ter Meulen
Biller-Andorno
Lenk · Lie

Evidence-based Practice in Medicine & Health Care

A Discussion of the Ethical Issues

Springer

Evidence-based Practice in Medicine and Health Care

Ruud ter Meulen · Nikola Biller-Andorno
Christian Lenk · Reidar K. Lie (Eds.)

Evidence-based Practice in Medicine and Health Care

A Discussion of the Ethical Issues

With 4 figures and 12 tables

Springer

Prof. Dr. Ruud ter Meulen
Instituut voor Gezondheidsethiek
Universiteit Maastricht
Universiteitssingel 40 (West)
NL-6229 ER Maastricht
R.terMeulen@IGE.unimaas.nl

Dr. phil. Christian Lenk
Abt. Ethik und Geschichte der Medizin
Universität Göttingen
Humboldtallee 36
D-37073 Goettingen
clenk@gwdg.de

Prof. Dr. Dr. Nikola Biller-Andorno
Charité – Universitätsmedizin Berlin
Professur f. Medizinethik
Klingsorstrasse 119
D-12203 Berlin
nikola.biller-andorno@charite.de

Reidar K. Lie, MD, PhD
Department of Clinical Bioethics
National Institutes of Health
10 Center Drive Building 10, Room 1C118
Bethesda, MD 20892-1156
Reidar.Lie@fil.uib.no

ISBN 3-540-22239-1 Springer-Verlag Berlin Heidelberg New York

Bibliographic information published by *Die Deutsche Bibliothek*.

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic information is available in the Internet at: http://dnb.ddb.de

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in The Netherlands

Editor: Thomas Mager, Heidelberg
Development Editor: Susanne Friedrichsen, Heidelberg
Production Editor and Typesetting: Frank Krabbes, Heidelberg
Cover design: deblik, Berlin

SPIN: 10905289 14/2109 fk – 5 4 3 2 1 0 – Printed on acid-free paper

Table of Contents

1 The Ethical Debate on Evidence-based Medicine – Introduction to the Volume

Ruud ter Meulen · Nikola Biller-Andorno · Christian Lenk · Reidar Lie

Among the intensely debated issues in medicine at the moment, is the role of Evidence-based Medicine (EBM) as the basis for decision-making in clinical care and health policy. The idea behind Evidence-based Medicine (EBM) is that doctors should only make use of therapies that have been proven to be effective, rather than making use of unproven therapies that may be ineffective or even harmful to the patient. According to the founding father of EBM, David Sackett, EBM helps make the best decisions for patients by providing health care workers with a tool to search for the best available evidence. The best evidence is evidence that is produced in randomised controlled clinical trials (RCTs), where the association between a specific intervention and its outcomes is researched within very strictly controlled conditions. While Sackett and his colleagues wanted to restrict EBM to the context of individual patient care, other advocates of EBM have extended its range to the area of health care systems and policy-making. Evidence-based Medicine is not only seen as an important means to improve the quality of medical care, but also as an instrument to control costs. Due to the scarcity of health care resources, decisions on the allocation of care will have to be made more explicitly and should be made more transparent and accountable.

According to Muir Gray (1997), health care decision-makers must practice evidence-based decision-making. As pressures on health care systems are increasing, the need for more explicit decision-making and rationalisation by using empirical evidence will become stronger. EBM now requires for a combined approach to improve the quality of clinical care, and controll the costs of care. This is achieved by using the best available evidence in order to weed out useless and possibly harmful treatments (Cochrane 1999).

Proponents argue that EBM will improve the quality of medical care and that its implementation is an obligation towards the individual patient. Medical practice is increasingly flooded with new methods and discoveries, rendering a careful reception and application of relevant information by the individual physician very difficult. Consequently, conventional therapies are still in use, and innovative modern approches require time to be intograted into practice. EBM could help by providing relevant, topical information in a time-efficient manner, improving the knowledge base (and job satisfaction) of health care providers. These additions will ultimately result in an improved quality of cqre for patients. EBM not only facilitates the effective assimilation of existing knowledge, but also helps to sort out 'real' knowledge from pseudo-knowledge. The result of this process will be more 'rational', and thus improved, medicine.

An additional argument in favour of EBM relates to the current and increasing pressure on health care systems to make allocation decisions. Evidence-based Health Policy (EBHP) will help to use available resources in an efficient way to make allocation decisions. Evidence-based allocation is thus a social responsibility. Moreover, evidence-based health policy could safeguard against an erosion of the quality of medical care that might be caused by purely economic thinking. This is because EBHP involves clinically experienced physicians, whose judgment and expertise can help to preserve an adequate standard of care.

However, other voices within the medical community have strong suspicions about the EBM

enterprise. An important aspect of these concerns has to do with the abstract nature of evidence, and of the guidelines that are based on this evidence. According to the critics, the data of Evidence-based Medicine is based on ideal populations, and has little relevance, if any, to doctor-patient relationship. Sackett defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients", which "integrates the best evidence with individual clinical expertise and patients' choice" (Sackett 1996). Sackett stressed that doctors should make 'judicious' use of the data of systematic reviews. Nonetheless, many clinicians have the view that EBM, and particularly the introduction of evidence-based guidelines, will result in a kind of 'cook-book medicine' that has nothing to do with the traditional skills and moral values of their professional practice. They fear that when guidelines are regulating medical practice, clinicians will lose their professional autonomy. Moreover, as guidelines cannot capture all characteristics of the individual patient, physicians will develop a blind spot for pertinent unique characteristics of the patient, and will consequently not fulfil the moral demand of beneficence (Willems 1995).

A further concern about EBM has to do with the nature of the evidence that is generated in the RCT. The argument is that the methodology of the RCT, which forms the core of the EBM movement, excludes many other kinds of research, such as observational studies and qualitative research. Although proponents stress the idea that EBM should be based on the best available evidence, there is an inherent tendency to focus on randomly controlled clinical trials and on services and treatments that lend themselves to such research, according to the critics. There is a fear that if the RCT is considered the gold standard for research and attainment to evidence, researchers and their sponsors will focus more on drug trials. This provides the 'natural environment' for the RCT model. If policy decisions, particularly reimbursement decisions, are based on the 'best available evidence', drug-based treatment regimes could derive an unfair advantage over other kinds of treatments. This could attect treatments such as psychotherapy in mental health care or physical therapies and caring in rheumatology.

Additionally there are doubts about the implementation of the results of RCTs in clinical practice. The Dutch psychiatrist Kaasenbrood (1997) points out three pitfalls in regarding the RCT as an ideal basis for practicing medicine. First of all, he states that RCTs are available at best in only twenty percent of the cases. Even if RCTs were available for all cases, the results of these RCTs would not automatically be valid in daily practice. In the RCTs, people are removed from their environment and are reduced to the specifications of the trial. Secondly, there are indications of bias in the way the knowledge is generated. Knowledge is a social process, in which subjective preferences and values are guiding the decisions about what should be researched and how the research should be conducted. Even if knowledge was objective, there would still be a concerns related to appropriate use of this knowledge. Physicians do not seem to be able to interpret experimental research correctly, let alone to weigh and apply it. Thirdly, in practice, only a few of physician's decisions are actually based on empirical research. Kaasenbrood comes to the conclusion that the RCT should not be seen as the ideal basis for medical practice, but that the value of professional experience of physicians needs to be restored.

In their report on the role of EBM in the German health system, Biller-Andorno and Karageorgiou (2001) mention, as one of the arguments against EBM, that EBM will foster the development of one mainstream medicine. Complementary or alternative approaches will be sorted out as "futile". Such therapies and concepts (like homeopathy) might not be accessible to evaluation by EBM standards, and thus could be excluded from the health system. As a result there will be less choice for patients, except for private patients, who can afford costly insurances. Patients who could not afford private care would only be provided a narrowly defined standard medicine for the rest. Therefor, the introduction of EBM and Evidence-based Health care (EBHC) might lead to injustices. What about groups for which there is not much evidence so far? Does it mean they will be considered less when allocation decisions have to be made?

Critics of the EBM movement argue that there are also some doubts as to whether the results

of the RCTs will reach the individual doctor (Biller-Andorno and Karageorgiou 2001). Even once accepted, the implementation of EBM in clinical medicine will pose a considerable challenge. It requires the willingness of practitioners to familiarize themselves in detail with epidemiology and statistics. This will not be possible for many of physicians so they will have to rely on external expertise. This will be problematic, as it might lead to a new 'secret science', instead of the originally intended integration of EBM principles into general medical practice. Moreover, for many cases, ´hard´ evidence is unlikely to be found. EBM might not be the quick and handy guide to the best therapy. Overreliance on competence in EBM methods, while clinical skills are being neglected, might lead to more insecurity and a worse quality of care.

Moreover it is questionable if EBM can initiate a change in the power structures within medical institutions. Some authors argue that, as a result of EBM, medical treatment could be subject to discussions with the staff involved so that all experiences and arguments could be weighed. Old hierarchical structures could be replaced by a more constructive atmosphere of teamwork and collegiality. EBM will empower patients, because they can ask for the knowledge base of some recommendation or even research it themselves. In this way, evidence-based medical practice might increase patients' confidence in the suggested therapy as well as in the physician, who is acting as the expert guide presenting the available therapeutic options. It might also improve compliance, since patients are involved in the process and discussions concerning their treatment.

Hope also put forward this argument (1995). According to Hope, Evidence-based Medicine can serve as a critical response to the inadequacies of traditional medicine, in part as a reaction to the (excessive) authority of doctors and other health professionals (Hope 1995). With acces to the the outcomes of systematic reviews, patients and their representatives could claim a more central role in medical decision-making. However, it is questionable whether EBM will result in a medical "democracy". There is serious concern among doctors and other health care practitioners that the increasing dominance of EBM might lead to a shift of power from doctors (and their patients), to managers and purchasers (who is meant by that term?). Policy-makers and other third parties like insurance companies, are often tempted to use guidelines and cost-effectiveness analysis to control clinical practice and health care delivery. They thus can control the costs of health care. In such cases, the professional autonomy of the clinicians and the autonomy of the patient will be seriously restricted (Biller-Andorno et al. 2002).

The ethical debate on the subject of EBM reflects different points of view concerning its desirability and anticipated consequences. Some of these viewpoints have been put forward in various publications (see bibliography). However, most of these publications address only a particular aspect of the debate. A further characteristic of the debate is that the ethical aspects of EBM are hardly discussed in a systematic way. This Volume is the outcome of the European Project "Ethical Issues of Evidence Based Practice in Medicine and Health Care" (EviBase), which tried to get more insight into the ethical backgrounds of the debate on the role of EBM in various areas of medicine. This included clinical practice, medical education, medical research, health policy and medical sociology, particularly in the context of European health care systems. Most of the chapters are based on papers that were discussed at the workshops and conferences of this project. Some chapters were added later, and a few other chapters are (elaborated) versions of articles that were published at an earlier moment (Health Care Analysis, Journal of Medical Ethics). This Volume tries to bring together as many viewpoints as possible regarding the contribution EBM makes and might make to medical practice and health policy, with a focus on the ethical issues that are at stake in this process.

The Volume opens with a section on the *historical and political role of EBM*. While many discussions on EBM take place *within* the context of EBM, this section tries to get more insight into the relationship between EBM and society. That is the role of EBM in the context of society. A few questions that are at stake here are: How has EBM developed from an internal medical issue to an instrument for health policy, how does EBM affect our lives and is there a role for the state or the public in directing or re-directing EBM.

Christine *Sepers* and *Ruud ter Meulen* offer a historical analysis of the role of guidelines in medical practice and in health policy. They describe the various stages in the development of guidelines, and of their expected role in clinical decision-making. Their chapter includes a reflection on the ethical debate on the role of guidelines, particularly on the social and ethical consequences of guidelines when they are used for allocation purposes.

The second chapter in this section, by *Rein Vos, Rob Houtepen and Klasien Horstman*, analyses the political role of EBM in our western societies. He evaluates how EBM affect our lives in society, and what kind of role EBM ought to play in our health care system and our society as a whole. Vos et al. argue in accordance to the citizenship theory, that the role of EBM in health care or society is not fixed, but depends on the relation between state and society. They conclude that EBM cannot be isolated as a kind of value-free technical rationality. It can only have practical relevance when it is tied to rather than separated from the real world, where values and interests of patients and professionals play a decisive role.

The last chapter in this section is by *Mariachiara Tallachini*, who discusses the relationship between scientific knowledge and the law. The legal view of science has changed through time, moving from a positivist, non-critical position of law towards science, to a critical view of science based on a more objective knowledge and of the role of society. Tallacchini explores some judicial cases that illustrate these attitudes, suggesting that reference to science (particularly to EBM) can be equitably made. This can be done only when it serves the cause of transparency and democratization both in science and in law. Like Vos et al., Tallachini argues that EBM should acknowledge the broader link of science and society, by the inclusion of citizens and patient groups. Only in this way can EBM contribute to the democratisation of medicine and health policy that its proponents often favour.

The second section of the Volume deals with the *role of EBM in clinical practice.* Improvement of clinical practice has been at the heart of the EBM movement since it's beginning. EBM is now shaping clinical practice, not only by teaching doctors how to interpret the existing medical literature and how to make individual treatment decisions. However, EBM has also been taken up by health policy makers as a tool to allocate resources. The ethical questions raised by the use of EBM thus concern not only quality of care issues but fairness issues as well. The four papers address one or both issues.

Porzsolt and Leonhardt-Huober present a six-step model of evidence-based decision making. This model adds to the formulation of appropriate questions and the searching, critical appraisal and integration of internal and external evidence a sixth step. This aims to evaluate the decision-making process and its outcome. Part of this last step is the realization that a large percentage of clinical decisions have to be made under conditions of uncertainty, and that value judgments play a particularly prominent role under such conditions. Ethical analysis can thus contribute to a comprehensive understanding of actual clinical decision-making processes. As ethical analysis, on the other hand, needs to be based on factual data, the relationship between ethics and EBM can best be described as "bi-directional". Understanding the relationship of fact and value, of EBM and ethics should already be part of medical education because they are a requirement for sound and self-critical decision-making in clinical practice.

The contribution by *Praetorius* describes the role of clinical guidelines in cardiology, a clinical area that seems particularly amenable to the conduct of RCTs and the establishment of guidelines. The process of developing guidelines, however, is not always a transparent and evidence-based one. In addition, there is an overwhelming flood of existing guidelines, which partly contradict each other. The chapter discusses a variety of preconditions for a more appropriate use of guidelines, among them the need for clarity and international harmonization, for methodological sophistication, and for a role of guidelines in individual decision-making. The latter use acts to enhance rather than decrease the autonomy and sense of responsibility of the physician who applies them. Whether EBM can keep its promises is an open question: the effect of EBM on cardiology is in fact experimental and should itself be subject to investigation.

Fabris, Scarafiotti and Maero focused on the applicability of EBM to frail elderly people in community and institutional settings. They emphasize the methodological issues that frequently preclude the inclusion of this population into RCTs. The scarcity of high level evidence, and the use of indicators like shorter life expectancy and reduced probability of recovery are likely to have an – albeit unintended – discriminatory effect on health care for the elderly population. The authors conclude with recommendations for adapting EBM strategies to geriatrics, including the use of non-RCT evidence like observational studies and clinical experience.

Reis, Lenk and Biller-Andorno investigated the implications of an increased reliance on EBM in psychiatry and psychotherapeutic medicine, as another area where particular challenges can be anticipated. The chapter describes how professional bodies and individuals in the discipline have taken up EBM. It analyses on what grounds attempts to put EBM to use in mental health have been criticized. The ethical reflections offered towards the end of the chapter focus on consequences for both quality and fairness of mental health care, and call for a broadening of the evidence base and methodological refinement.

The third section deals with *philosophical and methodological issues of research,* particularly research based on the RCT. Many commentators have pointed out that the preference for RCTs and formal meta-analyses within Evidence Based Medicine may exclude rival, equally valid knowledge claims. It is not the aim of this volume to re-examine these issues, as we have focused on the ethical issues raised by EBM. However, five chapters discuss how some of the epistemological and value issues intersect.

Richard *Ashcroft* exposes some of the assumptions made when we assert that we know a treatment to be effective. He emphasizes that such claims are often relative both to other beliefs we have and to the wider social context. Although he maintains that EBM and its reliance on RCTs does produce reliable evidence, an examination and questioning of these assumptions will enable us to improve the evidence base of clinical medicine further.

The paper by *Vos, Willems and Houtepen* provide concrete illustrations of Ashcroft's theoretical analysis for the field of orphan diseases. Some have argued that EBM methodology is not applicable to certain areas of medicine, such as chronic pain conditions or multi-symptom illnesses, because it is impossible to design rigorous RCTs. Vos et al. challenge this assumption, and prefer to analyse these types of cases by showing that they represent an intersection between two different spheres or worlds of norms, the world of science and the world of patient experience. It is not simply a matter of applying the norms of science to a different domain, or giving up scientific claims for the sake of patient values, but a matter of negotiation and adjustment of the two different perspectives.

The chapter by *Reidar Lie* in this section argues that the requirement to test new interventions by doing RCTs may, in situations of resource constraint, lead to a conflict with a key value in research ethics. This conflict is that patients in trials should receive a treatment that is at least as good as the best current therapy. Reidar Lie demonstrates that the familiar controversy over standards of care in developing countries is also a challenge under conditions of moderate resource constraints found in the rich part of the world.

The chapter by *Heiner Raspe* deals with the issue of benefits in clinical research. Much has been written on the concept, measurement and reporting of inconveniences, burdens and risks of subjects participating in controlled clinical trials. However, the category of "benefit" has been widely neglected, as if even the slight possibility of a remote and undefined "benefit" outweighs a wide range of burdens and risks. The qualification and quantification of benefits, however, requires equally careful consideration to that of risks and burden. According to Raspe, it is an essential task of researchers and ethics committees to safeguard the rights and integrity of clinical trial subjects. The weighing procedure required has to carefully consider both risks and possible benefits, be they primary or secondary, direct or indirect.

An important part of the EBM movement is to establish parameters for acceptable clinical practice, and a range of instruments for monitoring and enforcing these parameters. Some critics, how-

ever, have questioned the movement's 'sometimes exclusive' focus on one particular research design (i.e. the randomised controlled trial) as unnecessarily narrow. They also claim that it reinforces the cultural and political values of particular research groups. In the final chapter, *Alessandro Liberati and Paolo Vineis* state that in the current debate, there is a mixture of epistemological confusion about the proper definition of 'proof' and 'evidence' and resistance to cultural and professional changes from within the medical profession. They also argue that there are misplaced criticisms from EBM-sceptics and, to some extent, over-enthusiasm and reductionism from those who fail to recognise EBM's practical and methodological limitations. By clarifying the potentialities and the limitations of EBM, Liberati and Vineis want to present a more appropriate view on the expectations of EBM in order to prevent conceptual and practical mistakes.

The fourth section of this Volume turns to the discussion of the role of *EBM in Public Health Policy*. As already indicated, EBM has moved into the area of health policy where it has been ascribed the task of promoting the rationalisation of allocation and rationing decisions. An example of this rationalisation process is the effort to sharpen the indications for access to scarce medical treatments by the use of cost-effectiveness analysis. There is some concern, however, that the rationalisation of the rationing process will be at the expense of widely shared social values like justice, equity and solidarity with the elderly population. (Biller-Andorno et al, 2002) For example, patients who are expected to benefit to a low degree from particular treatments, such as the elderly and the chronically ill, or for whom there is no evidence available, such as minority groups, may be excluded from access to such treatments or may not benefit from them.

This section opens with a chapter by *Wendy Rogers* on the implications of EBM for the position of disadvantaged groups, like cultural minorities or socially deprived groups. According to Rogers, the data of EBM are produced under the very narrowly defined conditions of the RCT, with specific research populations. The concern is whether, and to what extent, the results of the RCTs are relevant for disadvantaged groups, who are generally not represented in these trials. In the case where cultural and social determinants of health are not taken into consideration in research, this might lead to serious issues of justice, when the outcomes of this research are the basis for decision-making in health policy. Another important question addressed in this chapter is, whether EBM is intrinsically inimical to the interests of weaker groups or whether its potential for these groups has not yet been realised.

As indicated above, there is a growing interest among health policy makers to use economic evidence, like cost-effectiveness studies, for decisions on priority-setting in health care. In the following two chapters, the role of formal outcome evaluations in health policy-making, particularly priority-setting is discussed. *Ole Frithjof Norheim's* contribution deals with the role of the normative perspective in such decisions. According to Norheim, there is a lack of awareness about the value judgements embodied in health technology assessments performed by institutions like NICE in the UK, where outcome evaluations are generally seen as a value-free exercise by scientific experts. In his chapter, Norheim analyses where value judgements come into play in technology assessments, and on what level these value judgements take place. However, when these value judgements, for example as indicated patient preferences, are taken into consideration, how should they be interpreted from principles or theories of fairness? One of the main problems of contemporary societies is the pluralism of values and the lack of one over-arching theory of justice. In view of this predicament, Norheim makes a plea for a deliberative process above the scientific outcome studies. In this deliberative process, citizens' preferences can be discussed and incorporated.

