases pressure to protect against blowouts.

Mud engineer employee who tests and maintains the properties of the mud as required by the drilling conditions.

Naphtha a volatile, flammable liquid hydrocarbon distilled from petroleum—may be used as a solvent or a fuel.

Olefin unit unit in a refinery that produces ethylene and propylene.

Petrochemical a chemical manufactured from petroleum and natural gas or from raw materials derived from them.

Petroleum a naturally occurring substance composed mainly of carbon and hydrogen but perhaps including sulphur, oxygen and nitrogen; may be in a gaseous, liquid or solid state.

Pig a device which is forced through a pipeline to clean the interior walls, to separate different products travelling in the same pipe, or to displace fluids in a gas pipeline.

Reforming a process where low octane naphthas or gasolines are converted to higher octane products.

Roughneck (rotary helper) a worker on the rig floor; usually a job requiring considerable physical effort.

Roustabout a labouring type job handling equipment and supplies.

Sour containing or caused by hydrogen sulphide or another acid gas.

Tongs large wrenches used for joining or taking apart pipe or casing.

Toolpusher employee in charge of the drilling crew and the rig, a supervisory job.

Turnaround a period during which a plant is shutdown for repair, inspection or modification.

Workover performing one or more of a variety of operations on a producing oil well to try and increase production.

REFERENCES

1. Clothier JG. Medical aspects of work in Arctic areas. *Practitioner* 1974; **213**:805–811.
2. American Conference of Governmental Industrial Hygienists. *Threshold Limit Values for Chemical Substances and Physical Agents in the Work Environment and Biological Exposure Indices with Intended Changes for 1984–85*. Cincinnati OH: ACGIH, 1984.
3. Bracken WM, Cupperage F, McLaury RL et al. Comparative effectiveness of topical treatments for hydrofluoric acid burns. *J. Occup. Med.* 1985; **27**:733–739.
4. McCulley JP, Whiting DW, Petitt MG et al. Hydrofluoric acid burns of the eye. *J. Occup. Med.* 1983; **25**:447–450.
5. Edmonds C, Lowry C and Pennefather J. *Diving and Subaquatic Medicine*. San Pedro CA: Best Publishing Co, 1983.
6. Rushton L and Alderson MR. An epidemiological survey of eight oil refineries in Britain. *Br. J. Ind. Med.* 1981; **38**:225–234.
7. Hanis NM, Shallenberger LG, Donaleski DL et al. A retrospective mortality study of workers in three major US refineries and chemical plants. *J. Occup. Med.* 1985; **27**:283–292.
8. Wong O, Morgan RW, Bailey WJ et al. An epidemiological study of petroleum refinery workers. *Br. J. Ind. Med.* 1986; **43**:6–17.
9. A.I.P. Triennial Report 1980–83. *Health Watch*. University of Melbourne Department of Community Medicine 1984.
10. Rushton L and Alderson MR. A case-control study to investigate the association between benzene exposure and deaths from leukaemia in oil industry workers. *Br. J. Cancer* 1981; **43**:77–84.
11. McCraw DS, Joyner RE, Cole P. Excess leukemia in a refinery population. *J. Occup. Med.* 1985; **27**:220–222.
12. Berlin M, Gage JC, Gullberg B et al. Breath concentration as an index of the health risk from benzene. *Scand. J. Work Environ. Health* 1980; **6**:104–111.
13. Bell J. Paraffin epithelioma of the scrotum. *Edinb. Med. J.* 1876; **22**:135–137.

14. CONCAWE Report Number **4/85**. Investigations into the relationships between chemical composition of mineral oils and their ability to cause skin cancer in mice.
15. HSE Occasional paper series: OP3. *Managing Safety* London: HMSO, 1983.
16. HSE Occasional paper series: OP9. *Monitoring Safety* London: HMSO, 1985.
17. National Safety Council. *Work Injury and Illness Rates*, 1983.
18. HM Factory Inspectorate. *Manufacturing and Service Industries Report* 1984.
19. Smith FR, Gutierrer RR and McDonagh TJ. Exxon's health information system. *J. Occup. Med.* 1982; **24**:824–826.

13. OCCUPATIONAL HEALTH IN THE MOTOR MANUFACTURING INDUSTRY

S. E. Brill

INTRODUCTION

Throughout the world 41·3 million cars, commercial vehicles, tractors and buses were built in 1985. Until the mid 1970s most manufacturing was centred in the dozen or more progressive industrialized nations. The less developed countries imported completed vehicles or simply finished the very final stages of the build, which needed little engineering expertise or skills in the labour force. As countries have emerged into more advanced technology they have aimed at making a greater proportion of the product.

BACKGROUND

Motor vehicle manufacture is a conglomerate of many individual and diverse industries and processes. It can be thought of as a pyramid with the final assembly at the tip where the body shell is painted, fitted to the complete drive unit, and fixed with the seats, trim and wheels. Below this the body shell is constructed from the many sheet metal pieces pressed out in the stamping plants. In the next layer the engines and transmissions are built. The whole is supported by a mass of component manufacturers responsible for the glass, tyres, batteries, fastenings, plastics, washers, screws, nuts and bolts. No car manufacturer makes the entire sequence. Some are concerned only in the narrow area at the top while others produce a wider range.

Component suppliers account for the major proportion of those involved in the industry. A number are engaged in small workshops with just a few people where there may be little consideration for their health and safety. The rest of the workforce are employed in large companies with familiar logos. The majority of cars are manufactured by large concerns instantly associated with the generally recognized badges displayed prominently on their products. Working conditions are usually of a high standard, with finance allocated to maintaining health, safety and welfare amongst their workers. Experience shows that they rate very favourably with other major engineering concerns. A further group distribute, sell and service the vehicles. The automobile industry has been a focal point for industrial development and economic activity in the EEC. It consumes 20 per cent of the output of the steel and machine tool industries, 15 per cent of rubber production and 5 per cent of glass.[1]

When Henry Ford set up a moving assembly line in 1913 a new era in industrial history was opened. Many and continued improvements were made and the stage was reached in November 1922 when almost a quarter of a million cars were produced in that month. The vehicles were more basic than at present and required considerably less man hours for their completion. The moving track was criticized as boring and repetitive, dampening and depressing the human spirit.

Much has been said against mass production but Brown felt that the solution to such problems lies in the ability to adapt the new technological organization of industry so that it corresponds more closely with human needs.[2]

In 1974 at Volvo's Kalmar plant in Sweden dramatic changes were made to the working method. Several years of detailed planning went into the adoption of group production unconcerned with the fixed pace assembly line. The car bodies are individually transported on computer-controlled battery-driven carriers to some 20 team zones. At each zone 15 or so assemblers work on an entire system, the electrics as an example, for up to 40 minutes. The team members can interchange jobs, vary the rhythm and the content and take collective responsibility for quality and completion. Initial high labour turnover, increased absenteeism, low output and disruption of production had to be overcome. After ten years their experience was surveyed and progress assessed. Assembly costs were the lowest of all Volvo plants with lower man hours per car. Output increased, working conditions improved and the majority of employees now enjoy membership of a work team with a varied job content. But this brave and interesting experiment has not been taken up by others. Volvo has been concerned with a labour force of between 630–730 reaching a capacity of 32 000 cars per year in 1983.[3] Many other industrial enterprises engaged in mass production have considered this working method but they have kept their distance from it. The Volvo experiment has excited admiration but no other car manufacturer producing up to 500 000 vehicles in a year has felt able to adopt it.

Since the late 1970s dramatic changes have quickly and subtly taken place. The common conception of the large dull monochrome factory with myriads of employees standing almost shoulder to shoulder rushing to complete each task is outdated. The modern vehicle construction plant is being transformed into light-coloured, airy, ordered places with fewer workers. Advanced technology has both greatly increased productivity and altered production methods. It is mainly computer-based, controlling expanding automation and the increased use of robots. Driverless automatic guided vehicles, which receive their instructions through cables sunk in the floor, transport their components to the assembly lines. New materials, lighter, cheaper and more durable, are being substituted for the old. The most recent scientific discoveries are quickly considered for application and adopted within the industry.

The revolution is in its early phase. The traditional practices, processes and plant are intermingled with the latest advances. Similar processes occur in all types of vehicle manufacturing, the component industry, service centres and body repair garages. This chapter concentrates on the hazards found in car manufacturing.

ACCIDENTS AND INJURIES

Few statistical reports have been published on the extent of injuries in automobile manufacturing or their causation. Comparisons can be made only at an informal level between different companies within the industry and between car manufacturing and general engineering.

Press injuries

The risk of crush injuries and amputations is high in stamping and pressing operations. Legislation to ensure guarding of these machines varies greatly throughout the world. In countries where mandatory mechanical guarding of the power presses is enforced and where they are examined at the beginning of each period of work the serious injuries are considerably fewer. In other countries, the standard ranges from presses without any guards to sophisticated photosensitive devices where reliance is placed on the interruption of a beam of light to halt the press instantly. Although these have been in use for many years, they are less effective than mechanical guards. New developments are currently taking place in photosensitive technology which need to be further assessed and evaluated; the conclusions may indicate improvements that allow these systems to be considered acceptable. While mechanical barriers are effective in preventing direct injury to the hands, there may be risk to the trunk and upper arms from the guards rising in front of the press at the beginning of each cycle of operation.

Hands

About a quarter of all the injuries in automobile manufacturing involve the hands and arms. In a study of hospital in-patients and out-patients Smith et al. found that more serious hand injuries requiring admission to hospital occurred in the home than at work, and only a few were caused by industrial machinery.[4] The razor-sharp edges of sheet steel can cause severe and highly damaging lacerations. The front of the wrist is particularly vulnerable and adequate hand and arm protection is essential (*Fig. 13.1*). A fine wire mesh between layers of felt and covered by pvc can be incorporated into a wrist/arm cuff and affords excellent protection. This equipment is lighter and more flexible than rigid plastic and permits easier wrist movements. It must be used in conjuction with an appropriate glove.

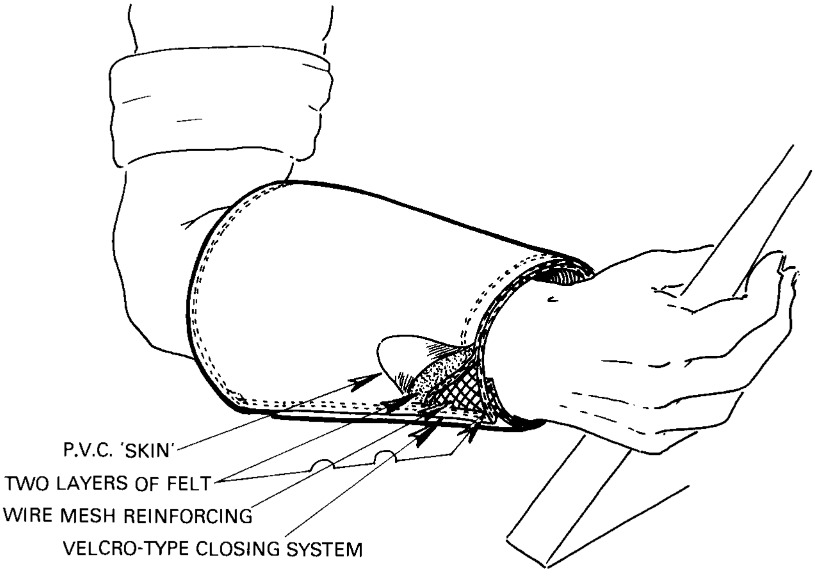


Fig. 13.1. Flexible protective reinforced armlet for handling sheet metal.

Accidents involving other machinery are relatively fewer than in the engineering industry in general and do not present a major problem, provided care is taken to ensure safe systems of work especially during maintenance.

Eyes

Eye injuries account for a small percentage of the total but there is a constant threat of deep penetrating eye wounds and embedded corneal foreign bodies for workers involved in cutting and grinding metal. These workers, and arc welders, are convinced of the need for wearing eye protection. Since the outcomes of dust, paint flake and spatter from spot welding are seldom so serious, these employees are less assiduous in protecting themselves unless motivated by enthusiastic supervisors.

Slipping

About 40 per cent of all accidents are due to falling, tripping and striking against objects. Manning et al. investigated 1153 employees attending the Occupational Health Department for back pain and found that 66 per cent of true accidents occurring at work could be accounted for by underfoot slippage.[5] Initial research into footwear that could prevent accidents of this nature found that on oily surfaces polyester microcellular polyurethane showed better slip-resistance than nitrile rubber or pvc. Research into this subject is continuing.[6]

Musculoskeletal injuries

Cranes, lifts and mechanical aids have reduced the manual handling of crates, boxes and bins; but numerous parts need to be added during the vehicle build. Back problems are much in evidence from the requirement to bend into containers—often awkwardly shaped and deep—to retrieve the stock. Bending into the vehicle, stretching to reach distant corners and crouching to squeeze into small spaces compound the problem. Acute strain injuries affecting the trunk and limbs are common. Few deliberate attempts to fit the man to the job at pre-employment or pre-placement examination have been made and past high labour turnover discouraged employers from trying.

Automation and robotization are ideally suited to engine assembly, stamping and body construction. In an engine factory in Italy opened in 1985, 90 per cent of the work is automated and of the 78 work stations only two are manual. Final vehicle assembly is still highly labour intensive and further efforts to design out this problematic area without overreaching capital investment are yet to come.

Repetitive strain injuries

Repetitive or repetition strain injuries (RSI), also known as cumulative trauma disorders, are convenient and descriptive terms to group together a number of conditions affecting the tendons and muscles in the hands, wrists, arms and shoulders; included are carpal tunnel syndrome, tendinitis, tenosynovitis, epicondylitis and cervical brachial disorders. The lower limbs can also be affected. Rapid and repeated movements set up traumatic inflammation but other factors including posture, muscular load, leverage and personal characteristics play an important part in individuals. The exact mechanisms have not been clearly established.

Carpal tunnel syndrome (CTS) has been studied in food packers, printed-circuit board assemblers and keyboard operators. While jobs in the vehicle industry have strong similarities, they have not so far had detailed investigation. Caution must be exercised before attributing causation to occupation. CTS is two to ten times more common in women than in men, and use of forceful exertions and of deviated wrists and pinch hand positions are associated factors.[7] In a series of 507 patients with CTS, 34·7 per cent were work-related with more males (51·7 per cent) being affected.[8] In a case-control study at an aircraft engine manufacturing company the use of vibratory tools and a history of hysterectomy and oophorectomy were strongly associated with the onset of CTS.[9] An ergonomic analysis of individual operations in a car assembly plant shows that some specific elements associated with the onset of CTS are: support of the body by the left hand fully pronated, dorsiflexed wrist and sometimes concurrently in radial deviation; use of the palm as a hammer and direct pressure on the carpal tunnel; the hand in mid-supination performing rapid and violent dorsiflexion of the wrist; use of pneumatic tools while the wrist is in full pronation and ulnar deviation.

Tenosynovitis, though commonly seen in new employees and after changes in work rate, can also be found where there has been unchanged employment for many years. Uniformity of diagnosis on an anatomical basis would greatly assist in epidemiological study.[10] A close look at work method and layout can often determine the characteristics that can be corrected.

It is likely that repetitive strain injury is a significant problem and that its prevalence has been underestimated.

PHYSICAL HAZARDS

Noise

Noise is probably the biggest single hazard. The extent of the resultant disability from prolonged exposure and the number of people affected are not known. Noise levels in excess of 90dB(A) are only too evident on entering a large press shop. The massive machines exerting a force of 2000 tons can now produce a complete car body side which previously had to be built up from a number of smaller panels. A sheet metal panel passes through a series of presses and robot arms fitted with appropriate suction devices transfer the body side from one press to the next. Although these presses need few operators, the noise affects others in close proximity. Smaller component production can still result in equally high noise levels as parts eject at a rapid rate on to the collecting chute. In this example there are two main sources of noise: first the metal-to-metal contact and second the mechanical action of the presses themselves. No means have yet been found to lessen the deep thud as the die strikes the sheet steel panel. To prevent noise emission small presses can be totally enclosed incorporating the minimum practical entry and exit ports. Overlapping PVC strip may be constructed to form a curtain and reductions of 12–16 dB(A) can be expected. In other situations flexible barrier materials are suitable and can be drawn aside rapidly to permit maintenance and any necessary tool replacement.[11] The larger presses may be difficult to deal with when placed close together in line and the costs of noise suppression may be prohibitive. In this situation adequate hearing protection must be provided. Chutes and benches can be made of wood, rubber and ultra high molecular weight polyethylene, but their poor wearing properties can be expensive if they need to be replaced frequently. A sandwich of high density acoustic material between sheet steel may overcome this problem.

The hiss from pneumatic air release can be treated by either the reduction of air pressure or by the duration of the release. This usually results in noise reduction and in financial saving because the supply of compressed air is costly. Even more effective and inexpensive is the fitting of silencers. In another example a marked reduction was achieved by replacing the spur gear drive by a multidrive belt in a 750-ton forging press (*Fig. 13.2*).[12]

In the body construction areas the clamping of sub-assemblies into the jigs prior to welding and their passage down the production lines contribute to a noisy environment.

Another noise source is the closed wash tanks which remove the film of oil or grease from metal parts. Hydrocarbon solvents have been commonly used for degreasing in the past but in one plant eighteen manually operated solvent tanks have been replaced by detergent washes. This trend is typical and reduces the hazard from exposure to solvent vapours. Cleaning can be very efficient and rapid when parts are forcefully sprayed with hot alkaline detergent, but noise arises from the air intake fan and associated motor. *Table 13.1* demonstrates the reduction of 8·2 dB(A) after erection of an acoustic barrier.

This illustrates how an unsuspected hazard may be brought in when one process is changed for another and hence requires a different

Table 13.1. Reduction due to acoustic enclosure

	Sound Levels dB(A)		
	Intake fan on	Background level	Calculated level of intake fan/motor
Before acoustic enclosure	90·0	77·0	90·0
After acoustic enclosure	83·0	77·0	81·8
Reduction			8·2

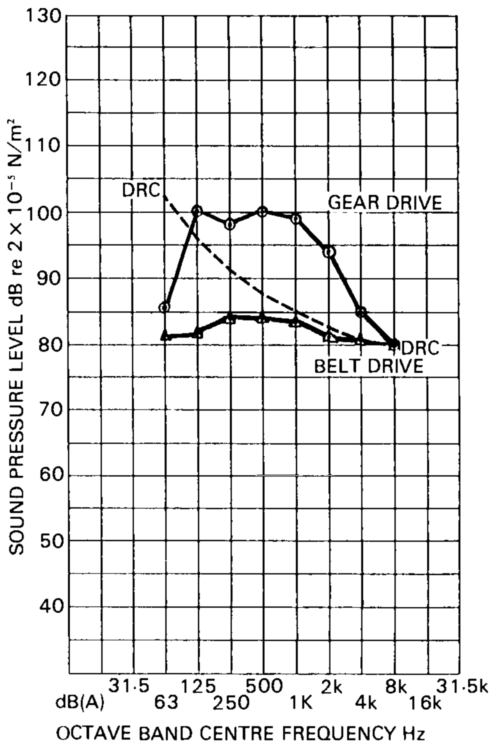


Fig. 13.2. Replacement of gear drive by belt drive. DRC = damage risk to criterion according to NR85.

form of remedial action. The opportunity to minimize risk must be taken should total elimination of a hazard not prove possible.

Metal finishing is the process of using hand-held rotating sanders to smooth the surface of the panels and rid them of irregularities such as weldspatter. The source of the noise is the action of the sander on the car body itself. The operative's ear can never be more than an arm's length away from the source and local attenuation is very difficult. There are additional potential hazards from this process (*see* later), which require the use of down-draught ventilated booths and these serve to contain the noise and prevent it spreading to adjacent areas. A similar situation exists with the spray painters who have two noise sources which cannot be practically reduced. One is the noise arising from the hiss of air-spraying guns and the other is from the inlet valves in the air-supplied hoods. Lead discers are at risk in the same way. Of course personal hearing protection must be worn here but it is rarely possible to use ear muffs, the most effective type of protection, while wearing the hood; airfed visors do permit muffs to be worn. Operators have a preference for glass down/wool or expandable foam ear inserts but the attenuation claimed by the supplier should result in an exposure of less than 85 dB(A) and, following an examination of the ear canal and drum, there are no clinical contra-indications to using plugs.

Large quantities of air are moved in a manufacturing plant for both local and general inlet and exhaust ventilation. The fan motors may be the cause of considerable noise, particularly when they operate at high speeds or move air at high velocity. Noise levels may be increased if maintenance is neglected.

The well-established principles in any hearing conservation programme are: (1) noise measurement and analysis, (2) reduction of the noise at source where possible, (3) erection of a barrier between the source and the ear to interrupt the acoustic pathway, (4) provision of hearing protection, and (5) performing hearing tests (audiometry). The first three call for the expertise of the fully trained and experienced industrial hygienist together with the assistance of the acoustic engineer. The last two principles are the province of doctors and nurses. It is not enough simply to hand out hearing protectors or to make them readily available. Conventional health education raises general awareness and its benefits should not be dismissed, but younger people in particular are not easily convinced of a future impairment of hearing. It is disappointing that compliance with the advice to wear hearing protection is not sustained for longer than a few months after a general campaign. Individual counselling together with a demonstration of personal hearing loss, if any, on their own audiogram can be very effective in conveying the long term and insidious effects. Hearing conservation is complex and complicated and can best be achieved by the close co-operation of experts.

Vibration

Vibration exposure is frequent from the common use of pneumatic tools such as sanders, chisels, screwdrivers, nutrunners and from pressing an object against vibrating surfaces as in pedestal grinding. Vibration syndrome caused by road-breaking drills and forestry chainsaws is well documented and explored by Taylor (1985).[13] Little investigation has yet taken place with the lighter weight tools. Bossche and Lahaye showed a high frequency of radiological anomalies, mostly localized to the carpal bones, in workers exposed to vibration by light tools.[14] However, no significant increase in the prevalence of bone cysts of the hands and wrists of a group of miners was found and exposure to chain-saw vibration did not seem to be a determinant in the development of vacuoles in the carpal bone.[15] Further investigation is required into the outcome of prolonged use of light tools, the factors which may give rise to any resultant conditions and the work methods that may have to be altered to minimize symptoms.

RESPIRATORY PROBLEMS

Spray painting

In undercoating the body shell by the electrocoat method, the shell is positively charged electrically and totally immersed in a tank of paint which is negatively charged. This permits an even thickness of paint cover to reach the otherwise inaccessible box sections. There is little human involvement here. Motor car sales are often made on immediate eye-catching appeal, the final colour and surface finish, so that paint formulation and the way in which it is applied are of immense importance. Robotic spraying of the final coat is now common but by no means universal and it has to be interspersed with the manual method to reach the deep recesses. In crash repair garages all spraying is by hand. Paints, varnishes and enamels have consisted of: pigments (compounds of titanium dioxide, iron, lead, chromium, cadmium, aluminium, complex organics); solvents (alcohols, esters, xylene, toluene, petroleum distillates, glycol ethers); additives (cobalt and lead naphthenate driers, phthalate and adipate ester plasticizers); binders (acrylic, alkyd, cellulose, polyester polyurethane). Spray application either by atomizing the material or by the 'airless' method creates fine droplets in a form very readily capable of being inhaled. Adequate ventilation must always be provided. Except for small surfaces, spray painting should be in a purpose-built down-draught booth. Exhaust air volume must be sufficient to dissipate all paint spray emitted immediately for the maximum number of spray guns that will be in continuous operation. A vertical water wash curtain will additionally capture paint droplets prior to discharge to atmosphere. Complete fire protection is essential. These paint booths are capable of holding a number of vehicles and 2 or 3 pairs of operatives working simultaneously on each side. In applying the paint to the horizontal surface of roofs, bonnets and boot lids overspray can be considerable. The efficient booth can maintain the respirable volatiles to below currently acceptable control limits. But paint and final surface coating are developing to the extent that 30–40 individual chemical substances can be used in a single proprietary product. Classically hazardous metals such as lead and chromium are being replaced by complex organic chemicals.

Conducting an epidemiological study in a group of car painters Chatterjee found that race has to be considered when comparisons are made between occupational groups.[16] It is well known that acute exposure to high doses of organic solvents leads to narcotic effects.[17] n-hexane and methyl n-butyl ketone may cause peripheral neuropathy. Low level exposure to organic solvents in Swedish house painters has been reported to result in behavioural difficulties and personality changes, impairment of memory and neurological changes.[18] The symptoms are variously referred to as neuropsychiatric disease, psycho-organic syndrome, painters' syndrome and organic solvent disease. A cross-sectional study on car and industrial spray painters revealed statistically significant differences from a control group in a number of psychometric tests when exposure levels to organic solvents were considerably lower than the TLV.[19] Husman and Karli found a decrease in light touch and pain and an increase in vibration threshold in car painters.[20] Critical examination of the investigative methods has cast doubt on the neurophysiological and neuropsychological effects.[21] Maizlish et al. have recently reported on 104 spray painters and 101 controls and their study did not show poorer behavioural test performance by workers chronically exposed to hydrocarbon solvents below the TLV.[22] Further evidence is needed before the long-term effect is clear.

Polyurethane paints are used to provide more durable and textured surfaces and for repair to the top coat of paint. They are more resistant to chipping and cure at lower temperatures so avoiding the need for high temperature stoving. These paints contain pre-polymer isocyanates. In the single component system the isocyanates are not given off until curing in an oven and so are said to be 'blocked'. Two separate groups of components can be mixed together immediately prior to application in which one contains the pre-polymer isocyanates and the other a polyol and amine/amide curing agents. This is known as the two-pack paint system and it aids chemical curing. The curing agents may give rise to skin problems. Inhalation problems of sensitization with asthma-like symptoms may arise from exposure to aerosol mists of the pre-polymer and respiratory protection is essential (*see* later).[23] Inhalation hazards are minimized by brush or roller applications.

Rigid polyurethane foams are produced in situ in vehicle cavities both to minimize resonance and to reduce vibration and noise. Diisocyanate monomers as opposed to the pre-polymers are used. MDI (methylene bis-phenyldiisocyanate) is fed to the injection gun nozzle where it is mixed and begins to react with propellant and polyol. MDI is employed as it is less volatile and has a lower vapour pressure than toluene di-isocyanate (TDI). Worker exposure and hazards do not arise from the contained production operation as little monomer is released. Careful consideration and procedures must be applied to deal with spillage, leakage and malfunctions at the spray gun nozzle and to decontaminate containers in order to minimize exposure to di-isocyanate monomers.

Medical surveillance is recommended for those involved in isocyanate paint spraying and foam injection. It should include pre-employment examination, routine periodic examination and re-examination following sickness absence.[24] Special emphasis should be placed on those with existing respiratory conditions such as asthma and existing air flow limitations. Detailed self-administered peak flow testing at two-hourly intervals is the most satisfactory way of establishing a diagnosis (*see also* Chapter 2).[25]

With the increasing complexity of modern paints the long-term effects of many of the newer chemicals are unknown, and mixtures may be interactive and synergistic. The current level of toxicological assessment by examination and extrapolation of the formulation is unlikely to give an accurate guide to hazards. It is therefore desirable that respiratory protection, preferably of the supplied air type, is used with spraying in a walk-in down-draught booth. When there are isocyanates this work method should be obligatory. Where small parts are being finished in a side-draught booth discretion can be exercised in the use of respiratory protection.

A mortality study of paint sprayers in the automotive industry did not demonstrate an excess of lung cancers.[26]

Dust

Prior to painting, individual metal components and the body shell are given a final smoothing by sanding with hand-held rotary tools. Large quantities of iron dust can evolve and inhalation could lead to benign siderosis. This is infrequently seen as engineering methods are readily available to keep the atmospheric levels to a small fraction of the occupational exposure limit of $10\,mg/m^3$ for nuisance dust. For pieces being finished on the bench and when the operation is mainly static, properly placed exhaust ventilation ducts can usually provide satisfactory control. This is impractical for the complete vehicle. Walk-in self-contained booths of similar design to the paint spray booth are required but without the water curtain and with less frequent air changes. Exhaust ventilation fixed to the rotary sander has been developed but the extraction force needed to be efficient, their increased weight, bulk and lack of mobility have prevented their widespread use. Disposable dust masks of cotton fabric with a fibre filter should also be used in addition to general personal protection.

The discing of soldered areas is a different matter. Solder consists of 92 per cent lead, 5 per cent antimony, 3 per cent tin. From the mid-1930s solder was used to fill in the gaps where one panel abutted the adjoining one. As much as 60 lb of solder was used in each car during the 1950s to gain the desired streamline effects. The very fine dust, mainly less than 5 μm in diameter, is readily absorbed and can quickly result in lead poisoning. Solder usage has fallen considerably with changes in design, with the ability to press larger sheet metal panels and with the development of other methods of joining metal. Solder areas may be filed by hand in the open factory or plant. Lead discing must take place in a down-draught booth specially designed and reserved for the purpose. Air must be filtered before discharge to the atmosphere. Strict observance of preventive measures is necessary for the discers themselves, janitors, maintenance workers and all who may come into contact with the dust. If the precautions detailed in the Code of Practice[27] are maintained there should be few difficulties in meeting the current UK standard of 70 μg of lead/100 ml of blood (3·36 μmol/litre). Workers in excess of this value must be removed from lead work.

Abrasive operations on painted surfaces can give rise to toxic and irritant dusts from the solid constituents of the paints. Dust suppression can be achieved by wet sanding or by vacuum sanders; the latter are more successful here than in metal finishing as the volume of dust is considerably less and lower capture velocities are sufficient for control.

Oil mist and fume

The distinctive bluish haze which is noticeable on entering engineering shops is oil mist and fume created when the cutting tools are cooled and lubricated. Mist and fume can also arise from the welding of oil-coated mild steel. Coolants can be primarily mineral-oil based, semi-synthetic with a 3–5 per cent mineral oil content or fully synthetic with no oil content. There has been little satisfactory evidence that exposure to oil mist results in respiratory illness. Recently, investigations have reported respiratory symptoms, pathological changes and occupational asthma.[28,29] It is not generally clear whether oil or chemical additives were involved or what the exposure levels were. In at least one case[29] pine oil or colophony additives were implicated in occupational asthma. Exposure should be minimized and a guidance value of 2·5 mg/m^3, which is half the current UK Occupational Exposure Limit, is appropriate.

Sheet steel has to be lubricated with oil when it is being pressed. It is sometimes applied as a fine spray to each individual panel before it is fed into the press. Although convenient, it is a source of oil mist in the atmosphere of stamping plants. Satisfactory lubrication can be made by depositing several blobs of thickened grease on to the panels or by passing them between rollers dipped in oil.

Exhaust fume

Inhalation of petrol engine exhaust fumes can lead rapidly to carbon monoxide poisoning. Hydrocarbons and oxides of nitrogen are also present. Diesel fumes are a nuisance and the particulates (soot) and hydrocarbons can be irritating. Exhaust fumes from any internal combustion engine also contain very low concentrations of polycyclic aromatic hydrocarbons. Exhaust fumes may cause respiratory functional changes, increased susceptibility to infections and they may increase the risk of developing respiratory tract cancer. Extensive studies are still in progress and evidence is inconclusive.[30,31] While engines are being tested on the bench and when cars are started up for final adjustments, direct connections to the exhaust pipe in order to extract fumes to the external atmosphere are practical. Good general ventilation is necessary in the dispatch area where cars are constantly being moved for dispersal to parks outside the factory.

Asbestos

Asbestos does not have the ubiquity it once had in the industry. Before the hazard was appreciated sacks of dry fibrous material would be freely available to be used, thoroughly wetted, as 'putty'. Prior to welding it would be placed on either side of a metal junction to prevent excess heat distorting the structure. After use it was allowed to drop on the floor, to dry out and become asbestos powder once more. Apprentices were often assigned to making the putty and descriptions of 'pillow' fights with combatants finally covered in grey 'flour', the floor inches deep in it, are recalled from the distant past and are sad reminders of the lack of recognition of the long-term effects of asbestos. Modern welding techniques can contain the heat to a more circumscribed area. Even when required a 'putty' of soaked newspaper does the job equally well.

In stamping plants the friction materials of the brakes and clutches of the presses may require drilling, filing and shaving before fitting. This practice must be carried out under local exhaust ventilation which prevents the asbestos particles from being inhaled by the operator. The effectiveness of the extraction should be checked at regular intervals, either by air sampling or by the use of a Tyndall beam. Filters must be provided to clean the extracted air before its discharge to atmosphere.

Asbestos is to be found in plastics, sealing compounds and underbody protection corrosion materials and does not normally present any hazard in the uncured state nor in the cured state, unless drilled, cut or sanded. Gaskets, washers, friction materials of brakes and clutches, insulation pads and various small parts may break up on handling to release fibres. Shipping containers may contain dust residues. That hazard is of a low order but may nevertheless be present and steps should be taken to eliminate the dust.

Crocidolite and amosite are now banned from nearly all vehicle applications. Parts and materials containing other forms of asbestos must be clearly labelled: '*Warning: contains asbestos*. Breathing asbestos dust is dangerous to health. Follow safety instructions.' These regulatory requirements follow from EEC Directive 83/478/EEC. Motor manufacturers increasingly avoid the use of asbestos for most applications—particularly in new car and truck models and in brakes and clutches. A wide variety of substitute chemicals is involved and a careful consideration of their toxicity and hazards is proving necessary.

Welding

The universal adoption of the monocoque design in the late 1940s dispensed with the need for a chassis in motor cars although it is still present in the heavier commercial vehicle. The strength of the modern vehicle depends on the rigidity of the body shell and the sub-assemblies forming it, which are held together by a series of welds. In a family four-seater there are some 8000 spot (resistance) welds and about 45 other points where arc or other types of welding procedures are used. Matthews (1979)[32] has described the hazards of welding and the general principles of their control. There should be little difficulty in their implementation as exotic metal or filler rods are very seldom used in vehicle fabrication, and working in confined space is uncommon.

Most forms of welding are well suited to robotization which can better cope with the repetition, provides higher quality and has less wear on the tools. This is particularly the case with spot welding. Many tasks remain which are best performed manually and some need special care. The sheet metal for fabrication of petrol tanks has an additional coating of lead to prevent corrosion by the fuel. Lead fume arises when the halves of the tank are seam welded or the filler neck is attached. Working within the deep recesses of the luggage or engine compartments can be compared to being in a confined space and we have seen cases of metal fume fever arising in this way. Organic vapours may evolve when heat reaches uncured sealers from welding near them. Welding may also need to take place on oil-free pieces which have been degreased by immersion in aliphatic hydrocarbon solvent vapour. If the metal is still wet with solvent at the time of welding, decomposition to highly toxic phosgene and toxic and corrosive chlorine and hydrogen chloride may take place. Perchlorethylene is often used as the solvent and trichloroacetyl chloride can additionally be formed.[33] Clearly, welding should only take place on dry pieces and in a solvent-free atmosphere. Other types of welding in use include plasma arc and metal spraying, each with their own hazards.