In his chapter, Erik Nord discusses the usefulness of formal outcome data for *priority decisions*. Nord agrees with Norheim that priority setting should be considered a complex deliberative process. However, this raises the question of which role may/should be attributed to numerical indicators of values and the relation of these values to costs. Nord tries to answer this question by drawing up a number of scenarios about the usefulness of summary value indicators for health policy makers. According to Nord, numerical estimates of the values a society places on different health outcomes can

be meaningful for decision-makers, and may facilitate the deliberative process and decision-making. However, it depends on their precision whether they can be usefully employed for deliberations and decisions in the area of priority-setting.

One of the concerns regarding the use of EBM in policy-making is that it may result in a shift of power from the clinician to external authorities, like insurance companies. Managed care is an example of such a system: in managed care, a third party tries to control the use of medical and health care services by influencing the decision-making of the patient as well as that of the service provider. Managed care often joins decisions about the reimbursement of care, and the delivery of care, together in one agency. In her chapter, *Marion Danis* analyses the role of EBM in Managed Care settings with respect to both quality of care and access to care. According to Danis, Managed Care Organisations should develop guidelines that simultaneously steer effective care for their population as a whole, and include permissible strategies for the individual provider to negotiate satisfactory patient-centered care. On the other hand, Managed Care Organisations are in a uniquely strong position to generate evidence. Hence these organisations have a responsibility to generate evidence that promotes population health, attempts to fill the gaps in needed data particularly for understudied segments of the population, incorporates patient utilities, and avoids undue influence from those who are marketing innovations or from those who wish to cut costs without concern for good patient outcomes.

The *final section* of the Volume contains the *Recommendations* of the EviBase Project, which were drawn together at the final conference of the Project in Maastricht. At this conference, policy-makers, clinicians, scientists and other involved persons discussed the best way to promote an ethically sound way of evidence based practice in clinical and policy settings, as well as a balanced approach to research priorities in order to protect the position of weaker groups in our societies. This section is followed by a selected bibliography, based on the extensive bibliography of ethical and related publications on EBM produced within the EviBase project (http://perseus.isi.it/bioeth/bioeth.htm).

EBM by itself can neither improve clinical practice nor provide solutions for allocation decisions. Scientific measures alone cannot and may not guide clinical or policy decisions. These decisions are inherently based on values about the benefit or harm of treatments and about a fair distribution of scarce resources. Besides, it is not only the use of evidence that is subject to values, the production of evidence on outcomes of medical treatments or health policy measures is guided by covert values and decisions about what counts as important values, and which groups should participate in research trials. This Volume wants to contribute to a better understanding of how decisions are being made (and by whom) in clinical practice, medical research and health policy, and how (hidden) values and moral assumptions are guiding these decisions. Knowledge of these assumptions and values is an indispensable precondition for an informed discussion on the status and desirable developments of health care in democratic societies, where the management of information and knowledge becomes increasingly important.

Acknowledgements

The editors wish to thank the participants of the project Ethical Issues of Evidence-based Practice in Medicine and Health Care (EviBase), and the Editors of the journal Health Care Analysis and the Journal of Medical Ethics, for their permission to publish a selected number of (extended versions of) articles published earlier by the EviBase group. They also want to thank the European Commission who sponsored this project within the Fifth Framework Program (Project QLG6-CT-2000-00964).

We specially thank our research assistants who helped us to edit this volume, Jason Sigurdson

and Anna Jakovljevich. Special thanks to our research assistant Christi Nierse of the University of Maastricht who edited all the references and checked the English spelling of all chapters.

References

1. Biller-Andorno N, Karageorgiou A (2001) Health Care and EBM in Germany. http://perseus.isi.it/bioeth/bioeth.htm
2. Meulen RHJ ter, Dickenson DL (2002) Into the hidden world behind evidence based medicine. Health Care Analysis 10: 231-241
3. Biller-Andorno N, Lie R, Meulen RHJ ter (2002) Evidence based medicine as an instrument for rational health policy. Health Care Analysis 10: 261-275
4. Cochrane A (1999) Effectiveness and Efficiency. RSM Press, London
5. Hope T (1995) Editorial: Evidence-Based Medicine and Ethics. Journal of Medical Ethics 21: 259-260
6. Kaasenbrood JA (1997) Evidence based medicine en de dagelijkse psychiatrische praktijk. Kwaliteit & Zorg 2: 50-60
7. Muir Gray JA (1997) Evidence-based Healthcare. How to make Health Policy and Management Decisions. Churchill Livingstone, Edinburgh
8. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. British Medical Journal 312: 71-72
9. Willems D (1995) Ethiek en richtlijnen. Kwaliteit & Zorg 4:157-65

2 Regulating Health Care. The Development of Guidelines in Medical Practice and Health Policy

Christine Sepers · Ruud ter Meulen

Introduction

Decision support techniques play a significant role in today's medical practice. The increase in initiatives in this field leads to guidelines, protocols, standards and algorithms. Althogh this multiplicity of outcomes can be confusing, all these techniques can be seen as a series of instructions that prescribe what should be done in a specific situation. These instructions can be more or less precise, or more or less structured, but they all guide medical workers through a certain sequence of steps (Berg 1996). The goal of this article is to offer a historical view on the role these instructions play in medical practice. To increase readability, the conglomeration of instructions will be referred to as guidelines.

Guidelines until the 1970s: medicine from art to science

Until the end of the 1960s, 'guidelines' or 'protocols' rarely appeared to in medical journals. After the Second World War medicine was confronted with an explosive growth in medical knowledge and diagnostic and therapeutic techniques. The public trust in medicine was unprecedented. Physicians were held in high esteem as keepers of the art of medicine: they used scientific principles and techniques in the care of an individual patient that were implented relative to a patient's unique needs. Problems in medicine were precieved as socio-economic such as the lack of time and money, the shortage of personnel, or obtrusive authorities. Inadequate medical treatments were not attributed to the physician's decision-making, but to external problems, such as time constraints and the limitations of medical technologies. It can be seen as the 'Era of Expansion' (Relman 1988).

The sparse guidelines discussed in the medical literature of the 1950s are therefore not aimed at supporting medical decision-making, but at issues such as the complex social legislation patients in the United States faced. Until the 1960s, the issue was to solve problems external to medicine.

From the end of the 1960s on, the medical profession, in West-European countries as well as in the United States, was increasingly criticised and the medical profession slowly lost its 'mandate' (Starr 1982). It was more explicitly held responsible for its doings, and for the economical, ethical and political consequences of its actions (Rothmann 1991; Schnabel 1988). This process was accompanied by the emancipation of the patient: the patient was no longer an object out of which as much information as possible should be obtained, but a subject with individual wishes, interests and goals (Mol and Van Lieshout 1989). The public demanded legitimisation from the medical profession for the large increase in costs, and the low quality of care. "The physician was knocked off his pedestal" (Klazinga 2001 P 79).

One of the consequences of this changing social position was that medicine was being conceptualised in a fundamentally different way. From an unspecified art or the individualized application of scientific knowledge, medicine was now redefined as a scientific activity. Alvan Feinstein (1967) and

Lawrence Weed (1968) were most explicit in describing practising medicine as a scientific process with a problem definition, description, unravelling, clarification of relations between sub-problems, standardised communication etcetera (Weed 1971). Practising medicine was structured as a scientific experiment- and 'exactly the same methods apply' (Feinstein 1967).

At the same time as the scientific 'upgrading' of medicine, the door was opened to a critical exploration of physician's actions. Where previously external factors were pointed to when problems occurred, physicians could now be held explicitly responsible. It so happened that since the separate components of medical practice were revealed, a measuring rod was created to evaluate the actions of physicians. This way, according to Feinstein and Weed, the inadequate performance of these tasks by medical personnel surfaced. Because medicine is not recognised by the practitioners as the scientific process it actually is, Feinstein (1967) argues, medical practice lacks the scientific qualities of valid evidence, logical analyses, and demonstrable proofs.

Guidelines were one of the techniques that could offer a solution for this newly developed gap between the scientific potencies of medical practice, and the real state of the art. Feinstein especially, made a strong plea for practically adapted, pragmatic, but scientifically well-founded algorithms (Feinstein 1974). At the time, medicine was considered to be a necessarily individual 'art' - guidelines for medical practice were simply unthinkable. In the new concept of medicine, guidelines could take a central position. Guidelines could support the scientific process of medical practice itself. Guidelines, according to Feinstein, offer a description of a logical series of successive steps in medical practice.

The guidelines published in medical literature at the beginning of the 1970s partly lived up to Feinstein's ideals: they standardised medical decision and treatmentroutes, and were founded in the needs and demands of clinical practice. However, they hardly focussed on physicians, only on assistant-physicians, or at the'physician-extenders' popular at that time in the United States (Farmer 1993). Such guidelines enabled specially trained 'health-assistants' to make a first selection 'at the gate', so the scarce time of specialists could be used more efficiently. These protocols were very strict. The physician-extender had to follow the instructions in the protocol step by step. A deviation from the protocol could only be made in consultation with the senior-physician (Greenfield et al. 1974).

Guidelines from the 1980s: three generations of consensus-guidelines

Strongly steering protocols, as described above, were unacceptable to physicians themselves. The guidelines for physicians in the 1970s were local initiatives, mostly directed at co-operation between the growing group of specialists and other health care workers who gathered at the bed of the patient. Information from experienced clinicians, local experience and peer consensus were central. As the social pressure on the medical profession augmented, the pressure on clinicians to make their own actions explicit increased. As Arnold Relman (1988) states in his editorial: "The Era of Assessment and Accountability is dawning." (P 1222)

One of the initiatives to come out of these developments was the rise of nationally organised consensus development. This influential development can be seen as an attempt to sort the eclectic pack of local, subjective and often contradictory guidelines at national level, and to based it in scientific concepts. One of the first American organisations that occupied itself with this was the National Institutes of Health (NIH). In 1977, this organisation introduced the Consensus Development Program, to stimulate the spread of scientific knowledge in medical practice (Everdingen 1988; Jacoby 1988; Kanouse et al. 1989). Since its introduction, the NIH organised consensus meetings in which the members of an expert panel, after a plenary session and an open discussion in a private meeting, assess the argumentation of experts and formulate a final guideline. The consensus method does not only aim at an account of the scientific state-of-the-art, but also makes visible where knowledge is lacking. This way, it becomes clear where scientific research should focus on in the future. This

method of consensus building was adopted by several countries and is partly still in use. In the UK, the King's Fund Centre organized national consensus procedures between 1984 and 1991 (Willems 1998). In the Netherlands the Dutch Institute for Healthcare Improvement CBO organises consensus meetings since 1982. In 2000 this institute had produced a total of 67 guidelines (CBO 2001; Klazinga 1997).

The consensus guidelines are far less detailed than the protocols to support assistant-physicians. They offer a framework, but do not coercively prescribe a certain course of action. This characteristic, as well as the role of the 'consensus', and the procedures through which this 'consensus' was reached, were initially confronted with a lot of criticism. It was argued that there was not enough use of explicit criteria, which leads to an unclear relationship between the recommendations and the quality of the scientific foundation (Woolf 1992). Moreover, the limited time in which a guideline needs to be formulated has the danger of the guideline being based on subjective and idiosyncratic compromise, rather than being a true consensus on the optimal way of treatment (Oliver 1985). In the UK, the King's Fund Centre was forced to stop the national consensus meetings, because funding was becoming increasingly difficult, and the support for, and impact of, the statements were considered disappointing (Willems 1998).

As a response to this criticism, the consensus methods used in the United States and Western Europe are more and more formalised and concentrated on scientific reasoning. This 'second generation' of guidelines is characterised by what was recently called the 'evidence based' approach of guideline development. This method focuses explicitly on the assessment of scientific evidence and is, among others, followed by the Agency for Healthcare Policy and Research (AHCPR). Since it was founded in 1987, this American governmental institution published a large amount of guidelines in which use was made of a combination of an evidence-based approach and the formal consensus method. The Dutch Institute for Healthcare Improvement CBO also adopted an evidence-based approach to develop guidelines. In the UK in 1999, the National Institute for Clinical Excellence (NICE) was founded. Among other things, this institute offers decision support to medical workers (NICE 2001). NICE puts the evaluation of clinical effectiveness firmly at the centre of its policy (Dickenson et al. 2001). In Germany, clinical guidelines by professional – national or international – societies (Leitlinien), are supposed to secure quality and "rationalize" clinical practice by establishing rules for good medical treatment, thus defining the medical standard. The Agency for Quality in Medicine in Cologne (Aerztliche Zentralstelle Qualitaetssicherung), which is a joint institution of the German Medical Association and the Associations of Health Insurances, was commissioned in 1996 to develop a German Guidelines Quality Program, and to serve as the clearinghouse for these guidelines (Biller-Andorno and Karageorgiou 2001).

In the evidence-based approach, the better research designs live up to the ideals of universality and objectivity, the higher they are valued (Kaasenbrood 1997). Randomised clinical trials are seen as most ideal (Sackett et al. 1991). The research concerning the effectivity of treatments is now so extensive, that review articles are sometimes no longer sufficient to get a good insight into the effect of an intervention. If different research outcomes come into conflict, statistical meta-analyses can generate a systematical assessment of the effectivity: in a meta-analysis, a weighed average effect-value is produced for each study separately and combined with other values found (Cuijpers 1996). The Cochrane Collaboration is an international co-operation of scientists that develops, maintains and disseminates systematic, current overviews of randomised effect research of medical treatments (Kleijnen et al. 1995).

The methodological oriented evidence-based approach to guideline development can be supplemented by an extra step of an analysis of the costs of treatment or diagnostic procedures. With the gradual undermining of the professional mandate since the 1970s, the economy is also an increasingly important factor in the medical field (Ashmore et al. 1989). From the mid-nineties on, the efficiency of care also became important. As a consequence, the effects and costs are given consideration explicitly in the formulation of guidelines (Casparie 1998). These guidelines can be con-

sidered to be the 'third generation' of guidelines. From the point of view of cost containment, the development of guidelines in which not only quality and effectiveness play a role, but also the costs of care, is a promising possibility.

Quantifying being right

One consequence of this development towards 'cost-effectiveness based medicine' is that the development of guidelines becomes more and more expensive. The development of a state of the art guideline presupposes exhaustive literature research, meta-analyses, cost-effectiveness studies- however one needs to keep in mind that the guideline should be ready before it is based on outdated information.

Another important aspect is the statistical foundation of guidelines. Statistics are not new to the medical field. Since the 1830s, statistics contributed enormously to the public health measures that can undoubtedly be seen as the most important medical contribution to longevity and general health in the world (Gigerenzer et al. 1989). Nevertheless, one can note an increasing pressure on the statistical foundation of medicine in randomised clinical trials, in meta-analyses, in cost-effectiveness analyses and other forms of decision support. This growing importance of quantifiability cannot be seen as a simple increase in the scientific content of medicine. From the view of the sociology and history of science, the shift to quantifiable information is a way in which a domain of science can legitimate its character (Porter 1995). For Alvan Feinstein, this shift is not a logical consequence of the process of rationalisation he wanted to start. On the contrary: Feinstein considers the emphasis on statistically significant correlations and the use of statistical calculation models to be 'mathematical quantophrenia'. Feinstein (1974) claims that the object of rationalisation is the preservation of the vitality of clinical reasoning and enhancement of its scientific effectiveness. According to Feinstein, this clinical vitality can be threatened by too much emphasis on statistical considerations. The pragmatic considerations that characterise clinical work should outweigh the formal-mathematical considerations that characterise the models of decision theory (Feinstein 1977). Moreover, the 'soft data' in the clinic, such as visual impressions or psychosocial information, are of crucial importance. This kind of data is too 'vague', however, to be incorporated in statistical models or formulas, which carries the danger of dehumanisation of medicine.

Finally, to Feinstein, science is understanding how a relationship functions, for example by explicating pathophysiological mechanism and not by simply accepting that there is a relationship.

Feinstein's position makes clear that 'scientific character' can be defined differently. The shift from a more qualitative notion of objectivity to an explicitly quantitative notion, as took place during the last decennia in medicine, does not stand by itself. The historian of science, Ted Porter (1995), points out that this shift is typical for scientists or professionals who are in a weak social position: in a strong position this demand for numbers is unnecessary. Scientists and professionals who are faced with less critical attention from the public, do not need to prove their objectivity and reliability by formal methods, that in their view are mechanical and missing the essence of their expertise. Where actors are highly vulnerable outsiders numbers grow in importance: they provide an aura of reliability, scientific character and neutrality that supplement the lack of professional or political legitimacy (Porter 1995). Professionals can no longer justify themselves by an appeal to 'character', but need to appeal to method-rules and procedures, substantiated by objective numbers (Abbott 1988).

From co-ordinating action to mental support

It is clear that scientific character is a plural notion. An important shift in the notion of 'scientific medical practice' took place from the 1970s on, under the influence of the cognitive revolution in psychology. From this period on, medical practice was more and more conceptualised as a scientific thinking process. As Feinstein and Weed spoke mostly about medical practice as a collective scientific practice, during the 1980s more and more people described medical practice as an individual activity of logic reasoning. One can no longer trace the steps of the scientific process through the individual actions of different health care workers, but rather in the head of the individual physician.

The increasing importance of this new view on medical practice, created a new perception of the problems of this practice, and again generated a different positioning of guidelines. Compared to statistical and symbolic cognitive models, the physician-brain appeared to have fundamental shortcomings in the area of interpretation, combination and integration of medical information. Physicians, it was now argued, were unable to contain the increasing complexity of medicine. The research into inter-doctor variance in different countries was connected to this image.

Through this perception, guidelines again get another meaning. For a long time they were mainly seen as a regulating link: as a way to make the common task of scientists and clinicians more efficient and more scientific. In contradiction to this collective and process-like function, they are now regarded as a self-evident and necessary addition to the limited capacity of the physician. A guideline is the 'missing link' which enables the physician to optimally perform the decision making process.

The ethical debate on guidelines

As already illustrated in this chapter, the rationalisation of medical practice by means of guidelines and cost-effectiveness analysis has raised a number of professional and ethical concerns. An important part of these concerns has had to do with the alleged gap between abstract measures and guidelines on the one hand, and the situation of the individual patient on the other. The data of guidelines was based on ideal populations, and has little relevance in the encounter between the physician and the patient (Sweeney 1998; Van Weel and Knottnerus 1999). In clinical practice, the patient plays a central role: it is the patient who decides whether a treatment is beneficial or appropriate by assessing his or her own values and beliefs. Such situations are often ambiguous: what are the complaints of the patient, what diagnosis can be made, what therapeutic alternatives are available, what benefits and harms are to be expected for the patient, what harms can the patient bear, what does the patient actually want (Culpepper and Gilbert 1999)?

One of the other criticisms of the rationalisation process has to do with its abstractness. In guidelines and cost-effectiveness analysis, medical work is seen as a series of distinctive moments, in which only a limited set of clear-cut criteria come into play. However, medical practice follows a totally different kind of rationality, in which there are no clearly demarcated steps (Berg 1997). Guidelines and Evidence-based Medicine, in general, can play a positive and supportive role, by offering scientific information about diagnosis and therapy. However, guidelines can only do so by abstracting from the complexities of clinical practice. As a result, they have a tendency to become too formal and lose sight of the complexity of clinical practice, and the medical and social constraints of this practice (Berg et al. 2001).

Rationalising medical care and health care by evidence-based guidelines has further implications beyond improving the quality of care. As already indicated, guidelines are increasingly used to promote the rationalisation of rationing and allocation decisions. An example of this rationalisation process is the effort to sharpen the indications for access to scarce medical treatments by the use of cost-effectiveness analysis. Patients who are expected to benefit in a low degree from particular treat-

ments, such as the elderly and the chronically ill, may be excluded from access to such treatments, particularly when such treatments are expensive. There is a serious concern that the rationalisation of the rationing process will be at the expense of widely shared social values like justice, solidarity, equity and solidarity with the elderly population (Dickenson 1999; Maynard 1997). Access for the elderly will be jeopardised when access to treatment is based only on cost-effectiveness analysis (or cost-utility analysis), since in general, younger patients will have a more positive response to medical treatments (Biller-Andorno and Karageorgiou 2002).

There are more problems with guidelines than the possible exclusion of weaker patient groups. Policy-makers, and other third parties (like insurance companies), are often tempted by guidelines and cost-effectiveness analysis as an ideal way to control clinical practice and health care delivery, and thus to control the costs of health care. Managed care is an example of such a system: in managed care, a third party tries to control the use of medical and health care services by influencing the decision-making of the patient as well as that of the service provider. Managed care often joins decisions about the reimbursement of care and the delivery of care together in one agency. Such a system will have a profound influence on the professional autonomy of the physician and the autonomy of the patient (Berger 1996). Managed care already dominates U.S. health care and raises tremendous ethical problems for health care providers in respect of access to care. In European health care systems that are based on health insurance (like Germany, The Netherlands and France), managed care is expected to gain a strong influence in the coming years. In The Netherlands and Germany, for example, insurance companies will have to bear higher financial risks in the coming years, which is the reason why they are exerting a strong pressure to gain more control over the way in which medical care and health care are delivered.

Professional organisations have rather mixed feelings about the increasing emphasis on evidence-based guidelines, particularly those intended to ration scarce or expensive medical treatments. In Germany, for example, the health care system has undergone major reforms in the last years, which have attempted to respond to the increasing need to allocate limited resources. One example of these ongoing changes is the recent development of a list of medications that will be paid for by insurance companies. However, within the medical community, there is fundamental disagreement on how and to what extent economic considerations can and should influence health care policy and medical practice. The majority of clinicians are very sceptical about these developments, as they fear they may lose their therapeutic freedom. On the other hand, the Royal Dutch Physicians Organisation has conducted a research project into possible ethical problems of evidence-based guidelines in the field of cardiology and psychiatry (Berg et al. 2001). The idea behind this project was that if clinical decisions increasingly come to be guided by evidence on their cost-effectiveness, it is better to have insight into the ethical assumptions and problems behind the use of this kind of evidence. In that case doctors, policy-makers and patients could reach a common ground for the implementation of such a policy. Nonetheless, the fear of losing their professional autonomy to external forces is widespread among doctors.

Guidelines, and particularly evidence-based guidelines, are increasingly expected to guide clinical decision-making and the allocation and rationing process in health care systems all over the world. However, when the values and assumptions behind this process remain hidden and implicit, the rationalisation of medical care and health policy may be a serious challenge to the values of our societies in respect of the quality of delivery of medical care and health care, as well as of access to medical and health care services. This transparency is needed to enable a professional and political debate about the contribution evidence based tools and systems can make to the quality and allocation of health care. Such a debate will help us to accept on a clinical level, as well on a societal level, the positive contribution these rationalising instruments might make.

References

1. Abbott A (1988) The system of professions. An essay on the division of expert labor. University of Chicago Press, Chicago
2. Ashmore M, Mulkay M, Pinch T (1989) Health and Efficiency: A Sociology of Health Economics. Open University Press, Milton Keynes
3. Berg M (1996) Problemen en potenties van het protocol. Medisch Contact 51: 366-370
4. Berg M (1997) Rationalising Medical Work. Decision-Support Techniques and Medical Practices. The MIT Press, Cambridge
5. Berg M, Burg M van den, Meulen, RHJ ter (2001) Guidelines for appropriate care. The importance of empirical normative analysis. Health Care Analysis 9: 77-99
6. Berger JF, Rosner F (1996) The Ethics of Practice Guidelines. Archives Internal Medicine 156: 2051-2056
7. Biller-Andorno N, Karageorgiou A (2001) Health care and EBM in Germany. http://perseus.isi.it/bioeth/projects/evibase-project
8. Biller-Andorno N, Lie R, Meulen RHJ ter (2002) Evidence based medicine as an instrument for rational health policy. Health Care Analysis 10: 261-275
9. Casparie AF (1998) Ontwikkeling en implementatie van kosteneffectieve richtlijnen. Onderzoeksprogramma 'Bevordering Doelmatig Handelen' van start. Medisch Contact 53: 88-90
10. CBO (2001) http://www.cbo.nl/richtlijnen/lijst.htm
11. Culpepper L, Sisk J (1998) The development of Practice Guidelines: A Case Study of Otitis Media with Effusion. In: Boyle PJ (Ed.) Getting Doctors to Listen. Ethics and Outcomes Data in Context. Georgetown University Press, Washington, P 71-85
12. Culpepper L, Gilbert TT (1999) Evidence and ethics. Lancet 353: 829-831
13. Cuijpers P (1996) De effecten van psychologische behandeling van depressie. Tijdschrift voor Psychotherapie 22: 402-415
14. Dickenson DL (1999) Can Medical Criteria Settle Priority-Setting Debates? The Need for Ethical Analysis. Health Care Analysis 7: 131-137
15. Dickenson D, Ashcroft R (2001) Health care and EBM in the UK. http://perseus.isi.it/bioeth/projects/evibase-project
16. Everdingen JJE (1988) Consensusontwikkeling in de geneeskunde. Bohn, Scheltema en Holkema, Utrecht
17. Farmer A (1993) Medical practice guidelines: lessons from the United States. British Medical Journal 307: 313-317
18. Feinstein AR (1967) Clinical Judgment. Krieger Publishing Company, Huntington, New York
19. Feinstein AR (1974) An Analysis of Diagnostic Reasoning. III. The Construction of Clinical Algorithms. Yale Journal of Biology and Medicine 21: 482-496
20. Feinstein AR (1977) Clinical Biostatistics XXXIX. The haze of Bayes, the aerial palaces of decision analysis, and the computerized Ouija board. Clinical Pharmacology and Therapeutics 21: 482-496
21. Gigerenzer G, Swijtink Z, Porter T, Daston L, Beatty J, Krüger L (1989) The Empire of chance. How probability changed science and everyday life. Cambridge University Press, Cambridge
22. Greenfield S, Bragg FE, McCraith DL, Blackburn J (1974) Upper-respiratory Tract Complaint Protocol for Physician Extenders. Archives of Internal Medicine 133: 294-299
23. Jacoby I (1988) Evidence and consensus. Journal of the American Medical Association 259: 3039
24. Kaasenbrood AJA (1997) Evidence based medicine en de dagelijkse praktijk. Kwaliteit & Zorg 5: 52-60
25. Kanouse DE, Winkler JD, Kosecoff J, Berry, SH, Carter GM, Kahan J, McCloskey L, Rogers WH, Monroe Winslow C, Anderson GM, Brodsley L, Fink A, Meredith L, Brook RH (1989) Changing medical practice through technology assessment. An evaluation of the NIH Consensus Development Program. Health Administration Press, Ann Arbor, Michigan
26. Klazinga N (2001) Van algemeen naar bijzonder. Vijf variaties op een misverstand. In: Bossuyt P, Kortenray J (Eds.) Schaatsen op dik ijs. Boom, Amsterdam, P 79-89
27. Klazinga N (1997) Quality Management of Medical Specialist Care in the Netherlands. An Explorative Study of its Nature and Development. Belvédère, Overveen
28. Kleijnen J, Vet HC, Rinkel GJ, Keirse MJ (1995) De Cochrane Collaboration: systematische overzichten van kennis uit gerandomiseerd onderzoek. Nederlands Tijdschrift voor Geneeskunde 139: 1478-1483
29. Maynard A (1997) Evidence-based Medicine: an incomplete method for informing treatment choices. Lancet 349: 126-128
30. Mol A, Van Lieshout P (1989) Ziek is het woord niet. Medicalisering, normalisering en de veranderende

taal van huisartsgeneeskunde en geestelijke gezondheidszorg, 1945-1985. SUN, Nijmegen

31. Oliver MF (1985) Consensus or Nonsensus Conferences on Coronary Heart Disease. Lancet 332: 1087-1089
32. Porter TM (1995) Trust in Numbers. The Pursuit of Objectivity in Science and Public Life. Princeton University Press, Princeton
33. Relman AS (1988) Assessment and Accountability. The New England Journal of Medicine 319: 1220-1222
34. Rothman DJ (1991) Strangers at the bedside: a history of how law and bioethics transformed medical decision making. Basic Books, New York
35. Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) Clinical epidemiology: a basic science for clinical medicine. Little Brown & Co, Boston
36. Schnabel P (1988) De gezondheidszorg: van immuniteit tot publiek domein. In: Kreukels, AJM, Simonis, JBD (Eds.) Publiek Domein. De veranderende balans tussen staat en samenleving. Boom, Meppel, P 172-200
37. Starr P (1982) The social transformation of American medicine. Basic Books, New York
38. Sweeney KG, MacAuley D, Gray DP (1998) Personal significance: the third dimension. Lancet 351: 134-136
39. Weed LL (1968) Medical Records that Guide and Teach. New England Journal of Medicine, 278: 593-600, 652-657
40. Weed LL (1971) Medical Records, Medical Education, and Patient Care. Year Book Medical Publishers, Chicago
41. Weel C van, Knottnerus JA (1999) Evidence based interventions and comprehensive treatment. Lancet 353: 916-918
42. Willems D (1998) Outcomes, Guidelines, and Implementation in France, the Netherlands, and Great Britain. In: Boyle PJ (Ed.) Getting Doctors to Listen. Ethics and Outcomes Data in Context. Georgetown University Press, Washington, P 153-163
43. Woolf SH (1992) Practice Guidelines: A New Reality in Medicine, II, Methods of Developing Guidelines. Archives of Internal Medicine 152: 946-952

3 Evidence-based Medicine and Power Shifts in Health Care Systems[1]

Rein Vos · Rob Houtepen · Klasien Horstman

Introduction

Over the past two decades, Evidence-based Medicine has attracted more and more attention in the medical community. Local, regional, national and international networks of clinical review groups have evolved. One is the Cochrane Collaboration, which produces, maintains and disseminates systematic reviews of the evidence about the prevention and treatment of health problems. The Cochrane Collaboration operates on such an international scale, that it has been conceived of as "an enterprise that rivals the Human Genome Project in its potential implications for modern medicine" (Naylor, derived from Muir Gray 1997 P 223).

What is the reason for considering power shifts in health care systems from the perspective of Evidence-based Medicine, as discussed by Biller-Andorno, Lie and Ter Meulen? A natural response would be to state that Evidence-based Medicine deals with (scientific) evidence and nothing but evidence. The accumulated body of medical evidence is necessary to support diagnostic and medical interventions by physicians. At best, one could say that Evidence-based Medicine is advocated to counterbalance societal power, be it from patients and patient groups, insurance companies, governments, or other parties involved in health care. How could one object to the ambition and endeavour of proponents of Evidence-based Medicine to critically assess scientific evidence to support decisions in medicine and health care? What, if anything, could be wrong with 'doing the right things right', as Muir Gray (1997) formulates the new management agenda of evidence-based health care?

There is, however, an interesting way to question the relationship between Evidence-based Medicine and power shifts in health care systems. Although definitions of EBM are phrased as a scientific approach to medicine (Sackett et al. 1996; Muir Gray 1997; Offringa et al. 2000), EBM is a normative concept. The concept of EBM aims to improve medicine and health care, that is, it encompasses a view on five issues: the best source of medical evidence, the best method to gain medical evidence, the best medical result to be achieved, the best expert to make medical decisions and the social function of medicine? Regarding the last issue, Sackett and his colleagues (1996) argued that EBM is concerned with making evidence-based clinical decisions in the care of individual patients. Other proponents of EBM, however, have extended the range of EBM to evidence-based health care (EBHC): "In the 21st century, the health care decision-maker, that is, anyone who makes decisions about groups of patients or populations, will have to practice evidence-based decision-making" (Muir Gray 1997 P 1).

Within the scope of this paper, it is not possible to analyse the pros and cons of the various definitions and interpretations provided either by proponents or opponents of EBM. Our contention is that both proponents and opponents use normative concepts on the five issues identified above. More particularly, they provide particular views on positions, responsibilities, possibilities,

[1] This chapter is an updated version of an article that was first published in the journal *Health Care Analysis* 10, 2002: 319-328, published by Kluwer Academic Press. Reprint takes place with permission by Kluwer Academic Press.

norms and relationships between professionals, patient groups, governments and other parties in health care and society. From this perspective, we will analyse the role of EBM in modern western societies.

The power concept we want to employ is a Foucauldian one, based on the assumption that power effects are at the very core of the production and distribution of knowledge in the human sciences (Foucault 2000). Foucault has analysed this type of knowledge in terms of discursive systems, including ways of ordering human activities and defining human characteristics, especially subjectivity. Thinking of knowledge and power in terms of discursive systems enables us to focus on the distribution of roles and responsibilities between the various actors involved in the arena of EBM and health care. We amend this Foucauldian line of thinking with a normative perspective from citizenship theory, arguing that the role of EBM is not fixed, but depends on the relationship between state and society. We will first analyse the fundamental change in western societies during the past decades, from modern to post-modern societies. Then, we will present a fourfold model of possible relationships between state and society and discuss how EBM may fit in, by giving some examples of the practice of EBM in different European countries. Lastly, we propose to consider EBM as a public forum where proponents and opponents of EBM discuss diverse and possibly conflicting ways of changing medicine, health care, and health policy.

Power shifts in western democratic societies

Western societies are in a state of great flux. The roles of the state, the market and civil society are shaped and reshaped in new ways. In particular, the idea of regulating the market and society by a powerful nation-state is questioned. In order to appreciate this fundamental transition in western democratic societies, it is worthwhile to use the distinction between two modes of politics, as made by Van Gunsteren (*Figure 3-1*).

Figure 3-1
Two modes of politics (Freely derived from Van Gunsteren 1998 P 36)

MODERN SOCIETY	UNKNOWN SOCIETY
National unitary culture	Cultural diversity
Politics of emancipation	Life style politics
Equality	Difference
Hierarchy	Networks
Rationality	Rationalities
Identity	Identities
Representation guaranteed	Representation problematic
Analysis/Instruction	Variation/Selection
Command	Learning

The modern society of the second half of the 20th century had a typical landscape: "people and their differences could be fairly easily located on a map with coordinates for home, work, education, and religion. Their differences were to a large extent known, and their behaviour as voters or consumers could be fairly well predicted once one knew their position on the map. They would by and large behave as did other people in similar positions. Differences were part of an overarching unity of national culture and state policies" (Van Gunsteren 1998 P 36). The key features of political steering in modern societies are the central and top-down approach, which Van Gunsteren characterised as

analysis and instruction. An analysis is made of the problem situation, pros and cons of different options for acting are considered, the best rational solution is chosen, and subsequently bureaucratic organisations are instructed to execute the chosen policy. Policy makers, administrators and managers, both within governmental institutions and increasingly also in large health care organisations, needed transparent, accountable data and figures to support their policies and managerial decisions.

Within this style of thinking, power is relegated to the realm of the use of knowledge only, the realm of decision makers rather than scientists. This implies a specific view on the relationship between science and politics. Science provides objective, neutral and precise facts as the basis for social and political decision-making. This model of scientists who give the facts, politicians who make the normative decisions, and managers who execute the choices, has been very appealing and convincing. Public services such as health care, public housing and education, had to be uniform in their administration, universal in their provision, and administrated through stable bureaucratic hierarchies and fixed rules (Hirst 1997).

On this basis, modern societies created an extensive array of services in the areas of welfare, housing, education and health care. Modern societies are essentially intervention states and are implicitly steered by the normative view of the 'standard human being' (Van Houten 1999). 'Normal' people are healthy, well-educated, do their work, have careers and retire. If people get sick, unemployed or disordered, they should be normalised.

However, this practice is no longer adequate in post-modern societies. In some cases – fractures, acute diseases, and reparable disabilities – the model of the standard human being in modern society works perfectly well. A person gives up his or her citizenship, becomes a patient, subjects him/herself to the medical regime, is cured, and returns to societal life as citizen. In many other cases – chronic disease, disability or physical and mental handicaps – there is no clear separation between citizenship and sick role. Large groups of chronically ill and disabled people attempt to keep on their jobs and to participate in community and social life. Social exclusion and discrimination frustrate their personal and social lives and their access to jobs, insurance and housing (Kruijff and Schreuder 1999; Gehandicaptenraad 1999). More generally, people live pluralistic lives, they have multiple or mixed identities, and change identities unexpectedly. Many groups of patients currently growing older have different expectations from those of today's elderly. Such a pluralistic society requires a different mode of politics.

In terms of power, Foucault has analysed the advent of a rationalized and standardized style of thinking in the human sciences as a 'normalizing' discursive regime. By stimulating people to orient themselves on opting for a healthy lifestyle in purely rational terms, a certain standardized way of ordering practices could gain a self-evident aura. In contrast, today's societies can be called post-modern societies or, according to Van Gunsteren 'unknown societies': amidst a variety of transient identities, convictions, life styles, alliances, and rationalities of citizens, decision-makers in the centre of a pluralistic society lack the crucial information to take responsible and accountable decisions. The only viable option for rulers in unknown societies is "to accept proliferating and unpredictable plurality and make it an ally. Governing then becomes less a matter of analysis and instruction and more a matter of variety and selection – that is, of accepting variety and coping with it on the basis of selective values" (Van Gunsteren 1998 P 38).

Plurality in society: the triangle between the citizen, the state and the civil society

In response to the transition from modern society to post-modern society, different practices of governance and social communication have developed. Hilhorst (2001) elegantly classified the four

different responses of the past two decades to reorganise the role of the state, using two dimensions (*Table 3-1).*

Table 3-1
Two dimensions of the relationship between state and society (Freely from Hilhorst 2001 P 25)

	Top-down	Bottom-up
Non reflexive	New primacy of politics	Market orientation
	Key role	Key role
	Bureaucracy	Consumer
	Key role EBM	Key role EBM
	Passive, objective facts	Active market party
Reflexive	Managerial government	Responsive government
	Key role	Key role
	Professionals	Citizens
	Key role EBM	Key role EBM
	Technocratic	Normative judgement

The first dimension concerns the principle of political steering, either top-down or bottom-up; the second dimension concerns reflexivity, which is the capacity and competence to learn from faults and errors. This classification generates four models of the state-society relationship, and differentiates the potential role of medical professionals and EBM.

The strategy to regain a *new primacy of politics* applies to the central, top-down role of the government, and tends to formulate social problems as a law-and-order problem: the fabricated laws and regulations are correct, but the problem is supposed to be the improper execution by administrators, managers and bureaucracies. This approach denies that there are serious political and normative dilemmas at local and regional levels. In this approach, stronger government, normative in intention and firm in execution, is the default solution to all problems. Medical professionals are supposed to act as passive suppliers of the facts for governmental decisions on health care delivery and health care rationing. EBM will act as the referee for effective and deliverable health care services.

The *market-oriented* strategy strengthens the position of the citizen as a consumer. It is not governments who decide what kind of health care should be offered, but the market. The problem with this approach is that, it reduces the scope and impact of social and political evaluation of the outcomes of distributive processes: markets tend to increase, yet are insensitive to these inequalities. Medical professionals will be one group of actors in the health care market, and EBM is one of the competing mechanisms to adjust supply and demand of health care services asked for by consumers.

The *managerial state* approach is reflexive, by claiming that politics cannot devise and should not devise uniform standards for diverse and dynamic social practices. Politics can only decide what to do, but professional organisations and institutions such as schools, hospitals and housing corporations have the expertise, skills and information on how public goals should be realised. Organisations and institutions in various sectors of societal life should be free to develop their tasks, but they have to be forced into context-specific learning through benchmarking. This approach, however, assumes that political judgements are technocratic – political goals are quantifiable and judgements can be based on objective data -, whereas underlying and potentially conflicting values are very often the issue. In this approach, medical professionals are supposed to generate criteria

of medical success and relevance, while EBM has to deliver further criteria for benchmarking, e.g. how 'bad hospitals' can learn from 'good hospitals', or 'badly' managed care organisations from 'well' managed care organisations.

The *responsive government* strategy opposes the instrumental and engineering approach of social problems, and therefore attempts to involve social actors in the development of policies. The policy-making process must be interactive, all stakeholders must be involved, and the policy agenda must be open. This strategy has many problems, one of which is that it focuses particularly on the phase of policy-making, or the process part, rather than on the practice of execution and performance, or the outcome part. In this strategy, medical professionals constitute one of the parties who formulate the health care problem and define medical relevance and success. Depending on the problem definition, EBM needs to generate different types of evidence.

There is not much literature on how EBM is implemented in the various European countries, but some tentative remarks are useful to give some content to the presented classification scheme.

In Germany and France, the development of guidelines and EBM seems to be very much influenced by hierarchical and central government policies (Klazinga 2001). These policies fit the strategy of the primacy of politics. In France, a semi-governmental institute has been set up to develop guidelines and, in addition to evidence-based guidelines, 'refences medicale opposable' have also been developed. The latter are guidelines to instruct physicians in primary health care to refrain from the delivery of particular health services. By connecting these references to the financing of health care, the French government was able to make substantial savings (Klazinga 2001 P 81-82). In Germany, self-government ('Selbstverwaltung') is the basic principle of social insurance, which implies that associations of physicians and insurers have to determine, in accordance with German law, the budgets for medication, the payment of physicians, etc. This form of self-regulation, as applied to EBM, seems to be authoritarian, centralistic and top-down (Wigge 2000). The medical community, in both France and Germany, has resisted EBM (Comments workshop). Biller-Andorno and Karageorgiou (2001) report that 'whereas a few physicians greet EBM as a paradigm of new humanism and sincerity in medicine, the community of clinical practitioners remains largely sceptical, fearing their 'therapeutic freedom' might be curtailed.'