SKIN CONDITIONS

The results of constant handling are to be found in the numbers of accidental injuries to the hands and the attendances for skin complaints at the occupational health departments. Lesions also affect the face, neck, ankles and the trunk. Newhouse[34] in a survey found four times as many cases as were previously recognized in one department of a motor car factory. Ryecroft comments that many cases of severity equal to those noted probably go unreported.[35]

Adhesives are used as cost-effective alternatives to engineering jointing and, with glues and sealers, are frequently the cause of skin problems. It is not only the worker who directly applies a substance who may be affected by it, but other unprotected operators working further along the line of vehicle build may be contaminated by excess quantities of the offending material. Among those in frequent use are epoxy, formaldehye, urethane, methacrylate, anaerobics and cyanacrylates—the 'super glues'. Primary irritation and sensitization can arise from the acrylic acid content, epoxides, amines and methacrylates. Other glues can contain 2-butanone and n-hexane. As they have synergistic action with regard to neurotoxicity, skin contact as well as inhalation must be avoided. A comprehensive review of the effects on the skin of chemicals used in motor manufacturing appears in Cronin.[36]

The risks associated with the cutaneous hazards of lubricants as described by Hodgson have been substantially decreased.[37] In many machining operations neat oils have been replaced by synthetic coolants and by soluble oil. More automation and alternative methods have reduced direct handling of wet components and immersion of the hands during machining operations. There is less contact and less trauma. Anti-splash guard design has improved and there is widespread use of personal protective clothing. Caution must be exercised that hand protection is not advised if accidental injury might result from its use, and operators may have difficulties in wearing gloves where there is a need for manual dexterity and they have to handle awkwardly shaped parts. Although the incidence of dermatitis has reduced, it has to be borne in mind that new developments in cutting fluids can still have operators working in a hostile environment.

Mineral oil based cutting oils, along with other lubricants and greases, continue in wide use. Occupational vitiligo (patchy depigmentation of the skin) is still an occasional, if infrequent, problem associated with skin contact from substituted phenols and catechols. Allergic types reactions from skin contact with oil products containing pine oils and rosin acids also still occur.

THE EVERCHANGING PRESENT

The vehicle industry is innovative and aggressive. Researchers within it are developing their own ideas whilst at the same time investigating and adapting discoveries from any other sphere where there is a process that can be put to use. Like the computer industry on which it depends, no sooner are new methods understood and established than they are succeeded by the next generation of techniques. Each new process is overtaken by another at a furious speed.

Increasing automation and robotization have resulted in a decrease in the total numbers employed, particularly for unskilled and physically strenuous work. The jobs where traditional, now outdated, processes required great muscular effort and lengthy exposure to an unfavourable physical environment have virtually disappeared and with them the gross occupational diseases which they caused. Yet substantial numbers of people will still be needed to build the cars and to produce components in the immediate years ahead.

The horizons for new hazards have to change. Recently-established processes include the use of microwaves for the fine tuning of engine timing and lasers for cutting, etching and welding. The quantity of plastic parts has doubled in the past three years. The overall hazards with these materials are considerably fewer than those of sheet metal. Plastics, being light in weight and corrosive proof, can be easily moulded into complex shapes and are less expensive per unit volume than metal.

More than 7000 different substances consisting of 40 000 chemicals and mixtures are in use by vehicle manufacturers with a comprehensive product line. Where new materials are under consideration detailed toxicological assessments need to be performed before they are introduced to predict and control any potential harmful effects. An on-line data base system is the only practical means of dealing with this huge store of information. Materials engineers and purchase personnel will need access to the data base as well as the health professionals. Chatburn et al. have described an established system.[38]

A number of companies are developing comprehensive occupational health surveillance systems which store three groups of data. First, worker exposure to materials (chemical content and levels of exposure derived from environmental measurement); second, work history (length of time in jobs, content and job changes); and third, individual worker health (absence rate, accident types and frequency, clinical examination findings, results of personal monitoring, mortality). Surveillance systems can analyse exposure information, track events and patterns of illness and will assist in reducing risks to health.

Improvements in the health of car industry workers have already been achieved through the collaboration of all health workers including doctors, nurses, hygienists, safety personnel and toxicologists. No one health professional has the necessary breadth of experience or the extensive knowledge to deal with the complexities of the problems. The future health and safety of the workforce can be assured only through the integrated team approach.

THE FUTURE

Electrical wiring systems will disappear and the next series of vehicles on today's drawing boards will be simpler to operate with more sophisticated electronics receiving and relaying messages. Microprocessor control, for example, will sense the requirements for engine ratios and fuel supplies and reminders will be flashed up when the car needs to be serviced. By the end of this century cars will be smaller, lighter, more aerodynamic and with smaller engines that are more fuel efficient. Poor energy absorption and greater labour costs are the disadvantages of reinforced plastics for the upper body but research is in progress to resolve these problems. Vehicles will not be built up from separate components but complete modular sections will be fashioned from a series of injection moulding stages. These will be able to be snapped together to form an infinite variety of models.

We can only speculate how many people will be on the shop floor but it will be substantially less than today. The health of the smaller work force of the future will no doubt continue to improve. The focus of occupational health professionals must move towards scrupulous vigilance over the rapidly changing processes and the newly developed chemicals that are a predominant feature of modern industry.

REFERENCES

1. *Bulletin of the European Communities*, Supplement 28/1. Luxembourg: European Communities Commission 1981.
2. Brown JAC. *The Social Psychology of Industry*. London: Penguin Books, 1964.
3. Aguren S, Bredbacka C, Hansson R et al. *Volvo Kalmar Revisited*. Stockholm: Efficiency and Participation Development Panel 1984.
4. Smith ME, Auchincloss JM and Ali MS. Causes and consequences of hand injury. *J. Hand Surg*. 1985; **10–B3**:288–292.
5. Manning DP, Mitchell RG and Blanchfield LP. Body movements and events contributing to accidental and non-accidental back injuries. *Spine* 1984; **9/7**:734–739.

6. Manning DP, Jones C and Bruce M. Improved slip-resistance on oil from surface roughness of footwear. *Rubber Chemistry and Technology* 1983; **56/4:**703–717.
7. Armstrong TJ and Chaffin DB. Carpal tunnel syndrome and selected personal attributes. *J. Occup. Med.* 1979; **21/7:**481–486.
8. Tountas CP, MacDonald CJ, Meyerhoff JD et al. Carpal tunnel syndrome: a review of 507 patients. *Minn. Med.* 1983; **66/8:**479–482.
9. Cannon LJ, Bernacki EJ and Walters SD. Personal and occupational factors associated with carpal tunnel syndrome. *J. Occup. Med.* 1981; **23/4:**255–258.
10. Browne CD, Nolan BN and Faithful DK. Occupational repetition strain injuries. *Med. J. Aust.* 1984; **140/6:**329–332.
11. *100 Practical Applications of Noise Reduction Methods.* London: Health & Safety Executive, 1983.
12. Hickish DE. *Industrial Noise and its Reduction:* Proceedings of seminar at Leamington Spa. Wantage: BNF Metals Technology 1975.
13. Taylor W. Vibration. In: Ward Gardner A (ed.) *Current Approaches to Occupational Health—2.* Bristol: Wright, 1982.
14. Van Den Bossche I and Lahaye D. X-ray anomalies occurring in workers exposed to vibration caused by light tools. *Br. J. Ind. Med.* 1984; **41:**137–141.
15. Harkonen A, Riihinaki H, Tola S et al. Symptoms of vibration syndrome and radiographic findings in the wrists of lumberjacks. *Br. J. Ind. Med.* 1984; **41:**133–136.
16. Chatterjee DS. A health survey of car painters of different ethnic origin and the importance of racial factor. *J. Soc. Occup. Med.* 1984; **34:**86–89.
17. Browning E. *Toxicity and Metabolism of Organic Solvents.* Amsterdam: Elsevier, 1965.
18. Hane M, Axelson O, Blume J et al. Psychological function changes among house painters. *Scand. J. Work Environ. Health* 1977; **3:**91–99.
19. Elofssan SA, Gamberale F, Hindmarsh T et al. Exposure to organic solvents. *Scand. J. Work Environ. Health* 1980; **6:**239–273.
20. Husman K and Karli P. Clinical neurological findings among car painters exposed to a mixture of organic solvents. *Scand. J. Work Environ. Health* 1980; **6:**33–39.
21. Grasso P, Sharratt M, Davies DM et al. Neurophysiological and psychological disorders and occupational exposure to organic solvents. *Food Chem. Toxicol.* 1984; **22/10:**819–852.
22. Maizlish NA, Langolf GD, Whitehead LW et al. Behavioural evaluation of workers exposed to mixtures of organic solvents. *Br. J. Ind. Med.* 1985; **42:**579–590.
23. Belin L, Hjortsberg U and Wass U. Life threatening pulmonary reaction to car paint containing a pre-polymerised isocyanate. *Scand. J. Work Environ. Health* 1981; **7:**310–312.
24. *Isocyanates—Medical surveillance:* Guidance Notes MS 8. London: Health & Safety Executive 1983.
25. Newman Taylor AJ, Pickering CAC and Coutts II. Occupational Lung Disease. In: Ward Gardner A (ed.) *Current Approaches to Occupational Health—2.* Bristol: Wright, 1982.
26. Chiazze L, Ference LD and Wolf PH. Mortality among automobile assembly workers. *J. Occup. Med.* 1980; **22/8:**520–526.
27. Approved Code of Practice: (Revised June 1985). *Control of Lead At Work.* London: Health & Safety Commission, 1985.
28. Bennett EO and Bennett DC. Occupational airway diseases in the metal working industry. *Tribiology International* 1985; **18/3:**169–176.
29. Hendy MS, Beattie BE and Burge PS. Occupational asthma due to emulsified oil mist. *Br. J. Ind. Med.* 1985; **42:**51–54.
30. Cuddily RG and McClellan RO. Evaluation lung cancer risks from exposures to diesel engine exhaust. *Risk Analysis* 1983; **3(2):**119–124.

31. Cuddily RG, Griffith WC and McClellan RO. Health risk from light duty diesel vehicles. *Environ. Sci. Technol.* 1984; **18(1):**14A–21A.
32. Matthews G. Gases Vapours and Fume. In: Ward Gardner A (ed.) *Current Approaches to Occupational Medicine.* Bristol: Wright, 1979.
33. Andersson HI, Dahlborg JA, Wettsriem R et al. Phosgene formation during welding in air contaminated with perchlorethylene. *Ann. Occup. Hyg.* 1975; **18:**129.
34. Newhouse ML. Epidemiology of skin disease in an automobile factory. *Br. J. Ind. Med.* 1964; **21:**287.
35. Ryecroft RJG. Occupational Dermatoses. In: Ward Gardner A (ed.) *Current Approaches to Occupational Medicine.* Bristol: Wright, 1979.
36. Cronin E. *Contact Dermatitis.* Edinburgh: Churchill Livingstone, 1980.
37. Hodgson G. The hazards of cutaneous lubricant. *Trans. Soc. Occup. Med.* 1968; **19:**9–15.
38. Chatburn GR, Baker MP and Raffray M. Use of a data base management system for handling toxicity assessments of chemical materials in a motor company. *Med. Inform.* 1985; **10/2:**129–136.

14. OCCUPATIONAL HEALTH PROBLEMS IN HEALTH CARE WORKERS

J. A. Lunn

INTRODUCTION

The exhortation 'physician heal thyself' from St Luke, the patron saint of physicians, approximately 2000 years ago has by and large, gone unheeded. Doctors and nurses have been notorious in the neglect of their own health and that of their colleagues. Perhaps this neglect is a form of self-defence. In coping with illness in others it may be necessary to assume a mantle of bodily infallibility. Be that as it may, health care today requires a complexity of technology and expertise from many other professional groups and workers who, to a greater or lesser extent, are subject to occupational ill health. Their degree of risk, especially from infectious diseases, varies considerably according to the prevalence in the surrounding population. In most western countries the incidence and prevalence of tuberculosis for example are now so low that health care workers in these countries are virtually at no greater risk than their general populations. This contrasts with health care workers in Asian and some African countries where there is a high incidence of occupationally acquired tuberculosis.

No short review of this subject can be exhaustive. Infectious diseases remain serious risks and a cause of great concern. Rather more space has been devoted to them therefore. Hazards from new technology and methods of treatment are becoming increasingly serious threats to health. These are also discussed at some length. It would be impossible to obtain a perfect balance and sense of proportion for even the main health risks. These must inevitably vary from urban to rural environments and from country to country. The order in which the occupational health issues are discussed does not necessarily reflect their relative importance.

PSYCHOSOCIAL PROBLEMS

Doctors, nurses, dentists and many other health care workers suffer stress and anxiety which is often specifically related to their occupation. Psychological problems are probably the most prevalent form of morbidity. Precise measurement and observation are difficult. It is not possible to give figures for the overall prevalence. Certain statistics are available regarding suicide rates which for doctors, dentists and other workers are significantly higher than for the comparable occupations outside the health care professions. This is particularly true for doctors and dentists in the UK. Female physicians in the USA, however, although having an above-average suicide rate, do not have a higher rate than that for women in comparable professions.[1] It is necessary to look a little critically at the causes and to attempt to distinguish between stress and anxiety which is specific to occupation on one hand, and on the other, anxiety which relates to adverse factors generally found in many other organizations and occupations. The personalities of those entering medical, dental and nursing professions may be an important issue. There may well be a degree of self-selection of those with an introspective and potentially depressive personality. Ease of access to the means of committing suicide may be an equally important contributing factor. When coal gas was readily accessible in the UK it accounted for a high proportion of suicides. When it was removed there was a reduction in the total number of suicides thus suggesting that ready availability of the means of self-destruction will affect the number of suicides. Availability of drugs to health care professionals therefore may be an important reason for their high suicide rates. Sickness absence rates have some relationship to morale and motivation. Variable rates can be observed amongst student nurses depending on the type of ward or department in which they are working. Intensive care units often have high sickness absence rates. A superficial assessment might attribute the reason to the obvious stress of dealing with critically ill patients. In practice this is often found not to be the reason. The greatest cause of stress, tension and frustration for nurses in these units is often their lack of involvement in discussion and decision making about critically ill patients. Disparity between decision making and responsibility for carrying out the consequences is a common cause of stress. Many hospitals are large and have unwieldy administrative processes which are usually ill-defined to the majority of those who work in them. Many workers suffer much stress and frustration when they have major responsibilities with disproportionately little authority. There may be insufficient consultation about important aspects of work, especially plans for changing work patterns. Shift rotas may be arranged without adequate thought for minimizing the consequences of a disturbed circadian rhythm (*see* Chapter 7). Non-medical staff, whose work is essential for the overall well-being of health care institutions, often fail to obtain sufficient recognition of the value of their contribution. Although non-specific issues deriving from poor organization account for a great deal of stress, certain activities are specifically stress inducing. Student nurses may for example develop their first stress or depression when working in psychiatric wards. The effect of management processes on sickness absence rates has already been mentioned. It is also worth mentioning that individual attitudes to work can also play a significant role. A review of sickness absence amongst student nurses has shown that those who eventually gave up their training, before they did so, had three times the number of days of sickness absence compared with those who continued—although *bona fide* illnesses were not the reasons for them giving up their training.[2]

Alcohol abuse and dependency has an increased incidence amongst doctors. Studies in the USA suggest that up to 10 per cent of doctors may have an alcohol-related problem and that such dependency is the commonest cause of loss of license.

Whatever the aetiology of psychological problems in health care workers, sufferers may often wait too long before help is obtained. Other workers often fail to recognize the problem. The sufferers may sometimes be regarded as being just not interested in their work. This can sometimes lead to even greater pressures being put on them. It becomes more difficult to deal with stress and anxiety problems when the sufferer is a senior member of staff. Doctors frequently fail to obtain treatment early enough. In recognition of this, a National Counselling and Welfare Service for Sick Doctors has been set up in the UK.[3]

EXPOSURE TO INFECTION

In 1984 a working group convened by the WHO regional office for Europe reviewed occupational hazards in hospitals. The report of the convention emphasized particularly that *fear* of contracting serious infections at work remained uppermost as a hazard for most health care workers.[4] Sometimes the degree of anxiety may be disproportionate to the actual risk. This anxiety can be increased by inadequate knowledge. It is important therefore for occupational physicians and nurses to be well versed on the occupational health risks from infection. Sound knowledge will be the most effective way of minimizing fear. Some of the more serious infectious diseases, with high mortality rates, are fortunately not very contagious for those administering medical or nursing care. Some of the West African fevers and AIDS illustrate this point—although understandably the seriousness of these diseases causes considerable anxiety for those responsible for treating them. Diseases which are transmitted to staff by contamination or inoculation do not necessarily cause the greatest risk to clinical staff. The incidence of serum hepatitis has been shown to be greatest among laboratory workers and postmortem attendants.[5] Certain relatively minor upper respiratory tract infections are more prevalent amongst nurses early on in their training or when they work in children's wards.

Hepatitis B

The outbreaks of hepatitis following inoculations amongst conscripts to the armed services during World War II led to the realization that the virus could be transmitted from symptomless carriers by inadequately sterilized needles and syringes, The term 'serum hepatitis' was commonly used. In 1972 particularly severe outbreaks occurred amongst health care workers in renal dialysis units in Edinburgh and Manchester. There were several fatalities. It was then realized that careful screening of staff and patients was necessary. The unfortunate experiences of that time have subsequently been avoided. The prevalence of carriers of the hepatitis B virus varies from approximately 1 per 1000 in European populations to 1 or more per 100 in some African and Asian countries. Certain individuals are more likely to be carriers, especially drug addicts and male homosexuals. Although many body secretions and fluids have been found to contain the virus, the most common mode of transmission is by blood. Inoculation injuries by contaminated needles and scalpels create a significant risk as does contamination of skin abrasions and wounds by infected blood. Contamination of intact skin by blood or other body fluids does not appear to create a significant risk of transmitting the virus.

The World Health Organization review of hepatitis B showed amongst hospital staff prevalence rates 3–6 times higher than those of the general population. Medical laboratory technicians in the United Kingdom have the highest incidence for health care workers with a rate for the years 1975–1979 of 0·36 per 1000 followed by 0·31 per 1000 for staff working in mentally subnormal institutions. Rates for surgeons and dentists for these years were 0·19 and 0·17 per 1000 respectively. The majority of countries reviewed by the World Health Organization showed that laboratory workers had the highest risk of contracting the disease.[6]

Prevention

Careful screening of patients in renal dialysis units has done much to reduce the hazards. It is usual practice to screen staff in these units for hepatitis B antigen at 3-monthly intervals although viral transmission by staff has not been identified as a likely method of spread.[7] Meticulously careful clinical practice is important at all times but especially so when dealing with patients identified to be in high-risk categories. Laboratory specimens from such patients must be flagged with biohazard labels. Contamination by blood from carriers must be avoided. In the event of an accident occurring, either by splashing on to broken skin or by inoculation, substantial protection can be given by specific immuno-gammaglobulin within 7 days of the risk. Hepatitis B vaccine has proved effective and should be offered to all health care workers who are at greater risk from infection. A course of vaccine can be started when the initial gammaglobulin is given. When this is done it is not necessary to give the usual second injection of gammaglobulin (*see* Immunization Guidelines and Schedules, later). Laboratory technicians, doctors, dentists and nurses in institutions for the mentally subnormal, sexually transmitted disease units and drug addiction clinics should have the vaccine made available. Unfortunately early unfounded concern that the vaccine might transmit AIDS has resulted in a low uptake of this valuable preventive measure. In the United Kingdom hepatitis B is a prescribed industrial disease for health care workers. Many other countries give compensation for this disease when occupationally acquired.

Acquired immune deficiency syndrome, AIDS

It is generally accepted that this is an entirely new disease. It was first recognized as such in the United States in 1981, although it was evident that cases had been occurring there since 1978. In the United Kingdom the first cases were diagnosed in late 1982. The dominant feature of the disease is the loss of a major element of the normal immune system responsible for conferring protection against infection. There is also a reduction in resistance to certain tumours—Kaposi's sarcoma and lymphoma in particular. The aetiology remained unknown until 1983 when a virus—termed lymphadenopathy virus (LAV) was first isolated in France from AIDS patients. An identical virus has since been isolated in the United States and has been termed Human T-cell Lymphotropic Virus type III (HTLV III). This is now known as Human Immunodeficiency Virus (HIV). The majority of sufferers have been male homosexuals, but haemophiliacs and recipients of blood transfusions have also contracted the disease from contaminated blood or blood products. Intravenous drug abusers are also at risk. Transmission by heterosexual relationships has also been recognized. However, epidemiological evidence indicates that the greatest risk exists amongst male homosexuals, bisexuals and drug abusers. Out of a total of 241 cases of AIDS reported up to October 1985 in the United Kingdom, 231 were male of whom 214 were homosexual or bisexual. There were 134 deaths.[8]

The issue of major concern to health care personnel is what infectious risks arise from dealing with AIDS patients or their pathological specimens. The publicity given to the disease by the world's newspapers, radio and television has added enormously to the anxiety and concern of all staff who have responsibility for the care of these patients. Fortunately there is increasing evidence that the virus is of very low infectivity and that staff are at minimal risk from patients. Certain facts are reassuring to this end. In the United States for the 2–3 years before it was known that the causal agent was a virus no particular precautions had been taken by health care workers when dealing with AIDS patients. There is, however, no evidence of the disease being transmitted. In addition, a review of 1758 health care personnel in the United States who had been exposed to blood from AIDS patients showed that 26 (1·5%) were seropositive for HIV antibody, and that all but three of these belonged to groups recognized to be at increased risk socially. Unfortunately, no pre-exposure tests were available for the three workers who may have had conversions after exposure.[9]

Precautions and protection

For health care personnel, unlike the majority of workers in nearly all other occupations, potentially serious occupational hazards have to be accepted because of the commitment to caring for patients. There is no alternative process or product to substitute. However, as has proved so effective in hepatitis B risks, meticulous application of good clinical practices ensures the elimination of most of the significant circumstances in which transmission of infection could occur. Clear guidelines have been laid down which emphasize the need for wearing gloves and plastic disposable aprons when blood is taken from an AIDS patient.[10] It is recommended that needles are not resheathed in their protective cap. This ensures reduction in the incidence of puncture wounds to fingers. If puncture wounds do occur they must be dealt with immediately by encouraging bleeding and by liberal washing with soap and water. These accidents and contamination of broken skin or mucous membranes must be reported as soon as possible. A sample of blood should be taken from staff sustaining these accidents and the serum stored. Some authorities recommend, in the light of the present knowledge and understanding of this condition, that further blood samples should be taken at intervals of 3 months up to 1 year after exposure. Because there is not at present a uniform interpretation of the finding of HIV antibodies, not all virological experts at present recommend testing these samples. However, in view of the low infectivity of the AIDS virus, testing is likely to result in reassurance that infection has not occurred. In addition, epidemiological evidence can be gathered only by acquiring more precise information. In the event of HIV antibodies being formed in a health care worker it is essential that a suitably experienced counsellor is available for the individual to consult.

A general recommendation for those involved in caring for AIDS patients, is that a serum sample should be permanently stored when they first come in contact with AIDS patients or pathological specimens from them. The occupational health records of those staff should be permanently retained. It is important that those who do volunteer to have their serum stored are guaranteed that no testing of the serum will be carried out without their written permission. Should HIV antibodies subsequently be found following an occupational risk of infection, the availability of the initial serum for testing will afford the opportunity to demonstrate an occupationally required seroconversion.

Tuberculosis

In many developed countries the incidence and prevalence of tuberculosis has declined steadily since the latter half of the 19th century. The introduction of BCG immunization and antibiotics which are effective against the tubercle bacillus has accelerated a trend already established in those countries as the result of improved social and economic factors. In the 20 years from 1961 to 1981 the annual incidence of pulmonary tuberculosis in the UK population fell from 0·42 to 0·12 cases per 1000. Unfortunately such dramatic improvements have not yet been achieved in countries in Africa and Asia. The risk to health care personnel is significantly greater where the incidence of open pulmonary tuberculosis is high. A recent review of this disease in hospital workers in the United Kingdom suggests that there is now very little added occupational risk of contracting tuberculosis in the United Kingdom. Nearly all cases identified were in workers who had recently joined the National Health Service and who had therefore almost certainly become infected elsewhere. This contrasts with the very real risks of infection to hospital workers in Africa and Asia.

Prevention and surveillance

Workers liable to be exposed to tubercle bacilli should be tuberculin positive. The World Health Organization criterion[11] for this is a reaction of 5 mm or more of induration from 5 tuberculin units given by the Mantoux method. Tuberculin-negative subjects should be given BCG and should not be allowed to be exposed to tuberculosis patients, pathological specimens and so on, until a satisfactory BCG reaction can be demonstrated approximately six weeks after the initial inoculation. It is important to protect not only doctors and nurses, but also laboratory workers, postmortem attendants and morticians. Dentists, because of their close proximity to patients, may be at greater risk. However, a review of the incidence of morbidity and mortality amongst American dentists failed to show any increase in risk. It should be borne in mind, however, that the low incidence among American dentists probably reflects the low incidence in the general population. Figures for other countries where tuberculosis is more prevalent are not unfortunately available. Those who are regularly exposed to tuberculous patients or pathological materials and fluids should have a chest X-ray annually. In the event of a worker who is tuberculin negative and who has not had BCG being exposed to infection, a repeat tuberculin test should be carried out six weeks later. Conversion to a positive response during this interval suggests the possibility of early infection and is usually an indication for starting antituberculous treatment.

Tuberculosis is a prescribed disease for health care workers in the United Kingdom and compensation for tuberculosis as an occupationally-acquired condition is given in many other countries.

Meningococcal meningitis

This serious and often fatal condition requires prophylactic antibiotics for staff who have been in close and prolonged contact with the patient. Rifampicin 600 mg b.d. for two days should be given within 48 hours of exposure. Alternatively, a highly purified vaccine containing mucopolysaccharide extracts of meningococcal strains A and C will confer virtually immediate protection for staff at risk. A vaccine against strain B has not yet been produced, so that in countries where strain B predominates—including the United Kingdom—the vaccine is inappropriate. Fortunately, in Africa, South America and Asia strains A and C predominate and the vaccine in these areas is particularly valuable as a prophylactic.

African haemorrhagic fevers—lassa, marburg-ebola

These serious viral diseases have a differential degree of infectivity for health care workers. There is no known prophylaxis. Strict isolation is imposed upon patients with suspected or confirmed disease. Secondary spread is low for lassa but much greater for marburg-ebola fevers.

Crimean Congo haemorrhagic fever

This viral infection, endemic in Africa and parts of Asia, poses a serious risk to staff. There is no known prophylaxis. A recent report from Africa indicated the nature and degree of risk. An initially undiagnosed case admitted to hospital with sore throat, fever and myalgia resulted in seven members of the staff contracting the disease, one of whom died.

Rubella*

Staff of child-bearing age should be given rubella vaccine if found to be seronegative. All staff, whether male or female, who work in antenatal units should be seropositive to avoid transmission of the virus to susceptible patients during the first two to three months of pregnancy.

Poliomyelitis*

Although now rare in most temperate climates, it is still relatively common in tropical areas. The simple and effective oral vaccine should be given to all staff.

Typhoid*

This food- and water-borne infection should not constitute a risk to staff if their clinical and laboratory techniques are good. However, it is generally recommended that laboratory staff should be given typhoid monovalent vaccine.

Tetanus*

It is probably good sense for all staff to be protected, but maintenance workers, plumbers and gardeners should have priority.

Herpetic whitlow

These viral-induced lesions will occur on nurses handling upper respiratory tract secretions containing the virus. Postoperative care of neurosurgical patients is a common source of infection. It is important to recognize these characteristically vesicular lesions because resolution without treatment is the rule, and interference usually creates complications with secondary bacterial infection.

Chronic paronychia

A low grade but persistent form of paronychia occurs in nurses. Trauma often from repeated scrubbing up seems the commonest aggravating factor with a superimposed chronic infection with *Candida*

* *See* Immunization Guidelines and Schedules, later.

organisms. It is commonest amongst theatre staff and may require the sufferer to be transferred to alternative work.

Scabies

Although this condition may not strictly be classified as an infection it is convenient to mention it here. It often goes unrecognized in patients admitted to hospitals for general medical or surgical conditions. Staff who come into direct contact with patients are at risk.

CHEMICAL HAZARDS

Cytotoxic drugs

Antineoplastic drugs as a group include substances which are chemically unrelated. They are capable of inhibiting the growth of tumours by killing or arresting the growth of living cells. During the last 10–15 years these drugs have played an increasingly important part both in the treatment of some neoplastic conditions, and as an immunosuppressant role for organ transplantation. Although often used in specialist oncology units, there has been an increasing trend for them to be used in wards and even in domiciliary situations. They are more commonly given by injection although some are administered in tablet form. It is during the preparation and giving of these drugs by injection that staff may be at risk from inhalation, ingestion or by skin or eye splashes. The potential health hazard derives from the carcinogenic and mutagenic properties of these drugs. Doubts have been raised as to whether these drugs are a health risk in humans when used in therapeutic quantities. A review by the Health and Safety Executive in the United Kingdom failed to find evidence of absorption of these substances by health care workers.[12] However, a note of caution should be sounded because cyclophosphamide has been detected in the urine of nurses administering the drug, and a recent case/control study from Finland has shown that fetal loss is twice as frequent in women exposed to cytotoxic drugs in the first trimester.[13]

Precautions

Cytotoxic drugs should be prepared by trained pharmacists only, in properly ventilated exhaust cabinets. Staff who administer the drugs must be experienced and specifically advised of the hazards. PVC gloves, goggles, long-sleeved gowns and face masks must be worn for appropriate protection. Spillages must be removed immediately and contamination of the skin and eyes must be copiously washed away with water. It is recommended that a permanent occupational health record is kept for all who prepare and administer these drugs. The cells of a developing fetus may be unduly sensitive to cytotoxic drugs and in the absence of adequate proof to the contrary, it is recommended that pregnant women, or those wishing to become pregnant, should not work with or administer the drugs.

Anaesthetic gases

The commonest inhaled anaesthetic gases, halothane and nitrous oxide, have been shown to produce background levels in operating theatres in the United Kingdom from 0·1 to 60 and from 30 to 3000 parts per million respectively. A study of Russian anaesthetists revealed an increase in spontaneous abortion, although the specific relationship to anaesthetic gases as opposed to operating theatres in general was inconclusive. Later studies in North America, Great Britain and Scandinavia produced evidence to suggest that there is a specific association between increased spontaneous abortion and exposure to anaesthetic gases in female anaesthetists and anaesthetic nurses. High concentrations of nitrous oxide are probably responsible. Interference with vitamin B_{12} metabolism has been shown to occur in subjects exposed to high nitrous oxide levels. Reports of subacute combined degeneration of the cord in 15 subjects, 14 of whom were dentists, who had been abusing nitrous oxide suggests inactivation of vitamin B_{12} as the probable cause. Studies of dentists have shown direct evidence that occupational exposure to nitrous oxide levels ranging from 159 to 4600 parts per million may cause measurable changes in bone marrow secondary to impaired synthesis of deoxyribonucleic acid as a result of depression of vitamin B_{12} activity.[14] Dentists are often exposed to much higher nitrous oxide levels than staff in most operating theatres. The American National Institute of Occupational Safety and Health has recommended[15] a maximum level in operating theatres of 2·0 parts per million for halothane and 25 parts per million for nitrous oxide. These levels are unlikely to be reached without the use of effective scavenging machines. If these levels are not exceeded it is unlikely that there is an increased risk of spontaneous abortion. However, it is only good sense to recommend that theatre nurses or female anaesthetists who are pregnant or wishing to become pregnant should not work in an environment containing anaesthetic gases.

Ethylene oxide

Ethylene oxide is a colourless gas at room temperature. It is used for sterilizing medical materials and equipment. The penetrative power of this gas makes it particularly useful in this respect. Its less serious effects on human subjects are local irritation of eyes and respiratory tract. Of far greater importance, however, are the mutagenic and carcinogenic properties of the gas. Because of its value for sterilizing in hospitals there has been a reluctance to use substitutes. However, its continued use presents a serious potential hazard and it should not be used unless there is no other satisfactory method available.

Formaldehyde

Many studies have been carried out to determine whether or not formaldehyde is carcinogenic. Concern arose because of experimental induction by formaldehyde of nasal cancer in rats. A recent study from Germany[16] is especially reassuring in finding no evidence that formaldehyde exposure induces a risk of cancer in humans. Local skin irritation and dermatitis can occur. Asthma may be induced in susceptible subjects and attacks can persist for some time after removal from exposure. Medical students who wear contact lenses usually are forced to wear ordinary glasses when dissecting bodies preserved in formalin because of severe eye irritation from fumes.

Acrylic cements

These are used in orthopaedic surgery for joint replacement and in dentistry. Common constituents are methyl methacrylate and N,N-dimethyl-*p*-toluidine. Exposure to high concentrations of the vapour in poorly ventilated areas for long periods may result in drowsiness, headaches, or mental confusion. Local skin irritation may occur from careless handling. Carcinogenic and mutagenic hazards from long-term exposure have been suggested but experimental studies in animals have not confirmed these risks.

Mercury

Instrument repair technicians in hospitals may be exposed to mercury vapour if adequate care is not exercised when repairing sphygmomanometers. Scattering of mercury globules on bench surfaces and floors constitutes the usual process of exposure. The main group of workers at risk from mercury, however, are dental practitioners and dental mechanics. Mercury has been widely used in restorative amalgams. There can be exposure to vapour during the preparation of these and to particles of metallic mercury during the drilling out of old fillings. Various studies have shown clear evidence of mercury absorption by detection of raised levels in blood and urine. The method of working and general standards of housekeeping have also been shown to make a significant difference to the amount of mercury absorbed.[17] Dentists and their technicians appear to be at minimal risk when good standards of housekeeping are maintained. A basic requirement, in addition to good clinical technique, is the avoidance of floor surfaces which contain cracks and crevices in which spilled mercury can accumulate. Mortality rates for dentists interestingly enough have shown a significantly reduced rate for nephritis and nephrosis (SMR-38) which must be reassuring evidence against chronic toxicity.

RADIATION

Ionizing radiation

Closed sources

Diagnostic radiology is the main closed source of ionizing radiation. Although of great potential hazard the safety standards of modern diagnostic departments ensure protection from radiation. The expertise of a medical physicist is usually necessary to confirm whether safety standards are adequate. It is essential that such advice is sought to confirm that the lead lining of walls, doors and protecting screens is adequate. For individuals, lead aprons and gloves must be available to avoid scatter from X-rays. All staff working in diagnostic units are required to wear personal dose meters so that possibility of accidental exposure can be monitored. Radiological safety committees play an important role and it is essential that these are established to ensure proper practice. All diagnostic radiological departments must have safety guidelines and local rules. There are specific officers who must be appointed in these committees. The radiological safety and supervisory medical officers are two important posts. The latter usually falls to the occupational health physician, who can play a very valuable role in ensuring that good practice is carried out and that safe and logical policies are made regarding staff protection. The doctor who is appointed must be ready to seek advice from medical physicists. The greatest danger to staff usually arises from old and deficient equipment. Portable X-ray machines can be potential hazards especially when used in emergency circumstances. Image intensifiers are often used for emergency operations outside routine hours. Staff, other than radiographers, may be used at such times to operate these machines. It is important, however, that only properly trained radiographers should operate image intensifiers.

There has been a tendency to carry out routine annual blood counts on radiographers and radiologists. It seems good sense to carry out an initial one, but routine annual checks are not necessary. Occupational health records of those who work with ionizing radiation should be permanently retained. There may be a differing opinion among the experts about pregnant women working in X-ray departments. In well regulated units there should be no excess radiation to which a fetus could be exposed. However, pregnancy often generates anxieties because of known and unknown factors which could cause fetal damage. It is probably wise to accept that pregnant women should not be required to work in diagnostic units. Clear and detailed guidelines for medical surveillance of workers against ionizing radiation are given in the Ionizing Radiations Regulations 1985.[18]

Open sources

Radiopharmaceuticals used in clinical physics laboratories and metabolic units are the common open sources of ionizing radiation. Their use for diagnostic purposes gives rise to minimal risk to staff and in these circumstances patients can be placed in open wards. When used for therapy, radiopharmaceuticals give rise to much greater quantities of radiation. Patients receiving radiopharmaceuticals for therapeutic purposes must therefore be appropriately isolated in screened side wards. All body fluids must be dealt with and disposed of correctly. It is essential for staff to know the local rules which govern safe practice. Provided staff comply with the required safe practices, health hazards will not arise. It is the responsibility of the supervisory medical officer to be aware of all areas where radiopharmaceuticals are used and to ask the radiological protection adviser to inspect and report on any previously unknown areas. Radiopharmaceuticals should be available only from one central source so that the issuing and use of them can be recorded and controlled. Although cleaners and portering staff should not be at risk, some training sessions for these groups in hospital may be necessary and helpful. Domestic staff in wards where patients are receiving radiopharmaceuticals should be fully conversant with the safety rules.

Non-ionizing radiation

This is used in physiotherapy departments in the form of shortwave diathermy and ultrasound, and as laser beams in many applications in medical practice. Although there have been suggestions that some of the radiation used in physiotherapy departments may have adverse effects there has been no evidence to justify concern. Ultrasound therapy has been considered a potential cause of hearing loss, but it does not appear to cause this in physiotherapists or in workers in sterilizing departments, where ultrasound may be used to clean instruments prior to sterilizing.

Laser beams

The use of laser beams for medical purposes has been a major development following their invention in 1960. The areas of medical practice where lasers are being used include gynaecology, otolaryngology, neurosurgery, urology, dermatology, and ophthalmology. The greatest risk from laser beams is to the retina from inadvertent exposure. Beams classified in categories 3B or 4 are sufficiently powerful to cause permanent destruction of the retina from an exposure of even a fraction of a second. Exposure can occur not only from a direct beam, but also from scatter when the main beam is reflected from shiny surfaces. In practice lasers used for medical purposes are well controlled and safe for the operator. It is important to realize the risk of a reflected beam to an assistant. In gynaecology there is the theoretical possibility of scatter from the surface of a speculum. In practice there is little or no possibility of this. It is important that assistants and nurses are aware of the hazard and that they do not put themselves at risk. A laser safety committee must be established, with the appointment of a laser safety officer and supervisory medical officer. Only a nominated person who is skilled in the use of the equipment must use a class 3B or 4 laser. A register must be kept of those nominated and of all users of the equipment. Suitable premises must be made available where outsiders cannot inadvertently enter when laser operations are in progress and where scatter from laser beams cannot occur. Medical and nursing staff should not be exposed to retinal damage if adequate safe practice is ensured. It is recommended that all those who use lasers should have an initial eye examination by an ophthalmologist.[19] The result should be recorded in permanently retained occupational health records. This will form a base record for reference should any untoward event occur in the future.

BACK INJURIES AND TRAUMA

In industry as a whole, back pain and disabilities occur predominantly in occupations where there is heavy lifting. In nursing there is not only heavy lifting but often, in addition, a major mechanical disadvantage when patients are lifted or turned in bed. Figures for the incidence and prevalence of back disabilities amongst nurses are not available on a national scale. Various studies in individual health authorities in the United Kingdom have been carried out, however. The subjective nature of the problem itself and the inaccuracy and lack of standardization of morbidity and sickness absence statistics between studies make comparisons unreliable. Back injuries which result in clear symptoms of an intervertebral disc lesion usually require absence from work and in this respect some statistical assessment is possible. Less serious back problems, on the other hand, may or may not result in absence from work. It is a very subjective matter how any individual responds to such symptoms. Morale and motivation obviously influence responses to less acute symptoms. Nurses are aware from their clinical practice that some back problems are serious and may require an operation. They are not always aware that the majority of problems get better with simple treatment, and there may be unnecessary anxiety about the potential seriousness of the symptoms. Experience shows that the risk of developing back pain can be reduced by training in lifting techniques. Improved tone of back muscles and spinal mobility will result in fewer injuries. Classes for back exercises are will worth considering as an additional measure for reducing injury. When nurses of unequal height lift patients there is some evidence to indicate that undue strain occurs on one or both of them. It has also been suggested that the usual uniform worn by nurses restricts correct positioning of the body in some crucial acts of lifting. If nurses wore trousers and not dresses this would probably reduce back problems. Mechanical lifting aids should be available whenever practicable. A recent survey suggested that lack of proper training in the use of lifting aids was a major reason why they were not always used. It should not be forgotten that porters and ambulance personnel may develop back problems. Careful pre-employment selection and training will help to reduce back injuries in these groups.

Patient assaults

The level of violence in many societies today has risen and staff in hospitals have not been exempt from this trend. Casualty departments experience most violence in general hospitals, but nurses in psychiatric units also suffer assaults by patients. It has been shown that assaults on staff by psychiatric patients increase when there are frequent staff changes. Patients clearly become calmer and less aggressive with staff they get to know and trust.

Sharps injuries

These commonly occur from hypodermic needles. Nurses sometimes stab themselves when replacing a used needle in the plastic cover. Because of the risk of puncture wounds many control of infection committees recommend that resheathing of needles should be avoided. Other sharps injuries occur to domestic and portering staff when needles are incorrectly discarded into waste bins or sacks. Workers in central sterilizing departments are also at risk from scalpel blades left in used dressing packs. Sharps boxes should be provided wherever needles are used. These boxes should be carefully evaluated. Not all afford adequate protection. Boxes which are made of cardboard can allow needles to penetrate their walls when they are packed too tightly, or if the cardboard becomes damp and softened.

Other injuries to health care workers arise from slipping and falling on wet floor surfaces—often to be found in hospitals. Heavy gas cylinders falling on inadequately protected feet are a source of injury mainly to portering staff.

SUMMARY

Advances in methods of treating diseases have in some instances introduced new health risks to staff. The development of cytotoxic drugs, radiopharmaceuticals and laser beams for treating malignant disease are examples which emphasize this point. Improvements in treating and preventing some infectious diseases have on the other hand reduced risks to staff. Effective use of vaccines has given substantial or complete protection against previously serious or fatal infections—poliomyelitis for example. Research is at present progressing on the production of a vaccine to afford protection against AIDS. The present state of knowledge suggests it may be at least five years before it will be available. Psychosocial problems and stress are difficult to quantify, but remain at times important consequences of caring for sick people. There may be insufficient training and help for staff who care for incurable patients. Some doctors and nurses develop an undue sense of failure when their patients do not recover. They may need help to understand that the *quality* of the care they give to terminally ill patients is the measure of achievement in such cases. There are many varied challenges for occupational health nurses and doctors not least of which is being under the critical eyes of their fellow health care professionals. There is a need to develop occupational health services to meet the many challenges. Clinical practices and procedures require to be co-ordinated and consistent and more training facilities are needed. These must be designed to meet the requirements not only of full-time doctors and nurses, but also those of the many part-time practitioners.

IMMUNIZATION GUIDELINES AND SCHEDULES

Introduction

Health care workers may require protection against infectious diseases in addition to that given to the general population. The degree of risk will vary and local assessment will be necessary to decide on the need for any particular additional immunization.

It is important to note some general principles. Immunizations should not be given to anyone who has an acute febrile condition. Live viral vaccines should not be given to those with impaired immune responses, or to pregnant women especially during the first three months of pregnancy. Two or more live viral vaccines as a rule should not be given at the same time but spaced by an interval of at least three weeks. If it is imperative to give them at the same time they should be given simultaneously in separate sites. They should not be given within three weeks of BCG. Anaphylaxis from vaccines is extremely rare, but adrenaline 1 in 1000 should always be readily available. Proper records of all immunizations must be made to include manufacturer, batch number, site, method and date. A personal immunization record card should be provided.

Cholera

This vaccine contains heat-killed cholera organisms. It is indicated in many hot climates. It is probably the least effective of the available vaccines, and to avoid risk good hygiene is essential. The primary course consists of two injections at an interval of preferably one month. If time is limited the second injection may be given 7 days after the first. Booster injections must be given every 6 months where there is a continuing risk of exposure. The vaccine may be given subcutaneously or intramuscularly using 0·5 ml, or by the intradermal method using 0·2 ml of the vaccine. Reactions after the intradermal method are usually significantly less and it is well worth remembering this when there is a history of a marked reaction to 0·5 ml of the vaccine given subcutaneously or intramuscularly.

Diphtheria

This vaccine is recommended only for staff exposed to cases of diphtheria. It should be given to those who have no immunity as demonstrated by the Schick test. This test consists of an intradermal injection of 0·2 ml of Schick test toxin into the anterior surface of the left forearm and 0·2 ml of Schick test control into the equivalent site on the right arm. Readings should be made at 24–48 hours and at 5–7 days. No response at either site indicates immunity. An erythematous response at the test site and not the control indicates susceptibility. An equal response at both test and control sites indicates a false positive response and therefore immunity. If both sites show a reaction, but the one on the test arm is significantly larger, a true positive response has occurred in addition to a false one indicating that the subject is susceptible.

The primary course consists of three injections of 0·5 ml of the adsorbed diphtheria vaccine, the second injection being 1 month after the first and the third injection 6 months later. A booster injection of 0·5 ml is usually sufficient to restore immunity (as checked by the Schick test) after an initial primary course.

Hepatitis B vaccine

This vaccine costs at present approximately £70 for a course. Use should be limited to staff at particular risk. This will include laboratory workers, and those working in drug dependency and sexually transmitted diseases units. If an individual is known to be hepatitis B surface antigen or antibody positive it is not necessary to give the vaccine. However, it is not medically indicated or financially justifiable to screen potential recipients in order to eliminate the 1 in 1000 who would not need the vaccine. The primary course of vaccine consists of three injections each of 1 ml. The second should be given 1 month after the first and the third injection 5 months after the second. It is important to check whether adequate antibodies have been formed at approximately 3 months after the third injection. If the antibodies have not risen to a satisfactory level a further injection of 1 ml of vaccine should be given. Booster injections of 1 ml should be given at approximately 5-yearly intervals after a satisfactory primary course.

Hepatitis B immunoglobulin

When unprotected staff receive an inoculation injury or contamination of a wound or conjunctiva from body fluids from an infected patient, a considerable measure of protection can be given immediately by injections of specific immunoglobulin. The first injection of 500 mg should be given as soon as possible but within 7 days of the risk. The second injection, also of 500 mg, should be given a month later. On account of their volume (5 ml), injections must be given intramuscularly into the buttock and the patient kept under observation for 20 minutes afterwards because of the possibility of an anaphylactic reaction to the injected protein. Specific immunoglobulin is very expensive and at times difficult to obtain. It should be given only to those who have had a clearly identifiable risk of infection. The material is obtainable from Public Health Laboratories from whom advice on indications for its use can be sought.

There is no contraindication to starting hepatitis B vaccine at the same time as the first injection of immunoglobulin. In fact when a vaccine course is started at this time it is not necessary to give the second injection of immunoglobulin.

Poliomyelitis

The vaccine is taken orally and contains live attenuated strains of the three types of poliovirus. The vaccine contains traces of polymyxin, penicillin, streptomycin or neomycin. Those known to be hypersensitive to these antibiotics are recommended not to be given the vaccine. However, there are such small traces of the antibiotics present that in practice hypersensitivity reactions are unlikely. Where there is a definite risk of exposure to poliomyelitis the vaccine should be given.

A basic course consists of three doses, each containing three drops of vaccine which can be taken on a lump of sugar, on a spoon, or straight into the mouth. The interval between the first and second doses should be 6–8 weeks and 4–6 months between the second and third. A reinforcing dose at approximately 10-year intervals is sufficient to maintain immunity. A further reinforcing dose would be advisable for anyone about to be exposed to a known case.

All health care workers would be immunized.

Rabies

This safe and effective vaccine should be considered for health care personnel who intend working in areas where rabies is endemic. The basic course consists of two doses of 1 ml each given intramuscularly at an interval of 4 weeks. A reinforcing dose should be given a year later and thereafter every 2 years.

Satisfactory immune response has been shown to occur when the vaccine is given intradermally using a 0·1 ml dose. This results in considerable cost saving. The vaccine must, however, be used within 2 hours of being reconstituted and therefore the appropriate number of people to be immunized must be organized to attend within this period.

Rubella

The vaccine contains a live attenuated virus. Occasionally, 9–10 days after inoculation, a generalized reaction can occur consisting of pyrexia, joint pains and glandular enlargement. Less commonly a rash will develop. The vaccine should be given only to seronegative subjects. The Cendehill strain should not be given to anyone known to be hypersensitive to rabbits or neomycin. The RA 27/3 strain should not be given to those who are hypersensitive to neomycin or polymyxin. All staff who could infect susceptible patients in early pregnancy should have their rubella antibody levels checked. Those who are negative, whether male or female, should be given rubella vaccine. All the female staff of child-bearing age should be immunized if seronegative. A single dose of 1 ml confers immunity in the majority of cases.

Tetanus

This vaccine contains adsorbed tetanus toxoid. The basic course consists of three injections each of 1 ml. The interval between the first and second should be 6–8 weeks and 4–6 months between the second and third. One reinforcing dose should be given 5 years later and thereafter at 10-yearly intervals. Too frequent immunization will cause hypersensitivity reactions. If hypersensitivity to this vaccine is suspected on the history a test dose should *not* be given. Even this volume may be sufficient to cause a severe reaction in very sensitive subjects. The decision whether or not to give more tetanus vaccine must be made on a general assessment of the history. All staff should be protected against tetanus.

Tuberculosis

BCG is a live vaccine containing attenuated strains of tubercle bacilli. It should be given only to tuberculin-negative subjects. A negative tuberculin reaction is indicated by a response of less than 5 mm of induration to 0·1 ml of 1 in 1000 tuberculin given intradermally, or by a grade 0 or 1 reaction from the Heaf test. It is important to remember that the tuberculin response may be suppressed during viral illnesses, especially glandular fever. Tuberculin testing should also not be done within 3 weeks of giving a live viral vaccine.

Freshly reconstituted vaccine, 0·1 ml, should be given intradermally in the left arm at the site of the insertion of the deltoid. It should not be given if the skin in this area is infected or eczematous. The injection site must be inspected 6 weeks later to confirm that the response has been satisfactory and to identify any adverse and persistent reaction. A scar from a normal reaction measures approximately 4 mm in diameter. With careful selection of subjects for BCG, adverse reactions should occur very rarely. When adverse reactions do occur, the technique of giving and the accuracy of tuberculin testing must be carefully checked. Immunity will have developed when a satisfactory reaction can be demonstrated 6 weeks after the initial inoculation.

Typhoid

This vaccine is now in monovalent form and the paratyphoid contents are not included. The latter gave poor protection and accounted for much of the adverse reactions. The basic course consists of 2 doses each of 0·5 ml given deep subcutaneously with an interval of 4–6 weeks. A reinforcing dose should be given at 3-yearly intervals It is recommended that workers in microbiological laboratories should be immunized against typhoid.

Acknowledgement

I would like to record my gratitude to Sonja Drummy for her help in preparing the typescript.

REFERENCES

1. Pitts FN, Schuller AB, Rich et al. Suicide among US women physicians, 1967–1972. *Am. J. Psychiatry* 1979; **136**:694–6.
2. Lunn JA. *The Health of Staff in Hospitals*. London: Heinemann, 1975; 44–6.
3. *National Counselling and Welfare Service for Sick Doctors*, 3rd floor, 7 Marylebone Road, London, NW1 5HH.
4. *Occupational Hazards in Hospitals*. Copenhagen: World Health Organization Regional Office for Europe, 1983; 13–15.
5. Dienstag JL and Ryan DM. Occupational exposure to hepatitis B virus in hospital personnel: infection or immunisation? *Am. J. Epidemiol.* 1982; **115**:26–39.
6. Pattison C. Epidemiology of hepatitis B in hospital personnel. *Am. J. Epidemiol.* 1975; **101**:59–64.
7. *Hepatitis and the Treatment of Chronic Renal Failure. Report of the Advisory Group 1970–1972.* London: Department of Health and Social Security, 1972.
8. *Communicable Disease Report*. London: Communicable Disease Surveillance Centre, 1985; 48.
9. *Recommendations for preventing Transmission of Infection with Human T-lymphotropic Virus Type III/Lymphadenopathy-associated Virus in the Workplace. MMWR* 1985; **34**:683–5.
10. *Acquired Immune Deficiency Syndrome. General information for doctors*. London: Department of Health and Social Security, 1985.
11. Control and prevention of tuberculosis: a code of practice. *Br. Med. J.* 1983; **287**:1083–1156.
12. *Precautions for the Safe Handling of Cytotoxic Drugs. Guidance Note MS21.* London: Health and Safety Executive, 1983.
13. Selevan SG, Lindbohm ML, Hornung RW et al. A study of occupational exposure to antineoplastic drugs and fetal loss in nurses. *N. Engl. J. Med.* 1985; **313**:1173–1177.
14. Sweeney B, Bingham RM, Amos RJ et al. Toxicity of bone marrow in dentists exposed to nitrous oxide. *Br. Med. J.* 1985; **291**:567–569.
15. Sharer NM, Nunn JF, Royston JP et al. Effects of chronic exposure to nitrous oxide on methionine synthase activity. *Br. J. Anaesth.* 1983; **55**:693–700.
16. Triebig G and Valentin H. Occupational health and the risks of formaldehyde: an assessment of the present position. *Pathologe* 1985; **6**:64–70.
17. The occupation of dentistry: its relation to illness and death. Bureau of Economic Research and Statistics. *J. Am. Dent. Assoc.* 1977; **95**:606–13.
18. *The Protection of Persons against Ionising Radiation arising from any Work Activity. Approved Code of Practice*. London: Health and Safety Commission, 1985.
19. *Guidance on the Safe Use of Lasers in Medical Practice*. London: Department of Health and Social Security, 1984.

15. NEW TECHNOLOGY IN THE OFFICE

D. G. Davies

INTRODUCTION

It is a striking characteristic of the human race that we come in such different shapes, sizes and colours, and that our personalities, attitudes of mind and behaviour also vary so much. One has only to look at attitudes towards the future to see this variation vividly illustrated. At one extreme we have the person who is excited at the future and all its unknowns and cannot wait to discover its secrets, whilst at the other extreme is the person who is full of foreboding and anxiety and would gladly remain in the present if they could. And most of us probably fit somewhere between the two.

The anxious view of the future has often been a hallmark of situations where new technology has been introduced into the office. The people affected may have read of possible health hazards and fear for their own health, or imagine that new technology means redundancy and fear that their own livelihood is at stake. Or again, they may have heard that the new technology will remove all the skills from their job and leave them with a mindless round of boring tasks. Worse, they may know someone who has experienced new technology and who has told them it is every bit as bad as they imagine. Any organization contemplating the introduction of new technology will ignore this phenomenon at their peril, and even more so the issues which give rise to it.

Health hazards—someone may say 'No such thing, they are a figment of anxious imaginations'. It is very difficult to make precise and generalizable comments about this problem area, because very little rigorous scientific research has been carried out on the mammoth scale that would be required to provide such statements. However, such work as has been done points to the following sort of picture. (In every case, unless stated, the comments refer in particular to visual display terminals and their associated equipment and visual display terminal (VDT) related tasks because concern about new technology has tended to focus on these units.)

COMMON CONCERNS AND THE ISSUES INVOLVED

1. Radiation

Many people believe that VDTs emit harmful levels of radiation. The UK Government's Health and Safety Executive conducted a radiological survey of VDTs, which came to the conclusion 'The radiation normally emitted from a VDT does not pose a hazard to operators either in the long or the short terms'. Cox's paper[1] is a helpful exposition on this matter.

2. Skin irritation

There is a fear that VDTs may cause skin irritation and allergies. There have been a small number of such cases, mainly facial rashes, reported and scientifically investigated. The net conclusion has been that VDTs in themselves do not appear to cause such problems, see Ch. 3, 4 in Pearce.[2] However, what may be a possibility is that some workers using VDTs in the presence of a great deal of static electricity may acquire contact dermatitis from airborne dust influenced by the static, see Ch. 4 in Pearce.[2] Dissipation of the static electricity is the simple solution in such situations. (*See also* Ch. 1 in this volume.)

3. Cataracts

It is claimed that using VDTs can cause cataracts. It is a fact that some radiant energy of the electromagnetic spectrum can cause cataracts in people.[3] However, the amount of energy needed to produce a cataract is very high indeed, far above that emitted by a VDT. Zaret[4] has reported cases of cataracts in VDT workers which he has studied and which he concludes were probably caused by using the VDTs. However, he states 'accumulating meaningful data about radiational injuries is extremely difficult. Indeed, to this date (1984) it has never been accomplished in a completely satisfactory manner'. The conclusion has to be that the issue is by no means clear.

4. Visual discomfort

It is felt by some people that whether or not VDTs cause damage to the eyes, they certainly may cause eye strain. Eye strain is a somewhat unfortunate expression; it is more appropriate to speak of 'visual discomfort' since the sort of symptoms which people complain of are eyes which water excessively, aching eyes, headaches and difficulty in focusing. Since these are actual conditions which the sufferer is experiencing, they cannot be denied! A number of studies have been carried out assessing the visual condition of VDT operators and other office workers doing more traditional work.[5,6] They have consistently shown a higher incidence of eye strain amongst the VDT operators.

What is not easy is diagnosing the cause of eye discomfort. It is caused by an interaction between factors relating to the people, and factors relating to the working environment. The personal factors include refractive error, convergence insufficiency, accommodation fatigue and binocular instability. These are helpfully discussed in Stone,[7] and at more length in Duke-Elder.[8] Factors in the working environment affecting visual discomfort include the task, the screen, the keyboard, the document and the lighting. *Fig. 15.1* provides a diagrammatic summary of these and the personal factors.

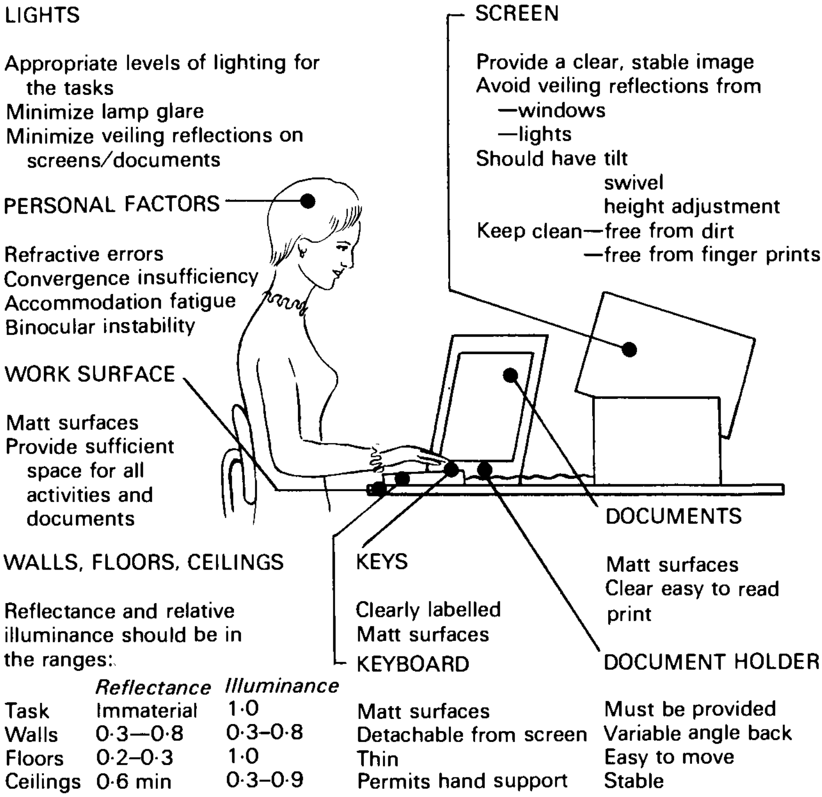


Fig. 15.1. Factors affecting visual performance, comfort and fatigue, indicating some desirable characteristics.

If the VDT operator has some form of disability, as mentioned above, of which they are unaware, this may become apparent when they first begin to use VDTs. This is because they are using their eyes to do a task which they have not done before, and the task has uncovered the existing but unnoticed weakness. This has proved to be the situation in many cases which Human Sciences and Advanced Technology Research Centre (HUSAT) has seen. Such events demonstrate the importance of providing eye tests of all people who are to use such equipment, to ensure that their eyesight is adequate, or, if not, to provide them with the necessary correction by means of spectacles or contact lenses to carry out the job concerned. In HUSAT's experience of these situations, it is very rare (less than 1 per cent) for anyone to require a special prescription to work at a VDT.

The visual well-being of the VDT user is affected by the task which they have to carry out, which depends critically on the VDT equipment they are provided with. The screen which is to be viewed must provide a clear, stable image. If there is flicker, jitter or other image instabilities, the possibility of visual discomfort or fatigue is greatly increased. Also the screen needs to be kept clean and free from dust and fingerprints. These all degrade the image which the person has to view. Care must also be taken when locating the screen so as to avoid reflections such as windows and light sources. These create veiling reflections, masking the image to be viewed. If possible, position screens at right angles to windows, between lights, or introduce a barrier between the screen and them, such as other furniture. Failing this, filters can be obtained to fit over the screen which greatly diminish the effects of these unwanted reflections.

The keyboard can create visual discomfort if the keys create mirror-like reflections on their top surfaces, or when they are not clearly labelled and cause the user to search the keyboard to find the required key. The keyboard and keys should have a matt surface, and all keys should be clearly labelled.

Sometimes the source document that a person is using can cause more problems than the VDT equipment. Paper with a glossy surface, for example, can be very difficult to read, due to veiling reflections, as can documents with low contrast print or poor character definition. Paper with matt surfaces should be used, with clear, easy-to-read print.

The visual environment is also important. To carry out reading tasks we require a good level of light: if the levels are low, then the reader will be presented with a more fatiguing task. If the levels are too high, then there may be glare which arises, either directly by means of the light sources concerned, or indirectly by high reflected light from paper, work or other surfaces. It is also probable that the contrast of the screen image will be seriously degraded, making the visual task more difficult and fatiguing. The reflectance values of the work surface, walls, ceiling and floor are significant. The work surface should have a matt surface and have low reflectance. The reflectance values for the other surfaces should approximate to ranges provided in the VDT Manual.[9]

5. Aches and pain

It is believed that the use of VDTs can cause aches and pains around the head, neck, shoulders, arms, hands, back and buttocks. Studies into VDT operators, for example Hunting[10] and Maeda,[11] have shown that operators have more complaints about these symptoms than do people carrying out more traditional office work. However, the cause of this higher incidence of complaints is not necessarily the new technology, but an interaction between the factors relating to the people and factors relating to the working environment. The personal factors relate to the fact that the body is built for movement, and that movement and use should be within certain physical constraints. Factors in the working environment include the chair, work surface, keyboard, screen and document and its location.

When designing or planning a VDT workstation the aim so far as working postures are concerned must be to provide the range of people involved with the facility of a variety of postures which are comfortable and permit efficient operation. This can be done by considering the critical elements in the workstation mentioned earlier and ensuring these are appropriately selected and correctly positioned.

The chair should provide the occupant with the support they require to the appropriate point. The seat should support the bulk of the body weight, principally through the buttocks. If the seat is too high and/or the cushion too soft then pressure will be exerted on the thighs, and this will restrict blood flow to the legs, as well as affect the nerves. If the seat is too low and/or the cushion is too hard this will cause excessive pressure on the bony points in the buttock. Unpleasant sensations may be felt in the legs or buttocks.

The simple way of knowing the correct seat height for any individual is that it should be approximately equal to the distance from the

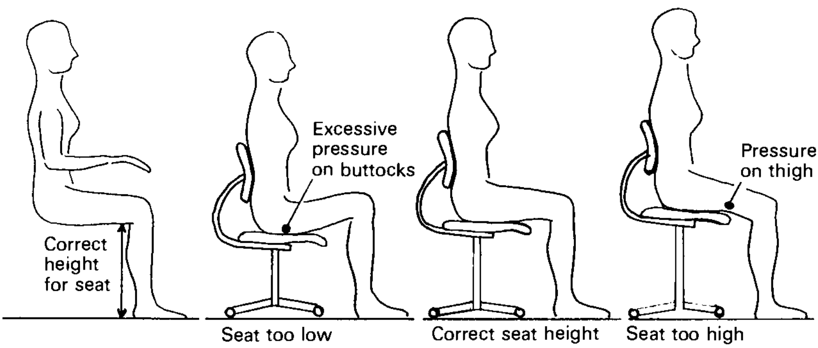


Fig. 15.2. Correct seat height.

back of the knee to the floor when they are sitting (*Fig. 15.2*). This means that it is essential that the chair has vertical height adjustment. It may also be necessary to provide a foot rest for those of shorter stature. This is necessary in order to stabilize their body, and allow the hands and arms to move freely with the minimum of fatigue. Whenever movements are made with the hands and arms, the rest of the body has to be kept stable. This is achieved by muscles around the spine, neck and shoulders remaining taut whilst the activity with the hands goes on. This is static muscular work, which cannot be undertaken for long periods without fatigue and unpleasant sensations.

When the seated person is operating a VDT, the back-rest should stabilize the body, thus relieving the muscles around the spine, neck and shoulder of this job. If it does not, then those muscles, being statically loaded, will inevitably become fatigued, and so the aches and pains will follow. If the back-rest is to do its job properly for everyone it too should be adjustable both vertically and for rake and should be padded and shaped to fit the full range of the body sizes to be encountered.

Just as the body muscles become fatigued with static work when they have to stabilize the body, so the muscles of the shoulder, arm and hand can similarly suffer when they have to position the hand to operate a keyboard. This means that support for the arm can be very helpful as can support for the hand (more of this later). The arm support can be provided by means of arm rests. When deciding whether to have arm rests on a chair or not, take into account other factors, such as does the person need to rotate in their chair to access other parts of the workstation, and can they do that if they have arms; can they position their chair under the work surface, or will the arms prevent this, and so on. The arm rests should be padded since the bones around the elbow region are not well padded with surrounding tissue.

The work surface on which the person is working should provide enough room under it to accommodate the person's legs. If it is lacking in depth, the person will be forced further back from the work area and have to lean forward causing static loading of the back and neck muscles which will be fatiguing and results in aches and pains if they have to work in that way for extended periods. The surface should also be thin. If it is not then the top of the work surface will be too high and the hands and arms will have to be lifted up, again causing static loading with its consequent effects. Or the person may sit sideways, or at a distance from the work surface, again with undesirable effects. Taking into account the size of the VDT equipment, the work surface should provide adequate space for documents and any other items which the person needs to use in their task. Failure to do this can mean awkward and fatiguing postures in order to accommodate the activity required. The keyboard should be thin, so that the hands can be operated comfortably for typing and be rested and supported whenever possible. The keyboard should be detached from the screen, so that each person can adjust it and the screen separately to their personal requirements for the task concerned. If the keyboard is thick the hands will have to be raised by flexing the wrist, causing static loading of the shoulder, arm and hand muscles. This is very fatiguing. If the keyboard is not detachable it is impossible to obtain optimum positioning of the screen and keyboard so the position of one of them will be compromised with the attendant postural problems.

The screen should be at right angles to the line of sight of the user. In a seated position the natural position of the head is inclined slightly forward, and the line of sight is slightly downwards from this. This means that the screen should be below the vertical height of the eyes, and tilted upwards. So, the screen should permit tilt adjustment and if possible height adjustment since the eye heights of people will vary. A swivel mechanism is also desirable because variation in the horizontal location is necessary.

If the screen is too high the user will raise their eye level by 'sitting tall', lifting their head upright from its natural forward inclination. This involves static loading of the neck and shoulder muscles and will cause fatigue. If the screen is too low the user will bring their eye level down by inclining the neck and dropping the shoulders. This again involves further static loading, and produces fatigue.

Any documents which the person is working with in addition to the VDT equipment will also affect the working postures. In some tasks a document may be used to provide source information for input to the system. In these cases the primary visual task is reading that document, and it should be positioned in front of the person at right angles to the line of sight. On other occasions infrequent reference only may need to be made to the document in which case the screen may be in front of the person and the document to one side. This means that there needs to be a document holder, with variable angle back, which can be moved easily about the work surface. If documents have to be placed on the work surface because a document holder is not supplied it appears likely that people will sit at a lower level so that their reading distance is reduced. This then means that their hands are too high for comfortable operation of the keyboard and this produces more aches in the hands and arms due to the additional static loading.

6. Repetitive strain injury

It has been reported recently in the press that actual damage to tissues may occur to people who suffer the musculoskeletal aches and pains over a long period particularly where intensive work at VDTs is involved. The symptoms of the condition referred to as repetitive strain injury (RSI) involve pain in any or all parts of one or both upper extremities, with or without neck pain. The condition has been most common in Australia, for reasons which are not apparent. In 1984 the World Health Organization sent an eminent epidemiologist, Hadler, to Australia to study RSI. He presented his findings to a congress in Sydney in May 1985, and reported on part of the results in the *Journal of Hand Surgery*.[12] Hadler asks the question 'What is the disease, the patho-anatomical abnormality, that underlies this incapacitating discomfort?' His answer is 'the inescapable conclusion is that the pathophysiology of the upper extremity use-associated discomfort is indeterminate and the symptom-complex defies current nosology.' He goes on to suggest that 'perhaps, as a consequence of differences in the biomechanics of performance, the discomfort is best conceptualized as a form of fatigue'.

The UK Health and Safety Executive commissioned a large research project on Repetitive Strain Injury in 1985, and it is expected that the results of this and other research work will throw further light on the phenomenon.

7. Boring, dissatisfying jobs

There is at present a widespread belief that new technology does mean de-skilled, boring and repetitive tasks. This sometimes does happen, but it is by no means a foregone conclusion. When computer systems are being written, the tasks to be carried out by the end users are being developed. Jobs will be constructed by building together series of tasks. The major problem is that this process of job design is not conducted in a systematic and structured way, involving the users or potential users. As a consequence, a common problem is that jobs are produced on a very narrow range of criteria, which result in them being short cycle, highly repetitive and dedicated. They are the office equivalent of the worker doing just one activity frequently on a production line. These types of jobs are the ones which suffer most from complaints about visual discomfort, aches and pains, and so on. It is understandable that these jobs produce more complaints since the longer the exposure to the situation, the more likely it is that problems will arise. Also the repeated use of the same faculties in the same way promotes fatigue. So when the person concerned is looking at the same visual field (the screen, the document) for many hours, they are more likely to encounter visual fatigue. Similarly, a person sitting at a workstation all day in more or less the same posture may suffer muscular fatigue in the neck, shoulders, and so on.