In the UK, the situation seems to be quite different. The National Health Service increasingly orientates its priority setting and research commissioning activities around an Evidence-based Medicine agenda (Dickinson and Ashcroft 2000). In addition, regional health authorities implement the development of clinical guidelines, and recently the National Institute of Clinical Excellence (NICE) was set up. Klazinga notes that the practice in the UK should be seen as an attempt to encapsulate medical practice within the framework of government's priority setting, rather than a practice of self-regulation. However, in contrast with France and Germany, the development of EBM in the UK seems to be much more similar to the practice of managerial government. This implies that an important role should be played by professional organisations, and British medical associations have in fact been very sympathetic to EBM. It is no wonder that this policy fits the ideology of EBM proponents in the UK as noted by Dickinson and Ashcroft, namely that "their ideology is largely a simple form of Enlightenment rationalism: following Leibniz and Bacon, the idea is to resolve differences through measurement, calculation and comparison".

In the Netherlands, the development of EBM has a different structure, fitting best the practice of responsive government. Professional medical associations, such as the Dutch General Practitioner Association, have taken the lead to develop EBM (Klazinga 2001). Dutch government has stimulated by law and with other measures, such as the funding of patient organisations, the contribution of patient organisations to the national and regional development of clinical guidelines and EBM (Sepers and Ter Meulen 2001).

The classification scheme has been used to differentiate the various ways in which EBM can be implemented in health care policy practice, albeit in a general way. The four models of governance should be conceived of as ideal types. In reality, mixes of such models can be seen, and different

European countries exhibit different mixes. Even national and regional policies may exhibit different mixes. The Dutch government, for example, has used the market approach in the area of occupational medicine, the care for mentally disabled people and home care, however, the new minister of health is currently known to be a proponent of the managerial state approach (Hilhorst 2001).

The public debate on EBM: neorepublican theory of citizenship as a new ethical approach

No single model of organising the incorporation of EBM is fully satisfactory. Being reflective of power effects implies that one cannot straightforwardly and consistently opt for one particular approach. Furthermore, a Foucauldian power concept is incompatible with straightforward attempts to predict specific power effects. It does, however, guide our focus to the question of whether a self-evident split between power and knowledge is assumed, or whether a variety of inclusive relationships between knowledge and power is acknowledged.

From a normative perspective, however, the strategy of the new primacy of politics is conservatively holding to the analysis-instruction practice of modern societies. It relegates EBM to the realm of pure knowledge and implies a rather voluntaristic view of power, to be originated from this one restricted and highly visible political area. The other three approaches, on the other hand, at least force EBM to acknowledge that epistemology is inherently related to the ordering of the public realm. They are innovative attempts to implement the model of variety and selection, yet each of them is beset with difficulties and problems. Progress towards a solution of these problems by any of these policy strategies requires the incorporation of the perspective of citizens and their social networks, professionals with practical and tacit knowledge, and diverse public views on what counts as a 'good life', which is not the same as a 'normal life'. It remains to be seen how EBM is able to cope with the diversity of health care practices in post-modern societies. Inasmuch as EBM is expected to be practically relevant, it ought to be tied to, rather than separated from, the normative world of emancipated patients and diverse health care practices. From the perspective of the neorepublican citizenship theory, medical-professional groups and health care organisations have to be considered as 'citizenship-practices', that is, as practices that shape the public debate and take responsibility for the creation of links with the public. It is the task of the government to stimulate plurality. One central authority no longer governs post-modern societies, and therefore plurality is needed. Instead, there are several, partly overlapping organisational networks with their own perspectives, language and rationale. These organisations should participate in the public debate, and explicate the various value orientations and normative assumptions of the development of EBM. The citizenship theory implies that public interference should not be discussed in terms of 'guidance' but in terms of 'learning with an open agenda'. Proponents and opponents of EBM ought to be prepared to defend the normative claims that are inherently tied to any presentation of evidence. They should do so with the awareness that, in practice, there are different defensible ways of dealing with evidence claims. If both parties participate actively in the ensuing deliberative and decision-making practices, such as those on guideline development and the reimbursement of treatments, the organisation of plurality will hopefully contribute to an open development of diverse power regimes, rather than a single shared or enforced view of the way health care should be organised.

This would at least imply that EBM could not stay entirely separated from questions concerning the relationship between power, and the production and distribution of scientific knowledge of human persons. Whether consciously or not, the introduction of EBM affects the way we think of our bodies, health and disease, and thereby it affects the way we order the practices involved. In this respect, advocates of EBM typically adhere to the specifically 'modern' discourse of either separating knowledge from power, or assuming that both fall under one closed system of rationality. The virtue of a 'post-modern' citizenship theory, on the other hand, is that it does not attempt to supplant one

specific power/knowledge regime with another. The consistent emphasis on plurality and citizenship does not deny power/knowledge associations, but delegates a responsibility for a multiplicity of these associations to a multiplicity of actors involved. The issue is neither sidestepped, nor is there any pretension of theoretical or practical 'solution'. It is straightforwardly put on the agenda for each and all, and thus opens the discursive regime of EBM into a multiple array of citizen's responsibilities.

References

1. Biller-Andorno N, Karageorgiou A (2001) Country Report on Evidence-Based Medicine in Germany, EVIBASE. Maastricht
2. Biller-Andorno N, Lie R, Meulen R ter (2001) Evidence-based Medicine as an instrument for rational health policy. Discussion paper EVIBASE. Goettingen
3. Comments by Prof. D. Dickinson, Prof. M. Danis, Prof. R. ter Meulen. (2002) Workshop EVIBASE. University of Bergen
4. Dickenson D, Ashcroft R (2001) Country report on Evidence-Based Medicine in the United Kingdom, EVIBASE. Maastricht
5. Foucault M (2000) Power: Essential Works of Foucault, 1954-1984, Volume III. (Trans. Faubion JD et al.) New Press, New York
6. Gehandicaptenraad (1999) Strategisch beleidsplan Gehandicaptenraad. Gehandicaptenraad, Ede
7. Gunsteren HR van (1998) A theory of citizenship: organizing plurality in contemporary democracies. Westview Press, Boulder
8. Hilhorst P (2001) De wraak van de publieke zaak. Uitgeverij de Balie, Amsterdam
9. Hirst P (1997) From statism to pluralism. UCL Press Ltd, London
10. Houten D van (1999) De standaardmens voorbij – over zorg, verzorgingstaat en burgerschap. Elsevier/De Tijdstroom, Maarssen
11. Klazinga N (2001) Van algemeen naar bijzonder – vijf variaties op een misverstand. In: Bossuyt P, Kortenray J (Eds.) Schaatsen op dik ijs – Evidence-based Medicine in de praktijk. Boom, Amsterdam, P 79-89
12. Kruijff AF, Schreuder RF (1999) Chronisch zieken en scenario's genetica. Rapport STG. Zoetermeer
13. Muir Gray JA (1997) Evidence-based health care – how to make health policy and management decisions. Churchill Livingstone, Edinburgh
14. Offringa M, Assendelft WJJ, Scholten RJPM (2000) Inleiding in Evidence-based Medicine. Bohn Stafleu Van Loghum, Houten
15. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence-based Medicine: what it is and what it isn't. BMJ 312: 71–72
16. Sepers C, Meulen R ter (2002) Ethical aspects of EBM and allocation in the Netherlands: selected issues from expert interviews. Workshop EVIBASE. University of Bergen
17. Wigge P (2000) Evidenz-basierte Richtlinien und Leitlinien – Qualitätssicherungs- oder Steuerungsinstrumente in der KGV? MedR 18: 574-585

4 Beyond Legitimate Science: The Case for Policy-Related Science[1]

Mariachiara Tallacchini

Introduction

Though philosophy and sociology of science have argued for the non-neutral character of scientific knowledge and the social connotation of the scientific community, and have changed the perspective on scientific knowledge from validity to credibility, until recently, legal theorists and political scientists have looked at science as providing neutrality and objectivity, which appeared irremediably lacking in political and legal systems. Moreover, the enduring conception of the "republic of science" has shown science as an outstanding example of democracy (Ezrahi 1990; Merton 1968; Polanyi 1962).

However, impasses have emerged in the legal regulation of science. In fact, dealing with scientific knowledge, law has to fill a number of cognitive gaps, since scientific data prove uncertain, insufficient or value-laden, and the question arises as to which methods law should use to evaluate and validate "good" science. Increasingly, the relationship between law and science appears as a multifaceted, complex one (Smith and Wynne 1989).

Evidence-based Medicine (EBM) has been introduced as offering a unique chance for science and justice to meet, as far as it harmonises the often-divergent aims of objectivity, equity and patient preferences. The authors' position towards this easy view of EBM is quite critical (industry-driven research, medical paternalism, etc), though they acknowledge that deeper awareness in implementing EBM could improve both professional practice and respect for patients.

I shall briefly explore this problem from a different perspective, namely the role that science and experts play in the courts. In fact, while the legal vision of science is undergoing a dramatic paradigm shift, both in the US and in Europe, legal reliance on scientific standards and methods is endowed with ambiguities.

Referring to well-grounded, widely accepted scientific standards and methods, may result in a more self-reflexive attitude towards the mix of scientific and legal facts-and-values that merge in concrete decisions – namely, it may give rise to a reduction of scientific and legal arbitrariness, and to substantially more legitimate legal decisions.

In contrast, referring to scientific standards as a shortcut to legitimising opaque legal choices means treating science and law as powers for manipulating people, and not as open and transparent forms of knowledge and agreements on values in democratic societies – namely, betraying science and law by black-boxing them.

In this perspective, I would also argue for what has been called the democratisation of science (extended scientific expertise and public participation in science-based decisions).

[1] This chapter is an updated version of an article that was first published in the *Journal Health Care Analysis 10*, 2002:329-337, published by Kluwer Academic Press. Reprint takes place with permission by Kluwer Academic Press.

From a naïve to a critical legal view of science – Some leading cases

Differences exist between the American and European governance of science – the former being more science-based, rigorously and objectively informed by scientific knowledge (Jasanoff 1990, 1995; Jones et al. 1996), the latter being more preventative and in many ways precautionary (Commission of the EC 2000a; Shepherd 2000). Despite this, the view of science as traditionally informing legal decisions, both at the legislative level and at judicial levels, has undergone dramatic changes on both sides of the Atlantic Ocean in the last few decades, shifting from a more positivist image of scientific knowledge to a more critical position towards science (Jasanoff 1995; Freeman and Reece 1998).

Roughly speaking, the positivist stance conceives law as being neutral towards science, and considers science as basically certain knowledge. The Italian *Di Bella* case (1998) is a good example of this attitude: both science and the law are treated as black boxes, as a separate and secret form of knowledge that only experts may legitimately use.

A different, fairly recent legal trend – apparently influenced by the latest philosophy and sociology of science – has contested the idea that law should simply acknowledge scientific opinions and defer to science as the ultimate word on science-based social decisions. Indeed, this represents an important attempt to open the box of science and fully acknowledge that scientific knowledge is socially embedded.

As it is, the shift from the former to the latter paradigm has not yet been fully completed, and a variable oscillating relation exists between the system of science and the system of law. As I shall try to argue, this oscillation is due to the underlying fight between the power of law and the power of science, both aiming to present themselves as "speaking the truth" (Wildavsky 1987), and therefore deserving the ultimate word on social decisions.

The best example of this shift can be found in the US judicial system, where the traditional legal attitude of deference to science has been abandoned for a more critical account of the links between science and law. The *Frye* (1923) and *Daubert* (1993) cases illustrate this passage.

Two other leading cases, the UK *Bolam* (1957) and *Bolitho* (1998) cases, witness a very similar shift in the scientific-legal paradigm also taking place in Europe, though it is worth noting that this has happened first in a common law country, where the judge-made law system – compared with statutory systems – makes the legal approach to science more flexible and responsive (Jasanoff 1995).

The Di Bella case

An Italian leading decision, delivered in 1998, the *Di Bella* case, is an outstanding example of how the ideology of reciprocal neutrality between science and law can be used for black-boxing – or obscuring – the deeper reasons of scientifically based legal decisions. I discuss this judicial case only as an example of the wrong way for science and law to validate themselves.

The *Di Bella* case concerned a trial on the validity of an alternative therapeutical method based on somatostatin, and the alleged constitutional right to "freedom of therapy". In 1998, the Italian physician Di Bella discovered a new method for treating cancer, even in patients with a terminal disease condition. A group of Di Bella's patients claimed, before a court, for the right to free, autonomous choice of a therapy beyond official medicine -, as a broader interpretation of the constitutional right to health (Art. 32 Cost.). According to this interpretation, the right to health involves an obligation for the State to provide for free whatever therapy citizens ask for. The lower court acknowledged the right to free therapy, but the Constitutional Court reversed the decision, denying both the therapeutical effect of somatostatin and an unconditional right to health.

What is interesting in this context is how the two different courts argued for and against the

right, shifting from the scientific to the legal domain, adopting law to validate science and applying science to interpret law.

According to the lower court, since a right to health is granted, it refers both to official and to alternative medicine. Legal certainty supplies scientific uncertainty. In the opinion of the Constitutional Court, the right to health exists as far as it is supported by mainstream science (Pretura di Milano Ordinanza n°247/98, 26 gennaio 1998). The very content of a constitutional right is deferred to scientific expertise. Individuals have a right to receive free medical treatment (paid by the State) as far as a scientifically accepted therapy exists, which has been formally approved by governmental scientific committees. What is "therapeutical" has to be "objectively" defined by mainstream science, and is not open to "subjective" interpretation by lay people (Corte Cost, 185/1998, 26 May 1998; Legge 8 April 1998, n.94).

When the responsibility of determining to what extent a right is a right is deferred to science, reference to neutral and objective standards becomes a tool for black-boxing the social dimension of scientific and legal choices.

From Frye to Daubert and beyond

The most traditional relationship between science and the law was defined in *Frye v. United States* (54 App. D.C. 46; 293 F. 1013; 1923 U.S.), a case dating back to 1923. Mr. Frye, convicted for the crime of murder, asked the District of Columbia Court of Appeals to admit evidence that was based on a forerunner of the modern lie detector test.

The Court refused, on the basis that the lie detector test (based on blood pressure) had not yet gained such standing and scientific recognition, as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made.

Thus, the so-called *Frye* standard emerged as the rule for admitting scientific expertise in difficult cases, when a novelty in science was involved:

- "When the question involved does not lie within the range of common experience or common knowledge, but requires special experience or special knowledge, then the opinions of witnesses skilled in that particular science, art, or trade to which the question relates are admissible in evidence. (…) Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs".

Here, the existence of uncertainty in science is acknowledged, but only in order to defer it to the judgement of experts belonging to mainstream science. Scientific uncertainty seems to exist only as a matter of discussion among scientists and not as a reason for broader social involvement.

In this perspective, scientific knowledge exists as far as it undergoes the "general acceptance test", which does not consist in objectivity, capacity of producing data and of quantifying probability (Porter 1995), but in the self-relying statement of the scientific community.

More than half a century later, in 1993, with *Daubert v. Merrell Dow Pharmaceuticals* (509 U.S. 579 (1993)), the U.S. Supreme Court abandoned the well-established *Frye* standard (Bernstein 2001).

The case concerned the alleged teratogenic effects of Bendectin – a prescription against morning sickness – that had been proved by defendants as devoid of statistical evidence, as many peer-reviewed publications established. Petitioners – minor children with severe birth defects represented by their parents – responded with the testimony of eight experts, who provided (an unpublished)

"re-analysis" of previously published epidemiological studies. Reversing the opinion of the lower court, who declared inadmissible an expertise which did not meet the Frye standard, the Supreme Court decided that the consolidated criterion of the scientific community's general acceptance constituted only one of the possible elements that qualify hypotheses as scientific, and allowed the dissenting scientists to testify.

- "(T)he 'general acceptance' test of Frye v. United States (…) is not a necessary precondition to the admissibility of scientific evidence (…). (…) a federal trial judge must insure that any and all scientific testimony or evidence is not only relevant but reliable; and in a federal case involving scientific evidence, evidentiary reliability is based on scientific validity".

The Court argued that judges are free to allow anyone to testify as an expert who, albeit without the official acknowledgement of the scientific community, demonstrates possession of scientific knowledge and methods (arguing, that hypotheses are falsifiable and susceptible to testing).

The new criteria invoked by the Supreme Court are multiple and heterogeneous in nature, but at first glance they all seem to aim at a more objective and precise definition of the reasons for scientific method to be rigorous. Eventually, they are all shaped in order to allocate the ultimate power to validate science to judges.

Firstly – the Court wrote – Rule 702, governing expert testimony, places no limits on the admissibility of purportedly scientific evidence. Secondly, an essential element in determining whether a theory or technique is scientific knowledge is the testability of hypotheses, which have to be falsifiable. Thirdly, another pertinent consideration, is whether the theory or technique has been subjected to peer review and publication, though "publication (which is but one element of peer review) is not a sine qua non of admissibility; it does not necessarily correlate with reliability". Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, and the existence and maintenance of standards controlling the technique's operation. Finally, "general acceptance" can yet have a bearing on the inquiry. The Court notes, however, a "reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community".

Moreover, in 1999, the US Supreme Court added a new chapter to Daubert's criteria, in *Kuhmo Tire Co. v. Carmichael (*526 U.S. 137 (1999)). The Supreme Court ruling in *Carmichael* established the reliability criteria of Daubert will apply to all types of expert testimony, even non-technical testimony primarily based upon experience and training. If science can be codified in well-established methods, this knowledge is open to everyone, first of all to the courts: testability and falsifiability are universal methods in order to assess rigorous knowledge. The codification of scientific criteria, elaborated by the Supreme Court, is the legal mean for judges to deliberately shape their own autonomous, undeclared space for free decision.

From Bolam to Bolitho

Bolam v. Friern Hospital Management Committee (1 W.R.L. 582) was delivered in the UK in 1957. The plaintiff, who underwent electro-convulsive therapy without assuming a relaxant drug, suffered a fractured hip as a consequence. The experts were divided in describing the case as one of professional negligence and in stating that the patient had a right to receive information about the potential risk. The court's decision shaped what became famous as the Bolam test. This consists in requiring, as far as professional knowledge is involved, an ordinary level – and not the highest level – of competence. But what can be properly considered as ordinary skill?

The question was not univocally addressed and further interpretation of the Bolam test ap-

peared quite problematic. In fact, according to Brazier and Miola, *Bolam* has been interpreted "to allow judgment by colleagues to substitute for judgment by the courts" (Brazier and Miola 2000). Other professionals can define what represents ordinary skill, but this means that physicians become judges in their own cause.

Bolitho v. City & Hackney Health Authority (A.C. 232, H.L.), decided in 1998 by the House of Lords, has been seen as a correction to *Bolam*, though opinions are not unanimous.

Patrick Bolitho, aged two years old, suffered brain damage after a cardiac arrest for not having been intubated at the proper moment, after two episodes of breathing difficulties. Here the case was primarily about scientific causation, that is – could intubation have prevented the cardiac arrest? Hence it was only indirectly about medical negligence.

Again, scientific expertise was uncertain, because two opposed views were presented, both representing a responsible body of professional opinion.

In writing the opinion, the Court abandoned the *Bolam* test, assuming that judges are not bound to experts' knowledge, though acknowledging that usually distinguished experts' opinion is synonymous with responsible behaviour,

- "(nonetheless) if, in a rare case, it can be demonstrated that the professional opinion is not capable of withstanding logical analysis, the judge is entitled to hold that the body of opinion is not reasonable or responsible".

Doctors, the Court argued, cannot be self-referential, and, if scientific methods and logical analysis are open tools that anyone can apply, judges are legitimated to decide on the topic.

Co-production between science and law, and medical responsibility

Though the UK decisions deal with medical negligence, and only indirectly question what is sound science and who is entitled to legally validate scientific knowledge, the analogies between *Frye-Daubert* on the one hand, and *Bolam-Bolitho* on the other hand, are strong.

They are similar in their way of reasoning and in ultimately focusing on the relation between the power of scientists and of the courts.

The old legal-scientific paradigm expressed in *Frye and Bolam* treats science as a black box – a secret power and not a shared knowledge – that only scientists are allowed to disclose. The new paradigm, represented by *Daubert* and *Bolitho*, makes an attempt to open science to lay people by declaring its methods, and empowers judges – and through them civil society – to disagree with scientists.

The particular thesis argued in *Daubert* has been widely criticised both because it is poorly and heterogeneously articulated (Farrel 1994), and because it does not cease in supporting peer review as a valid mechanism, adding to existing criteria the discretionary power of judges.

Bolitho has met with very similar criticism, because it is controversial whether its effect will actually be that of improving patients' conditions.

Nonetheless, as Jasanoff observes commenting on *Daubert* (Jasanoff 1995), the role claimed by courts is appreciable and shows how a potential for self-reflection exists. Courts may represent deliberative arenas in which all interests have the possibility to emerge, because no privileged language, authority or group pre-exists. In using science, judicial systems – though with a rough generalisation which would need to take into account differences existing among legal cultures – have evolved toward a deeper awareness about co-production mechanisms between science and law.

What is still lacking is a more explicit self-criticism towards the law itself, its assumptions, choices and values, which need to be openly declared and discussed and not concealed through a misleading reference to the objectivity of science. The role of gatekeepers claimed by courts may

acquire the meaning that society as a whole has to make the final decisions about the evolution of technoscience.

In the complex interplay of the social spread of science and existing legal rules, scientific-legal knowledge appears as an original, hybrid form of knowledge, which establishes itself through reciprocal production and modification: a situation that has been called, according to Sheila Jasanoff, "co-production" (Jasanoff 1995; 1996).

Understanding science and law as co-production of knowledge implies a commitment to making the whole decision process more coherent with democratic rules and values.

Lack of public trust in mainstream science cannot simply be labelled as irrational, but depends on a more general mistrust towards both scientific and political institutions (Tallacchini 2002b). In this respect, the *Di Bella* case is not an example of the demagogic implications of public participation to scientific decisions, but shows how mainstream science has been too distant from civil society and how the need exists for more (and not less) participation and dialogue.

Reaching and involving citizens both as interest groups (e.g. in the formation of guidelines) and individuals (e.g. in the physician/patient-citizen relationship) represents, at the same time, the challenge and the chance for implementing/improving EBM. In this way, reference to scientific standards in law may become a call for responsibility and not a missed social responsibility grounded in an irresponsible science.

Toward a more democratic relationship between science and society

A common argument is animating present reflections on the relationship between science and society, in particular with respect to environmental and biotechnological issues. This is the conviction that the last word in choosing science policy rests, in a democratic society, with the public. No expert judgement may replace wider public discussion and the clarification of the concrete choices of values involved.

This argument underlies numerous studies that openly speak of a new agora (Nowotny et al. 2001), of hybrid institutions (composed of experts and members of the public) (Irwin and Wynne 1996), and of policy-related science (Funtowicz et al. 2000; Shepherd 2000). The general meaning of these expressions, which tend to recognise the new role of the public in the biotechnological age, has been summarised by Sheila Jasanoff in the idea of a civic epistemology.

The changes which have taken place in the relationship between science and society are, in fact, greatly influencing institutional set-ups, and the set of rights which refer to the notion of social contract and, in particular, to the idea of a State under the rule of law. The wealth of guarantees that come within the definition of State under the rule of law has, so far, scarcely touched the specific guarantees as regards the knowledge-power of science. The new ways of governing science need to be integrated with the (still desirable) notion of the state under the rule of law (Tallacchini 2002a). Today, the concept of democracy itself requires re-examination in the light of greater involvement on the part of the public in technoscience issues. As I said, the scientific community has long represented itself as a community of peers, open to dissent and devoid of external influences. Faced with ever more frequent situations of scientific uncertainty, as in the relationship between science and the market, not only has this political-juridical ideal of the scientific community lost credibility and the need to give voice to the minority fringe of science is widely augured, but above all this model is no longer a reference point for social institutions.

The ethical need to make the relationship between science and society more democratic therefore takes two paths. Firstly, there is a need to widen consultation with scientists (Gerold and Liberatore 2001), where there are differences of opinion as to the possibility of potentially harmful events occurring. Secondly, there must be greater involvement on the part of the public in decision-making on scientific issues that directly touch civil society.

The governance of science in Europe: policy-related science

Over the last few years, many of the questions linked to the concept of scientific knowledge have attracted the attention of European institutions (Commission of the EC 2000b). Emergencies resulting from inadequate or ineffective legal regulation (e.g. in the field of food safety) have certainly shaken European countries at their foundations.

The phenomenon that Europe is now experiencing and seeking to remedy is a crisis among citizens and institutions. This crisis has grown particularly evident in the lack of confidence in the capacity of the institutions to regulate science, and the reliability of experts to provide sound knowledge. Recurring events in which experts have appeared incapable of taking control of situations of scientific uncertainty, and in which errors of evaluation and interests in conflict with the safety and health of citizens have sometimes emerged, have made the problem of public distrust towards science crucial, whenever science is directly involved in public policy decisions.

Nonetheless, the innovative character of the European reflection on the epistemology that has to inspire the governance of science, is not only a pragmatic response to the political need to create sufficiently homogeneous and standardized decision-making processes in scientific and technological questions characterized by high uncertainty. Beyond this, it is possible to note the theoretical effort to elaborate an epistemological position in which it is possible to recognize science policy in Europe.

What is now taking shape is the foundational aspect that the interpretation of the relations between science and law may acquire in the construction of the European identity. Many European countries and EU institutions are elaborating a model of governance of science incorporating a conception of science, institutions, society and law to regulate these relations.

The term governance alludes to a system of government that actively pursues, among other things, the concrete involvement of citizens to make up for the lack of democracy.

The new epistemology that European institutions are seeking to build and implement, substantially joins the need for a more extended expertise and participatory democracy.

Policy-related science must be conceptually distinct and have different aims from both pure and applied science. Pure science is mostly guided by the researcher's curiosity, whereas applied science is oriented by a project and looks for particular practical outcomes. Policy-related science, on the other hand, has to help define questions that, in so far as they have to be applied socially, are tied to broad evaluations and demand a political choice – even when they present themselves as scientific-technical problems.

Policy-related science is still waiting to have an adequate epistemological statute. The definition of this statute certainly entails a hybridization of scientific knowledge and political-legal choices. The tangible interplay of the scientific evaluation of a problem, the weight to be attributed to uncertainty, the contribution of a multiplicity of players in both science and social evaluation, and, finally, the translation of all these elements into legal-political decisions and norms – this is the challenge, and also the intellectual appeal, of the construction of new epistemological-normative concepts.

As far as the determination of the meaning and scope of the scientific contribution in policy-related science is concerned, a pluralistic scientific expertise is needed. In the case of risk assessment, this extends to all experts in the field. Pluralistic expertise is necessary for a variety of reasons. As it has been noted:

- "Pluralistic scientific expertise is, therefore, needed essentially for three reasons. First of all, it is required in order to make scientific decision-making more responsive. To be clear, it is not a question of holding scientific rationality hostage to irrational fears and unfounded concerns. Rather, it is a matter of reconnecting science and society as a means of coping with such fears and concerns. Secondly, such a pluralistic approach is needed in order to help transcend the boundaries of segmented scientific expertise. (…) The aim, therefore, is to improve communication between disciplines. Whether between such

disciplines or between science and society, the third reason that pluralistic scientific expertise is needed is precisely to encourage the systematic exposure of unspoken or even unexamined assumptions and uncertainties underlying both expert and lay opinion" (De Schutter et al. 2001).

It appears advisable not only to review the concept of expert opinion, but also to rethink the notion of "expert" itself, which once referred to a narrow conception including only insiders in scientifically defined disciplines, but now extends to a multiplicity of differentiated forms of knowledge and practices. It is also necessary for experts to openly acknowledge situations of scientific uncertainty, even where this may create a situation of inconclusiveness that makes decisions more difficult. As Shepherd has observed,

- "in the provision of information for the policy purposes, science must simply do its best (…). Policymaking cannot wait for conclusive facts, in science-related fields any more than in others; and part of the art of policy is to make do with inadequate, confused and contradictory information. (…) Instead of providing 'scientific facts', the objective of the extended peer review process is to deliver uncertainties, error-costs, and also the social and ethical dimensions" (Shepherd 2000).

The way in which scientific knowledge, thus predisposed, has to be translated into a legal-political choice is what connects work on expertise to the need to elaborate new decision-making processes for science-based decisions. Attempts to reach civil society in a more diffused way involve, for example, trying to go beyond the simple and now limiting identification of stakeholders as official representatives of public interests. As far as healthcare is concerned, the Cochrane Collaboration published an interesting study focussing on involvement of the public in science-based decisions:

- "(A)n understanding of the perspectives of the people relying on health care services is vital to balance the judgments of those sharing the research and provider cultures. Specialists' views of the world are by definition unbalanced and incomplete. Professionalism creates distances from the rest of the population, and vests professionals with a variety of interests which may or may not be consistent with public interests. Processes which acknowledge this bias, and seek to redress it through public participation are necessary to help build a relevant body of work that is guided by the realities of people's lives" (Bastian 1994; Coulter 1999).

These proposals for the legal regulation of science have been defined by some authors as a "cognitive proceduralisation of law", namely "the putting in place of mechanisms which enable learning processes to be generated at collective level to manage the uncertainty linked to contexts of bounded rationality" (De Munck and Lenoble 2001).

This legal and cognitive process tends to open and set up an institutional discussion space, in which scientific knowledge may find forms of social stabilisation that are more critically and democratically screened and always open for review. Dissemination of knowledge, no longer isolated in a single social component – the scientific community – but ascribed to many different players, and no longer univocally conceived as the sole form of knowledge, but disaggregated and re-aggregated in different epistemic cultures of different importance, redistributes the needs for credibility and accountability among all the parties involved and in their reciprocal relations.

The theoretical merging of "philosophy of science" and "philosophy of science policy" aimed at more democratic public policy-making, shows a change not only in the contract between science and society, but also in the very bases of the "social contract".

References

1. Bastian H (2000) Allies or enemies? Evidence-based Medicine and consumer choice. MJA 172: 5-6
2. Bernstein DE (2001) Frye, Frye, Again: The Past, Present, and Future of the General Acceptance Test. George Mason Law & Economics Research Paper No. 01-07. http://ssrn.com/abstract=262034
3. Brazier M, Miola J (2000) Bye-Bye Bolam: A Medical Litigation Revolution? Medical Law Review 8: 85-114
4. Commission of the European Communities (2000a) Communication from the Commission on the precautionary principle. http://europa.eu.int/comm/dgs/health_consumer/library/pub/pub07_en.pdf
5. Commission of the European Communities (2000b) Governance and Expertise.http://governance.jrc.it/
6. Commission of the European Communities (2001) European Governance. A White Paper. http://europa.eu.int/eur-lex/en/com/cnc/2001/com2001_0428en01.pdf
7. Coulter A (1999) Paternalism or partnership? Patients have grown up-and there's no going back. BMJ 1319: 719-20
8. De Munck J, Lenoble J (2001) Transformations in the art of governance. A genealogical and historical examination of changes in the governance of democratic societies. In: De Schutter O, Lebessis N, Paterson J (Eds.) Governance in the European Union. Office for Official Publications of the European Communities, Luxembourg, P 29-51
9. De Schutter O, Lebessis N, Paterson J (Eds.) (2001) Governance in the European Union. Office for Official Publications of the European Communities, Luxembourg
10. Ezrahi Y (1990) The Descent of Icarus. Harvard University Press, Cambridge Mass
11. Farrel MG (1994) Daubert v. Merrell Dow Pharmaceuticals Inc.: Epistemology and Legal Process. Cardozo Law Review 15: 2183-2217
12. Freeman M, Reece H (Eds.) (1998) Science in Court. Aldershot, Dartmouth
13. Funtowicz S, Shepherd I, Winkinson D, Ravetz J (2000) Science and Governance in the European Union: a contribution to the debate. Science and Public Policy 27: 327-336
14. Gerold R, Liberatore A (2001) Report of the working group "Democratising expertise and establishing scientific reference systems". http://europa.eu.int/comm/governance/areas/group2/report_en.pdf
15. Irwin A, Wynne B (Eds.) (1996) Misunderstanding science? The public reconstruction of science and technology. Cambridge University Press, Cambridge
16. Jasanoff S (1990) The Fifth Branch. Science Advisers as Policymakers. Harvard University Press, Cambridge Mass
17. Jasanoff S (1995) Science at the Bar. Law, Science, and Technology in America. Harvard University Press, Cambridge Mass
18. Jasanoff S (1996) Is Science Socially Constructed – And Can It Still Inform Public Policy? Science and Engineering Ethics 2: 263-276
19. Jones M, Guston DH, Branscomb LM (1996) Informed Legislature. Coping with Science in a Democracy. CSIA Harvard University, Cambridge Mass
20. Lebessis N, Paterson J (2001) Recent developments in institutional and administrative reform. In: De Schutter O, Lebessis N, Paterson J (Eds.) Governance in the European Union. Office for Official Publications of the European Communities, Luxembourg, P 259-305
21. Merton RK (1968) Science and Democratic Social Structure. In: Merton RK (Ed.) Social Theory and Social Structure. Free Press, New York, P 604-615
22. Nowotny H, Scott P, Michael G (2001) Rethinking Science: Knowledge and the Public in an Age of Uncertainty. Polity Press, Cambridge
23. Polanyi M (1962) The Republic of Science. Minerva I: 54-73
24. Porter T (1995) Trust in Numbers. Princeton University Press, Princeton NJ
25. Shepherd I (Ed.) (2000) Science and Governance in the European Union. A Contribution to the Debate. http://governance.jrc.it/scandg-eur.pdf
26. Smith R, Wynne B (Eds.) (1989) Expert Evidence. Interpreting Science in the Law. Routledge, London
27. Tallacchini M (2002a) The Epistemic State. The Legal Regulation of Science. In: Mazzoni CM (Ed.) Ethics and Law in Biological Research. Kluwer, Dordrecht, P 79-96
28. Tallacchini M (2002b) Epistemology of the European Identity. The Journal of Biolaw & Business, Supplement Series Bioethix: 60-66
29. Wildavsky A (1987) Speaking Truth To Power. The Art and Craft of Policy Analysis. Transaction, New Brunswick

5 Teaching Evidence-based Medicine

Franz Porzsolt · Heike Leonhardt-Huober

The goals of Evidence-based Medicine

The primary goal of Evidence-based Medicine (EBM) is to apply the best available evidence to the care of individual patients. This process requires decisions, and EBM provides tools to support physicians in making these decisions. In order to make well-balanced decisions, the physician has to match her/his "internal evidence" with the patient's preferences and the best available "external evidence".

Since individual physicians, the scientific community, health insurance companies, and patients have different perspectives, a different knowledge base, different expectations, and different opinions about the best solution for a particular health problem, it is rather difficult in most situations to find a common denominator that satisfies all of the involved partners. EBM can help address conflicts and ultimately satisfy the needs of a range of actors.

In this contribution, we describe the six steps that are recommended in the practice of EBM, and demonstrate the bi-directional relationship between EBM and ethics in medicine. Using examples from day-to-day practice, we show that EBM needs a strong foundation in ethics if it is to make an effective use of the six-step procedure. In addition, evidence is provided to support the suggestion that ethics needs EBM to substantiate the basic assumptions of ethical considerations and recommendations.

The concept of Evidence-based Medicine

The basic concept of EBM proposes to base medical decisions on a synthesis of "internal" and "external evidence". Instead of using the traditional way of making clinical decisions, i.e. to derive clinical decisions only from "internal evidence" (with its three components: knowledge acquired at the university, experience acquired during daily practice, and information derived from an individual doctor-patient relationship), external evidence should be incorporated in the process of making clinical decisions. In this model, external evidence is newly accessible information from an external source. It is the explicit use of valid external evidence (e.g. in case of treatment questions, evidence mainly from randomised controlled trials) combined with the prevailing internal evidence that defines a clinical decision as "evidence-based". If valid and important external evidence is available, its integration into the decision-making process will lead to either confirmation of the existing internal evidence or to the synthesis of new internal evidence, which is then the new basis for the pending decision. In this sense, internal evidence is a dynamic concept that may change each time new external evidence is considered. In order to realize this concept in day-to-day clinical practice, the Evidence-based Medicine Working Group proposed a 5-step strategy: 1. Formulating the question, 2. Searching for information, 3. Appraising the (external) evidence found, 4. Integrating the external

evidence into one's own experience and the values of the patient, 5. Evaluating the process (Sackett et al. 2000).

In teaching this 5-step approach we encountered several difficulties. We, and others, noticed a growing reluctance on the part of students to accept this strategy the further they advanced in their medical training (Porzsolt and Sellenthin 2000; Porzsolt et al. 2003). In the presence of well-established methods of treatment or diagnosis, this resistance arises even more, regardless of the level of training. We assume that this barrier is associated with the process of socialization into the medical profession. Socialization means that students are virtually "trained" throughout education to make decisions under conditions of uncertainty. It is recognized that each of the reputable medical schools has its own teaching profile. The opinion and beliefs of individual physicians are usually strongly associated with the teaching profile of the medical school from which they graduated. There is increasing evidence that physicians who are convinced of a particular treatment strategy, and communicate this conviction to the patient – and thus provide hope and a feeling of assurance for the patient – achieve better treatment results than physicians without these attitudes and skills. To address this psychological principle, which can be considered as a member of the "placebo family", we used the term "knowledge framing" (Porzsolt 2004). We consider the concepts, but not the effects, of "knowledge framing" and "placebo" to be different, and this may be expressed in four assumptions. First, a placebo effect is considered to be an "as-if-therapy" effect, while the effect of knowledge framing is accepted as one of several components in the overall effect of a health care intervention. Second, a placebo is not thought of as a specific physical effect; while knowledge framing is considered a specific effect of the information provided. There are indeed data that support the view that placebo effects – not the placebo itself – are organ-specific (Meissner 2000), which is consistent with our concept of knowledge framing. Third, the placebo effect is thought to lie below the threshold of standard therapy, while the effect of knowledge framing is assumed to lie above that threshold. Fourth, the use of placebos is limited to clinical trials, whereas knowledge framing is not; knowledge framing is part of any doctor-patient encounter.

The advanced student, and more so the clinician, lose, to a certain degree, the ability to differentiate between scientific evidence and what seems to be evident. We recently organized an intensive training seminar at the University of Jena, Germany, where 140 medical students participated in a weekend course on "The principles of Evidence-based Medicine" (Porzsolt and Strauss 2002). At the end of this course, the students concluded that the sessions were very informative and helpful to develop skills for critically appraising not only the content of scientific literature, but also the validity of information that was offered in their lectures and seminars. They were afraid, however, to loose their confidence in medicine when continuing this education in critically appraising the information provided. In an editorial, we addressed the importance of including the critical appraisal as a fundamental part of Evidence-based Medicine. Based on our reading of the students' evaluations, it was considered impossible for the individual physician to trust in a particular treatment, as the attending doctor must, and to critically analyse the scientific basis of the same treatment, which is essential for the clinical researcher.

If we intend to implement evidence-based clinical practice more efficiently in the education of medical doctors, we need to modify the way students and clinicians learn to make decisions. To accomplish this, an additional step was introduced into the original EBM teaching program resulting in the following six steps. The original five steps, plus the additional second step, are shown in ❯ *Table 5-1*. It can frequently be observed, that physicians who change from strictly clinical work, to the side of critical appraisal and evaluation of the scientific basis of clinical decisions, lose their interest and ability to continue the clinical work. As a possible consequence of this phenomenon we suggested two branches of medical socialization, which start from a common clinical background. Physicians of the clinical branch complete their clinical attitudes, skills, and knowledge and acquire confidence in information that is carefully selected and taught at their medical schools. Physicians of the scientific branch acquire a solid clinical background – which is essential for the later inte-

Table 5-1
The six steps of evidence-based decision-making

Step	Action	Explanation
1	Transformation of the clinical problem into 3- or 4-part question	a) relevant patient characteristics and problem(s), b) leading intervention, c) alternative intervention, d) clinical outcomes or goals.
2	Additional step: Answer to the question based on „internal evidence" only	Internal evidence: acquired knowledge through professional training and experience (in general and applied to the patient) and evidence as integrated at that stage. Should be documented before proceeding to step 3.
3	Finding „external evidence" to answer the question.	External evidence: obtained from textbooks, journals, databases, experts etc. The value of the external evidence will be highly variable, see step 4.
4	Critical appraisal of the external evidence	Should answer 3 questions: 1) Are the results valid? 2) Are the results clinically important? 3) Do the results apply to my patient? (or is my patient so different from those in the trial that the results do not apply?)
5	Integrating external and internal evidence	The 2 sources of information (external and internal) may be supportive, non-supportive or conflicting. How the decision is made when non-supportive or conflicting will depend on multiple factors.
6	Evaluation of decision-making process	Once the decision has been made, the process and the outcome are considered and opportunities for improvement are identified.

gration of scientific data – and acquire, in the second part of their socialization, the attitude skills and knowledge to critically appraise medical information. Both groups of professional doctors are needed, one for the selection of adequate scientific contents and the other for applying the selected contents to the patients (Porzsolt and Strauss 2002). Both groups have to communicate. We learned that medicine needs specialists for internal medicine, paediatrics and surgery. Later on, anaesthesiology was separated from surgery. While scientific progress may require the continuation of this process of specialization, dialogue is essential.