If, on the other hand, the jobs are designed in such a way that the persons concerned do not have repetitive dedicated tasks, then many of these problems will be avoided. For example, simply getting up from a workstation and walking to another office or office block will promote the dissipation of muscular fatigue, and provide the eye muscles with a similar opportunity. However, the aim should be not merely the avoidance of undesirable effects but the positive promotion of worker satisfaction. To do this, it is important from the job holder's perspective that various tasks relate together to form a coherent job; that the job makes a significant contribution to the system as a whole; that there is a variety of methods available to the job holder and within which they may exercise discretion; that they are responsible for the exercise of discretion and its outcomes, and that they are able to receive feedback on performance.

When designing or acquiring a system it is important that there should be a systematic consideration of job design. Every opportunity should be given to elicit a variety of possible jobs and ways in which they may be organized. The job design literature provides possible alternatives for the latter—where tasks may be grouped according to function, product or service and, of course, there are many combinations of these. These choices, when enumerated, have to be evaluated and suitable criteria must be used in that process, relating to the people and their well-being. Eason[13] has a helpful summary of criteria which consider organizational effectiveness and employee health and well-being, directed to VDT jobs. He says:

—include preparatory and consequential tasks in the jobs of VDU operators
—include in the job both data input and tasks that involve using the data
—avoid loading a specific job with high volumes of data input
—avoid pacing by the computer or by the flow of work to the operators
—provide control over software sequences
—if the job has to be system-oriented, include related system tasks
—avoid over design—leave the job holder some discretion.

A mechanism, or mechanisms, needs establishing in order to ensure that the job design issues are properly considered, at the right time, by the right people. The latter must include the people who will be most affected by the decisions to be made. Eason[13] refers to the best known mechanisms for involving users, being those mechanisms developed by Mumford and Henshall.

8. Adverse reproductive outcomes

For some years there has been a contention that pregnant women can be adversely affected by working with VDTs, and that this manifests itself in an increased rate of adverse reproductive outcomes amongst the people so affected. This may take the form of spontaneous abortions or birth abnormalities amongst those affected. Initially it was feared that radiation could be the cause but there has not been any evidence to support that theory. Another idea was that the general levels of stress associated with working with VDTs were higher than manual office work, and that these were having this effect. Yet another suggestion was that the fixed postures of VDT work put stresses on a woman's body that are inimical to a successful pregnancy, particularly in the early stages. Also the fact of discussing the issues and of creating awareness of them has in the minds of some people merely increased their 'apparent' incidence. There are as yet no authoritative scientific studies reported, and so this problem area remains an unresolved one. A helpful summary of current thinking on the subject is contained in the publication of a recent International Conference on the subject published by Humane Technology.[14]

9. Unsuitable people

It is thought by some people that older workers are not suited to change and that the adaptations involved in using new technology will be too much for them. It is true that the older worker will have suffered the 'normal' ageing effects, so that their eyes will not be what they were twenty years before, and their capacity to learn new things may not be as great, but this in no way renders them unfit to use new technology. A highly motivated older worker can often do a better job than a lethargic young worker, and that without any undesirable effects to the person concerned. There are, however, certain health disorders which can mean that using a VDT produces very undesirable effects on people. These are very few in number, and the people suffering from them are a tiny part of the population. Anyone in any doubt about a worker should ensure that they seek medical approval for VDT operation by the person concerned.

Photosensitive epilepsy is one such example—some epileptic fits are brought on by pulsating light sources. The VDT is in effect a pulsating light, 50 pulses a second usually, and this may bring on a fit.

STEPS IN INTRODUCING NEW TECHNOLOGY

Having considered some of the concerns about new technology, and the ergonomic and human factors issues associated with them, a systematic process for taking full account of them is now described. It is suggested that when any new technology is to be acquired, this process should be followed, even if some of the stages can be executed in outline only.

Step 1—Initial decision-making process

It is essential that the objectives for the technology being considered are clearly identified and specified—productivity improvement, increased speed of a service, improved timeliness of information. This enables all the people who will be concerned with it to understand its purpose in the organization.

The implications and ramifications of having the technology and introducing it need to be worked out, the impact on jobs, pay rates, training, organizational structures, office accommodation and environments, and so on. Thus it is vital to identify the other human-related issues which must be taken care of and to have a thorough appreciation of the 'knock on' costs of acquiring the technology, which sometimes can be considerable.

The decision to acquire should therefore be made in an informed way, being fully aware of the wide ranging impact it will have, its capital and other costs and its business objectives. This, in my experience, is very rarely done, with inevitably disappointing results.

Step 2—The procurement process

Potential suppliers of the technology concerned have to be identified, and placed on a short list, or be invited to tender. Companies should be selected by criteria which include the companies' awareness and knowledge of the ergonomic and human factors issues involved in introducing new technology, so that they can assist intelligently in this process. Also, the technology itself must be evaluated by criteria which include ergonomic/human factors criteria. The VDT manual provides check lists which can assist this process.

If at all possible, involve the potential users in this activity, providing them with direct contact with the systems on the short list by means of:

—demonstrations of equipment
—visits to installations of similar equipment
—pilot systems use if the systems are going to be very large and very expensive.

Feedback to the potential users should be provided when the procurement decision has been made, with as much information as possible as to why the particular decision was made in that way.

Step 3—The preparatory phase

During this period all the necessary activities that were identified in Step 1 have to be spelt out in detail, and carefully planned. This should deal with all the subsequent steps that are described, and would include:

—consideration of all the personnel issues by relevant staff, management, trade unions, and so on with agreements on issues requiring this
—consideration of all the organizational issues by relevant personnel with relevant actions
—programme for erection/modification of buildings
—programme to provide all necessary services—for example telecommunications
—identification and purchase plan for all furnishings, office equipment, and so on.

Step 4—The introduction process

This should have a series of stages, each with clear objectives. It should begin before the technology ever appears anywhere in the organization. It must include:

—an initial exposure of staff to the business objectives for the technology, the reasoning behind the acquisition, and an opportunity for the staff to express their anxieties and fears, and for these to be honestly and reassuringly dealt with
—a familiarization stage, where specific staff are made aware of the impact of the technology on them and the information which is specific to them. As in the initial decision-making process, it may include visits to similar installations, demonstration and even pilot systems. Staff will want to know how their job will change; and its organizational effects, implications for their status, pay and future career; impact on their office layout and environment; how they will be trained and supported
—a training phase, which needs to be co-ordinated carefully with the delivery, installation and commissioning of equipment. The training may well have different modes—off-site courses, in-house courses, self-teach packages, and so on. These should be carefully selected, fully aware of the user's needs, for example self-teach packages, although they are not supposed to, often need supporting with human tutors nearby
—check points in the process to ensure that things are moving according to plan, or not—in which case remedial action can be instigated.

Step 5—The support process

The users of the system must be supported in their use of the system at each and every point of their need. This means that support is available whenever and wherever it is required. It will include:

—documentation
—within system aids and help facilities
—human assistance via telephone or via person.

This is a permanent ongoing step. It is closely linked to the next step in that the support which is being provided should be evaluated and, where it is found to be deficient, it should be suitably upgraded and improved.

Step 6—The review process

This activity should be carried out in order to identify where remedial action is required to deal with unsatisfactory aspects of the present situation. It should also assist the organization in improving its policy and practices towards acquiring and introducing new technology. It should include an assessment of:

— the technology or the system:
  Is it meeting its business objectives?
  Is it meeting the user's requirements and needs?
  Is it liked/valued by the staff?
  Is it well supported by the supplier?
  Is it reliable?
— the process of introduction:
  Any weaknesses or gaps?
  Good points
  Include full range of user groups
—the various areas of impact: were the issues properly identified and have they been handled appropriately and sorted out, for example job content/skills, work organization, personnel policies, office buildings and environments, and so on.

Finally, what of the future? What changes are imminent in this area? In the near future, say the next 2–4 years, the major changes that we can expect to see will relate to the products which manufacturers will be selling and the extent of their use. There will be the advent of flat panel screens with high quality resolution. These will permit new and novel arrangements of equipment at workstations. It will then, for example, be possible to have the screen lying flat, resembling paper on a desk top. Also new devices for interacting with the systems will start to appear—for example second generation mice, and the like. The systems themselves will be easier to use, should meet the genuine needs of users more closely, and should have aids to learning how to use them.

Increasing use will be made of the new technology by people such as secretaries, clerks, middle managers who have not been involved in its use before. They will represent a large population of people going through a major process of change as far as their jobs are concerned. In the period of transition they may suffer from varying degrees of stress and adjustment and will need understanding and support in this process. No doubt some installations will not receive the care and attention they require in their design, installation and introduction. The staff concerned may well suffer undesirable working postures and suboptimal visual environments with the consequential effects on their health and well-being.

In the next decade revolutionary change may begin to appear. The office may become a very different entity to the one we know today. Instead of office blocks, people's homes may well become their office as we see a new generation of 'outworkers' emerging, both part and full-time. It will be important that they are provided with advice and help in designing their offices to ensure that they provide good working environments for themselves and engage in good working practices. Vehicle interiors will become offices as cars, vans and trucks not only make use of cellular and radio phones but also of new types of computer terminals which are mobile. In short, the technology which at present is located in selected and special places will become far more pervasive and we shall have to learn to live and work with it in places which we have never dreamed of before.

Whether or not our learning to live with it all will be reasonable and agreeable or not will depend to a considerable degree on whether or not we implement the good ergonomic practice which this chapter is all about.

Acknowledgement

I acknowledge gratefully the help given to me by my colleagues at HUSAT in preparing this chapter.

REFERENCES

1. Cox EA. *Displays*. London: Butterworth, 1983.
2. Pearce BG. *Health Hazards of VDTs?* Chichester: Wiley, 1984.
3. Duke-Elder S. The pathological action of radiant energy upon the lens. *Lancet* 1926; **1**:118.
4. Zaret MM. Cataracts and Visual Display Units. *Health Hazards of VDTs?* Chichester: Wiley, 1984.
5. Ray P and Meyer JJ. Visual impairments and their objective correlates. In: *Ergonomics Aspects of Visual Display Terminals*. Chichester: Wiley, 1980.
6. Laubl T, Hunting W and Grandjean E. Visual Impairments in VDU Operators related to Environmental Conditions. *Ergonomic Aspects of Visual Display Terminals*. Chichester: Wiley, 1980.
7. Stone PTS. Issues in Vision and Lighting for Users of VDUs. *Health Hazards of VDTs?* Chichester: Wiley, 1984.
8. Duke-Elder S. *System of Ophthalmology, V*. London: Kimpton, 1970.

9. Cakin A, Hart DJ and Stewart TFM. *Visual Display Terminals*. Chichester: Wiley, 1984.
10. Hunting W. Postural Loads at VDT Workstations. *Health Hazards of VDTs?* Chichester: Wiley, 1984.
11. Maeda K. Occupational cervicobrachial disorder and its causative factors. *J. Human Ergo.* 1977; **6:**193–202.
12. Hadler NM. Illness in the workplace: the challenge of musculoskeletal symptoms. *J. Hand Surg.* 1985; **4:**451–6.
13. Eason KDE. Job Design and VDU Operation. *Health Hazards of VDTs?* Chichester: Wiley, 1984.
14. Proceedings of International Conference on Allegations of Reproductive Hazards from VDUs. Loughborough: Humane Technology, 1985.

16. ACHIEVING SAFE BEHAVIOUR

A. R. Hale

The achievement of safe behaviour is one of the oldest problems of health and safety and, in the light of such disasters as Bhopal, Seveso, Three Mile Island or Chernobyl, the most modern. It used to be seen as largely a matter of motivation to counter what was thought of as the inherent carelessness of human beings. If an accident happened and it was discovered that the victim had ignored or flouted existing safety rules, the case appeared open and shut; it must have been negligence or recklessness which made him do it. This reflex attribution is still with us today in many quarters, but it is being slowly replaced by a more sophisticated picture of where the problems really lie.

This development has occurred thanks partly to the far greater attention paid to human factors in the wake of disasters in the nuclear and chemical industries. It is also due partly to the information technology revolution. People have begun to develop and build computer-based 'expert systems' to carry out decision-making tasks in complex and uncertain situations, such as medical diagnosis. In trying to model the way in which human experts arrive at judgements in such circumstances two things have happened: researchers have begun to understand the full complexity of those decisions and to realize that even decisions which look simple once they have been taken are extremely complex in prospect; the programmes which they build have begun to make the sort of errors which are characteristic of the human errors which can lead to accidents. In particular the very nature of *rules* which govern decisions has come under scrutiny. A new methodology called *fuzzy logic* has come to the fore to describe mathematically how rules govern decisions under conditions of uncertainty. This recognizes explicitly that rules do not have hard edges which distinguish right from wrong, or in the case of the subject of this chapter safe from unsafe. The result of these new insights has been to focus attention on the cognitive, reasoning aspects of human error and health and safety instead of on its emotional, motivational side.

In this chapter I want to set out a model of human behaviour in the face of danger, which stresses this cognitive aspect. It is applicable to the analysis of any situation in which people are confronted with dangers, either to their short-term safety or their long-term health, but, in order to anchor it to the real world I shall use as a running example the issue of personal protective equipment and what can be done to encourage people to wear it correctly in the appropriate circumstances.[1]

PERSONAL PROTECTIVE EQUIPMENT

Control of dangers at source is always the preferred strategy in health and safety, but its complete achievement is utopian. Vast numbers of people at work rely for their safety on the use of personal protective equipment (ppe). The emphasis in the provision of such equipment used to be only on its technical specification, the degree of protection offered under laboratory conditions. With the development of better and more portable field measuring equipment it became clear that the protection actually given in practice by such equipment was often only a fraction of that theoretical protection factor. The difference was accounted for by the way in which the equipment was selected, fitted, worn, cleaned and maintained—all tasks which involved active co-operation of the wearers and of other people in the organization. In other words ppe was not a passive barrier to be hung on the long-suffering worker, but a tool which he must use actively. As such it epitomizes safety provisions and rules and can serve as the example both for the model I wish to present and for the factors which determine how people respond to such provisions and thus the scope for improving the overall system.

MODELLING SAFE AND DANGEROUS BEHAVIOUR

All situations contain an element of danger, and thus no action can be said to be entirely safe. The task of the human is to steer a course through the shoals, maintaining control over a situation which is at times complex and always has elements of uncertainty in it. Much of that uncertainty comes from the actions of others, and the individual is thus only able to influence some of the factors which will determine how things turn out. It is never a question of choosing a *safe* course of action, only a *safer* one, and the degree of safety may differ depending on whether it is looked at as a short-term or long-term decision. The task of the individual can be represented as the recognition of the danger and what can be done to control it, the processing of the alternative courses of action, and the taking and implementation of a decision to act. That task is set out in more detail in *Fig. 16.1*.[2]

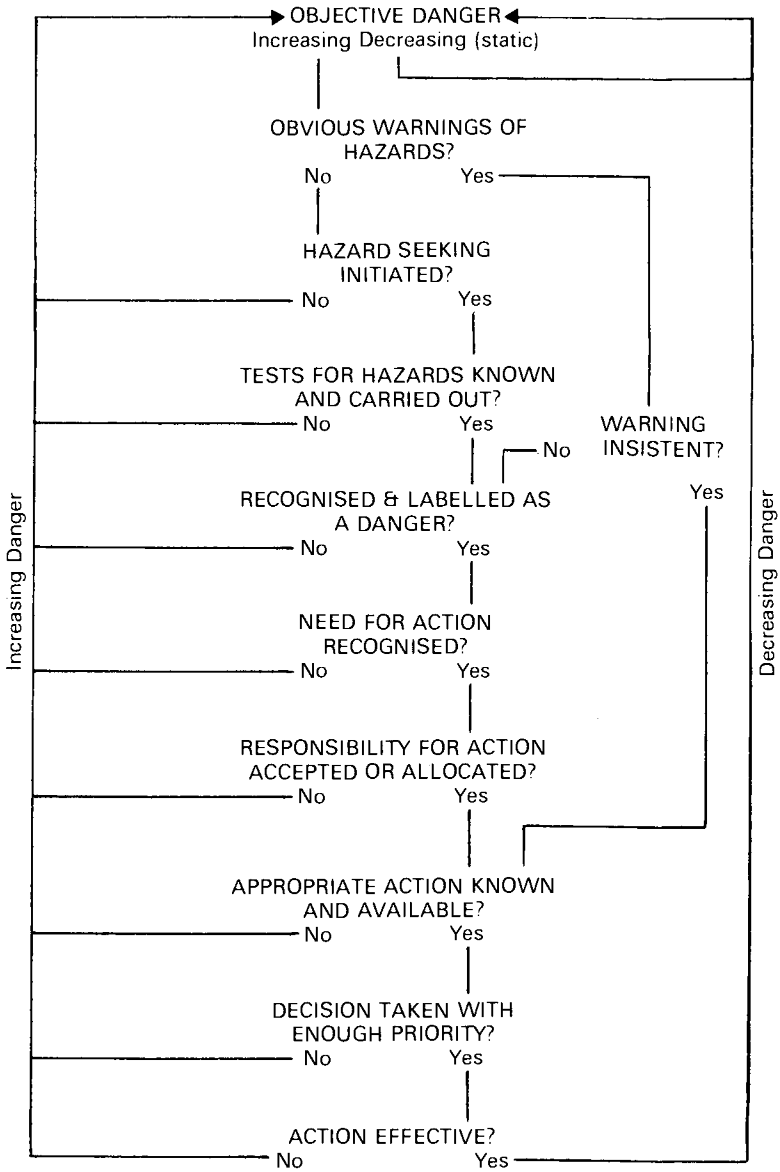


Fig. 16.1. Classification according to the stages of perception of and response to danger.

NATURE OF DANGER AND WARNINGS

Only in the case of dangers with both obvious and insistent warnings (fast moving objects hurtling towards one, a cloud of noxious gas, smoke pouring under the door of your room) is the process of reaction simple. Then it is only a question of the speed and co-ordination of response. In such circumstances there may still be problems with ppe, that people fail in their panic to put it on at all, or at least not correctly, but these are far from the usual problems of industrial life. However, it can be said that reduction of the obviousness of danger does bring with it a series of other problems which make safe behaviour more complex. Thus reduction of noise or airborne contaminants at source, which does not go so far as to eliminate the need for ppe, means that more rather than less thought has to be given to implementing successful ppe programmes.

A small study of asbestos workers[3] found some who only considered that danger was present when they could see dust suspended in the air, despite the fact that the most dangerous particle sizes are invisible in most lighting conditions. This is typical of problems with invisible hazards. People must learn that there are dangers, and they must learn how to tell that the danger is present.

HAZARD SEEKING

If warning signs on their own are not compelling enough to trigger an awareness of danger, the process of danger perception can only be initiated by some active decision to seek out the signs that danger may be present. This may be prompted by a warning sign or by an instruction or suggestion from another person. Where hazards are continuously present there is less of a problem; a warning notice and a strictly enforced rule can make a clear distinction between a danger zone and the rest of the factory. If the danger is only intermittently present, or if the hazard is dependent upon a time-weighted exposure, the position is more complex. People learn rapidly that the warning sign alone is not always a sign that it is dangerous for them. Danger zones which are too loosely drawn around continuous dangers can fall into the same trap, that the warning is seen as a false alarm. People then seek other indicators (are other people wearing ppe? is the process working? am I going to stay long?). It is vital that such grey areas are clarified as much as possible through layout of danger zones, consistent use of warnings, training in the nature of the hazard and clear rules, for example for short-term visitors to noisy areas. Lack of clarity allows scope for misinterpretation.

HAZARD TESTS AND RECOGNITION

Unless noise is loud enough to be physically unpleasant or to produce a noticeable temporary threshold shift (TTS) the short-term effects of exposure are not perceptible. Even where TTS occurs people may learn (incorrectly) that, since it goes away after a while, it is not something either to worry about or to treat as a warning of harm.

This is not the only problem in recognizing that noise is a danger. The subjective assessment of noise differs in crucial ways from the objective measure of energy immission which governs the degree of damage.[4] *First*, the neural mechanism of habituation to constant noise leads to underestimation of loudness, while the aversion to distracting, masking or disturbing noise may lead to overestimates of damage. *Second*, while both noise energy and subjective loudness follow a logarithmic scale—the doubling of energy is every 3 dB—the subjective loudness doubles every 10 dB. So the subjective assessment of the probability of damage does not keep pace with the actual danger. The same mechanisms will make it difficult for the wearer to detect, from the subjective noise experienced, any slow process of deterioration of the ppe. Finally, those who are not experts in the subject find the time-weighted nature of the dose and the effect of that on the protection afforded by a given duration of wearing surprising and counter-intuitive; they expect a linear relationship between time worn and protection afforded (*Fig. 16.2*, Else[5]). A similar problem arises

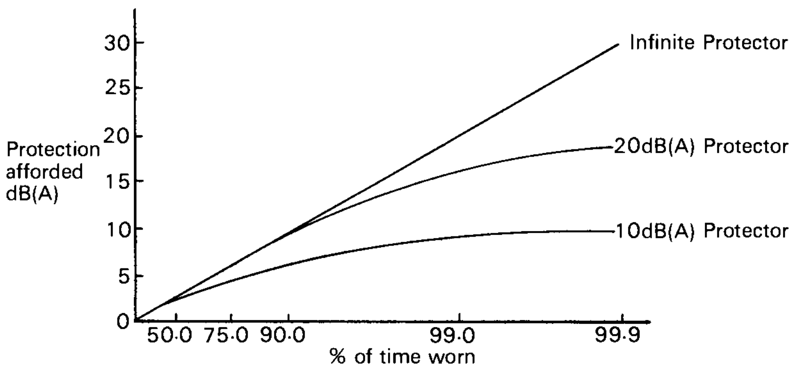


Fig. 16.2. Protection in practice by time worn: hearing protection.

for respiratory protection. As with many long-term hazards people do not naturally have a mental picture of how the harm occurs. If they are not given a mental picture which is suitable, they will construct their own. This may give incorrect predictions in certain cases. All of those problems must be tackled by training.

Problems with the recognition of dangers requiring respiratory protection are of comparable complexity, since many dangers can be recognized only with the aid of sophisticated measuring equipment. This means that the users of the ppe have to rely on others to carry out the tests. This brings the question of trust between the measurers and the wearers into the issue of recognition.

It is not only the mental models of the *wearers* which are important. Those who decide to use ppe as a preventive strategy in preference to other strategies, those who select it and, to a lesser extent, the other supervisory, maintenance and back-up personnel also have their own ideas of what factors influence the success of a ppe programme. Managers may not realize that protection in practice can be much lower than the theoretical protection factors specified by manufacturers, and sales literature unless they spend considerable time and money on training, maintenance and supervision of the ppe. In that case they may think their problems are solved once they have paid for the original purchase of equipment. The additional costs of the necessary backup over the life time of the ppe which is needed to make it live up to its theoretical protection factor may totally alter the cost/benefit of ppe when compared with control at source (which may be more expensive to install, but less expensive to maintain).

The selector of ppe must also realize in his 'mental model' of the ppe that comfort will affect the time worn and hence may be as important to protection in practice as the technical specification of the equipment. Finally, if the storekeeper is not aware that different types, grades and sizes of ppe can provide different protection in practice he may simply issue the first ones that come to hand and so inadvertently subvert the best laid ppe scheme.

In general the training which people need in order to adjust their behaviour to the true nature of a hazard is considerable, especially for long-term or contingent hazards. Simple statements that 'X is dangerous and so you must follow these rules' will store up problems for the future, since they give *carte blanche* to each person to build up their own model of the problem.

LABELLING AND NEED FOR ACTION

Research by Perusse[6] shows that the most important factors influencing people in their assessment of the importance of dangers, and hence in the urgency with which they will try to combat them, fall into two clusters. The first and most influential are beliefs about the degree to which, and the way in which, the danger can be controlled. The second cluster concerns the severity, likelihood and nature of the harm produced. Other research[7] produces a slightly different factor structure which emphasizes the familiarity of the hazard or technology and the numbers of people at risk from it, in addition to the aspects named by Perusse.

Dangers which are regarded as unknown or uncontrollable by the individual exposed to them are considered to be much more in need of regulation than those where the person at risk can choose whether to subject himself to the risk and where he also controls the methods of prevention. The crucial issue is whether the individuals' assessment of their degree of control is accurate. Overestimates of it will lead to complacency, underestimates to unwarranted fears. There are indications from research[8,9] that workers generally regard their control over safety hazards as very high and seem to ignore the very real constraints of time, production, hardware and work layout which may push them into dangerous behaviour. Their assessment of health risks is less optimistic. However, as I have indicated under Hazard Tests and Recognition *above*, there are a number of factors which can lead people to overestimate the efficacy of ppe once worn. These may be outweighed by the widespread cynicism about ppe which is commonly found among work forces. The problem is a difficult one for the trainers who, if they are true to the principle that ppe is a temporary or last line of defence, can hardly avoid giving the impression that the control it provides is fragile. They must point out and underline the importance of the user's own actions in keeping the efficiency of the equipment up to an acceptable standard, and yet they must convince the user that the danger really will be controlled if the equipment is worn properly. Failure to tread this tightrope and to ensure that all concerned maintain that relatively sophisticated understanding of the role of ppe will result either in overconfidence in the use of ppe which is only giving token protection, or rejection of it and strengthened agitation for alternative and better forms of control.

The understanding of the seriousness of the consequences of the harm which will result if ppe is not worn also presents problems. The concepts that people have of occupational disease and what it would be like to suffer it are frequently inaccurate.[3,10] Workers may believe that deafness is merely like turning the volume down on their hearing, simply correctible by standing nearer to people and by a 'booster' hearing aid, not appreciating the distortion and muffling effects, and not considering the social isolation which results from poor hearing. Others may not be aware that the process of harm has started, and may ascribe hearing difficulties to 'people not speaking so clearly as they used to.'[1] Training needs to combat these misconceptions. Techniques available include:

—use of tapes of conversation as it would be heard by someone

suffering from occupational hearing loss
—film of interviews with sufferers from deafness, asbestosis etc, and
—demonstrations through the use of audiograms,[11,12] lung function tests or other information feedback that the harm process has already started in an imperceptible way.

Again, however, a tightrope has to be walked between engendering a suitable degree of concern to encourage people to take protection seriously and producing such a high level of fear that people refuse to work with the hazard. Judging that balance presents ethical as well as training design problems.

ALLOCATION OF RESPONSIBILITY

Implicit in the concept of control are ideas about responsibility. People cannot be expected to act to remove dangers they do not believe that they control. Strong biases have been shown to exist in three aspects of this area.[13] *First*, people are inclined to attribute things which happen to other people to those peoples' own actions. When considering things which happen to themselves, however, they attribute much more to the influence of external circumstances. *Second*, people look for consistencies in situations and attribute on that basis; thus two accidents happening to two people in the same circumstances would be attributed to the circumstances, whilst, if they happened to the same person, they would be attributed to the person. *Third*, people overestimate with hindsight the degree to which they predicted the outcome of a situation. These biases add up to the belief that my accident was an unfortunate set of circumstances, but the other person's accident was their fault, especially if they have had more than one, since they must obviously have been able to see it coming. This sort of bias leads supervisors to deny their responsibility for controlling safety and health among their (careless) workers,[14] and to the popularity of 'accident proneness' as a dismissive explanation for accidents.

The volatility and distortions of allocation of responsibility make it a prime target for clarification in any successful preventive programme. It is vital, for example with ppe, that the responsibilities for detecting that the equipment is no longer functioning, for ensuring that it fits, is fit to wear, and that it is worn are clearly allocated to people who can and do perform them. If they are not specifically allocated it is highly likely that the general tendency to abdicate responsibilities which are seen as peripheral will ensure that they fall down the cracks in the system.

KNOWLEDGE OF ACTION

This part of the model is the one which people normally identify most readily as an area for training. Despite its obvious relevance it is frequently not well done.[2] Knowledge and skill must not only have once been learned, they must be available when needed. Most routine or well practiced tasks are represented in the worker's brain as a set of subprogrammes which are selected and put into practice in a highly efficient, but largely unconscious sequence. This selection is under the control of the mechanisms of attention which are strongly influenced by the availability of the subroutine. If the correct one is not immediately available, another similar but inappropriate one may be substituted in its place in order to preserve the smooth flow of the task. Alternatively, the person must switch from the efficient automatic control of behaviour to very slow, error-prone conscious decision-making.[15] A small study of users of self-contained breathing apparatus[16] found many who could not put it on rapidly and correctly and carry out adequately the face seal leakage test as they would need to do in an emergency.

It is not simply a matter of training people to use the safety equipment itself. Provisions such as ppe can significantly alter the whole task to be done. Safety goggles restrict the field of vision;[17] ear defenders reduce the ability to detect the direction of sound[18] and may cut out sounds relevant to the performance of the task.[19] So the whole task must be relearned or adapted, which disrupts the smooth automatic control described above, and causes frustration and sometimes a loss of self-esteem in an experienced worker who can no longer effortlessly outperform the beginner.

DECISION TO ACT

Here we come to the motivation which can be brought to bear to encourage people who know *how* to act appropriately so that they *choose* to do so. This is partly a matter of influencing the costs and benefits of different courses of behaviour.

The findings of research into health education,[20,21] and the work of the behaviour modification school,[22,23] indicate that the use of high levels of fear motivation in any training are of limited value. The emotion tends to swamp the message, and people do not learn what to do to avoid the hazard. They may also suppress the fear by suppressing all conscious thought about the whole problem. This message is also contained in reviews and studies specifically related to ppe.[24,25] Training which stresses that the individual can control the hazard, and that ppe is an effective means of control is much more successful. Another important factor is positive feedback from supervision and an indication of how well the group are doing in meeting targets e.g. the proportion of people wearing ppe. Other motivation which is relevant is the degree of social control which can be exerted by the establishment of group norms, and the extent to which 'role models' (group leaders, safety representatives, supervision, trusted experts) support, praise and back up the use of the equipment.

A dilemma in this area is that the incentive to wear ppe will be greater the higher is the level of noise or other danger, so that reductions in noise level produced by control at source may be offset to some extent by lower use of ppe, unless that tendency is checked by appropriate training. It must be stressed that the decisions which people take, even when they are ones not to wear the ppe which is provided, are rational ones based on the costs and benefits as each person sees them. Failure to persuade people to wear ppe can therefore be seen as a failure on the part of trainers and managers to appreciate and influence this balance, not a failure of workers.

The objective which is being aimed at is to overcome the initial barriers to wearing represented by the discomfort, need to adjust any previously learned skills, social pressure from ridicule etc. This requires an intensive effort, but in the long run the objective is to arrive at a state where the use of the ppe is an automatic habit which does not need to be subject to the uncertain process of conscious decision-making each time. The problem then becomes one of monitoring for any regression from the habit so that it can be corrected before it threatens the established consensus.

ACTION SEQUENCE AND FEEDBACK

This part of the model stresses the importance of supervised practice and feedback both to learn and maintain the skills necessary and to carry out each of the tasks.

In several of the sections above the need for feedback to the individual carrying out a task has been stressed. It is an established principle of training theory that successful learning is achieved only when the individual has been equipped with the means of assessing whether and when the behaviour in question has been carried out to a satisfactory standard. This applies to teaching users to detect inadequacies of fit or state of repair of the ppe as much as to teaching selectors of ppe how to assess whether or not the equipment selected was appropriate and achieved the desired level of protection in practice. Above all, people will modify their own behaviour based upon the success or failure that *they* perceive, or that is fed back to them. Thus, behaviour which does not appear to be punished and which yields positive results will be reinforced. Since accidents, injuries, disease, and even near-misses are rare occurrences, other substitutes must be found to indicate what is unsafe behaviour. Moreover, safe behaviour has little or no intrinsic positive reward attached to the fact that it is safe. The most work, therefore, needs to be done to build in substitute rewards to make up for that lack.

CONCLUSION

Even such a brief sketch of some of the issues involved in the study and influencing of human behaviour in health and safety indicates that it is a large and complex field. The importance of cognitive factors is also clear. It matters greatly that people understand in appropriate detail the mechanisms by which they or others can be harmed. Biases built into the way in which people reason can account for many of the apparent occasions of 'risk-taking'. Such biases are functional in other areas of decision-making, and make humans unparalleled in their flexibility and ability to cope with complexity and uncertainty. Hence people cannot be easily trained or motivated out of their biases. Finally, people adapt their behaviour continuously, based upon the way they perceive the feedback from previous experience; so, even in motivation there is an important element of cognitive processing.

The field of accidents, injuries and occupational disease is a field of small probabilities, where most accidents require a number of things to go wrong together, and where most diseases take a long time for causes to manifest themselves in effects. It is therefore a field which taxes the cognitive capabilities of humans severely. We are constantly inclined to simplify complexity in order to make it manageable (this chapter is no exception to that observation), and so to treat small probabilities which we do not want to happen as equal to zero. The implication of this is that the prevention of accidents, injuries and occupational diseases will always be uphill work, because it is working to counter in-built limitations and to supplement the individual's capacity to look after his or her own safety.

The advent of 'intelligent' computer support systems is opening up new possibilities for this in relation to some tasks such as the control of chemical and nuclear processes, and the avoidance of incorrect medical diagnosis (two closely related skills). The other recourse is to design organizational systems which act in the same way as supports to the individual, taking from him tasks, such as many aspects of hazard assessment and inspection, which he cannot effectively perform at the same time as all his other responsibilities and objectives. The history of health and safety has been a progressive realization that an individual's safety must rest upon the activities of many others.[26] The most recent addition to that list has been the designer of both products and workplaces, who is now increasingly held responsible legally not only for the way in which an object *should* be used, but for the way in which it *can* be used and *misused*.

If that message is applied to the topic of ppe which I have used as my example in this chapter, the large burden of planning, organization, monitoring and supervision, and, above all, of training involved in the introduction of a successful ppe programme is obvious. That should give organizations reason to pause before embarking upon one as an apparently cheap alternative to control at source. Ppe as a strategy of control radically shifts responsibility from management to the workforce for their protection. If all of the steps outlined in this paper are not taken to back up that shift the choice of ppe as a control measure is tantamount to an exploitation of that workforce and an abrogation of management responsibility.