The process starts with the selection of a clinical problem, e.g. a therapeutic problem, which has to be transformed into a structured four-part question. These parts describe the relevant patient characteristics and problem(s), the intended goal, the main intervention and finally the alternative intervention(s). It is essential in this first step to develop a definite concept about the intended goal. Therefore, we request that our students define the dimension that will be used to assess the attainment of the intended goal. Without definition of the dimension of goal attainment, it will be difficult to complete the subsequent steps.

The second step requires an answer to the four-part question based on "internal evidence" only. By documenting the internal evidence explicitly, the individual experience and knowledge of the student and physician is respected. They feel that they are integrated into the process of decision-making (Porzsolt and Sellenthin 2000; Porzsolt et al. 2003). The integration of the individual view of the physician – who ultimately has to make a decision – is essential for the process of decision-making for several reasons. Since the physician will be responsible and liable for the final decision, it is hard to believe that she/he will not base her/his decision on "internal evidence". In addition, there is increasing evidence that the confidence of the physician and patient in the effectiveness of the therapy influences the outcome of treatment.

We recently revised the interpretation of our systematic review (Porzsolt et al.), which suggested that the extended survival in renal cell cancer was related to immunotherapy (Coppin et al. 2004).

The stringent application of epidemiological criteria, lead to the conclusion that the extension of survival originally ascribed to a pharmacological effect of interferon treatment, may indeed have been caused by a psychological effect associated with the application of "immunotherapy". We prefer the term "knowledge framing", instead of "placebo effect", for describing this non-pharmacological effect. We consider this effect to be an essential component of any doctor-patient encounter. This component induces hope and provides a positive outlook, which most desperate patients expect to receive from their physician.

As a third step, the best available external evidence has to be found to answer the four-part question. We would like to emphasise that although ideally we would hope to find a systematic review or high quality randomised trials as best available external evidence, in 80% of all clinical situations we will have to rely on data which are considered weak evidence from an epidemiological point of view, for example, non-randomised or uncontrolled trials, case reports or textbooks of variable quality.

The fourth step, the critical appraisal of the external evidence identified, includes finding answers to three questions: Are the results valid? Are the results clinically important? Do the results apply to my patient, (Or is my patient so different from those in any trial found that the results do not apply)? This step, the assessment of the validity of scientific data (considering a predefined range of methodological issues), which is the most time consuming step, is usually not formally completed in day-to-day practice. When bypassing this step, data of questionable validity may be accepted as a basis for clinically important decisions. The risk of flawed decisions underlines the importance of the fourth step.

The fifth step is the most difficult one in EBM. In this step, the external evidence that has been identified and critically appraised has to be integrated into the existing internal evidence, followed by a definite decision. Agreement between the patient's preference, internal evidence, and external evidence varies. Completing the full process could result in finding evidence that confirms the internal evidence, validating and strengthening the physician's confidence in the decision. The process could also reveal that little evidence exists to support the decision, or that the available evidence is equivocal. In such cases other factors, such as cost or inconvenience to the patient, may need to be given greater consideration. Possibly, the best external evidence found is not in agreement with the internal evidence. This represents a particularly valuable experience for the physician, because it may avoid an ill-advised decision being made. It also shows the fallibility of making decisions on uncertain grounds based on internal evidence alone. This in turn will hopefully promote the routine assimilation of external evidence in clinical decision-making, which should include ethical, medical, economic and legal aspects.

The final and sixth step refers to the evaluation of the decision-making process (i.e. steps one to five), and finding ways to improve the process next time.

The relationship of Evidence-based Medicine to medical ethics

Practical experience with the concept of EBM suggests a bi-directional relationship between EBM and ethics: the application of EBM to day-to-day practice needs ethics, and the application of medical ethics to day-to-day practice needs EBM.

One aspect of ethics is to accept people's values and opinions and to keep an open mind. The application of EBM constantly requires the acceptance and critical appraisal of others' opinions, values and beliefs. Scientists, who are used to evaluating pros and cons, will have fewer problems accepting the principles of EBM than people who represent distinct views, like medical opinion leaders. These people are expected to know answers to open questions. As they usually meet this expectation, it is more difficult to disseminate EBM among those individuals than to teach EBM to people who are not opinion leaders. These people will be more likely to support the dissemination of EBM than the former. This aspect of the problem of acceptance is called the consumer's barrier.

At the same time, there is a producer's barrier that occurs when EBM is defended too dogmatically. If nothing is valid unless it rests on solid evidence, EBM will run into problems of acceptance. In our own experience, problems of acceptance occur more frequently the more involved our students are in patient care, and the further they are removed from research. We assume that this problem of acceptance is not a specific problem of EBM, but rather a general problem of communication. It can be explained by the different points of view of two professionals. It is a common experience that the same topic will be evaluated differently from different professional positions, e.g. the position of the researcher without clinical experience and the position of the clinician without research experience. The difference in their evaluations is explained by differences in their socialization. In the process of socialization, the individuals acquire their points of reference, which they need for later evaluations. The acceptance problem may therefore be explained by using different reference points, which were acquired during specific professional socialization.

In the field of quality of life research, it is accepted that considerable differences are observed between self-rated and proxy-rated assessments, explained by different reference points. In the case of self-assessment, the individual compares his or her own expected and observed quality of life. In the case of proxy-rating, the expected quality of life of the assessed person is unknown, and assumptions have to be made for comparison with the observed quality of life. Therefore, it can be expected that self- and proxy-rating of quality of life will not produce the same results.

We have learnt from this experience that conscious and formal integration of "internal evidence" into the process of decision-making facilitates the acceptance and dissemination of EBM.

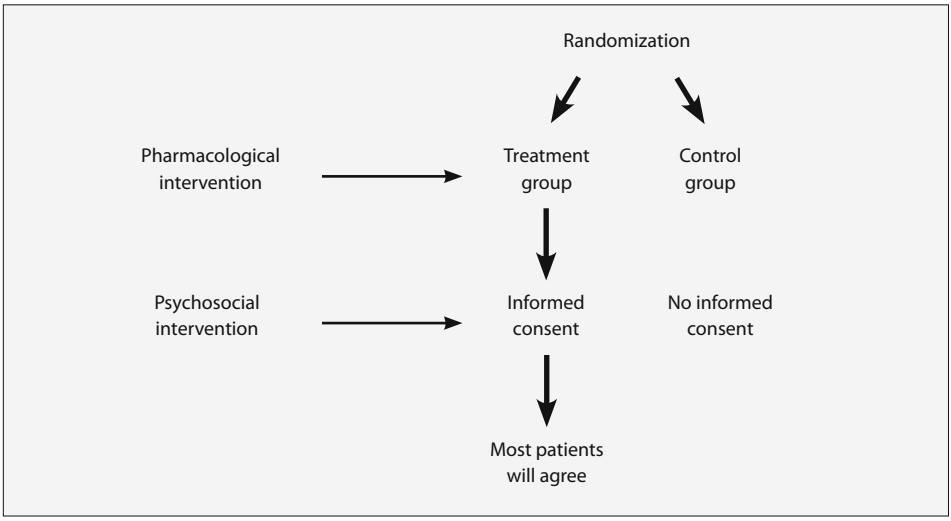
A similar problem of acceptance occurs when publishing results that contradict "main stream thinking". An example of this is the revision of our systematic review on immunotherapy for renal cell cancer mentioned previously. Our first interpretation of the data suggested that the observed prolongation of life was caused by a pharmacological effect, i.e. interferon therapy. It was much easier to convince our peers to accept this interpretation of the data for publication than the second interpretation of the same data, which was more critical than the first version, and suggested that the observed prolongation of life may not have been caused by a pharmacological effect of interferon, but rather by a psychological effect of the information ("knowledge framing") associated with any dose and any type of a promising therapy, e.g. immunotherapy (Porzsolt et al.; Coppin et al. 2004).

A further example of an acceptance problem is related to clinical studies with a non-treatment arm. This study design is occasionally used to test a new treatment in a clinical situation for which so far no treatment had been available. The ethical requirement of providing complete and appropriate information before obtaining informed consent would make it extremely difficult, if not impossible, to recruit patients to a study in which only half of the patients are treated. In our opinion, studies with a non-treatment arm cannot be carried out using current rules of good clinical practice, since patients have to be informed about a promising therapy that subsequently is not given to 50% of the patients included in the study. Experience with this study design confirms that many patients refuse to participate in such a trial.

To solve this problem of patient recruitment, Zelen suggested in his "post-randomisation, informed-consent design" (Zelen 1979) randomisation of patients before informed consent, and subsequently only inform patients who were randomised to the treatment arm. Patients randomised to the non-treatment arm were not informed about their inclusion in the clinical trial. This design was used to study adjuvant therapy using the monoclonal antibody Panorex® in patients with colorectal carcinoma (Riethmueller et al. 1994) (*Figure 5-1*).

The ethical aspects of this design, which induced some controversy, are related to the lack of information given to patients of the control arm who did not receive the adjuvant therapy. These patients did not know that they were part of an experimental trial. They were included in the trial without obtaining an informed consent. Due to the lack of information, they could not refuse their participation in this trial. We address two additional epidemiological problems that are caused by differences in the patients' management (apart from the experimental treatment), which in a well-

Figure 5-1
Zelen II design



designed trial should be identical in both study arms. First, only patients who knew that they were included in a clinical study could refuse their participation in the study (and patients agreeing to take part in a study may be different from those who do not). An asymmetrical distribution of patients in the two study arms will impair the quality of the study. Secondly, patients who knew that they were receiving a new therapy that appeared promising in pre-clinical trials also received psychological intervention in the form of the informed consent. This information may induce in the informed patients similar effects as discussed in the renal cell cancer patients who received immunotherapy together with the appropriate information.

As a result, an unusually large and statistically significant difference in survival between the groups of informed / treated and not informed / not treated patients was observed. This difference could have been caused by a pharmacological effect and/or a psychological effect. The possibility cannot be excluded, that the psychological effect was caused by being informed about a promising therapy, and by the associated induction of hope and a perspective for life. As the untreated group received neither the pharmacological nor the psychological intervention (information), it is impossible to determine which of the components of the intervention was responsible for the difference in mortality observed. Contrary to expectation, no difference in mortality could be shown in a second experiment. In this experiment, both groups had received chemotherapy and one of the groups had received additional immunotherapy. The essential difference between the first and second experiment may be the information (informed consent). In the first experiment, in which the difference in survival was found, only one group of patients was informed. In the second experiment, in which no difference in survival was seen, both groups of patients were informed.

These examples support the assumption that EBM needs ethics. In this context, we want to address the need to realize that our points of view, and the derived assessment of values, are the product of our socialization. We have to accept differences in the socialization and differences that are the product of different socializations. Different socializations introduce different reference points – the researcher accepts the result of an experiment and the clinician believes in her/his observations. There are many situations in which the results of experiments and of clinical observations do not agree, i.e. different reference points will lead to different results of assessments. Ethics can support EBM to accept different opinions that are generated by different evaluations and reference points.

Evidence has been provided to support the possible influence of informing the patient on the observed outcomes of treatment, and problems with concepts and results that are not in agreement with "main stream thinking" have been addressed. These examples indicate an acceptance problem that might be resolved if the principles of medical ethics were integrated into the formal process of medical decision-making. Eugen Bleuler recognized the main problem 80 years ago when he described "the autistic-undisciplined way of thinking in medicine" (Bleuler 1919). His books seem to be resistant to aging.

The second part of the bi-directional relationship assumes that ethical decisions need EBM. The mammography controversies among scientists are useful examples to demonstrate the importance of understanding the epidemiological controversy behind the ethical discussion. Kaplan recently summarized both topics. (Kaplan 2003).

In January 1997, the U.S. National Institutes of Health convened a panel to make recommendations about the use of mammography screening for women between the ages of 40 and 50 years. The panel's review shocked the American Cancer Society by concluding that the benefit of screening was estimated to be marginal, and that every woman should decide for herself whether to be screened or not. The panel suggested that women consider the costs, and the possibilities of being frightened by benign tumours, before deciding whether or not they should have a mammogram.

Richard Krausner, the Director of the National Cancer Institute, decided to disregard the report of the expert panel. He announced that he would ask another panel appointed by the President to look at the same question. The American Cancer Society called the panel's conclusion outrageous. The controversy regarding mammography screening applies in a similar manner to other malignant diseases. The advantages of screening are in no case unequivocal enough to allow a clear recommendation. Ethical considerations that rely on epidemiologically unambiguous results will, however, look different from considerations that are merely based on inconclusive evidence.

This example demonstrates different aspects of dealing with information. It must be clarified whether more benefit or more harm is produced by gathering and passing on information. The criteria for distinguishing between useful and harmful information should be discussed, and it is unclear whether the same criteria should be applied to passing on useful and harmful information. Doctors who provide information to patients should be aware of the immense power of their communicative intervention. When considering this powerful aspect of the provided (or withheld) information, the type and quality of the doctor-patient communication deserves more than the attention presently paid. The reaction of scientists in the case of mammography screening demonstrates the need for discussion that can be derived from of society's uncertainty in dealing with information.

Consequences

These examples clearly demonstrate the considerable difficulties that have to be managed in the process of medical decision-making. Scientific evidence is an important base for this decision, but cannot be the only source of information. All of us depend on previously accepted information that formed our individual "internal evidence". The validity of this previous information cannot be checked retrospectively. Whether correct or incorrect, this internal evidence is the basis for each new process of decision-making, which compares the existing internal evidence with the new external evidence. The effect of new external evidence on the formation of new internal evidence will, therefore, strongly depend on the pre-existing internal evidence.

The pre-existing internal evidence may be based on many different information and influential factors such as public opinion, established standards, and generally accepted procedures. The PSA (Prostate Specific Antigen) screening for early detection of prostate cancer is a frequently quoted example. Although extensive health technology reports from eight different countries concluded that the disadvantages of screening are not balanced by advantages of early detection, PSA screen-

ing is an accepted standard procedure in the health care systems of most industrialized countries. This example demonstrates that the pre-existing expectation of patients and doctors makes it almost impossible to change established "standards". Almost all partners of the health care system are convinced that the previously formed internal evidence was based on valid data. One can imagine how convincing new "external evidence" has to be in order to change the existing internal evidence. The "convincing power" of self-experienced internal evidence is much stronger than the influence of external evidence adopted by reading. The emotional impact of information gained through experience makes a far deeper imprint than the rational power of "external evidence".

Finally, new research strategies have to be developed to quantify the influence of the provided information on treatment outcomes. According to ethical standards, informed consent is an essential part of clinical experiments. Exceptions have to be made when the type of information provided to the patients is investigated. According to present standards, it will not be possible to randomise patients without obtaining informed consent. However, it will be impossible to inform patients about a study design that randomly allocates the patients to groups which get the same physical treatment but different information. Possible exceptions from a mandatory informed consent have been discussed (Tenery 1999).

Conclusions

EBM cannot be understood if the principles of ethics are not accepted and applied. If EBM is not understood, it cannot be applied, even if its application is demanded. It is time consuming to identify claims that supposedly, but do not actually, rely on scientific evidence. It is easy to assess whether ethical principles like transparency, patient-centered decisions and descriptions instead of valuations are applied in medicine. The close connection between ethics and EBM therefore suggests that there must be doubts about the application of EBM when the principles of ethics are not observed.

We have shown that ethical considerations require a solid epidemiological base. Valid data allow consequences that are different from those based on uncertain data. The assessment of the validity of scientific data is therefore essential for any ethical discussion.

So far, the two areas of ethics and EBM have been regarded as separate and independent of each other, both in medical education and in daily practice. The considerations presented here could support discussions aiming to create a closer link between these two scientific areas.

The reality in medicine shows that the vast majority (~ 80%) of decisions have to be made in uncertainty, i.e. in the absence of valid data. This does not mean that the attitude, skills and knowledge of Evidence-based Medicine can be applied to only 20% of our decisions. It means that in most cases the external evidence can only have a limited influence in the process of decision-making. The lack of formal evidence is a challenge to medical ethics. The conflict between limited economical conditions, and the increasing demand for health care service, is generating problems of resource allocation. Checking the validity of data cannot solve these problems. Ethical decisions will gain importance the less valid data are available for decision-making. This aspect should be included in a student's teaching program of Evidence-based Medicine. Otherwise the students will either reject Evidence-based Medicine or lose their confidence in clinical medicine.

References

1. Bleuler E (1919) Das autistisch-undisziplinäre Denken in der Medizin und seine Überwindung. Springer, Heidelberg
2. Coppin C, Porzsolt F, Kumpf J, Coldman A, Wilt T (2004) Immunotherapy for advanced renal cell cancer (Cochrane Review). In: The Cochrane Library. John Wiley & Sons, Chichester
3. Kaplan RM (2003) Zuviel des "Guten". In: Porzsolt F,

Williams AR, Kaplan RM (Eds.) Klinische Ökonomik. Effektivität und Effizienz von Gesundheitsleistungen. Ecomed, Landberg, P 45-49

4. Meissner K (2000) Gibt es organspezifische Placeboeffekte? Placeboeffekte anphysiologischen Parametern und ihre mögliche Vermittlung über kortikale-Organrepräsentationen. Shaker, Aachen
5. Porzsolt F, Sellenthin C (2000) Der sechste Schritt in der Anwendung der Evidence-based Medicine. Z Aerztl Fortbild Qual Sich 94: 619-620
6. Porzsolt F, Strauss B (2002) EBM – der Konflikt ist lösbar. Deutsches Ärzteblatt 99: 761
7. Porzsolt F, Ohletz A, Gardner D, Ruatti H, Meier H, Schlotz-Gorton N, Schrott L. Evidence – based decision making. The six step approach. Evidence-based Medicine. Accepted for publication
8. Porzsolt F, Schlotz-Gorton N, Biller-Andorno N, Thim A, Meissner K, Roeckl-Wiedmann I, Herzberger B, Ziegler R, Gaus W, Pöppel E (2004) Applying Evidence to Support Ethical Decisions: Is the Placebo really Powerless? Science and Engineering Ethics 10: 119-132
9. Porzsolt F, Kumpf J, Coppin C, Pöppel E. Stringent application of epidemiologic criteria changes the interpretation of the effects of immunotherapy in advanced renal cell cancer. (unpublished paper)
10. Riethmueller G, Schneider-Gaedicke E, Schlimok G, Schmiegel W, Raab R, Hoeffken K et al. (1994) Randomised Trial of Monoclonal Antibody for Adjuvant Therapy of Resected Dukes' C Colorectal Carcinoma. Lancet 343:1177-1183
11. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB (2000) Evidence-based Medicine: How to Practice and Teach. Churchill Livingstone, Toronto
12. Tenery RM (1999) Is informed consent always necessary for randomised, controlled trials? New England Journal of Medicine 341: 450
13. Zelen M (1979) A new design for Randomised Clinical Trials. New England Journal of Medicine 300:1242-1245

6 Evidence-based Medicine and Clinical Guidelines in Cardiology. Promoting Science, Practice, or Bureaucracy?

Frank Praetorius

Evidence-based Medicine (EBM), a movement which came into existence one decade ago, aims at advancing the quality of both medical science and medical practice – two dimensions of the profession, each with their own problems in the ethical field. As for science, the German Scientific Organisation *(Deutsche Forschungsgemeinschaft, DFG)* supports research in EBM, with the aim of improving so-called patient-oriented research, mainly in the form of Randomised Controlled Studies (RCT) (Deutsche Forschungsgemeinschaft 1999). This kind of research is a "step-child" (Raspe 2001) of the prevailing academic structures, which is the reason why the DFG also calls for steps to reduce structural deficits in medical academies, and for a stop to widespread *pro forma research*.

In clinical practice, the development and implementation of evidence-based guidelines are the predominant tasks. Here, separating versus reconciling the scientific and economic rationalities are the major issues (Davidoff et al. 2001; Schulman et al. 2002). In general, EBM tends to generate an increase in medical professionalization by means of making progress in scientific rationality (Vogd 2002). The underlying intention is to transform the results of Clinical Epidemiology in order to make better use of research evidence in clinical practice (Perleth and Raspe 1998). On the other hand, when science overwhelms practice with a large number of partly contrary guidelines, it may result in schematic and bureaucratic thinking by the individual physician, and thus lead to deprofessionalization by overregularisation (Praetorius and Sahm 2001). Cardiology, as the discipline with presumably the greatest number of guidelines, may provide an example when studying these developments, especially because trials in cardiovascular diseases (CVD) have a stronger tradition of independent input by the steering, as well as the data monitoring, committees than trials investigating most other diseases (DeMets and Califf 2002).

Evidence-based cardiology

During the past two decades, the progress of therapeutic techniques in CVD was exceptionally great, more than in the diagnostic sector. Many RCTs and several mega-studies (with more than 10000 patients each) were performed (for actual discussion see (DeMets and Califf 2002). Before creating guidelines, a meta-analysis of several clinical studies usually took place (basic problems of meta-analyses are not discussed in this paper). There is a tendency towards the internationalisation of guidelines, mostly formulated on a national basis, but with similar contents and sometimes-identical graphic displays (Von Schacky 2002). In Germany, possible legal implications may develop from the new *guideline clearing procedure* (1999), which is going to substitute the former consensus method (Dierks 2003).

Providers of guidelines

It is impossible to count all cardiologic guidelines worldwide, but one may reasonably estimate that there are more than 1000 guidelines operating in the discipline today. Accordingly clinics and practices are forced to select the appropriate ones under their own criteria. The actual list of guidelines from the 1st Medical Clinic in Offenbach/Germany (Dir. H. Klepzig; http://www.klinikum-offenbach.de/med1/) may serve as an example.

The German Society of Cardiology (DGK) is listing 38 so-called guidelines, with different statements, comments, reports, and 11 genuine guidelines. The German „Ärztliche Zentralstelle Qualitätssicherung" gives the following list of international providers of guidelines (› *Table 6-1)*:

Table 6-1
International providers of guidelines in cardiology (8 January, 2004); http://www.leitlinienanbieter/fachspezifisch/

ACC (American College of Cardiology)	USA	English
AHA (American Heart Association)	USA	English
American Healthways	USA	English
BCS (British Cardiac Society)	Great Britain	English
Canadian Cardiovascular Society	Canada	English
Cardiac Society of Australia and New Zealand	Australia	English
Deutsche Gesellschaft für Kardiologie	Germany	German
European Society of Cardiology	Europe	English
Heart Failure Society of America	USA	English
Heartfoundation of Australia	Australia	English
Heartfoundation of New Zealand	New Zealand	English
Nederlandse Vereniging voor Cardiologie	Netherlands	England
NHLBI Cardiovascular Information (The National Heart, Lung, and Blood Institute)	USA	English
NHLBI Clinical Guidelines	USA	English
Schweizerische Gesellschaft für Kardiologie	Schweiz	Germ., French
Société Française de Cardiologie	France	French
Österreichische Kardiologische Gesellschaft	Austria	German

The quality of guidelines

The American College of Cardiology (ACC) and the American Heart Association (AHA), have jointly developed a system of standards that was adopted (and modified) by many international cardiologists (› *Table 6-2)*. Two dimensions are considered: The levels of evidence (Grade A, B, C) and a classification of recommendations (Class I, II, III).

Table 6-2
American College of Cardiology/American Heart Association classification of recommendations and levels of evidence

Classification of Recommendations	Levels of Evidence
Class I Conditions for which there is evidence and/or general agreement that a given procedure is useful and effective.	**Level of Evidence A** Data derived from multiple randomised clinical trials.
Class II Conditions for which there is conflicting evidence and/or divergence of opinion about the usefulness/efficacy of a procedure or treatment.	**Level of Evidence B** Data derived from a single randomised trial or nonrandomized studies.
Class IIa Weight of evidence/opinion is in favor of usefulness/efficacy.	
Class IIb Usefulness/efficacy is less well established by evidence/opinion.	**Level of Evidence C** Consensus opinion of experts
Class III Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective, and in some cases may be harmful.	

A recent statement and recommendation from the AHA and the Centres for Disease Control and Prevention (CDC) may serve as an example (Pearson et al. 2003). The panel seeks to consider best available evidence for an association between inflammatory markers, especially high-sensitivity C-reactive protein (hs-CRP) and cardiovascular disease. The writing group not only used the AHA/ACC standards of Table 6-2, but also defined criteria for inference of causality in examination of the evidence. Namely the strength, temporality, dose-response relationship, biological plausibility, and consistency of the evidence were reviewed. All participants, including institutes and sponsors, were listed.

It is worth reading the AHA/CDC recommendations (access see (Pearson et al. 2003)), including a critical statement of hs-CRP in comparison with other parameters of inflammation, and of the restriction on precisely defined groups of patients and situations. Actually no indication was found for aggressive therapies depending on hs-CRP (e.g. interventions in the coagulation system). The statement also includes an ethical inference not further pronounced: Evidence-based therapeutic decisions should never depend on a laboratory test, whose therapeutic benefits are not evidence-based so far – although it might be most interesting in the context of current pathophysiological thinking. In cardiology, it is very rare to introduce laboratory tests or devices as a new method of clinical diagnosis and observation in order to control introduced as well as future therapies by means of EBM.

Different countries

The appropriate application of diagnostic coronary angiography in its different indications was never tested by means of EBM, but was nevertheless enthroned as a gold standard. Instead of RCTs, there is only some consensus opinion of experts, which is located in the periphery of scientific perception (Perleth and Raspe 1998), and perhaps may have a Grade C level of evidence, and a classification of Class IIb. The problem became apparent in the recently published critique of the large

number of diagnostic coronary angiographies in Germany (the largest in Europe), and the following discussion in LANCET 2002 (Dissmann and De Ridder 2002). The critics, as well as the defending DGK, have to face the same reproach: They do not discuss the effect of an increased use of invasive procedures on morbidity and mortality (Dissman and De Ridder 2002)). Instead of discussing the expedience of the method for patients, the numbers of examinations are counted and the costs calculated. This is a phenomenon very similar to what is found in the world of politicians – thus it is not surprising that coronary angiography becomes a target of public criticism, and a so-called example of "uneconomical medicine".

The guidelines and recommendations of the German Society of Cardiology (DGK) are jointly devised with the „Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften" (AWMF; http.//www.awmf-online.de). Several years ago, Lauterbach (Lauterbach and Redaèlli 2002) criticised the unsatisfactory or lacking classification of evidence in German and European cardiologic guidelines. His evaluation of more than 100 guidelines in the year 2000 did not reveal (with few exceptions) what should be normal today: The classification of recommendations and levels of evidence. In addition, the author missed understandable algorithms or decision support systems to translate the scientific results for primary care physicians. This is in contrast to the aims of the German Evidence-based Medicine, which calls for "short and precisely formulated summaries" (Kunz et al. 2000), handy for a quick look up before clinical decisions.

The actual DGK guidelines for the management of chronic heart failure (Erdmann and Hoppe 2001) have defined criteria, in contrast to the AHA/ACC Classification. Level „B" of the DGK, for example, reads "positive statement derived from a single randomised trial or from clinical experience" (AHA/ACC: Data derived from a single randomised trial or nonrandomized studies). Whose "clinical experience"? This quality is lacking a precise definition. Does it mean the AHA/ACC's Level C "consensus opinion of experts"? The DGK levels of evidence have a Grade "D", which contains "negative statement derived from single or multiple studies". This may have been derived from the Canadian and US-Task Force (CTF/USTF) guidelines grade D – but those guidelines are made exclusively for Periodic Health Examination, not for clinical use (Kunz et al. 2000). Thus, the critical specialist, as well as the primary care physician, hesitates to follow such otherwise well formulated guidelines.

Critique directed at Evidence-based Medicine. The role of outcome research registries

Some authors argue against the term EBM, but words and expressions can easily be replaced by others. For example, EBM could be replaced by the expression "research enhanced health", which may better reflect the *updated model for evidence based clinical decisions by R. B. Haynes* (Haynes et al. 2002): Clinical considerations must include the patient's clinical state and physical circumstances in the first place, tempered by research evidence (secondly), while taking into account the *patient's preferences and likely actions* (thirdly), before, eventually, *clinical expertise* is needed in order to compile all considerations and give a recommendation to the patient. Providing evidence to patients in a way that allows them to make an informed choice is a most challenging issue, which is "in many cases beyond our current knowledge of doctor-patient communication – very much a problem awaiting the generation of new evidence" (Haynes et al. 2002).

This leads to criticisms which argue in a more philosophic manner, like in the German debate about *Evans' evidence-based* versus *"Evidence biased medicine or: the treacherous certainty of evidence"* (Rogler and Schölmerich 2000; Raspe 2001). This distinction could be a serious issue, if EBM asserts itself to be a discipline of philosophy or any theory of sciences. It may well be that some ideologists believe so, and some politicians follow them willingly. However, the majority of clinicians, as well as specialists in the use of EBM, regard it as a purely functional support for their practice,

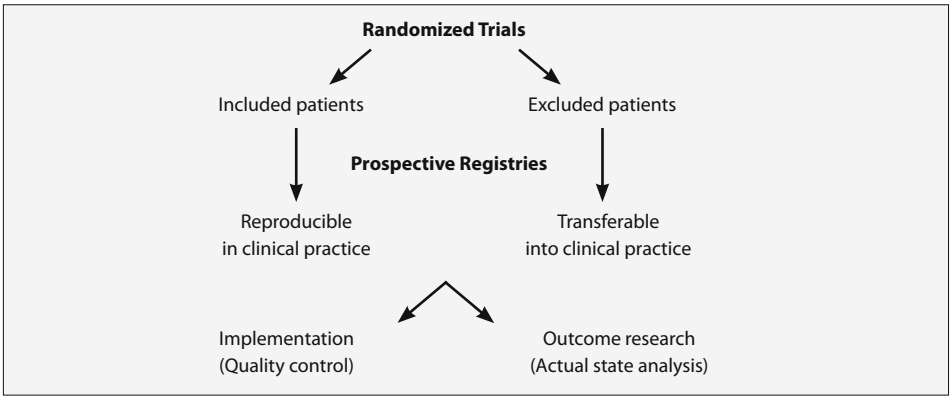
trying to demand more convincing evidence for medical decisions by application of the methods of clinical epidemiology.

Excluded patients in randomised controlled studies: One of the target groups for registries

The randomised clinical trial has emerged as the principal research tool for developing evidence. RCTs need to mark off groups of patients in order to get convincing results. The exclusion of elderly or younger patients, the restriction on men or women, and restrictions according to ethnic background or special clinical conditions (who will be treated and who will not) is often necessary. One can add to this the limitations created by the diagnostic criteria and laboratory tests. The difficulty of guaranteeing adequate participation of underrepresented minorities in RCTs seems to be insurmountable. Moreover, new tests may create new classifications of diseases and groups of patients to be included or excluded in trials. For example in actual cardiology, modern enzyme tests allow the differentiation between myocardial infarctions with or without ST-Segment elevation (see below). There are also the typical cases at risk for exclusion, such as multimorbidity, shock, reanimation, etc …

In multiple studies it became apparent that patients enrolled in trials tended to be different from the total group possibly treated. The enrolled patients are estimated to be rarely more than 50% (Kunz et al. 2000). In EBM literature, this is the main reason to differentiate between the terms *efficacy* (i.e. effective in the ideal conditions of a trial or in a laboratory) and *effectiveness* – effective in the real world of clinical practice (Kunz et al. 2000; Haynes et al. 2002), which is the true field of outcome research. The German cardiologist J. Senges emphasizes the supplementary character of evidence-based guidelines and prospective registries, the latter, for example, as observational studies of clinical effectiveness or postmarketing surveillance studies. Controlled registries, if they are prospective like the ALKK-Registry (Praetorius 1999) (see below), meet the term *nonrandomised studies* on the AHA/ACC Evidence B level (Senges and Gitt 2002). Prospective in registries means the inclusion of patients according to previously defined and controlled parameters, and also the intention-to-treat principle. Their advantage is not to exclude any case at risk, thus being representative of the general population to be treated. Therefore, prospective registries provide the control of the reproducibility of RCTs in clinical practice. They make the benefits of an RCT-controlled therapy for patients who could not fulfil the strong criteria to be included in the trial accessible *(❯ Figure 6-1)*.

Figure 6-1
Randomnized Trials and Prospective Registries (modified from Senges and Gitt, 2002)



The registry of German community hospitals revealed an unexpected dimension of the non–ST-segment elevation myocardial infarctions (MI), newly defined by the elevation of cardiac troponins, which replaced creatine kinase MB as the gold standard marker of myocardial cell injury. The registry of 9.947 acute coronary syndromes included MI with or without (None = N) ST-Segment elevation (STE), acronymised as STEMI and NSTEMI. 46% of the patients were classified as STEMI, 13% as unstable Angina, and 41% as NSTEMI infarcts – the latter more than doubling the expected amount (Senges and Gitt 2002): The actual state (cp. figure 1) has changed dramatically, allowing a better risk stratification than the previously used algorithms based on the ECG and creatine kinase MB. The results may provide a benefit from early invasive strategies, glycoprotein IIb/IIIa antagonists, low-molecular-weight heparins etc. (Bertrand et al. 2002). There is a call for randomised controlled studies and renewed guidelines in this field.

The Transplantation Study Group (Deng et al. 2000) of German Cardiac Surgeons (COCPIT) gives another example. The survival benefit of cardiac transplantation, as compared to conventional treatment in advanced heart failure, has not been tested in a prospective randomised trial. Since the introduction of ciclosporin in 1980, this benefit was seemingly evident, and any RCT would have been hampered by ethical concerns. The surgeons were taken aback by the result of their registry: Transplantation was not associated with reduction in mortality risk for the total cohort. There was only some survival benefit for the most severe cases with a high risk of dying on the waiting list, which, unfortunately, was restricted to the first months after transplantation. The findings claim an initiative for RCTs.

Participation in controlled registries includes a bureaucratic burden to obtain complete and correct data, as well as readiness to accept monitoring and audits. This is compensated by the possibility to compare one's own performance with the results of the total group (Praetorius 1999), and by obtaining scientific progress. To avoid competing interests, special attention should be given to manufacturers who want to support their products by financing outcome studies. For example, in a paper of the „Qualitätsregister ERICA" (Establishing Risk Reduction In Congestive Heart Failure through Add on therapy), which is designed to obtain better implementation of the treatment with beta receptor blocking agents, the betablocker Bisoprolol was not named (Antoni et al. 2002). The participating software company *mediNet*, however, revealed the source ("in cooperation with MERCK") and also the trade name of the medicament used in ERICA. These remarks did not want to put a moral point or any ethical reproach upon the publication, because there was no concrete suspicion to be charged with. However, since transparency is the basis of ethics in clinical science, naming the connection in the author's acknowledgement or conflict of interest statement would have avoided any suspicion.

RCT and guidelines – are they really up-to-date?

Large randomised controlled trials normally need a couple of years for design, planning and organization, and further months or years for the evaluation of the data. It may be allowed to estimate two years, in many cases it is up to five years for the complete procedure. In the contemporary era of EBM, an ethical problem arises: Is it permissible to withhold a new therapy from patients which has somehow proved effective – say level B of evidence, Class IIb – until years later when large RCTs are finished? Ideally, the choice of any therapy would be based on anticipated benefits in longevity, prevention of events, and improvement in quality of life at a reasonable cost (DeMets and Califf 2002). To overcome the problem, the use of a substitute or "surrogate" outcome measure in place of these clinical outcomes could allow shorter, less costly trials with fewer subjects. In the early-phase, research surrogates are inevitable because deductive, pathophysiological reasoning must be applied when developing new therapies (DeMets and Califf 2002). One must note the possibility of severe errors, however: For example, the surrogate of lowering blood pressure in hypertensive patients does

not guarantee a similar success in the clinical outcome of death, myocardial infarction or stroke (ALLHAT 2002).

Only about 30 % of the therapies available today are based on a high level of evidence. There is a tendency to introduce new drugs or treatment strategies on the smaller basis of the consensus opinion of experts (Level C), which is more readily obtained. For years this was the case in the field of coronary angioplasty (PTCA) and the application of stents. German clinical cardiologists therefore founded the „Arbeitsgemeinschaft Leitender Kardiologischer Krankenhausärzte" (ALKK; Registry of German Community Hospitals). The ALKK performed the first prospective registry on coronary interventions, in order to jointly get a critical survey of the practice in their cathlabs. Up until 1999, more than 200000 PTCA were logged (Praetorius 1999), and since 2001 the registry is part of the official German Quality Control System. One of the first outcomes for the participants themselves was very similar data in about 100 clinics, at a time when there was a lack of general regulations or evidence. It would seem that most of the participants performed the interventions on the basis of the same international literature. This closed group not only allowed for an answering of general questions related to mortality and complications, but could also perform actual registries like: *"Determinants of mortality after cardiac surgery: results of the Registry of the Arbeitsgemeinschaft Leitender Kardiologischer Krankenhausärzte (ALKK) on 10.525 patients"* (2000).

Interaction of drugs and interference of guidelines: The great issue for medical practitioners

The central task at the border between research findings and the care of individual patients, is translating the scientific data into the action of practitioners. At the present time, this field has not been adequately explored. RCTs tend to focus on the assessment of one treatment at a time, but most of the patients are treated with multiple therapies simultaneously. Unfortunately, the array of possible effects when combining two treatments, each of which is beneficial, is complex; the effects can be additive as well as synergistic, subadditive, neutral or negative (Califf and DeMets 2002). The interaction of medicaments, and similarly the possible interference of applied guidelines, may induce major problems for the patient. Those problems make the treating clinician or practitioner feel insecure, and therefore reticent to apply all the evidence-based guidelines to all patients with the same disease.

For example, there are disagreements on the relevance of interactions between acetylsalicylic acid (ASA) and angiotensin converting enzyme inhibitors (ACE inhibitors). Either drug is tested by RCTs and meta-analyses and generally accepted. They have different targets, coagulation and inflammation (ASA) and blood pressure or heart failure (ACE inhibitors), respectively. However, physiological experiments and observational studies since 1998 have provided ample reason to be concerned that aspirin may nullify the benefits of ACE inhibitors (Joost 2003). At the moment, the only solution is to look for retrospective analyses of older studies (Califf and DeMets 2002).

Generally, interactions in the chronic therapy of heart failure, as well as in the acute treatment of myocardial infarction and other fields, have not yet been explored by controlled studies ("does ASA nullify or attenuate the ACE inhibitors?"). Accordingly, insecurity is rising amongst clinicians and practitioners by whom all the guidelines should be "implemented". Sometimes they prefer to tell the patient that, because of side effects, one or two pills are better for him than seven or more daily (❯ *Table 6-3)*, though evidence based guidelines do recommend a combination of several substances or classes. Sometimes, the economic motive of sparing the budget may be prompting these tales, as has often been seen in Germany in recent years. The ethical principle of informed consent does not correspond with this kind of denied information (Kuhlmann 1999). However, it should be stressed that the origin of the ethical conflict is located in EBM itself, which up to now has not been able to transfer the results of medical science into the world of doctors and patients.

The mutual interferences of different guidelines, as well as the interaction of drugs, is an issue too complex to be delegated to GPs and local clinics, usually by calling for "better implementation" of evidence based guidelines. Implementation does not only consist of better education or well-controlled propaganda. Rather, EBM specialists should call for evidence-based guidelines about handling guidelines, as well as prospective studies on the interactions and mutual effects of recommended therapies. Of course, directly comparative trials are also needed, so therapies that do not provide sufficient benefit can be discarded. It may be difficult to secure sponsorship for such purposes, but it is an urgent issue for academic research, much more so than for EBM managers acting under the veil of "implementation deficits".

In Table 3, the conflict between EBM and primary care physicians is exemplified by the therapy of hypertension and heart failure. There are excellent guidelines, with a high level of evidence, for at least seven therapeutic regimes and drugs. The family doctor has the choice of the drugs listed in Table 3. He is not in the position of the specialist or clinician, who is accustomed to applying all possible drugs, connecting them with *Generic Names*, and without thinking of costs and the patient's compliance. For the representation of the classes in Table 3 fantasy names were chosen for *ACE-Inhibitors (Enalapril), Beta-blockers (Metoprolol), Calcium antagonists (Amlodipin), Angiotensin-II-Receptor-Antagonists (Losartan), Digitoxin, Chlorthalidon and Spironolactone*. The names may give some idea of the drug's effect and the current discussion. They can be replaced easily by well-known trade names. The problem of class effects can only be touched upon here. Especially in the group of ACE-inhibitors, clinicians are left in the quandary of whether to embrace the class effect, or to preferentially use one particular ACE inhibitor that has been tested in the actual RCT (5).