REFERENCES

1. Hale AR and Else D. The role of training and motivation in a successful personal protective equipment programme. Proceedings of the COPE Conference. Toronto: Canadian Centre for Occupational Health & Safety, 1984.
2. Hale AR. Is safety training worthwhile? *J. Occup. Acc.* 1984; **6**:17–33.
3. Wilson A. Attitudes of Workers to Hazardous Substances. MSc Dissertation. Dept of Occupational Safety and Hygiene. Birmingham: University of Aston, 1975.
4. Poulton EC. *Environment and Human Efficiency*. Illinois: Thomas, 1970.
5. Else D. A note on the protection afforded by hearing protectors—implications of the energy principle. *Ann. Occup. Hyg.* 1973; **16**:81–83.
6. Perusse M. Dimensions of the Perception and Recognition of Danger. PhD Thesis. Birmingham: University of Aston, 1980.
7. Slovic P, Fischhoff B and Lichtenstein S. Behavioural decision theory perspectives on risk and safety. *Acta Psych.* 1984; **56**:183–203.
8. McKenna SP. Accident attitudes. *Occup. Health* 1975: **27**:330–32.
9. Glendon AI. What does responsibility mean? *Occup. Health* 1981; **33**:245–50.
10. Williams JB. Knowledge and Attitudes in Relation to Toxic Substances. MSc Dissertation. Dept of Occupational Safety and Hygiene. Birmingham: University of Aston, 1976.
11. Zohar D, Cohen A and Azar N. Promoting increased use of ear protectors in noise through information feedback. *Hum. Factors* 1980; **22**:69–79.
12. Merriman RJ. The role of Audiometry in the Prevention of Occupational Deafness. PhD Thesis. Birmingham: University of Aston, 1977.
13. Hale AR. Attributing cause. *Occup. Health* 1980; **32**:519–22.
14. Abeytunga PK. The Role of the First-line Supervisor in Construction Safety: the Potential for Training. PhD Thesis. Birmingham: University of Aston, 1978.
15. Reason J and Embrey DE. *Human factors modelling of human error in abnormal conditions of nuclear and major hazards installations*. Dalton: Human Reliability Associates, 1985 (Contract ECI 1164-B7221-84-UK European Atomic Energy Community).
16. Holland JA. Self-contained Breathing Apparatus: a Review of Industrial Practice Relating to Equipment Inspection and User Training Procedures. MSc Dissertation. Dept of Safety and Hygiene. Birmingham: University of Aston, 1974.

17. Mitchell JA. The Effect of Eye Protectors on the Peripheral Visual Field. BSc Dissertation. Dept of Applied Psychology. Birmingham: University of Aston, 1978.
18. Else D. Hearing Protectors. PhD Thesis. Birmingham: University of Aston, 1976.
19. Dunn JG. Skills analysis used for the investigation of accident risk in power saw operators. AP Note 21. Dept of Applied Psychology. Birmingham: University of Aston, 1970.
20. Leventhal M, Singer R and Jones S. Effects of fear and specificity of recommendations upon attitudes and behaviour. *J. Pers. Soc. Psychol.* 1965; **2**:20–29.
21. Rogers RW and Mewborn CR. Fear appeals and attitude change: effects of a threat's noxiousness, probability of occurrence and the efficacy of coping responses. *J. Pers. Soc. Psychol.* 1976; **34**:54–61.
22. Komaki J, Barwick KD and Scott LR. A behavioural approach to occupational safety: pinpointing and reinforcing safety performance in a food manufacturing plant. *J. Appl. Psychol.* 1978; **63**:434–445.
23. Komaki J, Heinzmann AT and Lawson L. Effect of training and feedback: component analysis of a behavioural safety program. *J. Appl. Psychol.* 1980; **65**:261–270.
24. Pirani M and Reynolds J. Gearing up for safety. *Pers. Mgmt* 1976; **Feb**:25–29.
25. Tourigny P. Etude des facteurs incitant au port des protecteurs auditifs individuels dans l'industrie du meuble. Masters Thesis. Dept des Relations Industrielles. Quebec: Universite Laval, 1980.
26. Hale AR. *The Human Paradox in Technology and Safety*. Inaugural lecture. Delft University of Technology. Delft University Press, 1985.

17. THE PLACE OF LEGISLATION, REGULATION AND ENFORCEMENT IN OCCUPATIONAL HEALTH

Alun H. Jones

In 1795 Edmund Burke expressed the current philosophy of government which was to become untenable in the social upheaval consequent on the industrial revolution:

> The State should confine itself to what regards the State, namely the exterior establishment of its religion; its magistracy; its military force by sea and land; the corporations that owe their existence to its fiat; in a word to everything that is truly and properly public. . . .[1]

The Benthamite view that the dreadful social consequences of the expansion of industrial towns and the exploitation of child labour were indeed matters that were truly and properly public met continuing and bitter resistance throughout the early decades of the nineteenth century. Public concern—not to say terror—led to some early reactive intervention in matters of public health.

An outbreak of typhus among mill apprentices led to the formation of the Manchester Board of Health and a report by Thomas Percival on the working conditions of young persons eventually resulting in the first 'interventionist' factory legislation—the Health and Morals of Apprentices Act of 1802. At a much later date Victorian moral susceptibilities were deeply offended by the accounts and illustrations of scantily clad children in mines to be found in the monumental Report on the Employment of Children in Mines of 1842. The outcry resulted in the first Mines Act of 1842.

The new, developing philosophy was epitomized by the Poor Law Commission Report of that same year which largely bore the stamp of the Benthamite, Edwin Chadwick. It represented an approach to social problems—inquiry, report, legislation and enforcement—which characterized the deliberations of the Robens Committee more than a century later. Despite this evidence of rationality and objectivity in approach, health and safety legislation remained largely of a reactive nature until very recent times. The 1844 Factories Act required the appointment of Certifying Surgeons to pronounce on the age of young employees because the legislation on the certification of births, marriages and deaths was not by then fully effective. During the following hundred years legislation directed to the control of lead hazards was based on a *post hoc* reaction to each problem as it arose. This resulted in some seventeen sets of regulations applicable to different trades. By 1980 when the situation was rationalized many of these trades had ceased to exist while other lead exposures remained uncontrolled. This was despite the fact that the principles of the control of lead poisoning had been clearly stated by Thomas Legge at the turn of the century in a form which would have readily lent itself to legislation. Even primary legislation tended to be the result of a political response to circumstances—based on specific industries or groups of activities rather than an expression of a general philosophy of the maintenance of health at work.

Thus we have a Mines and Quarries Act, a Factories Act, a Nuclear Installations Act, and an Alkali and Clean Air Act originally created in response to the nuisance created by effluvia from quite a narrow range of industries. Furthermore, each Act was enforced by an Inspectorate unique to each piece of legislation. The development of further legislation over the years tended to be of a consolidating nature; the first radical examination of the situation and the first attempt to create primary legislation based on first principles and a rational approach had to await the deliberations of the Robens Committee in 1972.

THE ROBENS COMMITTEE

By 1972 the haphazard growth of legislation on Health and Safety at Work had resulted in the existence of eight directly relevant Acts of Parliament and some 390 Statutory Instruments (set of Regulations) made under these Acts. Furthermore, each Act had its own enforcing authority in general responsible to different government departments or to local authorities (*Table 17.1*). Despite this mass of legislation a large part of the working population remained unprotected by any statute law. Much of the legislation was, in any case, archaic and had not kept pace with the technical development of industry. Criteria of enforcement varied from situation to situation. The Committee on Health and Safety at Work under the chairmanship of Lord Robens[2] deliberated on the matter between 1970 and 1972. It concluded that:

1. The empirical approach to health and safety legislation had produced a body of law and a system of enforcement which was probably unequalled. Paradoxically, however, its piecemeal growth had led to a state of confusion. Furthermore, legislation was still reactive; the Aberfan disaster led to legislation on mine refuse tips, new fire

regulations followed eight fire deaths in a factory in the 1950s, and the control of drugs was strengthened after the thalidomide tragedy.

2. *The primary responsibility for preventing accidents and disease rests with those who create them and those who work with them.* Government action should be concerned with influencing attitudes to health and safety. It was important to involve those at shop floor and supervisory level directly involved in shaping the course of events in the promotion of health and safety.

3. There was an undue influence on punitive enforcement of the law rather than advice and guidance.

4. Actions for negligence at civil law often hindered objective assessments of causation of accidents and disease. The Robens Committee advocated a 'no fault' system of workers' compensation.

THE HEALTH AND SAFETY AT WORK, ETC. ACT 1974 (THE HSW ACT)

The outstanding characteristic of this Act is its universal applicability. It departs from the process of generating legislation in response to particular circumstances and events and takes as its premise that health and safety are the right of everyone gainfully employed (with the curious exception of domestic servants!). Equally, it is the duty of all employers to promote the health and safety of their employees and the duty of all those employees to co-operate in the furtherance of any measures taken by their employers and not to obstruct employers in matters of health and safety.

Employers are obliged to commit themselves formally to a statement of their intentions in the promotion of health and safety in a written statement of their Health and Safety Policy. Involvement of employees is ensured by a statutory requirement to establish safety committees which can advise management on relevant matters through appointed safety representatives. Unfortunately this latter provision is only applicable where a recognized trade union exists.

THE HEALTH AND SAFETY COMMISSION AND THE HEALTH AND SAFETY EXECUTIVE

The multiplicity of statutes in force in 1972 were administered by a corresponding variety of Government departments and agencies. Robens had suggested a new Authority to oversee the generation, administration and enforcement of health and safety legislation. This Authority achieved reality in 1974 as the Health and Safety Commission. This was constituted according to the favoured pattern of the time as a 'quasi autonomous non-governmental organization', a corporate body representative of workers (as trade unions), employers

Table 17.1. The main statutory provisions and central inspectorates concerned with safety and health at work

Act	Subordinate Statutory Instruments	Approximate number of establishments affected[a]	Enforcing Authority	Size (authorized establishment) of Inspectorate in *1971*	*1960*	Administering Department
Factories Act 1961	247	206 700[b]	Factory Inspectorate and others[c]	714[d]	448	Department of Employment
Offices, Shops and Railway Premises Act 1963	21	755 800	Factory Inspectorate and Local Authorities	See footnote[e]		Department of Employment
Mines and Quarries Act 1954	55	4946	Mines and Quarries Inspectorate	135	165	Department of Trade and Industry
Mines and Quarries (Tips) Act 1969	3					
Agriculture (Poisonous Substances) Act 1952 Agriculture (Safety, Health and Welfare Provisions) Act 1956	14[f]	112 660	Safety Inspectors and Field Officers of the Agriculture Departments	44[g]	70[h]	Ministry of Agriculture, Fisheries and Food
				11[i]	10[i]	Department of Agriculture and Fisheries for Scotland
Explosives Acts 1875 and 1923	60[j]	35 000[k]	Explosives Inspectorate and Local Authorities	9 (Local authorities have the major share of the work—see footnote [k])	4	Home Office
Petroleum (Consolidation) Act 1928	23[j]	40 000[l]	Local Authorities			Home Office
Nuclear Installations Act 1965 and 1969	5	29[m]	Nuclear Installations Inspectorate	46	15	Department of Trade and Industry

Radioactive Substances Act 1960[n]	20	3000	Radiochemical Inspectorate	6	8	Department of the Environment
Alkali etc. Works Regulation Act 1906	2	1700	Alkali and Clean Air Inspectorate	36	27	Department of the Environment
Alkali etc. Works Regulation (Scotland) Acts 1906 and 1951	2	200	Scottish Industrial Pollution Inspectorate	4	3	Scottish Development Department
Mineral Workings (Offshore Installations) Act 1971	—	40	Petroleum Production Inspectorate	9	—	Department of Trade and Industry

Notes:

[a] The figures of establishments affected overlap to some extent because some establishments are subject to more than one of the Acts listed.

[b] Includes factories, docks, shipyards and warehouses. Construction sites are excluded because it is not possible to put a reliable figure on their number.

[c] Certain provisions in the Act are enforced by local authorities. Fire prevention provisions are enforced partly by the Factory Inspectorate and partly by fire authorities.

[d] Including 40 assistant inspectors.

[e] Enforcement is undertaken by local authority inspectors and the Factory Inspectorate, in conjunction with their other duties.

[f] Includes the Agriculture (Avoidance of Accidents to Children) Regulations 1958 which apply to all agricultural holdings irrespective of whether persons are employed. The number of holdings covered by these regulations is 271 664.

[g] Full-time safety inspectors in England and Wales. In addition, in 1971 some 425 agriculture field officers undertook safe inspections in combination with other duties; they spend rather more than a quarter of their time on safety work.

[h] In 1960 there were no full-time safety inspectors: the 70 agriculture wages and safety inspectors spent about half of their time on safety work.

[i] Agriculture wages and safety inspectors who spend about one-third of their time on safety matters.

[j] A number of statutory instruments are concerned with dangerous substances other than in workplaces.

[k] The Explosives Inspectorate is responsible for 170 explosive factories and magazines. Local authorities are responsible for about 35 000 explosive stores and registered premises.

[l] Includes 27 refineries and 500 major distribution depots. The remainder are mainly petrol filling stations.

[m] The Nuclear Installations Inspectorate does not inspect or licence nuclear installations operated by government departments of the United Kingdom Atomic Energy Authority.

[n] The information given relates to England and Wales. In Scotland the Act is enforced by the Scottish Industrial Pollution Inspectorate of the Scottish Development Department.

(as employers' organizations) and others. In practice these others were largely local authority representatives. The Commission (the HSC) is responsible to Parliament through the Secretary of State for Employment. The functional arm of the HSC is the Health and Safety Executive, comprising all the pre-existing Inspectorates and a medical division, the Employment Medical Advisory Service, together with a number of new and inherited research functions within a Research and Laboratories Division.

The Commission is a peculiarly British organization; its function is to assume, as a non-governmental body, the delegated duty of generating national policy and legislation in matters of health and safety.

The Executive, on the other hand, is to all intents and purposes a Civil Service department, much more akin to the bodies administering these matters in other western countries. There are close parallels, for example, with the Office of Safety and Health Administration in the United States.

The HSW Act is a framework of reference within which the bulk of historic legislation continues. However, as will be seen, it allows for much greater flexibility in the generation of new legislation.

The enforcement agencies—the Inspectorates—still maintain their separate identities. Robens recognized that they differed in their philosophy and their expertise; nevertheless he envisaged an ultimately unified Inspectorate. Progress in achieving such unity has been slow.

Historically the role of the Inspector was that of enforcement; it was not for him to advise except on the means whereby an employer could ensure compliance with the law. This is still largely true; the Inspector draws attention to breaches of statute, but it is for the employer to seek whatever technical advice is necessary to put his house in order.

However, the statutory requirement in the HSW Act to establish safe systems of work is non-specific; it necessitates a hitherto uncalled for expertise in the enforcing authority. This need is met by a professional and technical group within the Executive which possesses expert resources in safety engineering, occupational hygiene and occupational medicine. The latter need is met by the Medical Division, a development of the Employment Medical Advisory Service (a title still applied to the Medical Division's services in the field).

THE EMPLOYMENT MEDICAL ADVISORY SERVICE (EMAS)

Among the Inspectors of Factories who historically administered the Factories Act were some with medical qualifications concerned specifically with the occupational health aspects of that Act. Those Medical

Inspectors of Factories were law enforcement officers; any advisory role they may have wished to fulfil was constrained by that fact, though to their credit they interpreted their duties with a degree of liberality that made them a major contributor to the development of occupational medicine both in Britain and internationally.

In 1972 Medical Inspectors, and the part-time Appointed Factory Doctors to whom certain routine statutory duties had been delegated, were replaced by a body of physicians whose role was no longer inspectorial but advisory. The Employment Medical Advisers were available as consultants on all aspects of health in relation to work to individuals, employers, trade unions, agencies concerned with the disabled and education authorities and others involved with the early stages of the employment of young people. The creation of EMAS was quite independent of the Robens developments, although both were the product of the same philosophy and, indeed, of the climate of 'consensus politics'. The Act establishing EMAS preceded the HSW Act, but was subsequently incorporated into that legislation.

Of recent years the Medical Division of the HSE has devoted a large part of its resources to generating policy on occupational health both nationally and internationally. Its originators saw it as the nucleus of a national occupational health service. To some extent the climate of political opinion has moved away from what is seen as an interventionist philosophy; the emphasis is now placed on the duty of those who create the risks to control them.

Employment Medical Advisers have statutory access to such premises, persons and information as they need to fulfil their duties. Increasingly, in the field, their role is reactive—responding to problems presented to them at short notice by the Inspectorates, by trade unions, individual workers and employers. Much of this reactive work leads to local and national studies of major topics in occupational health, but the greater contribution of the EMA is the day-to-day guidance and reassurance he or she is able to give to workers and employers in a limitless variety of occupations.

THE PATTERN OF LEGISLATION

The relevance of the Health and Safety at Work etc. Act 1974 is elegantly expressed on its first page in the description of the applicability of Sections 1 to 7:

1. General duties.
2. General duties of employers to their employees.
3. General duties of employers and self-employed to persons other than their employees.

4. General duties of persons concerned with premises to persons other than their employees.
5. General duties in control of certain premises in relation to harmful emissions into the atmosphere.
6. General duties of manufacturers etc. as regards articles and substances for use at work.
7. General duties of employees at work.

The essential philosophy of the Act is contained within these Sections. At its heart is Section 2. In Section 2 we find the basic requirement to maintain safe plant and provide *a safe system of work*.

Here, too, we find, for the first time, that constantly recurring qualification 'as far as is reasonably practicable'. The Act recognizes that the ideal is not always achievable; there comes a point where the cost of precautionary measures is grossly out of proportion to the benefits, or it may be that control of a hazard is simply not technically feasible. However, considerations of profitability or budgetary arrangements have no place in assessing 'reasonable practicability'. The onus of proof that a measure is not reasonably practicable lies with the employer.

This comprehensive Section also requires the preparation of a safety policy; it demands that workers be informed of risks and receive appropriate training; it sets forth the arrangements for employee consultation through Safety Committees and Safety Representatives.

Section 2 lays duties on *employers*. The duties of employees to co-operate in furthering measures for their protection and not to obstruct their employers' efforts is expressed in Section 7.

It is recognized that some factors in the maintenance of safe equipment and systems of work are outside an employers' control. For example, it may be necessary to use proprietary materials of which the constituents are unknown, or to use machinery designed and manufactured in circumstances over which the employer has no control. Suppliers and manufacturers, therefore, are obliged by Section 6 to provide safe equipment and materials (again, so far as is reasonably practicable) to ensure that users are made fully aware of any dangers presented by the materials and equipment which they supply.

Sections 3 and 5 are concerned with the protection of persons who, though not directly involved, may be harmed by work activities.

Acts of Parliament, including the HSW etc. Act, permit secondary legislation. This takes the form of Statutory Instruments—regulations applicable to particular activities or sets of circumstances. Historically, such regulations were made in response to public and Parliamentary concern about specific hazards, and this still remains true to some extent. For example, early sets of lead regulations were applicable to such exotic processes as the heading of yarn and the tinning of hollow-ware. In theory, at least, the contemporary approach is rather different, and it happens that the control of the lead hazard provides an excellent illustration of the new philosophy.

The Control of Lead at Work Regulations (Statutory Instrument 1980, No. 1248, to accord them a full title) are applicable to *any* circumstances in which a lead hazard may arise at work, and to absorption by any route—by inhalation, ingestion or through the skin. Where the risk exists the employer has a duty to *identify and assess the degree* of risk. Note that the onus is on the employer and not an enforcing authority to fulfil this duty. The employer must then take all practicable measures to control the hazard.

The Regulations say very little more than this; they are expressed in terms of general applicability. For guidance on how to achieve proper assessment and control an employer must turn to the Approved Code of Practice published in association with the Regulations.

This sets forth currently accepted criteria for assessing the degree of hazard, including recourse to the results of biological monitoring of workers. It specifies the degree of control and of environmental supervision and medical supervision required according to the level and nature of the risk. Lead exposure lends itself rather well to biological monitoring by reference to blood lead levels; the intensity of medical surveillance is also determined by the results of biological monitoring. In fact, the Regulations themselves demand medical supervision of lead workers by an Employment Medical Adviser or by an Approved Doctor; the Code of Practice specifies the criteria for such supervision.

This Approved Code is published by the HSC in accordance with Section 16 of the Act. It is the result of a prolonged period of consultation with many interested parties—employers, trade unions and professional bodies representing relevant expertise. Medical aspects of the Code would have been the subject of discussion with, for example, the Society of Occupational Medicine and the Faculty of Occupational Medicine of the Royal College of Physicians.

An Approved Code of Practice has a unique status in law. Failure to adopt its recommendations is not in itself an offence. An employer must be able to demonstrate, however, that his procedures at least fulfil the objects of the Code in ensuring the health and safety of his employees.

The Commission may supplement a Statutory Instrument and an Approved Code with a Note of Guidance which gives additional assistance in interpreting their requirements. Thus the Health and Safety (First Aid) Regulations 1981 are supplemented by an approved Code of Practice and Notes of Guidance. These Notes have no statutory force and are merely advisory.

Secondary legislation—Statutory Instruments—is subject to rigid Parliamentary procedures and examination before it becomes effective. These processes can take two or three years. Codes of Practice, on the other hand, can at least in theory be published fairly quickly after a much shorter procedure. This enables them to be amended from time to time in the light of new knowledge or the requirements of supra-national legislation—a point to be discussed later.

OTHER GUIDANCE

In addition to Codes of Practice and Notes of Guidance published in association with Regulations, the HSC and the HSE produce an abundance of advisory literature. The MS (medical series) Guidance Notes cover a variety of subjects and are constantly being added to. Similarly the EH (environmental hygiene) series comprise a storehouse of information on the assessment and control of toxic hazards. In this latter series is a Guidance Note of special significance: EH40, Occupational Exposure Limits, which is revised annually. The larger part of this list of exposure limits derives, at the present time, from its predecessor in Great Britain, the list of threshold limit values published by the American Conference of Government Industrial Hygienists. That document had been in use (or misuse) for many years.

EH40 in fact consists of two lists; a list of Control Limits in respect of a limited range of some common, more toxic, materials and a list of recommended limits in respect of a very much wider range of other chemicals. This latter list, in fact, is derived from the ACGIH list but will become less dependent on that source with the passage of time.

Control limits are applied to materials which have been the subject of special assessment and consultation by the Commission. They represent time-weighted exposures which must not be exceeded. Appropriate limits are set for long- and short-term exposures. In setting these limits the Commission takes into account the reasonable practicability of complying with them. Such statutory force as they possess derives from the requirement to maintain a safe system of work in Section 2 of the main Act. At the time of writing control limits are applied to twenty-three materials or groups of materials, but a rapid extension of this list can be expected. It is interesting to note, in this context, a reversion to the old reactive philosophy in setting priorities for consideration and action. It must be suspected that these priorities are to some extent determined by public reaction to a perceived hazard as well as by objective considerations of relative risk.

Recommended limits, on the other hand, are considered to represent good practice; some are well known to be supported by little evidence other than analogy and there has always been a temptation to regard threshold limit values as indicators of relative toxicity. The concept of control limits based on well-documented evidence is a welcome departure.

In addition to Guidance Notes a mass of informative material of many types, directed at a variety of audiences from the shop floor to senior management, is available. The extent and quality of guidance published by the regulatory bodies reflects the duty laid upon the Commission in Section 11(c) of the HSW Act to ensure that 'government departments, employers, employees, organizations representing employers and employees respectively, and any other persons concerned with matters relevant to any of those purpose are provided with an information and advisory service and are kept informed of and adequately advised on such matters.'

ENFORCEMENT

Enforcement of health and safety legislation is, in the first instance, a function of the Health and Safety Executive. The Secretary of State for Employment, however, can invest local authorities with powers of enforcement in areas which he considers appropriate. Thus inspection of premises to which the Offices, Shops and Railway Premises Act applies are normally inspected for the purposes of that legislation by Environmental Health Officers of District or Borough Councils; certain fire regulations are enforced by fire authorities. Robens had hoped for a simplification of health and safety law and a single body of enforcement officers. In fact, however, the greater part of the earlier primary legislation has survived together with the Inspectorates set up to enforce the main Statutes. All Inspectors (nowadays known collectively as Inspectors of Health and Safety) enforce the generalities of the HSW Act, but the historic bodies such as HM Factory Inspectorate, HM Mines and Quarries Inspectorate and so on still work more or less exclusively in the areas which their titles suggest. Contravention of health and safety legislation is a criminal offence for which offenders may be prosecuted in a court of summary jurisdiction or on indictment in a Crown Court (or their equivalents in Scotland). Prior to the HSW Act, indeed, the only action open to Inspectors was prosecution. The effectiveness of such action was limited by the fact that the 'occupiers' of premises to which, for example, the Factories Act applied were often corporate bodies—large companies with ultimate managements remote from the shop floor. The financial penalties exacted by the courts often meant little to them and only rarely could senior members of management be made personally accountable for an offence.

Two important aspects of the HSW Act have enormously increased the effectiveness of enforcement. The first is that an *individual* can be brought to account for failure to observe the general requirements of the HSW Act if he is clearly identifiable as the person accountable for seeing that those requirements are observed. He may, for example, be specified in the organization's written Safety Policy as being so accountable. In addition, Section 7 of the Act lays specific duties on individuals to co-operate in measures taken to protect their health and safety.

The second major contribution to the effectiveness of enforcement is the ability of an Inspector to serve Statutory Notices. These are of two kinds: A *Prohibition Notice* may be issued where an Inspector believes there to be an imminent serious risk to health or safety which can only be allayed by stopping the relevant activity forthwith, although in some circumstances an employer may be given a limited time to put things right. The more commonly used Statutory Notice is an *Improvement Notice*. This is used where there may be no imminent serious risk, but where there is a contravention of a statutory provision. The occupier will receive a Notice stating the nature of the contravention and stating a time during which corrective action must be taken.

Both types of Notice are subject to appeal before an Industrial Tribunal. Where some contravention of a statutory requirement has called for an Improvement Notice the matters under consideration are usually self evident, and, indeed, appeals against such notices are rare.

A Prohibition Notice, however, can have immediate and sometimes damaging implications for an employer; moreover they call for judgements of the degree and immediacy of a hazard based on an Inspector's skill and experience and any professional advice available to him.

Medical opinion may be crucial to the decision on whether or not to prohibit a particular activity. The probability of acute effects of exposure to a toxic material can usually be fairly simply assessed; more complex judgements have to be made in assessing risks having potential delayed effects. For example, is prohibition justified in the case of exposure to small doses of fibrogenic dusts? One view is that although ultimate disease may be a response to a cumulative effect of many such doses, each dose must be regarded as contributing to the eventual disability and therefore represents an immediate risk to health which cannot be allowed to continue.

So far there have been too few cases argued before tribunals to give firm guidance on these points.

The historic approach to the enforcement of legislation was punitive. Society expected, and still expects, retribution for breaches of the law which put human life and well-being at risk. Similarly, the

Common Law was, and is, used to exact compensation for injuries attributable to an employer's negligence. In both contexts the importance of apportioning blame for an untoward event takes precedence over the need to establish its basic cause. In the United Kingdom we have yet to overcome the problems created by our system of compensation at common law. The Health and Safety at Work Act does go some way to diminish the exclusive emphasis on punishment characteristic of earlier legislation and to replace it with a philosophy of what may be termed 'enforceable guidance'. On the other hand, there remains the power to punish those in persistent, flagrant or serious breach of the law. This has been strengthened by the concept of the accountability of individuals as well as corporate bodies for the maintenance of healthy and safe conditions of work.

It is interesting to consider developments in attitudes in the United States to law enforcement in this field. The American Occupational Safety and Health Administration was set up under an Act of 1970. It is a Federal body essentially concerned with ensuring adequate administration and enforcement of health and safety law by State governments. Its creation reflected social attitudes in the United States at the time. It now seems, however, that the current American political philosophy of 'less government' has resulted in a considerable diminution of resources available for law enforcement in this field, and a much greater reliance on the willingness of employers to comply voluntarily with the law and good practice.

In the United Kingdom the Health and Safety Commission has always supported a philosophy of 'self-regulation' in health and safety matters, although the term 'self-regulation' has been interpreted in various ways. The commonly understood meaning is that it is for the employer to ensure safe working practices and compliance with the law; he cannot rely on the constant availability of Inspectors or advisory bodies from central or local government to do this for him. Thus it is for an employer to provide his own occupational medical advice if it is needed. The Employment Medical Advisory Service will give immediate advice on a problem, and will give guidance on the level and type of future medical guidance needed. Where necessary it will ensure the adequacy of an occupational health service and any medical procedures required by statute, but it cannot itself provide such a service on a continuing basis.

There is some evidence of the US philosophy of 'less government' affecting British policy in recent times. In 1984 a major construction company was asked to regulate its own affairs for a trial period. All formal inspection would be withdrawn. A subsequent examination of accident records should show whether or not this degree of self-regulation was effective.

No conclusions have been published, but the indications are that the experiment was at best inconclusive. The company involved was chosen because it had a good safety record and a good professional safety organization.

In fact, HM Factory Inspectorate has to set priorities for the deployment of its limited force of inspectors. These priorities are based on a points system; the best employers attract less frequent attention than those with a poor safety record. In practice, therefore, the company concerned may well not have attracted the attentions of the Inspectorate in any case.

Self-regulation is an admirable goal; in the event, however, the ultimate constraint on self-regulation is the existence of Statute Law which must be enforced if it is to serve any purpose.

RESEARCH AND INFORMATION

Occupational health has long since ceased to be merely the practice of general medicine and hygiene in the workplace. Direct observation and anecdotal evidence sufficed to identify the major classical occupational diseases. Modern occupational health, however, necessitates the application of many academic disciplines and advanced technologies. Even more does it necessitate dissemination of the vast amount of information generated by research and the acquisition of new knowledge. This is recognized in British legislation.

Section 11 of the HSW Act lays a duty on the Commission 'to make such arrangements as it considers appropriate for the carrying out of research, the publication of the results of research and the provision of training and information in connection with those purposes and to encourage research and the provision of training in that connection by others.'

A large part of HSE's activities is devoted to fulfilling this duty. Policy making is preceded by the preparation of background papers and reviews, which often require original research, and which are usually published as definitive contributions to particular topics. Research in academic institutions and bodies such as the Medical Research Council is commissioned, and proposals originating outside HSE are considered for funding. HSE has its own resources for research, particularly in safety technology.

The Medical Division undertakes local and national investigations, largely epidemiological in nature, but also contributes by identifying topics worthy of investigation by other specialist bodies and, where necessary, by funding such investigations.

The promotion of health and safety in a highly technical society requires the dissemination of a huge quantity of information to many and varied audiences.

At the simplest level illustrated cards warn workers of particular hazards. Professional health and safety practitioners need more detailed advice to be found in the several series of HSE Guidance Notes, one group of which is devoted to occupational medical topics.

Employers and trade unions are given advice on the interpretation and application of policy and legislation.

THE EFFECTIVENESS OF LEGISLATION

An assessment of the effectiveness of legislation is fraught with difficulty. Reliable data on the incidence of accidents and ill health are usually gathered on a national basis. International comparisons are impeded by the fact that reporting methods vary from country to country, and even within any one nation the data cannot easily be related to reliable measures of exposure. Meaningful *rates* of accident and disease are difficult to establish. Examination of trends gives rise to the suspicion that much of the historic data in this field reflect levels of economic activity rather than the effectiveness of legislation or any other factor.

The Health and Safety Executive has recently attempted to review the effectiveness of its own legislative and enforcement activities.[3] The review recognized the influence of sometimes unmeasurable variables on attempts to assess effectiveness. One of the few clear conclusions to emerge was the immediate and continuing beneficial impact of Statutory Improvement and Prohibition Notices on employers' attitudes and behaviour.

Except in the case of Notice procedure it proved practically impossible to establish reliable measures of effectiveness in the field of occupational health. As is so often the case, exposure to lead provided the best documented example of the legislative impact on occupational

Table 17.2. Distribution of blood lead concentrations in workers at a lead refinery, 1981–83 (From *Measuring the Effectiveness of HSE's Field Activities*. Occasional Paper Series OP11. London: HMSO, 1985)

Blood lead concentration range (μg/100 ml)	% of levels recorded in each range		
	1981	1982	1983
<40	16·8	16·1	32·2
40–59	27·2	31·6	45·2
60–79	44·5	45·4	20·3
>80	11·5	6·9	2·3
Number of workers under supervision	416	421	345
Number of suspensions from work	17	9	2

Note: Decrease in lead levels 1981–83 and decrease in proportion of workers suspended from work 1981–83 is statistically significant at the 99% confidence level.

health. It was possible to demonstrate a clear change for the better in the distribution of blood lead levels following the introduction of the Control of Lead at Work Regulations and its associated Code of Practice in 1982 (*Table 17.2*).

THE EUROPEAN CONTEXT

The European Coal and Steel Community, the European Energy Community and the European Economic Community together constitute the organization known collectively as 'The European Communities'. This combination is usually referred to, somewhat inaccurately, by the initials EEC. Ultimate authority rests with the Council of Ministers, but the executive arm is the European Commission. Among the nineteen Directorates of the Commission is Directorate General 5 (Social Affairs).

The decisions of this body are promulgated as Directives, a near equivalent of British Statutory Instruments. Directives are binding on Member States.

In formulating a Directive the Directorate has access to advice from a standing Advisory Committee consisting of six representatives of each Member State; they, in turn represent the interests of government, employers and workers in each State. Somewhat confusingly there is also a Health and Safety Directorate within the Fifth Directorate General which has among its functions that of ensuring liaison with other divisions of the Commission. The situation is even more confused by the exclusion from the remit of the Advisory Committee of matters to do with mineral extraction and ionizing radiations. These are dealt with by the European Steel and Coal Community and Euratom respectively. The preparation of a Directive involves prolonged and complex investigations and consultations. Some of this work is undertaken by working parties and groups of experts concerned with specialist areas. After initial approval by the Council of the EC it is studied by other major institutions of the Communities and by governments. British comments and suggestions for amendments would be made by the Health and Safety Commission.

The process of formulating a Directive is inevitably influenced by factors other than those of a purely technical nature. It is important to ensure, for example, that the economic interests of an industry or group of industries within any member State are not unfairly prejudiced by difficulty in complying with a directive arising from economic or commercial factors peculiar to that State. Each member State has its own body of legislation developed over many years. There may be many cultural and legal constraints on any change in the law. Directives are the result of a compromise between many interests;

for all that, they generally succeed in achieving their objectives while being sufficiently flexible to allow interpretation in the light of differing national circumstances.

In 1980 the Commission issued a Directive of particular relevance to occupational health—this was Framework Directive 80/1107/EEC. It concerns the protection of workers against chemical, physical and biological agents at work. A Framework Directive initiates the generation of legislation on particular topics. So far there are specific Directives on lead and asbestos. These are to some extent already reflected in British Statutory Instruments and will be more fully incorporated in British legislation in the near future. It seems probable that benzene, arsenic, chlorinated solvents, nickel and noise will be among other subjects to receive early attention.

The Framework Directive is reflected in the proposed British regulations on the Control of Substances Hazardous to Health. This legislation will replace many of the reactive statutes relating to toxic substances made under the various Factory Acts.

A general interpretive Code of Practice will, of course, be appended to the COSHH regulations (*see below*). An important feature of this Code is that for the first time it gives additional status to control limits in establishing them as criteria of compliance with the Code. In addition, there is a specific Code of Practice relating to carcinogens and further Codes of Practice concerning vinyl chloride and fumigation are proposed.

HEALTH SURVEILLANCE AND THE CONTROL OF SUBSTANCES HAZARDOUS TO HEALTH (COSHH)

Existing health and safety legislation contains surprisingly few requirements for health surveillance. There are, of course, requirements in some sets of regulations for medical examinations in the case of exposure to certain substances such as lead or a group of substances or physical agents such as carcinogenic substances and ionizing radiations. The controversies that arise over the relevance of medical supervision in the case of these latter subjects illustrate the difficulty of laying down specific medical criteria for judging the effects of exposure.

The recognized objectives of statutory medical examinations are twofold: to ensure a worker's fitness for a task (the Diving Regulations are an example) or to detect early pathological changes in the worker. The Carcinogenic Substances Regulations are an example of the latter. The first objective is unexceptionable; the second, however, needs qualification. The detection of pathological change is valuable if the change is reversible or susceptible to treatment.

Thus the early detection of skin cancer and possibly bladder cancer is beneficial. Unfortunately this claim cannot, in general, be made for cancer of the lung. It would seem, unfortunately, that current European practice does not always consider the likely benefits of a routine procedure before enshrining it in legislation. Thus a requirement for the routine radiology of asbestos workers takes no account of even WHO criteria for routine screening: that a procedure should be capable of detecting disease at a stage where a fatal outcome may be avoided or the quality of remaining life improved. At the present time this is hardly true of lung cancer, and is very much open to debate in respect of bladder cancer.

Biological monitoring, strictly speaking, does not amount to medical examination. It is intended to detect the absorption of toxic substances at a stage prior to the onset of pathological changes or certainly before such changes are irreversible. It did not feature in historic British legislation but, as we have seen, features prominently in the Lead Code of Practice.

In fact, biological monitoring has proved to be less useful than had been hoped. The interpretation of measurements of toxic substances or their metabolites in body fluids depends on an understanding of the fate of these substances in body tissues, of the dynamics and the pathways of metabolic processes and the way in which individual and environmental factors can affect them. In all too many instances too little is known of these factors to permit relating biological measurements to exposure or to assess the probability of pathological changes. For this reason, with the notable exception of lead, requirements for biological monitoring do not feature in legislation to any extent.

Statutory Instruments are the result of a long process of drafting and Parliamentary procedure; they are, therefore, extremely difficult to modify in the light of new knowledge or changes in philosophy. Codes of Practice are much more flexible in application and ease of modification. Thus current criteria for the medical monitoring of lead workers is to be found in the Code of Practice rather than the Control of Lead at Work Regulations. Indeed, that Code has already proved to be readily amendable in response to the demands of the Lead Directive of the European Communities. The flexible approach is inherent in COSHH. 'Health Surveillance' is given the widest possible meaning.

At the simplest level, where the assessed risk is of a lower order, surveillance may amount to no more than keeping basic personal facts about an individual and the nature of any exposure. A greater risk invites routine assessment of health by a responsible person or, at a higher level by an occupational health nurse. As we have seen, regular surveillance by a doctor is only suggested in respect of a range of more toxic substances.

Even where the attention of a physician is required there is provision for the delegation of routine screening procedures to other professional staff such as occupational health nurses.

British practice in the routine surveillance of workers is very well expressed in a Health and Safety Executive guidance note *Health Surveillance by Routine Procedures*.[4]

European legislation

The European Communities can be expected to have an increasing influence on the development of health and safety legislation in Britain. Historically protection of the worker in the United Kingdom has depended on the Common Law; an injured person was always free to take action at civil law against a negligent employer. The development of the trade union movement made it financially possible for an individual to take such action, and it could be claimed that threat of such action has, in the past, had a more widespread and positive effect on the maintenance of safe working practices than statute law. This was even more the case in the United States, historically much less interventionist in this field and where employers' insurers have often taken the lead in promoting health and safety at work. To a large extent the Health and Safety at Work Act represents the incorporation of longstanding common law rights and duties into statute law, particularly in respect of the requirement to maintain safe systems of work. However, it should be noted that an individual cannot bring a *civil* action against his employer merely because he is in breach of the general provisions of the HSW Act expressed in Sections 2 to 7.

In many European nations law is based on the Napoleonic Code. This tends to result in the generation of detailed, specific regulations rather than broad enactments based on first principles. In France, Italy and Belgium codified legislation requires the setting up of occupational health services. These services have been characterized by the introduction of schemes of routine medical examination on a large scale whose value we might doubt. This over-reliance on the routine medical examination in preventive medicine can be detected in, for example, European Communities directives for the control of the asbestos hazard. Most British authorities would doubt the value of repeated chest radiology and physical examination of workers exposed to low environmental levels of asbestos.

Highly codified law is inflexible; compliance and enforcement become increasingly selective with the passage of time. The modern British pattern of short, simple sets of Statutory Instruments supported by Codes of Practice which can be readily amended has yet to gain general acceptance elsewhere.

THE WIDER INTERNATIONAL CONTEXT

Robert Murray has succinctly described[5] the way in which demographic, geographical, social and political factors affect the development of occupational health and its associated legislation in different communities. He summarized them in *Table 17.3*. It must be remembered that the clear separation of occupational health services and its associated legislation from other aspects of therapeutic and preventive medicine is peculiarly British. The existence of a comprehensive National Health Service which excludes preventive occupational health is an ultimate determinant of the pattern of occupational health services and occupational health legislation.

In many countries occupational health services provide total care of the worker, and sometimes of his family. This is certainly true of Eastern European countries and of large organizations in many others; the Philips organization in Holland, French State Railways and Indian Railways are examples. In France, as Murray has pointed out, comprehensive occupational health services were developed for transport, mines and industry during the German occupation. (Shades here, perhaps of Bismarck's surprisingly enlightened approach to matters of social welfare.) This survived after the Liberation. Occupational health services are paid for by the employer and must meet the requirements of legislation administered by the Medical Inspectorate of Factories. The French pattern was the basis of International Labour Organization recommendations which themselves determined the European Communities' approach to occupational health.

The United States of America inherited the British tradition of common and statute law. By tradition, however, its governments have been less interventionist in social affairs. The promotion of health and safety at work until recent years was regarded as the responsibility of employers. As the penalities of common law litigation were largely met by their insurers it was, as we have noted, insurance companies who often took the lead in insisting on adequate standards on their clients' premises.

In recent years there has been a considerably increased level of intervention at both Federal and State levels. Federal legislation now includes the Federal Coal Mines Safety Act of 1969 and the Occupational Safety and Health Act of 1970. The Occupational Safety and Health Administration (OSHA) of the Department of Labour administers the legislation, largely to criteria promulgated by the National Institute of Safety and Health (NIOSH). Each State is free to develop its own legislation and appropriate administration, but this must meet minimum Federal standards and is subject to Federal review. There is something of a parallel here with the requirement that each member state of the European Community must at least meet the minimum

Table 17.3. Factors governing patterns of occupational health administration (adapted from *Trans. Assoc. Ind. Med. Off.* 1963; **13**:89)

Factors	Traditional (e.g. UK)	Totalitarian (e.g. USSR)	Private initiative (e.g. USA)	Developmentary (e.g. India, Africa)
Historical development of industry	Over 100 years gradual development	Less than 50 years, rapid development	Less than 100 years, rapid development	Less than 20 years, very rapid development but still largely peasant communities
Geographical features and communications	Small countries, efficient internal communications	Extensive geography, slow internal communications except air transport	Extensive geography, good internal communications	Vast countries, poor internal communications
Administration	National rather than regional; ministry of labour responsible for occupational health administration	Revolution; republics with measure of autonomy; ministry of health responsible for occupational health administration	Loose federation of largely autonomous states; minimum of legislation; departments of health responsible as a rule. Recent changes in the USA	Legacy from former colonial status, weak at lower levels; department of health usually responsible
Industrial structure	Capitalist; large number of small factories; few large enterprises; some nationalization	Large units, state controlled	Many small factories but also many large enterprises	Town and country artisans; new modern factories, many with foreign capital
Social philosophy	Family is unit of social service and resources are deployed in interests of families	Worker is unit of social service	Laissez faire; occupational health services provided as fringe benefits	To achieve greatest good of greatest number with minimum resources; to give benefit to workers of such resources as are available to them in view of their economic contribution
Health resources	Complex health resources of high standard; 1 physician for less than 1000 population	Three medical faculties—therapeutics, preventive medicine, and paediatrics; 1 physician for 600 population	Very high standard; 1 physician for 800 population; medical-care cost is high	Poor health resources; 1 physician for up to 20 000; much endemic disease; employers obliged to provide occupational health services

requirements of a European Directive.

The level of enforcement of legislation in the United States tends to vary according to the views of the administration in office. At the time of writing it would appear that a philosophy of non-intervention is in the ascendancy and that enforcement has been weakened to some extent.

The International Labour Organization, a surviving agency of the League of Nations and now an agency of the United Nations is the major body devoted to stimulating legislation on occupational health and safety at the international level. It has a tripartite structure, representing employer, trade union and government interests in each participant country. The results of its deliberations are expressed as Recommendations which member nations are expected to ratify and adopt. In 1959 the ILO published a basic philosophy of health and safety at work as Recommendation 112, 'Occupational Health Services in Places of Employment'. More than a quarter of a century later that Recommendation is still the best expression of the fundamental tenets of health at work. To some extent it is reflected in the early deliberations of the European Communities.

Another United Nations Agency, the World Health Organization, has an influence on occupational health at the international level. Indeed, in 1950 a WHO/ILO joint committee produced a classic definition of occupational health which is undoubtedly a major basis of the framework within which legislation, at least in Europe, is created:

> Occupational health should aim at the maintenance of the highest degree of physical, mental and social well-being of workers in all occupations; the prevention among workers of departures from health caused by their working conditions; the protection of workers in their employment from risks resulting from factors adverse to health; the placing and maintenance of the worker in an occupational environment adapted to his physiological and psychological equipment and, to summarise: the adaptation of work to man and each man to his job.

The WHO, however, is primarily concerned with community health. Its contribution to occupational health is in the field of fitness to work rather than preventive aspects.

OTHER LEGISLATION

The Employment Protection Act 1974

Alleged unfitness for work is often a cause for considering the dismissal of a worker. An employee who considers himself to have been unfairly dismissed has recourse to an industrial tribunal. The tribunal will take professional medical evidence into account in its consideration of the case, and it is clear that the specialist opinion of an occupational physician, in addition to those of an employee's personal medical attendants, carries considerable weight. Indeed, one of the functions of an employment medical adviser is to advise an industrial tribunal on such matters. In practice he or she will usually be approached by the Advisory, Conciliation and Arbitration Service in such cases.

The Act also protects those workers who may be temporarily prevented from carrying out their normal duties as a consequence of the advice given by an employment medical adviser or an appointed doctor in the context of a *statutory* medical examination.

The Sex Discrimination Act 1975

Women have always been regarded as needing special protection by statute, historically because of their alleged lesser physical ability. We can suspect that, in addition, Victorian moral susceptibilities were offended by the thought of women working at night with men, as well as the degrading conditions of women working in mines so well illustrated in the 1842 Report on the Employment of Children in Mines. There was, too, a widely held belief that women were more susceptible to toxic materials than men. This belief was strongly held in respect of lead, and women were excluded from the worst exposures to it by several sets of regulations. Of recent years some of these restrictions have become increasingly difficult to justify on any objective grounds; legislation on sex discrimination, to a large extent, merely reflected contemporary attitudes.

Certain occupational hazards present a threat to the developing fetus; lead exposure is one example, ionizing radiations another. Furthermore, fetal damage may occur very early in pregnancy before a woman is aware that she is pregnant. One conclusion is that women of child-bearing capacity should be excluded from work which might threaten the fetus. This attitude is enshrined in the Code of Practice annexed to the Control of Lead at Work Regulations; women of 'child-bearing capacity' are excluded from work where there is a probability that blood lead levels will exceed a level currently regarded as representing the upper limit of that which may be found in unexposed workers.

This raises many philosophical and practical difficulties, not least in defining the term 'of child-bearing capacity'. The right of the State to override the personal decision of the woman in this context is also open to question. The matter is the subject of continuing debate; the Equal Opportunities Commission has understandable reservations about such restrictions. Such 'discrimination' against men cannot be found in Statute law, but there are such processes as the preparation of female sex hormones where men are peculiarly vulnerable to adverse effects.

A MODEL FOR LEGISLATION

This chapter has attempted to outline the development and current pattern of legislation for occupational health and safety. Inevitably it is based on the British experience. This is not merely because that is the pattern with which a British author is most familiar; such legislation is a characteristic of an industrial society. Great Britain has the longest history of industrialization, closely followed by Europe and the United States. These western nations also have in common a particular style of democratic government, a background in at least some of them of the Common Law and of legal practice based on Roman law and more recent European codified statute.

In Great Britain and the United States current debate on health and safety legislation has reverted, to some extent, to the early nineteenth century arguments on how much State intervention in such matters is desirable or justifiable—intervention or *laissez faire*.

At the same time 'advanced' societies are increasingly demanding of legislation to control perceived hazards created by modern industry. Some of these hazards are highly marginal in comparison to other dangers presented by modern society, road traffic accidents to mention but one obvious example. Because of this we see something of a reversion to nineteenth century thinking: the reactive approach to legislation. This is epitomized by the mass of recent legislation devoted to the control of the asbestos hazard both in industry and the community.

In Victorian times the power to influence events lay with a very small proportion of the population; it was the reaction of a relatively well-informed minority that, for good or ill, affected events. In modern times the extension of the democratic process and the development of means of mass communication has made social reaction much more powerful. Such reaction is rarely based on an objective appraisal of a problem but on a perception of risk determined by a host of complex social and psychological factors of which we as yet know little.

This has had an influence on legislative bodies such as the British Health and Safety Commission and Executive, the policy makers in the European Communities and the American Offices of Safety and Health Administration and the Environmental Protection Agency.

At the professional level these organizations represent a vast body of knowledge of, and ability to assess, environmental problems. They exist, however, by permission of democratic societies whose approach is anything but expert or objective. To an extent safety and health administrations have a vested interest in what *society* perceives as threatening. It is this perception, rather than the 'scientific' approach, that ultimately determines legislation.

These comments apply to the western nations. In more autocratic regimes political philosophy is clearly a determinant of attitudes, in theory if not always in practice. In less developed nations occupational health and safety, at least in so far as it is defined as a separate entity, can only be given very low priority in comparison to economic threats to the community or the effort that must be devoted to the control of endemic disease, malnutrition and infant mortality. In those circumstances it is inappropriate to consider legislation of a type developed in the rich nations of the world.

Unfortunately the poorer countries increasingly face the introduction of advanced industrial technology into contexts where the means of protecting workers and the community at large are rudimentary. There is something bizarre, as well as tragic, in the deaths of two thousand people in impoverished India as the result of exposure to an exotic chemical emanating from an advanced pesticide manufactory.

Perhaps such events are a contemporary equivalent of the typhus fever that killed the Manchester apprentices leading in time to an extension of the benefits of the protection of the law to many more of the world's workers.

REFERENCES

1. Young GM. *Victorian England, Portrait of an Age*. Oxford: Oxford University Press, 1936.
2. Robens (Chairman). *Safety and Health at Work, Report of the Committee 1970–72. Cmnd 5034*. London: HMSO, 1972.
3. Health and Safety Executive. *Measuring the Effectiveness of HSE's Field Activities*. London: HMSO, 1985.
4. Health and Safety Executive. *Guidance Note MS18: Health Surveillance by Routine Procedures*. London: HMSO, 1981.
5. Murray R. Occupational health in other countries. *Trans. Assoc. Ind. Med. Off.* 1963; **13**:89.

18. OCCUPATIONAL HEALTH PROGRAMMES

Tarek M. Khattab

INTRODUCTION

Occupational health services sometimes implement their general goal in the form of a working document designated as a comprehensive programme. More often, such services have only some of their specific objectives written down as working documents which are referred to as component programmes. A few of those have been extensively implemented, and in the process, have become more uniform. Examples of such popular component programmes include pre-employment examinations programmes,[1] hearing conservation programmes[2] and radiation protection programmes.[3]

COMPREHENSIVE OCCUPATIONAL HEALTH PROGRAMMES

The enormous variation in the range of industrial, commercial and agricultural activities renders it impossible to advocate a detailed model suitable for all undertakings.[4] Existing comprehensive programmes differ with differences in the availability of medical care in the area, the size of the work force to be served, the type of operations in which they engage,[5] and the geographical spread and organizational pattern of the enterprise.[6] Such wide differences between various comprehensive occupational health programmes, both in scope and components, may be responsible for the paucity of relevant literature. Thus, the most recent statements on the subject were issued by the American Medical Association in 1971 and by the American Occupational Medical Association in 1979,[7,8] while the last extensive review article was published in 1975.[9]

Comprehensive programmes' scope and documentation

Executive management must give its full support to the programme to assure the full cooperation of line management. It must be willing to give both the authority and the responsibility for carrying out the programme to the chief health professional in the organization.[10] The programme requires an organizational structure that fosters communication and coordination of activities between the various related skills.[8] That organization must report to a management level high enough to ensure that top executive management is continuously aware of any problems so that appropriate action is taken.[8] The programme document must state management's policy, define the organization's objectives, refer to utilized standards and describe individual component programmes. It will also standardize the use of all forms and procedures. Furthermore, it should include a procedure for programme evaluation such as that advocated by Webb,[11] or that formulated by the Occupational Health/Safety Programs Accreditation Commission.[12] The whole programme is put together in a single document enclosed in ring binders to facilitate updating.[9]

Automated comprehensive programmes

Several components programmes were partially computerized by the early seventies.[13] These included, among others, routine physicals in the form of automated multiphasic health testing,[14] computer-oriented hearing conservation programmes,[15] as well as industrial hygiene data.[16] The next logical step was computer linkage of relevant personnel, medical and environmental exposure data in what is known as occupational health surveillance or industrial health monitoring systems.[17]

COMPONENT OCCUPATIONAL HEALTH PROGRAMMES

The American Occupational Medical Association described the occupational health programme in terms of essential and elective components.[8] These were discussed as activities rather than component programmes. Konzen presented the concept of special (component) programmes, directed at specific hazards, as part of a comprehensive occupational health programme.[10] He mentioned as examples hearing conservation, eye protection, respiratory protection, and so on.[10] Recently, McDonagh highlighted wellness promotion activities such as smoking cessation,[6] which is frequently implemented as a programme, as well as employee assistance programmes.[6] Little else was written collectively about special (component) programmes, although literature abounds with articles discussing them individually.

Definition

It was mentioned, in the introduction, that both comprehensive and component occupational health programmes are working documents to implement stated goals or objectives. The implementation responsibilities must be assigned, to specific job-holders, in very certain terms. Implementation procedures, forms and standards must be described in detail. Additionally, periodic review and updating must be specified as an integral aspect of the programme. The following listing is a brief review of some of the common programmes, which are either implemented as components of a comprehensive occupational health programme, or are implemented independently in the absence of the latter.

Physical examinations programme

Pre-employment (preplacement) health examinations, were traditionally associated with occupational medicine.[18] The US Equal Employment Opportunity Act of 1973 mandated radical conceptual changes for such examinations. They provide that assessments be based on the ability to perform essential or critical job tasks,[19] rather than basing them on general health fitness. The employer must demonstrate to the satisfaction of the regulatory agencies that the standards employed are job-related.[20] Retirement physicals are less controversial and should be performed to establish a baseline in case of future compensation claims. Routine periodic health screening, including executive health physicals, may be a good public relation tool, but there is insufficient evidence to substantiate its benefits.[21] On the other hand, periodic examinations are essential for those employees who work with hazardous materials or under hazardous conditions, and those employees in jobs which may create hazards to others.[22]

Hearing conservation programmes

Big industry resisted government regulations to implement such programmes although they are already instituted in most western industrial organizations. Thus, the UK still has no such regulations, while the US implemented the Occupational Safety and Health Administration (OSHA) Hearing Conservation Amendment only in 1985.[23] The American Occupational Medical Association's approved guidelines state that the basic elements of a hearing conservation programme include: (1) exposure measurement; (2) engineering controls; (3) hearing protection; (4) audiometric testing; (5) education and training; and (6) programme assessment.[24] OSHA's 8 hours permissible exposure level (PEL) is 90 dBA, and its action level is 85 dBA at which level the programme must be instituted.[23]

Vision conservation programmes

The basic element of vision conservation programmes is eye protection from trauma, foreign bodies, welding, radiation and chemicals.

Employers have to supply protective goggles, face shields and welding or radiation protective lenses.[25] In areas with the potential for chemical splashes, warning signs and emergency eyewash facilities must be installed.[26] Employees requiring prescription spectacles must wear special safety lenses that can withstand high impact forces.[27] Another element in vision conservation programmes is vision screening. This is commonly done as part of general preplacement physicals, or sometimes with general periodic physicals. It may also be performed individually, on a periodic basis for employees in jobs requiring special vision standards. Special programmes for early glaucoma detection are implemented by an increasing number of US-based corporations.[28]

Pulmonary protection programmes

The medical surveillance aspects of such programmes require a medical and occupational history, a chest X-ray and an assessment of ventilatory capacity.[29] The 1966 British Medical Research Council's Questionnaire on Respiratory Symptomatology or similar questionnaires were advocated.[29] Interpretation of chest films is performed both clinically and semi-quantitatively. The latter is done in accordance with the ILO guidelines on the basis of type, profusion and extent of the opacities.[30] Assessment of ventilatory capacity is performed on a screening rather than clinical basis. The basic parameters which are usually tested are the forced vital capacity (FVC), the forced expiratory volume in the first second (FEV_1) and the percentage of air that can be expelled in the first second (FEV_1/FVC %).[31,32]

Radiation protection programmes

Personnel surveillance is commonly instituted, in those programmes, for monitoring the cumulative exposure of individuals to low level external radiation. Such exposures, as encountered in medical diagnostic radiology services, are monitored by wearing film badges or thermoluminescent dosimeters.[33] The accumulated dose is determined every month, and if it exceeds the calendar quarterly standard, the individual is removed from radiation work. He is allowed to resume his original job after one or more calendar quarters when his accumulated dose meets the respective quarter or annual standard. Most regulations require personnel surveillance when there is a reasonable probability of exceeding 25 per cent of the occupational dose equivalent limit.[34] Such limit for prospective occupational whole body exposure is commonly set at 5 rem in any one year.[35] This was extrapolated by US regulations to 1·25 rem per calendar quarter. It can be increased to 3 rem during any quarter provided that the total accumulated occupational dose does not exceed 5(N–18), where 'N' is the individual's age in years.[35] Other aspects of a radiation protection programme have been reviewed in NCRP Report No. 59.[34]

Employee assistance programmes

Management has to formulate a well-defined policy for alcoholism. It should be widely circulated as a written policy statement.[36] The 1974 Special Report to the US Congress, entitled Alcohol and Health, described the bases for occupational programming of the problem.[37] It advocated the 'troubled-employee' approach, later known as the 'employee assistance' approach, which is based on constructive confrontation.[38] Supervisors are trained to identify employees with impaired job performance without trying to pinpoint the cause or relate it to possible alcohol use. They are to confront such employees, with documented evidence of their impaired performance, and warn them about pending disciplinary action. This confrontation undermines the employee's rationalization that, as long as his work is not affected, he does not have a drinking problem. Furthermore, confrontation usually results in a crisis which can motivate the employee to act. He would often accept the help offered by the organization rather than face possible adverse action that shatter his central role in life.[37] Such occupationally-oriented programmes yield the highest reported rates of successful recovery from problems related to alcohol.[37]

Other component occupational health programmes

Numerous specialized component programmes are developed to meet specific needs of various organizations or to comply with certain government requirements. There is also a growing tendency to incorporate wellness promotion activities in the form of occupational health programmes. Such activities include health education, smoking cessation, hypertension detection, physical fitness, stress management, nutrition and health risk assessment.[6,39]

REFERENCES

1. Health and Safety Executive. *Pre-employment Health Screening.* HSE Guidance Note MS 20. London: HMSO, 1982.
2. American Occupational Medical Association's Noise and Hearing Committee. Guidelines for the conduct of an occupational hearing conservation program. *J. Occup. Med.* 1982; **24:**772–773.
3. Health and Safety at Work etc. Act, 1974. The Ionizing Radiations Regulations 1985 (SI 1985/1333). London: HMSO, 1985.
4. Murray R and Schilling RSF. Functions of an occupational health service. In: Schilling RSF (ed.) *Occupational Health Practice.* London: Butterworths, 1981: 145.
5. Health and Safety Commission. *Prevention and Health: Occupational Health Services, the Way Ahead.* London: HMSO, 1977: 26.

6. McDonagh TJ. Management of an occupational health program within an industrial setting. *J. Occup. Med.* 1984; **26**:263–68.
7. American Medical Association. *Scope, Objectives and Functions of Occupational Health Programs*. Chicago: AMA, 1972.
8. American Occupational Medical Association. Scope of occupational health programs and occupational medical practice. *J. Occup. Med.* 1979; **21**:497–499.
9. Howe HF. Organization and operation of an occupational health program. *J. Occup. Med.* 1975; **17**:36–400, 433–440 and 528–540.
10. Konzen JL. Design and operation of an occupational health program. NIOSH: *The Industrial Environment—its Evaluation & Control*. Washington, D.C. 20402: Superintendent of Documents, U.S. Government printing office, 1973: 693–701.
11. Webb SB. Objective criteria for evaluating occupational health programs. *Am. J. Public Health* 1975; **65**:31–37.
12. Occupational Health/Safety Programs Accreditation. Standards, Interpretations and Audit Criteria for Performance of Occupational Health Programs. Akron (Ohio), 1977.
13. Craig JL and Derryberry OM. Applied concepts of automation in an occupational medical program. *Ind. Med.* 1971; **40**:9–17.
14. Breckenridge RL. Experience with automated multiphasic health testing. *Ind. Med.* 1971; **40**:18–23.
15. Rightland N et al. A computer-oriented hearing conservation program. *J. Occup. Med.* 1974; **16**:654–658.
16. Snyder PJ, Bell JR, ZG and Samelson RJ. The computerization of industrial hygiene record. *Am. Ind. Hyg. Assoc. J.* 1979; **40**:709–720.
17. Rossi DA, Cox JD and Seger MJ. Computerization of medical and exposure records. *Am. Ind. Hyg. Assoc. J.* 1982; **43**:444–450.
18. Schussler T, Kaminer AJ, Power VL et al. The preplacement examination, an analysis. *J. Occup. Med.* 1975; **17**:254–257.
19. Hogan JC and Bernacki EJ. Developing job-related preplacement medical examinations. *J. Occup. Med.* 1981; **23**:469–475.
20. Hanks, TG. Medical standards for employment in relation to nondiscrimination laws. *J. Occup. Med.* 1977; **19**:181–187.
21. D'Souza M. Screening for all: Excellence or extravagance. In: Ward Gardner A (ed.) *Current Approaches to Occupational Medicine*. Bristol: Wright, 1979: 350.
22. Hipp LL, Jirak PD and Golonka EJ. Evaluation of an occupational health examination program. *J. Occup. Med.* 1977; **19**:205–207.
23. Occupational Safety and Health Administration (1983). *Occupational Noise Exposure; Hearing Conservation Amendment*. Fed. Regist. 48(46); 9738–9783.
24. American Occupational Medical Association, Guidelines for the conduct of an occupational hearing conservation program. *J. Occup. Med.* 1982; **24**:772–773.
25. American National Standards Institute. ANSI Z87.1: *Practice for Occupational and Educational Eye and Face Protection*. New York: ANSI, 1968.
26. American National Standards Institute. ANSI Z358.1: *American National Standard for Emergency Eyewash and Shower Equipment*. New York: ANSI, 1981.
27. American National Standards Institute. ANSI Z80.1: *Requirements for First-Quality Prescription Ophthalmic Lenses*. New York: ANSI, 1968.
28. National Society to Prevent Blindness. *Glaucoma Detection for Business and Industry*. New York: undated.
29. Make B. Medical surveillance for occupational respiratory diseases. *J. Occup. Med.* 1975; **17**:519–522.
30. International Labour Office. *Guidelines for the Use of ILO International Classification of Radiographs of Pneumoconioses*. Geneva: ILO, 1980.
31. Health and Safety Executive. *Lung Function*. HSE Guidance Note M55. London: HMSO.
32. Weill H. Pulmonary function testing in industry. *J. Occup. Med.* 1973; **9**:693–699.

33. Shapiro J. *Radiation Protection: A guide for Scientists and Physicians*. Cambridge, Ma.: Harvard University Press, 1981: 237–241 & 270–271.
34. National Council on Radiation Protection & Measurements. *Operational Radiation Safety Program*. NCRP Report No. 59. Washington, D.C. 1978.
35. National Council on Radiation Protection & Measurements. *Basic Radiation Protection Criteria*. NCRP Report No. 39. Washington, D.C. 1971.
36. Gardner AW. Identifying and helping problem drinkers at work. *J. Soc. Occup. Med.* 1982; **32**:171–179.
37. Chafetz ME. *Second Special Report to the US Congress on Alcohol and Health* (June 1974). DHEW Publication No. (ADM) 75–212. National Institute on Alcohol Abuse and Alcoholism. Rockville, MD, 1975: 129–139.
38. Trice HM and Beyer JM. Work-related outcomes of the constructive-confrontation strategy in a job-based alcholism program. *J. Stud. Alcohol* 1984; **45**:393–404.
39. Pearson CE. The emerging role of the occupational physician in preventive medicine, health promotion and health education. *J. Occup. Med.* 1980; **22**:104–106.

19. THE HEALTH OF AN ORGANIZATION

David T. H. Weir

Burns and Stalker describe the 'organismic' and 'mechanistic' systems of management.[1]

When we use the first sense we are emphasizing the organization as a body, likening it to an organic living thing, with a structure which is flexible and adaptive, possesses vitality, and in which the life of the whole is somehow more significant than the activity of its constituent parts: the emphasis is on the processual and dynamic elements of organizational life. Thus the separate purposes of departments within an organization are subordinate to the survival and maintenance of the whole. In the same way it makes little sense to talk of a healthy leg if it is detached from its parent body. It may be useful enough as a laboratory specimen it but it will never walk.

The mechanistic analogy sees the organization as a sort of structured environment. There are parts but they can indeed be taken down and used elsewhere, much as a block of tenement flats can be recycled into antique fireplaces, panelling, and art nouveau tiles. But this restructuring will never produce one 'single end'. So the whole can live or die separate from the sum of its parts. And the parts may become parts of some other structure.

For many reasons the second analogy is more helpful than the first. In science we often wish to decompose some larger structure in order to study precise patterns of activity or causation on a local level more precisely, and much scientific research indeed proceeds on the basis that only the narrowly conceived and the precisely specifiable can count as true scientific research.

In the field of organizational analysis there is much perfectly valid and useful knowledge which derives from the mechanistic perspective. Business schools likewise tend to be organized on departmental lines. Specialists in financial management offer different analyses to those of specialists in industrial relations. Operational researchers do not speak the same language as marketing and sales experts. None the less within the business school there is usually an approach to organizational study called 'corporate strategy' or 'business policy'. Typically this programme may be referred to as the 'capstone' course and be taught by the use of case methodology.

In medical science general practice seems to rank much lower down the hierarchy of status, and the great prizes and prestige go to those who know more about less rather than those who know something about everything. Cases are not as convincing as laboratory experiments.

Yet there is a parodox involved here for at the end of the day we are dealing with people, not with legs or eyes. Moreover, people spend much of their time within organizations which affect them in various specific ways, but which none the less have a separate life of their own which has to be understood in the round and as the product of a historical process. Typically the organization will survive the individual and thus have some claim to be considered as an entity in its own right.

A consideration of the way in which we learn about the effect of organizations on individual behaviour is instructive. Take studies of *labour turnover* and *absenteeism* for example. It is obvious to the man in the street that taking a day off working on an assembly line at General Motors may represent a quite different species of behaviour from taking a day off service in the pulpit as a priest. Yet studies of absenteeism and labour turnover typically do not make that kind of distinction. The part of organization which is significant is to these researchers the fact that it is 'work'. But it is 'work' in a different sense and the reality of the organization in which the work takes place is quite different, so absence or presence or leaving the job altogether has quite different social connotations.

At present there is a great vogue in organizational research for studies of 'corporate culture'. Since the publication of *In Search of Excellence*[2] it has been commonplace that successful organizations are those which have a 'culture of success'. But what is 'success' and how is it related to the healthy organization and how is the healthy organization related to the healthy individual?

Now while these questions may appear very complex for philosophers and scientists, they are readily understood by ordinary people. We do recognize success and are drawn to it; we sense failure and avoid it. And the same is true in the field of health. Thus an individual may not know what is wrong with him in the scientific sense but he can know perfectly well that he does not feel *good*. Individuals are normally quite wise about the operation of their own body and our everyday language embodies this wisdom. Thus individuals talk about 'being off form', 'being below par', 'not feeling too good'. These concepts may be relative as well as absolute. Thus a neighbour of mine, in the Yorkshire Dales, described the near-comatose state of his aged relative as 'wonderful well considering how badly she is'.

Similarly, individuals may move jobs for no better reason than that 'it seemed the right thing to do', 'I didn't feel happy there', 'they seemed a nice bunch'. The point is that although scientists may despair about such apparently unclassifiable judgements they none the less embody some quite meaningful insights into the reality of work in its totality in an organization which is seen as a whole and responded to as a sort of organism rather than as a collection of related but ultimately discrete parts.

When we are considering the health of an organization we have another problem to contend with which also relates to the way in which medical people come by their knowledge. Doctors typically see patients who believe themselves to be ill in some way. Likewise, organizations are very often only of interest to scientific researchers when something appears to be going wrong in them. Thus conventions of Cabinet confidentiality and collective responsibility become a prominent feature of everyday discourse only when a political crisis occurs. The vitality of a business organization is questioned by those who wish to acquire its shares, the sickness in industrial relations is more clearly perceived in a time of violent confrontation.

'Health' is not easily identifiable or specific. 'Disease' is more interesting and more recognizable. But although health is not easily defined, its absence is readily recognized. While a specific disease may not be easily diagnosed, individuals know what it is to feel unhealthy.

If asked to write a checklist of how we feel when we feel healthy the typical individual might respond something like this. When healthy we feel to be *successful*, we obtain *pleasure*, there is a sense of *well-being*, we feel *relaxed*, *outgoing*, *open* and *balanced*. We do not feel under the pressure of time. We understand ourselves. We feel we *can do* things. Above all, we feel to be physically and mentally *competent**.

When individuals are suffering from a disease they are in *pain*. They feel to be *failures*. There is a sense of *ill-being*. They are tense, vulnerable, inward looking. When the disease is mental we may describe them as *closed* or *blocked, unbalanced, rigid*. They may suffer from *anonymity* or the feeling that they *do not truly know* themselves. They feel harassed and *oppressed by time*.

* The latter adjective is particularly significant because it is widely, though unspecifically used in medical diagnosis. Thus for newborn children there are tests of basic functional competence. These tests relate at such an early age of course to what the infant can do rather than how he feels. But none the less clinicians know that individuals who attain such a definition of competence do appear relaxed and happy with themselves.

Can we transfer these epithets wholesale to the organization? Clearly it is not an easy task. Some clinicians and organizational analysts have sought to use the metaphor of stress to link the two levels of analysis. Thus Albrecht links the individual symptomatology of stress to feelings of anxiety which are derived from a perception that the environment lacks stability, permanence, and predictability.[3] These perceptions, in turn, derive from the mobility, crowding, pace and threatening change of modern life. In a certain type of organization the stress may be reinforced or even rewarded by the characteristics of the organization. This is quite a common occurrence in organizations which are governed by rigidly defined sets of rules and regulations. The functional tasks with which the organization confronts its problems are broken down in a specialised and differentiated manner. Each individual is tasked to perform according to an abstract specification of rules which specify certain behaviours which are functional in terms of the organization's goals. However, if these behaviours are pursued to excess they lead to the individual giving greater prominence to the realization of subunit roles rather than the goals of the organization as a whole.[4] It is in order to avoid the displacement of goals (which ultimately is dysfunctional for the organization) that the structure of rules enjoins that each hierarchical level controls the performance of the level below it. So it is the task of the immediate superior to reconcile deviant performances.[5]

The rights, obligations and technical methods attached to each functional role are very precisely defined. They are translated into the responsibility of a functional position. Within the hierarchical structure, knowledge is presumed to be located most *generally* at the top of the organization. Thus interaction between members of the concern tends to be along vertical lines of communication. Operations and working behaviour are governed by a superior. Within this overall framework as each individual sees only a part of the whole, there is an insistence on loyalty to the concern and obedience to superiors as a condition of membership. It is taken for granted that only those deeply embedded in the system have the appropriate loyalty and are to be protected from the contamination derived from exposure to a more general cosmopolitan type of knowledge, experience and skill.

In practice, changes in the market and the environment, in the basis of recruitment to the organization and in the goals and purposes imposed on the organization, can lead to a systematic tension between what the organization *needs* to do and what it is *capable* of doing. In these situations there is considerable potential for stress. As organizations lose touch with their market there is a tendency for the internal structures to become even more rigid. Thus the adaptive responses to difficult conditions becomes the organization's *real* way of working. I call this 'overcontrol'; Kets de Vries,[6] writing from a psychoanalytic standpoint calls it 'bureaupathology'.

Individuals are therefore led to cover up their behaviours because they are under dual pressures. There is a high degree of *formal* control, but a real requirement to operate the rules in a *discretionary* way. The oil between the cogs in the formal system is provided by interpersonal relations of a reciprocal nature between organizational participants of the same adjacent levels. As there is a network of mutual support and obligation, problems tend to find a solution at the level at which they occur because the consequences of taking a problem to a higher level may be very negative for the individual concerned. Thus superiors who have the formal responsibility of reconciling problems which occur at lower levels may not in practice know how lower level participants typically solve these problems. In time higher levels tend to regard ignorance as a positive virtue and to disregard or define as unimportant knowledge which lower participants do have.

Thus even problems for which explicit and possibly good solutions are available tend to be solved by direct negotiation between the participants. Thus in practice lower participants have as much or more discretion in playing their role and in taking decisions as do top decision-makers. Power thus flows downhill and the top decision-makers find in practice that their freedom of action is very much constrained by 'custom and practice'. The result is to build up a body of case law and precedent.

The way an organization works in practice can never be derived from the rule book which sets the framework within which individuals and groups work out their strategy. The organization's formal goals tend to be displaced and eventually subverted by those which derive from the interaction of lower participants who develop new targets which they can meet in practice without compromising those interests. Ultimately, both higher and lower participants become involved in passing to each other erroneous information on non-events and it becomes difficult if not impossible for the organization to monitor its own processes because real decisions can never be adequately identified. The organization becomes unhappy about the kind of information it produces and tends to withdraw from the outside world and to avoid investigation by outsiders. At worst this can degenerate into a determination to keep the outside world in ignorance of what is going on inside.

Organizations of this type are relatively common in declining industries and declining industrial sectors. Unless the organization declines altogether and goes out of business there comes a point at some level, either inside or outside the organization, when the requirement for change is felt very severely. Possibly an acceleration in the pace of competition or a determination on the part of a major investor, either the government or a main board, not to continue throwing good money after bad, poses the organization with a requirement that it change or die.

The 'subterranean'* nature of some of these processes is illustrated by the following example. In a large aircraft manufacturing plant in the United States, research uncovered a widespread pattern of behaviour which was in fact illegal or criminal in terms both of the rules of the organization and of the criminal law. The workers were engaged in joining previously assembled parts by a bolt screwed into a nut recessed in one of the previously drilled parts. However, the design of the whole assembly was so tightly constrained that it was practically impossible to align the parts in the required way. So the workers who had the task of inserting the bolt had to choose either to insert it knowing that the thread of the bolt and nut were not in alignment or to recut a thread through the original nut and the part to which the bolt would be joined. However this new assembly would not be consistent with the original design and might thus fail in operation.[8]

None the less this was the route usually chosen by the operators, rather than the approved behaviour of sending the assembly back to be remade, which would impose pressures on production and hit the individual's pocket as well as cause embarrassment to the company. This is a fairly typical manufacturing problem when design fails to take account of feedback from operations.

One of the interesting phenomena observed was that the cutting of the new thread in the previously drilled part thus weakening the structure was undertaken not by the inferior and incompetent but by the best craftsmen. Thus those at the apex of the hierarchy of skill (in the normal or approved sense) were selected by the foreman to undertake the task of drilling the illicit new threads.

The ultimate root of the problem was that the system was over defined and too precisely specified making it difficult to machine the parts according to a specification by rigorously following the approved instructions; so an organizational shortcut had to be invented. Once this was devised it was supported by a network of shared understandings and behaviours, in which those who bore the most formal responsibility for maintaining the legitimate system were those who also came to occupy crucial places in the informal or illegitimate system. The approved agents of control, the airforce inspectors, stood rather outside this system of shared understanding and behaviours and were the last to comprehend the reality of the situation.

Bensman and Gerver conclude that the totality of actions, both normal and deviant, together comprise an interconnected pattern of behaviours which can be disentangled only with difficulty. Similarly the human personality can be separated analytically into a number

* 'Subterranean' is used in the same sense here as Matza and Sykes.[7]

of specific behaviours and patterns and it may be possible to label these for convenience, 'normal', 'neurotic', 'pathological' and so on, but in any one particular person they go together and normality and pathology are close twins, if not inseparable.

There is no very simple transferability between the feelings which people have and the health of the organization of which the individuals form a part. Studies of prison camp guards, torturers, and civil servants involved in the extermination of fellow human beings show that the evil we perceive in the organization is not necessarily readily visible in the individuals who undertake the activities which we subsequently define as evil. Officially sanctioned murderers don't necessarily feel bad about their work, whether in Nazi Germany, Argentina or Uganda.

The depressing fact about the Nazi regime according to Hannah Arendt was 'the banality of evil', and it is possibly the bureaucratic correctness of Eichmann which is more representative of mass extermination than the rantings of Julius Streicher.

So the fact that people feel happy about their work does not entitle us to conclude that the organization itself is a happy one. But there is clearly a relation, though more complex than it at first appears, between the two. Some commentators seek to relate the organization and the individual through the concept of stress.

Stress itself is unlikely to take us more than halfway towards our goal of understanding just what it is about organizations that entitles us to call them 'healthy' or 'unhealthy'.

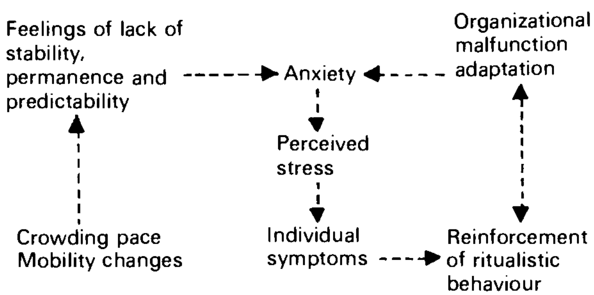


Fig. 19.1. Organizational stress (adapted from Albrecht[3]).

Stress is an essential ingredient both of competent performance and of the learning process which produces competence. Some stress does arise from change and it is easy to conclude from this that a change in itself is a cause of dysfunctional stress[3] (*Fig. 19.1*). Recent research calls attention to the stress caused by the dislocation of life experience which occurs when schoolchildren become university students. But no one would conclude from this that people should remain at school all their lives or that the stresses of undertaking higher education, leaving home, friends, family, established relationships and expectations were not compensated by the exposure to wider horizons and opportunities.

Some stress can be produced by circumstances which are inherently alienating. Individuals may become dissociated, rootless, anomic and fixated on means rather than ends. Other stress arises from purposive action, is time bound, can be controlled and is oriented towards ends and goals. Such stress, whether in organizations or individuals, may be indicative of positive health.

The distinction between the two is that in the first the individual appears to be out of control and in the latter there may be an element not merely of control but even of positive acceptance or introduction of seeking for stress. Thus athletes in training deliberately stress their bodies. Organizations attempting to grow by acquisition, not infrequently, bite off more than the market feels they can chew. But if they are successful, the experience teaches them lessons which lead to even greater opportunities for this enjoyable experience of stress.

Control is also important. Control has an external as well as an internal reference and can often refer to the successful operation of the organization on the environment. Organizations which are in control of themselves develop control over the environment, which of course largely consists of other organizations. This type of control can be described as 'proactive control', emphasizing the fact that the internal system needs to be organized so it is prepared for environmental change.

Eilon argues that operational research can have a major role in this proactive process if it can participate in 'defining objectives and determination of constraints . . . seeking new areas to investigate, and generating new ideas that would have a beneficial effect on the performance of the enterprise'.[9]

Operational research and other formalized rigorous approaches typically make use of models of cybernetic control in which certain assumptions are made. Typically it is assumed that it is possible to devise standards and measures then relatively precisely, to measure achievement of these standard performance trends and to feedback information on positive or negative variance. Moreover, it is often assumed that such feedback information will be sufficient to maintain control of the system. Not all of these requirements are met in practice, but the discipline of attempting them is generally useful.

We are close to a definition of the health of an organization in terms of two sets of factors. One relates the organization to its environment. Healthy organizations are those which successfully achieve conformance to certain goals. For business organizations these are normally measurable in terms of profit before tax, growth in sales or profit, or market share. There are a number of key ratios which summarize different aspects of these criteria of organizational performance. They are end not means oriented. The purpose of an organization is not simply survival. Organizations which elevate survival and longevity as goals in themselves are probably doomed.

Health may be measured also by criteria which relate to the development of organizational competence. Attitudes and studies of behaviour may provide insight into this but they are not themselves sufficient. As we saw earlier it is possible to have good feelings about inadequate or deviant performance. The essential thing is the development of organizational competence and this can only be judged ultimately in terms of external criteria. Thus if the Japanese consistently perform in a superior way in terms of manufacturing industry, it is no answer for Western managers to justify their incompetence in terms of such evasions as 'but the Japanese are more disciplined', 'they have high tariffs against imported manufactures', 'they are different from us'. The fact is that there is ultimately only one set of rules governing success of performance in the game we call manufacturing industry and the Japanese are now performing in that particular game better than we are.

The values of business enterprise are implied by the ground rules of the business enterprise itself. There are not really two measures of profitability any more than there are two measures of bankruptcy or ultimate business failure. Organizations in the business world which succeed develop a trajectory of successful performance. Organizations which fail cease to exist.

This also applies to the well-known cant phrase 'small is beautiful'. Small is *not* beautiful *per se*. Small organizations in fact are less likely to succeed than large ones and research does not confirm the absolute superiority of the small organization in terms of profitability or even the pleasantness of the experience of being employed. But when an economy consists very largely of old, tired organizations which have outlived their competence and effectiveness, then the competent and effective organizations may be more typically found amongst the small.

Kanter makes this point very well in her study[10] of corporate entrepreneurs. She contrasts the narrowly focused companies which specialize purely in high technology with those which are subsidiaries of diversified companies. As she points out 'many people find it easy to account for a high degree of innovation in the first instance, and even write it off as a function of the industry, organizational youth, recent entrepreneurial origin, or market pressure for technical achievements rather than internal organizational practices. But the more instructive case is really the second... where innovation flourishes within a large, much older, and more traditional business.' In examining the medical system subsidiary of General Electric she points out that 'Medsystem's leadership does not stem from technology alone, nor is technology its only realm for innovation. Indeed it is perhaps marketing and production efficiency, general business expertise and good management systems that contribute proportionately to the organisation's success.' This competent performance has its analogues in individual behaviour and attitudes. Medsystem's people demonstrate 'a remarkable open-mindedness, willingness to listen, non-defensiveness and ability to let go an investment in their own ideas in order to pick up on a different idea that might produce results.' '... This ability to switch gear and pick up on someone else's idea was true not just of the particular individuals I met but seemed to be the attitude encouraged at Medsystems in its product development as well as its treatment of people ... Medsystems seem to avoid the bureaucratic trap and the entrepreneurial trap.'

We have now discussed some of the ways in which organizations might be said to be healthy. But for the average individual an abstract definition is not in itself very rewarding. What is of more interest is to learn the ways in which one may become healthy or at least more healthy than one seems to be at present. So in the rest of this chapter we describe how an organization in almost a terminal state of decay developed new ways of handling apparently intractable problems and in the process came to feel better about itself.

We trace the process of a major programme of organizational change in a West of Scotland engineering company over a period of years and focus on one particular phase within this programme in which the process and modalities of introducing change were so handled as to increase competence and reduce stress. Managerial goals were more generally achieved and were adopted by the workforce as the basis for a more sophisticated understanding of the nature of the organization and its developmental requirements. The health of the organization improved.

Much work on organizational change and, in particular, change towards greater participation neglects the role of managerial strategy and plan in forming the policy judgements which lead to increased participation. As Buchanan and Boddy[11] put it 'The impact of technical change on the labour process is influenced by prevailing patterns of management objectives. These objectives are shaped by the wider social and organizational structures within which managers function.' But managers are equally prisoners of the prevailing organizational culture and their own social character as are employees. So the process of change has to change behaviours and performances over a wide front. To be successful it must become self-sustaining and become embedded in new structures. To the extent that these structures are rooted in old understandings but de-escalate stress, and increase competence, they are more likely to be successful. The successful overcoming of stressful situations reinforces the new patterns of behaviour.

Management-induced change thus depends for its effectiveness on management's willingness to implement it, to accept the burden of stress, and to manage the situation, by recognizing the reality of the likely culturally mediated response of the employees and go with it or round it. In part, management's perception of the benefits to them of introducing a major change are conditioned by their own perception of employee support and how this can be managed and their own perception of top management or board level support and how this can be sustained.

Care in implementation is thus essential because at each stage an inadvertent build up of pressure, hostility or threat may lead to a flashdown to previously dominant atavistic modes of behaviour.

Managerial goals are often complex as indeed are employee responses. In the typical West of Scotland heavy engineering company there may well be as many as a dozen trade unions, all of which may perceive advantage in either opposition or support depending on their own strategic positioning, *vis-à-vis* other unions and management. The behaviour of significant key individuals can also be crucial. Specific employee representatives, managers and customers are all strategists in their own right and may have to be dealt with separately. A major programme of organizational change has to take account of these factors and cannot presume that either managerial or employee reactions can be readily subsumed under some catch-all definition such as 'exploitative' or 'alienated'.

The precise history and traditions of a specific organization are built into its contemporary structure and functioning. Thus, both management and employees are predisposed to behave in certain ways and more importantly to perceive situations in terms of a 'recipe' for action in which certain opportunities are available and others simply not perceived. The term 'recipe' is a useful shorthand, which recognizes the specific as well as the generalizable elements in the culture of an organization or industry, and the way reality is structured by its members. The presentation of opportunities which are not perceived as opportunities but as threats will themselves be stress-inducing regardless of their objective potential for success. Thus, the management team, in inducing change, has to operate as a group of skilled clinicians coping with the texture of an organism of which they are themselves constituent and contributory parts.[12]

The successful introduction of organizational change has thus to recognize the political and cultural realities of the organization. Berg[12] comments 'a programme never occurs in a vacuum. It is implemented in a social system characterized by emotional and political bonds between the members that are not easily changed by "artificial" and human relations training.' He goes on to argue that 'the meaning of a change programme in relation to the network of power relations is of particular importance. Life in organizations is political and competitive and this is a fact which must be taken into account in the design of programmes of planned change'.

The recent history up to the mid-70s of this company was characteristic of the heavy engineering industry. The company had been in existence in the same location for nearly ninety years employing several thousand people, but in recent years the domestic and international markets for its products had been fiercely competitive. The 50s and 60s had been a period in which demand had expanded quickly, earnings and profits were high and there were few pressures to achieve significant productivity improvement. Considerable investment and development work took place during the 50s and 60s but with a deterioration in the market in the 1970s costs increased relative to competitors and new investment was urgently required. The company was thus faced with a situation in which risks were required to ensure survival. A programme of systematic recruitment to key managerial positions emphasizing technical competence and openness towards change and a development of a more open management style was the first indication that the company recognized the real requirements of the international competitive market.

In the initial period some key industrial relations, pay and work allocation problems were tackled. Despite these changes however a massive crisis faced the company in the mid-1970s. The shortfall in workload after 1977 made it more urgent for Government intervention to become necessary. A joint initiative involving managements, unions and middle management was seen as the means of bringing pressure on government to bring forward a key order to fill the gap in the forward plan.

The effect of this exercise was to start to break the mould of confrontation. There were meetings, television presentations, meetings with Ministers, visits to Whitehall and Westminster. All combined the management and employees in a situation of high visibility. The programme was successful and the order was obtained. In the process there was an opportunity for the mobilization of collective resources and the statement of traditional corporate and organizational values. Among these was the profound significance of the plant for the economic life of the local community. This exercise was important to the company not merely because it was successful but because it had gone some way towards setting up new structures.

By late 1978 and early 1979 it had become clear to the main board that a very substantial further investment was required to bring the facilities up to internationally competitive standards. It was decided to commence the reinvestment with the introduction of new technology into the machine shop, which would involve a complete break with traditional work practices and substantially increase capital investment per employee.

Substantial progress remained to be made on the structural problems of manning and work flexibility. The board believed that their proposal of a major investment programme consisting of three phases, a total of approximately £100m over a period starting in 1979, represented 'a commitment' to the future of the plant and the community. They therefore structured the decision in such a way that if a commitment was made by the board a corresponding commitment was to be expected from the workforce. In general terms such a commitment might well be forthcoming. But what did those general terms mean in terms of precise pointers for action and how could they be worked through the structures of management and union decision making?

Management first of all had to specify its own commitment in fairly precise and understandable terms. But in order for a countervailing commitment to be obtained for the workforce the nature, extent and detail of management's commitment itself had to be *understood*. This imposed a prior discipline on management to understand its own proposals. How could they be presented to the workforce in a clear unambiguous way which could, over a relatively short period of time, result in a countervailing commitment?

Management and the Joint Union Negotiating Committee agreed that the Rebuild proposals required the agreement of all the employees. In order that there should be a common basis for understanding and agreement, it was decided that management's message should be expressed over a short period of time and over a precisely similar format to all employees. The information should respect the group structure of the enterprise. So it became the responsibility of individual foremen and managers to understand management's proposals in such a form that they would be available for offering sensible and specific direction on what the proposals meant for their particular work area.

The programme was regarded by management and union leaders as the most significant major development in the company for several years. It was seen as an issue of *quality*. It should therefore be presented to the workforce in a quality medium. The standards of the very best television presentation were used. A presenter of a major TV programme was invited to introduce and present the management plan. Almost immediately he came to adopt something of a mediating and explanatory role, not taking all aspects of the management case for granted and operating in every way as if he were in the regular TV situation.

In the first video programme the management proposals were outlined. The programme started with a presentation of symbols of the company's industrial past successes. The challenge of competition was identified, and sophisticated and clear visual presentation supplemented a review of the main work areas of the plant and a presentation of the proposals for the initial phase of the Rebuild in the Machine Shop.

More significant, however, than the nature and quality of the production itself was the manner in which it was presented to the workforce. Each programme lasted 25 minutes and was shown to employees throughout the works in small groups of about 25. Each group had the benefit of the presence of one member of the management team to answer questions. All employees on the site saw the video over a 24-hour period. As well as oral response, each participant had the opportunity to present written questions which were then analysed and subjected to analysis.

Over one thousand questions and comments, often in considerable detail and specificity were collected and examined by a Joint Union Management Committee. The points most frequently raised were used as the basis for a second video programme filmed in the company's training centre. This programme adopted a rather different format. The same presenter introduced the proceedings and acted as chairman and compère. The management team were on the platform armed with visual and other material. The audience consisted of representative members of all the unions and work areas in the company. So each employee representative was charged to react, not merely in general and subjective terms but in accordance with the distribution of responses already collected from the workforce. In other words a script existed which served to codify and moderate the employee response. The resulting video programme was again shown thoughout the plant. Many employees responded again in writing and the issues they raised were in every case discussed with them by a senior manager. Following this phase of presentation, more conventional modes of communication were adopted. A Newsletter relating to the Rebuild was created and appeared regularly.

The issues of manning, flexibility, and work organization were referred into the conventional structures of negotiation and consultation.

In due course the commitment from the employees was forthcoming in the terms in which management had hoped and expected as follows:

1. The return on investment should be sufficient to justify moving to Phase 2.
2. A steady level of employment would be maintained.
3. Unions would not press for job retention and management would not seek redundancies.

4. Improvements would be sought in costs, deliveries and quality.

5. Capacity would be fully utilized and where necessary this could require transfers from one job to another.

6. Manning and job allocation issues would be resolved in accordance with domestic agreement.

After the decision to proceed with the first phase of the investment programme, a third videotape was prepared which went over the Phase 1 proposals in greater detail, dealing primarily with the commitments of management and unions. The company is now moving into a new phase and further investment is forthcoming. The forward order book is reasonably healthy. The company is moving to a new competitive situation.

In later aspects of the programme teams from each of the major work areas participated in a major internal consulting programme which had as its goal the achievement of productivity and output targets attained in a competitor Japanese company.

What lessons can be learnt from this brief case study? The first point to make is that at the beginning of this process the organization was under stress. As the major employer in a medium-sized industrial community, senior management felt real responsibility for maintaining employment. However, international pressures meant that the company was simply not competitive. The external pressures were resolved through a breaking of the mould which recognized the real interests of the organizational participants. A key element in the process was the continued and sustained commitment of local plant management. Without this commitment expressed in actions over a considerable period of time, the conditions for effective organizational change would not have existed. A study of the Tannoy company shows similar functions in operation.[14]

In his review of the factors affecting corporate turnaround, Bibeault[15] evidences a number of reasons for corporate decline, of which over 50 per cent of cases are caused by internally generated problems within management's control. 'Management', says Bibeault, 'is the principal cause of failure. Managers of companies which fail to turnaround demonstrate incompetence, narrow vision and displacement activity, a tendency to short-term control and letting the big picture go by. Such management makes errors of omission, failing to respond to changes in the external environment and becoming lax in developing or utilizing control information. Such management teams may be characterized by a strong leading entrepreneurial figure, but a lack of management depth and of succession and management change planning. Such structures lead to inbred bureaucratic ways of decision, with employees clinging to old ways of working and failing to define new goals. Action is typically taken without studied reflection and there is an air of 'institutionalized contentment'. The old wisdom is passed on to new people and there is a low tolerance for criticism. The Board of Directors tends not to participate in management detail.'

Arguably all these factors may have existed in this company before the crisis set in. However, during the crisis management *has* managed. This is by no means a lax or weak or low achieving management culture, in the present situation.

This period has also been characterized by industrial conflict of a conventional type. There is a multi-union situation and there has been a strong and consistent pressure from the trade unions for improved wages and conditions and for guarantees of job security.

The environment has become exceedingly competitive and the pressure from the Far East in particular has become very intense. None the less during this period major organizational change has been produced through programmes which bear the marks of the Rebuild programme we have been examining.

If we return to our basic model of organizational health and competence we can see several reasons for the success of programmes of this kind:

1. Management recognized the importance of guaranteeing security. Thus, feelings of lack of stability and permanence were minimized.

2. Known and predictable ways of working were not attacked or bypassed. They were incorporated and served as the matrix for the new modalities.

3. The group structure of the enterprise was recognized and incorporated. Thus work groups were the basis for the presentation of the message, and so a realistic level of information and communication became a basis for commitment.

4. While the pace was continuous it was never hurried. Considerable organizational change could be packed into a relatively short period with *latency* periods in between; thus two years lapsed between the Impact and the Rebuild programmes. A further three years had elapsed before the next phase which is presently underway.

5. Management and employee representatives explicitly kept a balance between *programmed* and nonprogrammed activities, thus recognizing the dangers of Simon's prediction that programmed activities tend to drive out nonprogrammed activities.[5] As additions to the strength, external consultants were brought in to support rather than to grossly remaster decisions which had emerged through the conventional structures.

6. *Progress* points were put in or developed naturally in the systems so that the characteristic problem, of so dealing with unstructured problems that progress is rarely evident, was overcome.
7. The natural *orderings* of groups and structures were respected. No new group was created, no large group, no mass meetings in the town hall, no high pressure, no overselling of the managerial message.
8. The natural *pace* that was adopted allowed feedback. Feedback indeed was forthcoming. It was specific and was responded to by management in the short term and was incorporated into the management agenda in the longer term.
9. The conventional *scale* of organizational activities was respected. Understandable information was available at the usual place of work.
10. Organizational *roles* were respected. Change consultants were used but in an advisory and expert role, supplementing decisions which at each stage were kept within the conventional structures.
11. While *information* about the future and about decline in market competitiveness was direct, it was not apocalyptic and shocking. The future appeared to be connected to the present and predictable.
12. The threat of a traumatic *disruption* of relationships such as that of redundancy or the failure of the firm altogether were minimized and controlled.
13. The media emphasized *tradition* and *success*; so the extreme sense of failure, defeat and disappointment leading to humiliation and loss of self-esteem which are characteristic of high stress situations were minimized by continually presenting the company as essentially a successful one.
14. The *mode of presentation* of potentially 'bad news' and the structuring of the management and union response emphasized the solid support which was available on the basis of common membership of the enterprise.

A key element in the generation of competition was management's ability to work through the uncertainties which it could control and to level with the workforce about the fears and uncertainties that were not under its control. This in turn has produced new agendas for action in dealing, for example, with the threat of international competition. But it also focused attention helpfully on areas which were actually or potentially under the control of the organization.

The case also indicates the crucial importance of the management of information in the reduction of stress. At each stage management was compelled to examine its own information bases because of the mode of presentation of information to the workforce. In this context the video medium was especially significant.

Action is itself stress reducing. The main organizational problems that face complex traditional enterprises such as this are indeed well known to both management and employees. Failure to surface them and recognize their reality can itself contribute to stress. The perception that action is underway contributes to its reduction.

The reduction of stress in this organization has been linked to the growth in collective self-confidence. Outside consultants and experts have been involved at the phase where the organization has made up its own mind at least in general terms as to what type of progress requires to be made. Thus change is perceived to be self-induced rather than imported from alien cultures and belief traditions.

The cultures and values of the organization are thereby celebrated. Change respects tradition and the tradition has been modified to include the inevitability of change.

This chapter has tried to demonstrate how an organization in a particular historical geographical context facing certain problems not necessarily all of its own making, learnt and adapted and consequently survived to fight another day. To what extent are we justified in interpreting this saga in terms of health and has the analogy been helpful?

At the start of the history, the organization faced certain problems which were likely to prove terminal. Indeed it had been decided at governmental level and in the main board of the parent company that this organization was unlikely to survive. Sentence of death or at least of limited life had already been pronounced. There was a feeling of malaise and individual workers and their organized representatives were voluble in their diagnoses. Moreover, at that stage the organization lacked the competence to diagnose its own problem, still less to produce its own medication. So as with the unhealthy individual, both self-awareness and the intervention of interested and skilled professional support were equally important.

The organization had developed a number of specific competences in areas of functioning in which good models of effective competent performance were available elsewhere. One of the symptoms of health noted by Kanter[10] was the rejection of the Not Invented Here (N.I.H.) syndrome. Organizations which are healthy learn from the health of others and learn to mimic their actions even without necessarily understanding the philosophy or culture which makes those actions credible. Similarly in medicine individuals may pattern their behaviour on that of others and learn the performance before knowing the rationale for it. Finally what counts as health or the absence of it for one individual and for one organization may be different from that which obtains for another. But the general criteria of health are observable enough: a glow on the cheeks, a healthy bank balance, and determination to face tomorrow's challenge with equanimity if not with enthusiasm.

Acknowledgements

This paper is based on a paper read at the joint meeting of the Faculty of Occupational Medicine and the Royal College of General Practitioners at the Royal College of Physicians and Surgeons of Glasgow on Friday 31 May 1985. I am grateful to participants at that joint meeting for comments and criticisms and in particular to Dr Anne Fingret and Dr F. G. Taylor.

REFERENCES

1. Burns T and Stalker GW. *The Management of Innovation.* London; Tavistock, 1966.
2. Peters T and Waterman RA. *In Search of Excellence.* London; Harper & Row, 1983.
3. Albrecht K. *Stress and the Manager.* Englewood Cliffs NJ; Prentice Hall, 1979.
4. Merton RK. *Social Theory and Social Structure.* New York; Free Press, 1968.
5. Simon HA. *The Shape of Automation for Men and Management.* London; Harper & Row, 1965.
6. Kets de Vries MFR. *Organizational Paradoxes.* London; Tavistock, 1980.
7. Matza D and Sykes GM. Delinquency and subterranean values. *Am. Sociol. Rev.* 1961; **26:**712–19.
8. Bensman J and Gerver I. Crime and punishment in the factory. *Am. Sociol. Rev.* 1968.
9. Eilon S. *Management Assertion and Aversion.* London; Pergamon, 1985.
10. Kanter RM. *The Change Masters.* London; Unwin, 1984.
11. Buchanan DA and Boddy D. The Impact of Technological Implications and Managerial Aspirations on the Organisation and Control of the Labour Process. University of Glasgow Dept. of Management Studies, 1984.
12. Spender JC. Strategy Making in Business. Manchester Business School, 1980.
13. Berg, Per-Olof. Emotional Structures in Organisations. Student Litteratur, 1979.
14. Weir M. *The Tannoy Story.* MCB Publications, 1985.
15. Bibeault DB. *Corporate Turnaround.* New York; McGraw-Hill, 1982.

20. ETHICAL PROBLEMS IN OCCUPATIONAL MEDICINE

W. Jeffrey C. Scott

'Life is short and the Art long; the occasion fleeting; experience fallacious and judgement difficult. The physician must not only be prepared to do what is right himself, but also make the patient, the attendants and the externals co-operate.'

Thus runs the first aphorism of Hippocrates, which has stood the test of time. And while doctors in occupational health have ethical problems peculiar to their speciality, their reaction to these situations must be compatible with the attitudes of their colleagues in other branches of medicine. Guidance on medical ethics has been available in the United Kingdom for many years. The British Medical Association had a Committee on Medical Ethics in 1849, and played a major part in helping to establish the General Medical Council under the Medical Act 1858, which provided not only for the registration but the regulation of doctors. Prior to this, much of medical ethics was medical etiquette, designed to protect the physician rather than the patient, much as craft guilds defended their members. Indeed, some of this attitude survives to this day.

One of the attributes of a profession is that it should regulate itself, and in the United Kingdom doctors are fortunate that unethical behaviour is not codified in national criminal or civil law, but instead is considered by our peers in the General Medical Council which, since the Medical Act of 1978, is more representative of the profession, and has a larger number of lay members. Thus it more truly reflects the morals and mores of present-day society. Since the end of the last war some other countries have felt it necessary to follow a different path. In France, for example, it is a criminal offence for a doctor to fail to stop and offer help at the scene of a road accident involving injured victims. The situation is very different in the United States of America, where, because of the contingency fee system operated by the legal profession there, doctors must be circumspect in offering advice or treatment, lest they be sued for malpraxis; the amount of damages awarded by the courts in these situations is notoriously high.

There is no shortage of advice, comment and guidance from a variety of sources, and certain publications are mandatory reading so that the obvious pitfalls may be avoided. The General Medical Council sends to all doctors on its register, at regular intervals, the 'blue pamphlet'—*Professional Conduct and Discipline: Fitness to Practice*[1] and the British Medical Association's *Handbook of Medical Ethics*[2] contains a chapter on occupational medicine. The Faculty of Occupational Medicine of the Royal College of Physicians has published a booklet *Guidance on Ethics for Occupational Physicians*[3] now in its second edition, which is short and to the point. The British Medical Association also has a pamphlet *The Occupational Physician* which contains ethical guidance.

In recent years there has been an increasing interest in medical ethics by consumer groups and individuals, and they have frequently questioned some attitudes of the profession which appear to them outdated and outmoded, and not in sympathy with the present climate of openness between patient and doctor. A lay member of the General Medical Council, Professor Ian Kennedy has argued most cogently that medical ethics are part of the general and moral order in which we all live, and they cannot be considered in isolation.[4] Another lawyer, Paul Sieghart has challenged occupational physicians to look again and closely at their ethos.[5] The editor of the Journal of Medical Ethics, Raanan Gillon has in a series of articles explored the philosophy of medical ethics,[6] and has defined four principles:

—a principle of respect for persons, notably for their autonomy.
—a principle of beneficence.
—a principle of non-maleficence.
—a principle of justice.

Adherence to these principles will stand the occupational physician in good stead.

CONFIDENTIALITY IN RESPECT OF THE EMPLOYEE

The doctor in industry has a duty to the employee as a patient, who must feel that details of his medical history are in safe keeping. The exceptions to this code of secrecy are clearly defined, and can usually be divided into statutory and civil areas of the law. Until recently the Health and Safety at Work etc., Act, 1974[7] required duties of employee, physician and employer, for example in the notification of an industrial disease or accident. In April 1986, however, the Reporting of Injuries, Diseases and Dangerous Occurrences Regulations[8] replaced the existing laws.

Medical examinations

The General Medical Council gives the following advice in its 'blue pamphlet':

'Special problems in relation to confidentiality can arise in circumstances where doctors have responsibilities both to patients and third parties, for example in the practice of occupational medicine. An occupational physician should ensure that any employee whom he sees in that capacity understands the duty of the occupational physician in relation to the employer and the purpose of the consultation. In particular, where an occupational physician is asked by the employer to assess the fitness to work of an employee he should not undertake such assessment except with the informed consent of the employee.'

An important part of the work of an occupational health service is giving an opinion on the fitness of an employee to perform his duties. The advice to management should always be phrased in general terms, without clinical detail, and should be restricted to ability and any limitation of function. Likewise, workers' representatives may have a statutory duty to be concerned about the fitness of their fellow-employees, and the same standards apply. They should receive only the information as described above.

Clinical examination may reveal a medical condition of which the employee is unaware, and it is felt that his personal physician should be informed. This should only be done if the employee agrees—see Consent *below*.

Biological monitoring

Many work-processes expose people to a variety of hazards which can be minimized by various forms of biological monitoring, and which may not require a specific clinical examination. The reasons for these tests must be explained to employees, and the results and their significance should be communicated *only* to the employee concerned. If it is felt necessary to inform the personal physician of the findings, the employee must agree to the disclosure.

Groups of results, without identifying the individual, may be given to employers and workers' representatives for appropriate action, and may also require to be notified to the appropriate statutory authorities.

Medical records

Access to written clinical records must be restricted to the doctor and the nurse, both of whom subscribe to an ethical code. Clerical staff and other members of the occupational health team may require to handle these records, and must agree to preserve their confidentiality. Information in the records may be disclosed to other physicians, but only after consent given by the employee.

The doctor and nurse are responsible for maintaining the confidentiality of the medical records, and for the safe transfer of these records to a successor, or to the occupational health department of another organization. If this is not possible then the records must be destroyed.

Consent

When an employee or prospective employee presents for a medical examination it must be made clear at the outset that the doctor is acting as a medical examiner, and that as a result of the examination a recommendation will be made to the authority which commissioned it. If this procedure is followed, written consent is not usually required, but if, in addition, a form of medical questionnaire has been administered it may be convenient to include a form of written consent to the examination and report being made.

It may be necessary to communicate clinical findings to the personal physician of the employee, and written consent for disclosure should always be obtained. Likewise, if in the course of his duty, an occupational physician requires information from an employee's own doctor, the consent for the approach should be in writing. It is inadvisable to ask for a 'blanket' authority; better to ask for consent on each occasion, and retain a note of the permission granted in the medical records.

Computers

In the United Kingdom, the Data Protection Act 1984 has important implications for occupational medical practitioners. The Act will be implemented in the following stages:

1. The establishment of a Registrar and Tribunal, which has already taken place.
2. A period for registration, which ended on 10 May 1986.
3. A final period of 18 months during which it will be an offence for a data user to operate unregistered, but the registrar will not be empowered to enforce observance of the data protection principles; this period will run from 11 May 1986 to 10 November 1987, leading to:
4. Implementation of the remaining provisions, including those on subject access, 2 years after the appointed day, i.e. 11 November 1987.

At the time of writing much discussion is taking place about patient access to personal health information. Under the Act this is concerned only with automatically processed personal data, and excludes indications of the data user's intentions in respect of the data subject.

An Inter-professional Working Group on Access to Personal Health Information, chaired by Sir Douglas Black, has advised that an acceptable mechanism should be devised for the exercise of a proper discretion by the responsible clinician or other health professional, to provide for subject access to the extent and in the manner judged most helpful by the responsible clinician or an appropriate colleague. Any patient dissatisfied with this arrangement should then have a right to seek access through an independent health professional of his choice, practicing in the same discipline or speciality as the responsible clinician or health professional. Finally there would be a legal right, exercised through the courts, to seek access to information which was still being withheld. It is to be hoped that this sensible advice of the Working Party will be incorporated in the Act.

It would be possible for there to be total exemption for personal health data from the subject access provisions of the Act. Under this option the data subject's right of access to personal health data held about them would be removed, and access to it would be at the discretion of each data user—in fact maintaining the current practice. However, any restriction on subject access to health data would have to be justified under the terms of the European Convention on Data Protection as being necessary to protect the data subject, or the rights and freedoms of others. As the United Kingdom intends to ratify the Convention it is unlikely that total exemption will be granted—and indeed it would be contrary to the spirit of the Act.

Registration

Data held in a computer by an occupational health department should be completely separate from any other system e.g. personnel records. It follows that the data user, i.e. the doctor, will have to register under the Act, as he is responsible for the confidentiality of the information.

Staff problems

The terms of the Act indicate that any unauthorized disclosure of information, or even error, by a member of the occupational health team will leave the doctor open to prosecution. Accordingly, all members of staff should have a clause in their contract along the following lines: *The position you hold has confidential aspects, and breaches of confidence could result in disciplinary action which may involve dismissal. In addition, a breach of confidence could result in a civil action for damages.*

CONFIDENTIALITY IN RESPECT OF EMPLOYERS

Doctors in the employ of major industrial concerns are often privileged to be involved in new processes and developments of commercial value to their employer, and more so to a rival. On occasion a new hazard to health may be introduced, and the doctor has a duty to point out to the employer that he has responsibilities under legislation, namely, that the hazard, or potential hazard, must be disclosed, not only to representatives of employees but to the appropriate government agencies. This permission for disclosure will rarely be refused, but if it is not granted the occupational physician may have to take independent action, according to the circumstances, in the knowledge that his prime duty is to the health and safety of the workforce; this takes precedence over his loyalty to the organization. Such a serious step should not be taken before he consults with professional colleagues and associations. Added difficulties may arise when a government is the employer, but the principles are the same, and compromise should be avoided.

At all times, however, management should be kept informed of the actions that are being proposed, in order that trust and respect may be preserved.

It should not need to be said that as a manager in his own right, the occupational physician must not divulge any information which can be regarded as a 'trade secret' if it has no implications for health.

RESEARCH

Occupational physicians may sometimes be involved in clinical research, and will be bound by the ethical considerations that apply to all such projects. There are numerous sources of guidance available, including those published by the British Medical Association[2] and the Royal College of Physicians.[9] Research in industry may involve healthy 'volunteers', and it is important that the volunteers are fully informed of the nature and purpose of the research. Individual informed consent must be given by each participant; it is not enough for a trade union to give consent. Individuals must be free to withdraw from the investigation at any time, and must not be subjected to undue pressure to remain within the project.

ETHICAL COMMITTEES

Usually an occupational physician will be conducting an investigation in association with other doctors who will be in the hospital service, and in this instance the protocol for the project will be submitted for approval to a local ethical committee for clinical research. Doctors should ensure if they are taking part in a *national* trial that the protocol has been approved by a national ethical committee. Information given to ethical committees should include details of the source of the funding of the investigation.

If the occupational physician is the clinician in charge of the project and he or she has received approval from a national ethical committee, it would be courteous and prudent to inform the local ethical committee of this. In certain circumstances it might also be appropriate to inform local consultants and general practitioners of the arrangements.

RELATIONSHIPS WITH OTHER DOCTORS

Occupational physicians may sometimes provide complete medical care for employees, and also their families, in situations where other health facilities are not available, usually in countries overseas. In some other countries the doctor in industry is precluded by statute from treating patients, and again other doctors may offer restricted facilities. Whichever is the case, it is most important that the occupational physician communicates with other doctors who have responsibilities for these patients, but should release clinical details only with the consent of the employee, as stated above. He should make a point of seeking out and meeting general practitioners and hospital colleagues, in order that potential difficulties and causes of friction may be avoided. Invitations to visit occupational health departments are usually accepted with alacrity, and do much to create and maintain goodwill. Offers can then be made to provide ambulant treatment at the workplace for those employees who are able to benefit.

Doctors who work part-time in industry must exercise care not to create a conflict of interest, by ensuring that while acting as occupational physicians they put their other duties on one side.

NURSES IN INDUSTRY

In its 'blue pamphlet', the General Medical Council states:

'*Delegation of medical duties to nurses and others.*

'The Council recognises and welcomes the growing contribution made to health care by nurses and other persons who have been trained to perform special functions, and it has no desire either to restrain the delegation to such persons of treatment or procedures falling within the proper scope of their skills or to hamper the training of medical or other health students. But a doctor who delegates treatment or other procedures must be satisfied that the person to whom they are delegated is competent to carry them out. It is also important that the doctor should retain ultimate responsibility for the management of his patients because only the doctor has received the necessary training to undertake this responsibility.

'For these reasons a doctor who improperly delegates to a person who is not a registered medical practitioner functions requiring the knowledge and skill of a medical practitioner is liable to disciplinary proceedings. Accordingly the Council has in the past proceeded against doctors who by signing certificates or prescriptions or in other ways have enabled persons who were not registered medical practitioners to treat patients as if they were registered.'

The British Medical Association[2] states:

'It must be reiterated that it is unethical for a doctor to delegate work unless he has satisfied himself of the competence of the person concerned. The competence of the nurse may vary considerably according to the type of training undertaken, and the experience gained during training'.

Doctors in industry are fortunate in having as colleagues nurses who have taken further training in occupational health, and in the United Kingdom the Royal College of Nursing Society of Occupational Health Nursing has a high standard of training, and its own code of ethics.[10] Occupational physicians should be aware of these documents, which are complementary to the ethical rules to which doctors subscribe.

Many nurses, qualified in occupational health, may work in relative isolation, and there is no reason why such a nurse should not provide a competent service in the absence of medical support, especially if the work is of an advisory nature. Medicolegal problems may arise in diagnostic and therapeutic areas; the law and custom will vary in different countries, and doctors should be clear in their minds where the bounds of responsibility are drawn.

THE TEAM APPROACH

The concept of health and safety, more so than ever, is dependent on a team approach, and as well as doctors and nurses, the occupational health team consists of other specialists, each experts in their fields and with important contributions to make. As physical hazards have been contained, chemical and environmental problems have become even more important, so occupational hygienists have taken their place alongside safety engineers in helping to make the workplace as safe and healthy as possible. These professionals have their own organizations and codes of conduct, and willingly exchange information with medical and nursing colleagues. Proper safeguards must be maintained, however, to ensure that confidential information about individual employees does not leave the occupational health department.

EMPLOYEES IN HEALTH SERVICES

Many hospitals and health authorities now have departments of occupational health, for the benefit of their own employees, and there may arise conflicts of interest and problems of confidentiality. It is imperative that these departments are completely separate from other organizations within the health service, and in particular the nursing administration—who administer usually what is the largest single group of employees in a health service. The medical records of the occupational health department must be distinct from, and kept physically apart from, other clinical records.

ADVERTISING

Doctors who are full-time employees in industry are not usually concerned with the problem of advertising their professional services. Those who have part-time appointments, or who are free-lance consultants, will require to observe the practices and customs of the countries in which they reside. Those living in some overseas countries are able to draw attention to their qualifications and ability in ways which would be unacceptable in others. They may advertise in newspapers or other publications and may generally adopt a high profile. In the United Kingdom, however, such exposure is regarded as unethical, and it is advisable to heed the advice of the General Medical Council as published in its 'blue pamphlet'. There is no objection to a doctor's name being included in a list of suitably qualified persons, maintained by a professional association, to which an approach can be made by an enquiring member of the public. In many cases contact will be made after the personal recommendation of a colleague. If there is doubt as to whether certain practices constitute unacceptable advertising, the doctor should consult his professional association or defence society.

Increasingly, doctors in industry are being asked to make personal appearances on radio or television, and these requests require careful consideration. If the purpose of the programme is to promote interest in occupational health matters generally, there usually is no cause for anxiety. However, in certain situations the doctor may have been invited to speak on a topic which is currently a matter of contention, and finds that he is confronting another medical person who holds a contrary view, with the result that the entertainment value of the programme is the spectacle of a public disagreement between doctors. In addition, the interview may be recorded, and then edited to suit the requirements of the producer, with the result that the context of the programme is outwith the control of the contributors. Doctors should be wary of *ad hoc* arrangements, and when deciding format should request written rather than oral agreements. It should be remembered that the broadcast can be recorded. Parts may therefore be incorporated in future programmes unless this is expressly excluded by the contract. It is mandatory to tell the employer (if any) before any programme is made, and if the physician has serious doubts, it is best to decline the invitation.

DOCTORS' DILEMMA

When all doctors were entrepreneurs, ethical considerations were relatively simple, as usually no third party intervened between patient and physician. In more complex societies not everyone can afford access to medical advice, so the State has to intervene, and pay or subsidize the doctor in a manner appropriate to the circumstances of the country. The doctor is usually able to maintain his professional integrity and act always in the best interests of his patients. Because the State employs a relatively large number of doctors, and because the State is paymaster rather than consumer, the doctors in a national health service are able to resist managerial pressures in respect of their medical practice. In occupational health, however, there are few employers of large numbers of doctors, and most of these are monopoly industries or government services. It is a sad irony that at a time when so many new hazards are being introduced into industry, occupational health services are sometimes regarded as a fringe activity which can be cut back in times of harsh economic climate, and only the minimum preserved. Doctors working in these situations sometimes find their loyalties and duties put severely to the test, and at such a time they will be glad to gather about them the strength of a professional organization. Yet the doctor—and nurse—in industry must always be impartial, even-handed and open, and demonstrate that in occupational health he who pays the piper does *not* call the tune.

WHAT OF THE FUTURE?

As in so many walks of life, consumers are now less ready to accept professional advice and decisions without demur, and nowhere more so than in the field of medicine. Occupational physicians deal with healthier populations than many other doctors, and these patients, both as individuals and in groups, are increasingly articulate and well-informed. Governments also are taking a more active part in imposing conditions to protect the health of the worker. The doctor in industry finds that he is much more accountable for his opinions and actions, and this trend will become even more pronounced. This is as it should be. People at work will co-operate much more readily if they feel that their medical advisers are not concealing any relevant fact or opinion.

Occupational health departments will find that much more of their time will be spent in *preventing* illness and disease. Legislation being introduced into several countries will require employers to alert governmental agencies if they suspect that one of their employees is suffering from an illness due to employment, and also if they are introducing new processes; detailed records of possible exposure to new substances will have to be kept. All this will increase the number of prospective studies that will be required, with all the ethical implications involved. In addition, there will be an increase in the amount of biological monitoring that will be necessary.

These developments will mean that the occupational physician, even more so than at present, will require to be seen as a completely independent expert, whose prime concern is the health and welfare of the workforce. Doctors are fortunate in being able to be judged by their peers, and they will retain this privilege by their example and their actions. If they lose the confidence of the State, the employer, and most of all the employee, then they must expect this right of self-regulation to be forfeit.

REFERENCES

1. General Medical Council. *Professional Conduct and Discipline: Fitness to Practice.* London: GMC, 1985.
2. British Medical Association. *The Handbook of Medical Ethics*. London: BMA, 1984.
3. Royal College of Physicians Faculty of Occupational Medicine. *Guidance on Ethics for Occupational Physicians*. London: RCP, 1982.
4. Kennedy I. *The Unmasking of Medicine*. London: Allen and Unwin, 1981.
5. Sieghart P. *J. Soc. Occup. Med.* 1982; **32**:4–14.
6. Gillon R. *Br. Med. J.* 1985; **ii**:1556–7.
7. Health and Safety at Work etc, Act, 1974. London: HMSO, 1974.
8. *The Reporting of Injuries, Diseases and Dangerous Occurrences Regulations 1985*. London: HMSO, 1985.
9. Royal College of Physicians. *Supervision of Ethics of Clinical Research Investigations in Institutions*. London: RCP, 1981.
10. RCN Society of Occupational Health Nursing. *Information Leaflet No. 13*. London: RCN, 1981.

21. AUDIT IN OCCUPATIONAL HEALTH

A. J. M. Slovak

Setting aside the financial sense of the word, audit in medicine has come to mean the systematic, critical assessment of function. This serves as both definition and objective. Many authors have unnecessarily circumscribed the meaning of medical audit by making it refer exclusively to that aspect of function which they happen to be addressing in a particular study. This piecemeal approach has led to widespread confusion about what audit is or is not. If it is seen as a hierarchy of review activities which encompass infrastructure, systems and impact, then all the piecemeal definitions are contained within the whole.

RATIONALE

That a discipline should become the subject of audit is a compliment to its importance: that is why audit started as a financial concept. That health care should increasingly become a target for critical analysis is inevitable. It is a consequence of the position it has assumed in postwar, global, societal values both as a proper and laudable objective for funding and as a marker of the success of nation states in providing for their populations. Added impetus towards audit comes in times of financial stringency, with the growing perception that the capacity of the health care industry to deliver useful services far exceeds the capacity of even the most affluent societies to pay. Choices have to be made and means need to develop to make the choices optimally. Given these circumstances, the approaches which evolve will reflect the dominant societal precepts. Currently these are financially oriented analysis and the scientific method, hence the origins of medical audit.

This may explain the feelings of strange familiarity which much of the rest of this chapter may arouse in readers but not the equally hard to define barriers to acceptance which many will instinctively feel. In my view, the barriers are explained by a series of dichotomies in the personal and corporate self-image of health care practitioners, further affected by generational differences. The first of these dichotomies lies between medicine as science and as art, the second between medicine as service to the individual as against service to the population and for occupational health, and the third is role ambiguity between medical and managerial responsibilities. Historical type-casting is also an influence, affecting the perception of what is proper for the professional to do or to eschew. These perceptions have been eroded by time but leave a series of generational strata with different views on how the dichotomies can and should be resolved.

MEANING

The substrate of medical audit has been outlined as the infrastructure, systems and impact of the relevant activities. This needs some amplification. Infrastructure refers to capacity to deliver a service and includes policy and organization, staff and their qualifications, facilities and equipment: it is the nuts and bolts of activity. Systems are concerned with the quality of delivery of services and incorporate remit, validity, effectiveness and efficiency: it is cost/effect. Impact describes outcome, thus the benefit obtained from the infrastructure and systems. This may be measured by a very wide variety of parameters with significance to different parts of wider society—for example mortality, morbidity, avoidance of disasters etc. as well as cost/benefit. These ideas are presented schematically in *Fig. 21.1.*

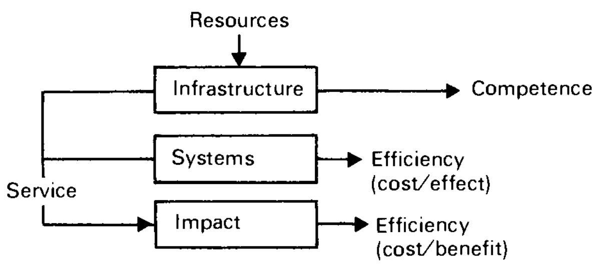


Fig. 21.1. Levels of audit and their inter-relationships.

The worldwide pattern of audit activity in occupational health is rather patchy and before considering it in detail it is extremely educative to examine the experience of other medical disciplines in a number of key areas relevant to industrial problems. The lessons learnt should be of value in avoiding unrewarding initiatives.

AUDIT IN MEDICAL PRACTICE

Hospitals

In North America where audit initiatives in health care have been most prominent, three main themes can be discerned. These have been in the areas of hospital management, medical services and nursing function. Undoubtedly, the impetus towards analysing hospital management has been mainly financial, a consequence of lean times for both private and state funding. Very typical of this approach is the example described by Brykowych and Maitland[1] where 'a comprehensive review of management policies and practices' was initiated in recognition of 'increasing pressure for accountability in the expenditure of public funds'. An established audit tool was used, namely the Program for Institutional Effectiveness Review (PIER) which was developed by the American Hospital Association. PIER is said to be based on the premise that 'the hospital operates as a series of inter-related, interdepartmental systems with clearly defined objectives and tasks that require specific management skills'. It addresses itself to governance, executive management, materials and facilities management. In terms of the model in *Fig. 21.1*, it deals predominantly with infrastructure and secondarily with systems.

Also underlying this type of audit activity has been the concept of hospital accreditation. Fundamentally a tactic to obtain satisfactory vocational and educational standards, accreditation carries with it connotations of quality of care and value for money which attract public funding and status, both of which are desirable goals. Typical of the newly broadened significance of such accreditation schemes is that associated with the outline published by the Canadian Council on Hospital Accreditation[2] which sets down a comprehensive set of standards and guidelines for hospitals to measure themselves against. Whilst the setting of standards may be an effective inducement to improvement among independant contractors, even here both the adequacy of funding and the political will to provide funding is a prerequisite. This is even more of a problem in countries such as the United Kingdom which have a national health service so that the State must assume the dual roles of standard setter and cash provider. This means that the data that are available from the governments' Performance Indicator review[3] tend to be underused for improving the service and that special reviews are likely to be facile and superficial.[4]

Medical services

The most popular of audit activities in health care relates to systems and is directly analogous to industrial quality control. Such activity in North America was reviewed recently by Van't Hoff.[5] The official basis of quality control programmes is the accreditation process which has been previously described. Audit is delivered in individual hospitals by audit committees which review hospital activity and set audit objectives year by year. Not unusually, special themes are picked out for formal study in a format similar to that of ordinary scientific studies. The only difference is that the outcomes are subject to peer review, action and follow-up. More routinely, deaths in hospital, medical records and validity of pathological indices are standard subjects of periodic review.

The main objection to audit of medical activities has been to external review. The bogeyman has been the hard, faceless administrator who 'does not understand'. This is probably a largely unjustified slur upon the intelligence and integrity of general administrators and managers and really has more to do with professional paranoia about independence. Nevertheless it is a fact of life. The corollary of rejecting external audit is the need to accept appraisal as the proper responsibility of an autonomous profession. Hence we have peer review.

The impetus towards quality control and improvement in the UK has come from individual enthusiasts who have recognized both the need for, and benefit flowing from critical analysis of their own work as well as the broader importance of improving health care delivery. To some extent this approach has been recognized, taken up and organized by professional bodies, notably the Royal College of General Practitioners, the Royal College of Pathologists and the Royal College of Physicians. In general practice, the extent of support for systematic audit activities is substantial. An example of extensive and systematic work is provided by Fleming and Lawrence[6] who followed up a previous review of various preventive measures—for example cervical smears, blood pressure measurements and so on—taken in a series of general practices and were able to demonstrate an increased uptake in all these services following educational feedback to those involved in the study. This study emphasizes the importance of positive feedback and continuing audit surveillance of the providers of health care if improvements in service are to be obtained. This is reiterated time and again in papers comparing and contrasting success and failure of audit initiatives.[7,8,9]

In the hospital setting, the best known examples of longstanding commitment to quality have been the various schemes set up for the validation of pathological tests. These procedures would be familiar in design to those operating 'quality circles' in production facilities. More recently, both physicians and surgeons have begun to undertake systems analyses which concentrate on both quality and also comparative cost/benefit. Thus Johnson's analysis of surgical practice in a community hospital[10] demonstrates that the results of selected surgical procedures do not suffer when compared with results in general hospitals and that the outcome, using a wide range of medical and social criteria, is beneficial. Similarly the work described by the Swansea Physicians Audit Group[11] reports their findings in relation to efficiencies obtainable with regard to diagnostic tests, drugs, length of hospital stay and follow-up outpatient attendances.

Nursing

The most comprehensive use of audit in the hospital scene has been in the development of nursing services in the USA. Curtis and Simpson[12] clearly identify and seek to use what they call structure, process and outcome audit in a hierarchical way to provide a rational means of assessing nursing practice. A seemingly necessary part of this activity is an unattractive form of pseudo-scientific jargon which does more to obscure than to clarify the line of arguments for the casual reader. However, the vocabulary is logical and established, at least amongst systems analysts, and therefore has to be learnt. Central to the development of audit in nursing is what is termed 'the nursing process'. This is simply good nursing practice systematized. It can then be broken down into a series of infrastructure, systems and impact activities which can be scored with regard to individual or departmental performance. The process of audit under these circumstances is time-consuming and cannot be rushed. Time constraints as well as innate conservatism probably explain the widespread resistance such concepts have met in nursing and in health care generally.

The direct benefits of audit may be seen as potential for the improvement of means of delivery, quality of delivery, and outcome of health care. More subtle effects may also be achieved by the impression that the audit process achieves on interested parties outside the immediate ranks of those delivering health care. For general practice in the United Kingdom and nursing in North America, audit has been an effective instrument for enhancing status and public standing. Within these specialties it has led to an improvement in the quality of career entrants and, within constraints, of service given. In democratic societies public standing and influence is of enormous influence in obtaining funding, a fact which is of immediate relevance for occupational health and safety in many countries.

AUDIT IN OCCUPATIONAL HEALTH

Infrastructure

Audit activities in occupational health have been piecemeal, mainly concentrating on infrastructure and impact, except in a few countries, principally Scandinavian, where the entire national occupational health programmes may be considered as audited procedures. Infrastructure has been most directly addressed in the legislation of a number of mainland European countries but audit of infrastructure is predominantly a North American activity, characteristic of many of the larger corporations' occupational health and safety operations both domestically and overseas. This inventory approach is summarized in the American Industrial Hygiene Association's publication *Standards, Interpretations and Audit Criteria for Performance of Occupational Health Programmes.*[13] Herein are addressed administrative, medical, nursing, occupational hygiene and health physics activities itemizing policy, staff, facilities, equipment, activities and records. The document asks the question, 'do we have the facilities to deal with problems?' It does not address itself to quality or outcome but that basic question itself is probably uncomfortable enough without further development!

At a more mundane level, a very practical example of infrastructural analysis is reported by Forsythe-Jauch[14] from the British Army. His concern was to obtain a minimum standard of compliance to regulations requirements in reserve army units (Territorial Army). His solution was to use a checklist requiring the making of returns which could be validated from the centre. Also in the United Kingdom, the Employment Medical Advisory Service (EMAS), the medical arm of the Health and Safety Executive, has begun a series of limited audits of recently introduced activities. These are mainly infrastructure oriented. An example is the working of the First Aid at Work Regulations, 1982. These laid down training, organization and manning standards and also a fairly elaborate administrative procedure. This procedure is to be reviewed as is compliance with the training schedules notified to EMAS. What is included in the scope of this audit is perhaps less interesting than what is left out. Thus, in terms of effective delivery of first aid, recent research suggests that the average first aider forgets the basics of resuscitation in three months and not the three years set down by the Regulations as the maximum training interval, but no research is likely to be initiated to validate either this finding or the more general effectiveness of first aiders.

Pre-employment and routine medicals

Among the most routine activities of occupational health functions remain pre-employment and periodic general medicals. Historically, these activities achieved their current prominence from the fusion of two themes in wartime and postwar medical thinking. The first of these was the obvious need to screen the severely debilitated populations of postwar mainland Europe for illness and their basic suitability for strenuous work. Subsequently there was also a need to ensure that health was maintained or improved. The other theme ran along a very much more idealistic and optimistic vein typified by the sentiments of the Beveridge report in the UK and the development of comprehensive health support functions in such war-important American industries as the Kaiser shipyards. The fundamental premise here was that if medicine was good for people then more medicine must be better. Both these themes have now been successfully institutionalized, perhaps ossified, and are invested with profound importance and faith in many societies today. In fact, the first theme's importance had greatly diminished in Europe by the mid-1950s as prosperity increased, the second has always been something both of an illusion and a delusion.

The 'hands-on' pre-employment medical is still widely performed. Although 'pass-standards' exist for many discrete groups—for example divers, drivers and nurses—there is often disagreement about what those standards should be and the means and vigorousness of the detection measures undertaken are seldom standardized. To some extent, the pre-employment medical is subject to modulation by social pressures. When employment is high the marginal candidate may be accepted; when the choice is wider, standards tighten, consciously or otherwise. Other social pressures, especially equal opportunity and the political climate on discrimination in general, may also be important.

With all these circumstances, there has come a tendency to accept people more as they are and to try and make sure that they are a comfortable fit with the job they intend to take up. Less regard has also been paid to the idea of 'suitability' or 'eligibility' for pension or benefit funds as the general health of populations has become uniformly improved and such funds have become more a universal benefit than a privileged perquisite. Recognition of the extensive self-selection of candidates for jobs—incidentally producing the healthy worker effect—has also contributed so that in practice, the level of detection of abnormalities and the level of medical rejection in the broad spectrum of jobs is now very low. Little work has been published on the critical evaluation of the pre-employment medical but what there has been can only be described as an instinctual drift in occupational health practice towards minimalist screening measures for the generality of jobs. Often, this new approach has been based on questionnaires biased towards suggestive past medical history as a preliminary sieve followed by predominantly nurse-operated screening if appropriate.

Slightly better documented in the literature has been the periodic medical or 'physical', although when looked at as a practice which has absorbed countless billions of occupational health staff workhours for several decades, the lack of validating, let alone supportive literature, is embarrassing. Perhaps the most comprehensive review of screening procedures, including the 'annual physical' was undertaken by a task force reporting to the Canadian Conference of Deputy Ministers of Health in 1979.[14] The task force reported that the annual check-up was 'inefficient and, at times, potentially harmful'. They suggested that it be substituted with a series of specific, targetted 'health protection packages'. A little pause for thought would suggest that this procedure, directly originating as it does from the traditional diagnostic examination of the sick person as practised in the early part of the twentieth century, is rather unlikely to be appropriate for the preventive needs of the healthy person at the end of the century. That we have regarded these procedures as pre-eminently useful says much for tradition, obscurantism and the very powerful wish for reassurance—or even perhaps for immortality—that acceptance of them by the lay public implies.

Qualitative improvements have been much more easy to produce in such predominantly numerate disciplines in occupational health as epidemiology and occupational hygiene. Hygiene, indeed, is an excellent example of an auditable discipline and there is now a substantial literature of systems and mathematical models of hygiene audit procedures in more or less complex enterprises. All of these models conform to a basic which is described by Mitchell and Fowkes[9] as the cycle of audit but which may also be called inductive feed back. This is presented diagrammatically in *Fig. 21.2.*

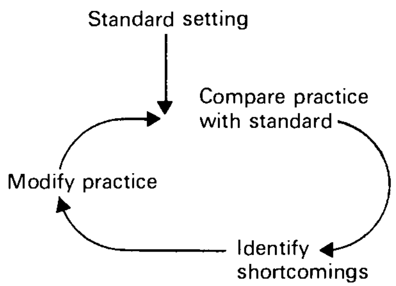


Fig. 21.2. Cycle of audit.

International aspects

Another interesting aspect of occupational health audit is how to look at questions of international comparison. Politically, this is a fraught problem riddled with chauvinistic prejudices. Setting these aside, so far as is possible, there are two distinct theoretically desirable objectives. One is the 'best possible'; that is the best standard obtainable in the most advanced or committed country. The other is the 'best practicable'; that is the promotion of the best standard achievable according to local conditions and perceptions; in different words, 'appropriate' or 'intermediate' technology. To further complicate things, both 'best possible' and 'best practicable' change with time. Therefore it may perhaps be best to look upon these approaches as the practical manifestation of two idealogically conflicting theories, absolutism and pragmatism. These philosophies may be seen in dispute elsewhere in occupational health such as in the setting of hygiene standards and with carcinogens policies. However, there are very painfully real outcomes arising from these theoretical conflicts when they are not resolved in the real world situation. This is evidenced by the Bhopal tragedy and the confused recriminations which have followed it.

A most illuminating contribution to international analysis is to be found in the ILO report on safety and health practices of multinational enterprises.[15] This study reports that the national prejudices and styles of delivery of the home country of the multinational tend to be exported to subsidiary parts of the organization. This, in practice, may vary from rigid centralist authoritarianism to complete *laissez-faire*. Also, whatever the regime of imposition, the standard of health and safety tends to deteriorate measurably in proportion to the distance from headquarters, whatever country that happens to be in. This is a thought-provoking finding, although maybe not a real surprise.

THE FUTURE

The Scandinavian model, in particular the Finnish approach, must be regarded as the 'best possible' at present. Here, a massive national effort has been mobilized to identify and eradicate known occupational disease and to improve the general health and safety provision and performance in a systematic and intensive way. This approach, at government level, has identified the resourcing levels deemed necessary, has ensured that they were made available and has measured the achievements in terms of a comprehensive range of outcomes. Less systematic, because they were on a steep learning curve, but nevertheless very apparent, have been the improvements actively pursued in investigative techniques in many of the occupational health disciplines where it has been found necessary to improve efficiency and effectiveness in order to deliver the required outcomes.

Most other industrialized countries have not committed themselves to the statements of policy and objectives which are a prerequisite of the 'cycle of audit', let alone the commitment of resources which implementation would necessitate. Instead, we have seen a gradual evolution of ideas and services which have achieved a variable level of penetration in different specialties, types of industries, sizes of industries etc., in different countries. Interesting examples of the results of this evolutionary approach are the relative numbers of occupational hygienists compared with occupational physicians in the United States (≃1:1) and the United Kingdom (≃1:20) and the relative importance ascribed to mechanical safety inspection and safety philosophies in those two countries. These evolutionary changes may amount to a progressive divergence between real balanced needs and the objectives of established professional interest groups.

With the avowed United Nations objective of 'health for all' by the year 2000, national occupational health and safety policies are probably necessary and would need to be enforced effectively in order to be successful. However, the absence of political will to implement such objectives at national level in most countries is an inescapable fact of life. In the United Kingdom, even the modest, sensible, but inherently pessimistic proposals of the recent Select Committee of the House of Lords are likely to be ignored.[16] Indeed, only recently, the Government has undertaken to investigate the 'unnecessary burden' of health and safety regulations on entrepreneurial initiatives.

However, inertia at international and national level is nothing new and is not necessarily a bar to progress. The influence of accreditation procedures in advancing standards of health care has already been described. A similar process is noticeable in professional bodies in occupational health and hygiene with regard to training and examination standards in a number of countries. The peer assessment or audit of facilities for training is a powerful motivator for more general improvements in standards which become a matter for professional pride. This route of audit is probably the most likely to obtain effective improvements in general standards in most countries not positively taking up effective national policies, although employer and trade union initiatives may also play an increasing part. To enhance the effect of such audit, the support of national executive bodies is important in the setting of minimum standards for qualifications and the basic level of provision of services. National bodies do, however, have another important responsibility and that is not to let the standards ossify and thus become irrelevant.

CONCLUSION

Failing the development of comprehensive and responsive national policies in many countries, audit becomes less concerned with outcome than with content and efficiency. The mildly cynical definition of occupational health as being what occupational health practitioners do is still painfully close to the truth. It should be possible for groups of professionals to agree what minimum standards ought to be available in particular industrial sectors along the lines of the AIHA programme.[13] Also a series of initiatives, necessarily piecemeal, need to be developed to assess the value of existing activities measured against needs and existing procedures measured against objectives. All of these are proper professional activities and their implementation is a measure of professional leadership.

REFERENCES

1. Brykowych G and Maitland B. Comprehensive adults—a managerial challenge. *Health Care* 1984; **26,** 6:48–49.
2. Standards of Accreditation of Canadian Health Care Facilities. Canadian Council on Hospital Accreditation Ottawa, January 1983.
3. Department of Health and Social Security. Performance Indicators: National Summary for 1981. London: DHSS, 1983.
4. NHS Management Inquiry. Griffiths Report. London: DHSS, 1983.
5. Van't Hoff W. Audit reviewed: Medical audit in North America. *J. R. Coll. Physicians Lond.* 1985; **19,** 1:53–55.
6. Fleming DM and Lawrence MSTA. Impact of audit on preventative measures. *Br. Med. J.* 1983; **287:**1852–4.
7. Freeling P and Burton RH. General practice and learning by audit. *J. R. Coll. Gen. Pract.* 1982; **32:**231–237.
8. Waters WHR, Kelly J and Lunn JE. Attitudes to audit. *J. R. Coll. Gen. Pract.* 1983; **33:**263–266.
9. Mitchell MW and Fowkes FGR. Audit reviewed: Does feedback on performance change clinical behaviour? *J. R. Coll. Physicians Lond.* 1985; **19,** 4:251–254.
10. Johnson DB. Audit of surgical practice in a community hospital. *Br. Med. J.* 1984; **288:**1293–1295.
11. Swansea Physicians Audit Group. Audit of clinical practice: Opportunities to rationalise the use of resources. *J. R. Coll. Physicians Lond.* 1983; **17,** 4:235–238.
12. Curtis BJ and Simpson LJ. Auditing: A method of evaluating quality of care. *J. Nursing Admin.* 1985; **15,** 10:14–21.
13. American Industrial Hygiene Association, *Standards, Interpretations and Audit Criteria for Performance of Occupational Health Programmes.* AIHA, 1978.
14. Periodic Health Examination. Report of a task force to The Conference of Deputy Ministers of Health. Minister of Supply and Services Canada. 1980.
15. International Labour Office. *Safety Health Practices of Multinational Enterprises.* Geneva: ILO, 1984.
16. House of Lords Select Committee on Science and Technology. Occupational health and hygiene services 1983. 1-Report, HMSO.

Index