Table 6-3

Hypertension and Heart failure: Choice of the Medicament(s)			Family doctor's calculation	Decisions		
Name (F)	**No. tablets**	**EUR/day***	**Take into consideration**	**Dilemma**	**Tablets**	**EUR/day**
ACEpril(F)	2/day	1,60	Guidelines?	Take all of them?	9	5,02
Betanolol(F)	2/day	0,81	Side effects? Interactions? Compliance?	Just one?	1	~ 0,56
Nonusodipin(F)	1/day	0,68		Any combination?	~ 2	~ 1,89
Reservosartan(F)	1/day	1,17	Budget?			
Digiforget(F)	1/day	0,09				
Aquatonal(F)	1/day	0,29				
Spironosubtil(F)	1/day	0,38				

(F) = fantasy (no trademark) * = calculated from the German „Rote Liste 2003"

Primary care physicians may want to put three issues before a consideration of evidence based multiple treatment recommendations: First, the possibility of interactions and side effects; second, the lack of convincing evidence supporting the increasing costs of added therapies; and last but not least, the declining compliance of patients which is proportional to an increasing number of pills. In the homes of his patients receiving care on an out-patient basis, the author found numerous boxes with unused medicaments, thus confirming the experience of many colleagues who think about the wasted billions in the health care system.

The problem of unanticipated negative interactions originates early in the design of a RCT, if a previous evidence-based therapy is contrasted with new drugs or devices. In this constellation, with respect to the ethics of clinical research, it is crucial to deny an established therapy in order to compare untreated patients with the new drug. This is the same, if the old and new drugs are to be

compared directly, except that there is a plausible, genuine uncertainty related to what medicament has more benefit for the patient. If the individual investigator is convinced of one therapy, he is not ethically allowed to give another drug just because of randomisation, and he may not join in the blind randomisation of the trial. The only possibility is, to extend the concept of genuine uncertainty to the uncertainty of the expert medical community (Freedman 1987) as a basis for clinical equipoise, but of course not to the uncertainty among pharmaceutical companies.

To minimize such moral problems, clinician researchers tend to produce combinations – with all the problems mentioned above. Typical examples are: testing the combination of a new drug with the established one against the prior drug alone, or creating more side arms with different drug combinations. In any case, a general tendency towards more medicaments and complex regimes will emerge. In patients at special risk, like those facing renal and hepatic failure or severe diabetes, there is a lack of systematic dose studies with respect to combinations of drugs. In these patients, the probability of affecting unintended targets for therapy is even higher than in "normal" patients.

Further, too few directly comparative or noninferiority trials are designed to find out whether a treatment is (or is not) worse than the other or has other advantages such as less toxicity, easier administration and/or lower costs. In patients with myocardial infarction, the ASSENT-3 trial demonstrated a novel approach in which 6095 patients were treated by different complex therapeutic "cocktails" in a randomised, open-label trial (The ASSENT-3 Investigators 2001). Weight-adjusted unfractionated heparin versus Low-molecular-weight heparins (enoxaparin) or platelet glycoprotein IIb/IIIa inhibitors (abciximab), was administrated in addition to fibrinolytic therapy. The goal was to get data short of a definitive result, but to decide whether enoxaparin would be an attractive alternative reperfusion regimen, that warrants the design of a definitive outcome trial. This kind of study could be a model for the mentioned problem of the practitioner. Despite the difficulties in interpreting non-inferiority trials and the possible lack of sponsors, more such trials are needed (DeMets and Califf 2002).

Side effects of randomised controlled trials by themselves

At the time of the Cardiac Arrhythmia Suppression Trial (CAST) the cardiovascular community believed that, because ventricular arrhythmias were predictive of sudden death, and because some drugs could suppress ventricular arrhythmias, these drugs would reduce sudden death and total cardiovascular death. However, the suppression of an arrhythmia was a surrogate outcome measure (DeMets and Califf 2002). CAST revealed a dramatic failure of the proposed outcome, as very early on two arms of the study had to be terminated because of a highly significant increase in both sudden death and all-cause mortality for patients on active treatment (Ruskin 1989).

The lesson learned by the medical community was to always have in mind the clinical outcomes longevity, prevention of events, and improvement in quality of life. Nonclinical outcomes (for example arrhythmias or blood pressure) can be accepted as a valid surrogate to allow shorter, less costly trials with fewer subjects for early-phase research – but they remain surrogates, and strict criteria should be met before they are accepted (DeMets and Califf 2002). This is a very essential point for any ethic review by the Institutional Review Boards (IRB).

Break offs in consequence of unequivocal differences between the groups are not the only problem. *P. Poole-Wilson* demonstrated the *Annual mortality in recent trials in heart failure* (Poole-Wilson 2003), with differences of 0.4 to 8.0 per cent deaths occurring in the groups *Control* and *Drug*. If thousands of patients have to enter a trial, one may well be able to calculate the relevance of the data presented by *Poole-Wilson*. Further, if *Control* means the prior standard treatment, and if break offs take place as early as possible (DeMets and Califf 2002), ethical consequences should be drawn: It is not only the IRB who review the ethical dimensions, but the clinicians should also ask themselves in

advance if the benefit of the clinical outcome, or the scientific gains, do justify the calculated mortality associated with the trial.

There is a great difference between trials testing new therapies and, for example, noninferiority trials with previously introduced therapies. If the intervention is a new drug or procedure, then less harmful or negative evidence is required to terminate the trial. Other situations may call for more substantial evidence before terminating a trial – as was with CAST (DeMets and Califf 2002).

On longevity of guidelines. What does scientific creativity mean?

Ideally, the patient's needs are the stimuli for new studies. Accordingly, trials are indicated if new drugs or therapeutic procedures arise and put into question the actual guidelines. There is also a great influence of possible financial interests. Large RCTs require substantial sponsorship, which can be given only by big companies. Of course, it should be recognized that companies have to calculate costs and returns. We see, on occasion, that new trials are required when patents are going to be terminated and generics are to appear – and prices are to decrease. Researching companies have so-called pipelines at their disposal, containing well-tested new substances, ready to come from the pipeline to the "study-line". An actual example could be read in the *Neue Zürcher Zeitung* on February 13, 2003: the blockbuster *Augmentan®* from *GlaxoSmithKline* is coming to end of the patent protected time, but for the shareholders there will be a *promising product pipeline*.

This may be sound from a business perspective, but a physician with scientific intentions may question the principles of what he or she does in such a research setting. Do they depend only on the rhythm of industrial companies? The problems come to the public fore when independent organizations bring out a trial about the treatment of a widespread disease like arterial hypertension, as was recently the case with the ALLHAT-Trial, which was supported by public funds (ALLHAT 2002). The study demonstrated an older and low-cost drug to be effective at a highly evidence-based level. This type of finding, of course, runs against the interests of the producers of newer and expensive substances. Accordingly, the scientific discussion has shifted to a more ideological level with labels like „Evidence-Based Medicine or Medicine that Suggests Evidence?" (Messerli 2002). The debate in this area is by no means finished, but one would also be well advised to consider the reason for the lack of studies about low-cost procedures in cardiology generally.

Areas of cardiovascular medicine underserved by EBM

Important parts of cardiology lack adequate evidence. Examples are pediatric cardiology, with less than thirty RCTs worldwide, the treatment of patients with valvular heart disease, endocarditis, and pulmonary hypertension (Califf and DeMets 2002).

This is also true for the area of diagnostics, the old pride of cardiologists. In the area of therapy, large-scale RCTs are able to detect small differences between therapeutic methods. In studies on diagnostic methods, there is not only a lack of interest in sponsoring, but also some basic scientific obstacles, especially the frequent problem of low prevalence (Praetorius 1992). Compared with specialists, prevalence may be even lower under the conditions of general practice (Praetorius 1992; Sox et al. 1990), and extremely low in the case of screening of a healthy population. The problem does cumulate in the so-called *mammography dilemma* as a crisis for Evidence-Based Medicine itself (Goodman 2002).

Diagnostic data are normally the basis for the inclusion of patients in RCTs, but do not provide adequate evidence in and of themselves. Sometimes they are too rough to define the pathophysiological core and the therapeutic target of a disease. A highly evidence-based therapy may then rely on inaccurate diagnostic measures. For example, the important risk factor obesity is defined by the

Body Mass Index, which though having level B (Yusuf et al. 2002), does not reveal anything about the amount of excess body fat and the important differences of regional obesity.

The definition of ST-segment depressions in the exercise ECG remains partly a matter of experience, and partly of standards given by the cardiologic societies (Fletcher 2001; Trappe and Löllgen 2000). These standards are not based on large-scale controlled studies, but on many references over decades of scientific cardiology. The choice of a discriminant value is complicated by the fact that exercise test responses do not have evidence-established values that separate normal subjects from those with disease. Statistically, judging ST displacements is based on the test's specificity and sensitivity. The predictive value of the test is greatly influenced by the prevalence of the disease in the group being tested (cp. Bayes' theorem) (Praetorius 1992). Angina induced by the exercise test is predictive of CAD, and even more predictive with associated ST-segment depression, and vice versa. In the same or contrary direction are the effecting risk-factors gender and age. In summary, a mixture of more or less soft factors may influence the diagnosis after testing – in combination with the clinical experience and professional situation of the doctor (Sox et al. 1990).

An example of inducing clinical uncertainty, are the guidelines of the DKG on Interventional Intracoronary Therapy (Erbel et al. 1997), which judge therapeutic indications by Grades A-D (unfortunately special categories of the DKG, including "consensus" to A), and grade stenoses with its own modification of AHA/ACC. However, the quantification of coronary stenoses remains based on subjective criteria nearly everywhere. Subjects (cardiologists) have understandable preferences: The audits of the ALKK – control of the coronary angiographies by independent specialists – revealed an average overrating of the stenoses before PTCA (diameter minus 4.5 %) and of the dilated vessels after PTCA (diameter plus 3.5 %) (Praetorius 1999). Efforts have been made to improve estimating techniques in diagnostic cardiology, but the diagnostic basis of therapeutic RCTs will remain underserved for some time to come.

Systematic distortions

The overwhelming quantity of published RCTs and also of guidelines ultimately produces inequality, however. The sheer number makes it impossible to distinguish important from disappointing papers, or to control all the methodological criteria found by reading each study. In many cases, the reader is confronted with concerns that economic or other interests have influenced the study or the decision to publish. This is defined as publication bias and has a positive and negative direction. In the case of positive results of a trial, for example, if the new drug is proven to be effective, the probability of its publication is higher, and to be published in the English language and in more journals. The danger of not being printed is proportional to the influence of the interested sponsor, who may be disappointed by the results of the study. The chance of high-profile trials with negative outcomes being published in leading journals is better because they are of interest to the cardiology community, and editors of peer-reviewed journals will usually publish them regardless of the results (cp. CAST; (DeMets and Califf 2002)). The problem is that negative trials, if less high profile, are less likely to be published. Accordingly, this situation also feeds publication bias.

A solution to the publication bias problem could be a registry of all clinical trials that are under way so that at least their existence is known. In the USA, the National Library of Medicine is presently developing such a registry (DeMets and Califf 2002).

General remarks. Possible consequences

The evidence of Evidence-based Medicine

Resuming the arguments, one could state that the health care of the individual cardiology patient has so far not been improved by means of clinical epidemiology. The claim of EBM to place scientifically validated medicine at the disposal of clinicians, as well as primary care physicians, has not yet been fully realised. Better communication between physicians, or between doctors and their patients, was not achieved.

A speciality of the European guideline commissions is to provide recommendations on the basis of highly sophisticated reviews and meta-analyses, but without offering decisive compilations and summaries for the consumer. In fact, a clear-cut staging in accordance with international standards is overlooked in many sets of recommendations, for example (Bertrand et al. 2002) or (Erdmann and Hoppe 2001). Guidelines are not always guidelines, in the most literal sense.

Among numerous trials there is no RCT about the outcomes of EBM itself and its effects on cardiology. In general, EBM does not simplify the conditions of medical science and practice as intended, but burdens it with increased complexity (Vogd 2002) – this can be seen in the problem of choosing the right drug via the right guidelines, navigating through numerous offers (cp. Table 3).

Transparency of interests is the basis of a new trust in EBM

Physicians are accustomed to receiving the results of RCTs and guidelines beforehand by commercial travellers. The feeling of being directed rises to consciousness if even guidelines of the AHA can come under attack, as happened recently, when a rather strange decision was made to declare alteplase a class I intervention (definitely recommended) for stroke, despite controversies about its safety and efficacy. This decision moved it up from "optional" (class IIb) to the top category. The action seemed less strange, however, when we learnt companies had donated $11 million to AHA, and six of the nine members of the AHA panel had undeclared financial ties to the manufacturers (Smith 2002). The criticised experts argue that "these relationships might indicate a laudatory effort by physician and organisational advocates" and call the claim producers of guidelines to be independent an „extreme financial correctness" (J.L. Saver (Lenzer 2002)).

In 1993 the International Committee of Medical Journal Editors published a policy on conflicts of interest. Up until now such conflicts have rarely been declared in journals, as several studies have shown (Smith 2002). In Germany, some journals include short declaration formulas, but without declaring the exact role of sponsors, as could be seen in the mentioned case of ERICA (Antoni et al. 2002). In 2002, the *British Medical Journal* performed a randomised trial on the effect of declaring competing interests on readers' perception of research: Among two groups of readers (n = 170) of the same publication but with (group I) or without (group II) the declaration of a special interest (the authors being employees or stock option holders of a fictitious company), readers of group I thought the same text was significantly less interesting, important, relevant, valid, and believable than those of group II (Chaudhry et al. 2002).

The AHA-stroke affair does demonstrate that even apparently independent guidelines may be suspect. How many sets of guidelines are affected? Recently a cross-sectional survey of 192 authors of 44 Clinical Practice Guidelines (CPG) from North American and European societies revealed that, overall 81% of authors per CPG had interactions with the pharmaceutical industry. It is clear that the appropriate disclosure of financial conflicts of interest for authors of CPGs, and a formal process for discussing these conflicts prior to CPG development, should be raised as an issue (Choudhry et al. 2002). In principle it is not wrong to have conflicts of interest that may be universal in medicine.

The problem with competing interests arises when they are left undeclared, thereby eroding the appearance of being a dispassionate scientist or guideline writer.

Publication affairs contribute to the current state of ethics, which is characterised by a dramatic decline in morale and motivation among providers of health care. In a recent survey of general practitioners (GP), two thirds of the respondents said that morale was low or very low, and also that morale was currently lower than it had been five years ago (*British Medical Association. National survey of GP opinion*. London: BMA, 2001; see also (Pendleton and King 2002)). One of the key reasons, in my opinion, is the failure of medical organisations and societies to declare the values which act as their guiding principles – and thereby *leave their members unclear about what the organisation stands for* (Pendleton and King 2002).

If values are left unstated they are inferred from observable behaviour (Pendleton and King 2002), as happened to the AHA in the mentioned controversy (Smith 2002). According to such situations, doctors form their own ethical understandings drawn from their own changing practice. Varying moral principles as well as ever passing and evolving guidelines form a kind of "anything goes" approach, which may reinforce the liberalism of the economy in the pharmaceutical sector. Any individualistic moral tends to lead not only into general deregulation but also into deprofessionalisation, which is working towards declining research outcomes of the medical science itself and, in consequence, decreasing commercial profits. Some companies are becoming aware of those consequences and beginning to ask for a new vision of moral leadership (Pendleton and King 2002). However, before having a vision, it is sometimes better to claim for transparency to know what the society stands for.

Transparency of guideline committees

Strong rules of controlling RCTs usually differ from the rules guiding the establishment of independent CPG committees (cp. (Lenzer 2002)). Expert clinical researchers serving as members of Guideline Committees should, as a matter of course, declare their entanglement with commerce (including research grants, lecture fees, consultancies, shares etc.). Choudhry's paper (Choudhry et al. 2002) demonstrates the need for [1] appropriate disclosure of financial conflicts of interest for authors of CPGs, [2] a formal process for discussing these conflicts prior to CPG development and [3] the complete disclosure of each author's potential conflicts to readers of guidelines.

Another formal process is to achieve a correctly developed consensus of experts in the CPG committees. The process of rating studies and review articles, as well as the way of making decisions, needs to be formalized in a way that allows complex groups to retain their own dynamics, of course. In Germany a scheme of developing guidelines is offered by the AWMF. The unsatisfactory or lacking classification scheme for evidence in German and European cardiologic guidelines mentioned above (Lauterbach and Redaèlli 2002), might be a result of the practise of informal and uncontrolled expert groups, which is typical in some European countries. Almost no one is ready to cooperate by giving up parts of his or her homemade scientific theories or laboratory practices.

The mentioned tendency towards the internationalisation of guidelines calls for better "rules of the game", which are defined in terms of competing interests, norms of payment for clinical research services and also obligations under ethical constructs of human experimentation (Topol et al. 1997). Those who are treating patients in RCTs must also be vigilant about payments exceeding standard reimbursement levels for the work completed, or that provide a bounty for enrolling specific patients (DeMets and Califf 2002). A detailed review of these evolving standards is available from the "Task Force on Financial Conflicts of Interest" of the AAMC (AAMC Task Force on Financial Conflicts of Interest in Clinical Research 2001).

Different governments in Germany have tried more than 14 "Reforms of the Health Care System" over several decades. All failed because too many interested groups and parties were taking

part in the game. The idea that one day a very last "Gesundheitsreform" will take place seems to be only a dream (Praetorius 2003). One of the ingredients of such a dream should be a fair balance between the interests of the patients and their family doctors on the one hand, and the specialists and commercial companies on the other, to improve confidence and adherence to ethical norms while not blocking the progress in research. If the dream failes, there would likely be a move to severely regularize practice, as has begun with the imported model of „diagnosis related groups (DRG)", more to strangulate than to reform the German medical community.

Evidence-based Medicine as a new paradigm?

Evidence-based Medicine as a programme to support medical research is oriented to the future of science. However, EBM is also declared to be the basis of daily practising, which is a more conservative and affirmative project, less a case of biomedical innovation (Raspe 2001; Rogler and Schölmerich 2000). The assertion of EBM being a new paradigm is controversial. Some say EBM may require a philosophical change analogous to "paradigm shifts"; others oppose that view, arguing that EBM is just a new view of old issues. According to Thomas Kuhn (Kuhn 1962), new paradigms, as historical processes, do arise from scientific revolutions while the former become a dying tradition. It is not possible to produce or bring about or wish for a new paradigm on purpose. A present paradigm and its validity may be a matter for discussion, whereas a possibly arising paradigm is something perhaps best dealt with through science fiction. Talking about future paradigms may sometimes divert attention from actual problems and means of addressing those problems in a practical way. Such exaggerated talks may get a touch of *staging knowledge*, which is also one of the risks of EBM itself, possibly resulting in an "*ideology of rationality*", as discussed recently by *Werner Vogd* (Vogd 2002).

The standards expressed in guidelines increase the demands placed on the individual physician, as do the present economic problems, with numerous methods of regulating medical services (cp. DRG). The individual doctor may respond by either increasing his autonomy or by resorting to a technocratic retreat (Vogd 2002). This would consequently include the tendency to abstract from the present patient, allowing *medical deresponsibilization* to progress (Sureau, see (Praetorius and Sahm 2001)). We should, however, admit that EBM is a real chance to control overwhelming commercial interests in science and clinics, as well as to provide better teaching and practice. Whether researching or practising in medicine, Evidence-based Medicine – just as any method of science – does need our individual honesty, independence and scientific engagement.

References

1. ALLHAT Collaborative Research Group (2002) Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs Diuretic. JAMA 288: 2981-2997
2. Antoni DH, Beuckelmann D, Hambrecht R et al. (2002) Therapie mit Beta-Adrenorezeptorantagonisten bei der chronischen Herzinsuffizienz. Deutsche Medizinische Wochenschrift 127: 2459-63
3. AAMC Task Force on Financial Conflicts of Interest in Clinical Research (2001) Protecting Subjects, Preserving Trust, Promoting Progress. Policy and Guidelines for the Oversight of Individual Financial Interests in Human Subjects Research. http://www.aamc.org/members/coitf/firstreport.pdf
4. Bertrand ME, Simoons ML, Fox KA et al. (2002) Management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. European Heart Journal 23: 1809-1840
5. Califf RM, DeMets DL (2002) Principles From Clinical Trials Relevant to Clinical Practice: Part I. Circulation 106: 1015-1021
6. Califf RM, DeMets DL (2002) Part II. Circulation 106: 1172-1175

7. Chaudhry S, Schroter S, Smith R, Morris J (2002) Does declaration of competing interests affect readers' perceptions? A randomised trial. BMJ 325: 1391–2
8. Choudhry NK, Stelfox HT, Detsky AS (2002) Relationships Between Authors of Clinical Practice Guidelines and the Pharmaceutical Industry. JAMA 287: 612-617
9. Davidoff F et al. (2001) Sponsorship, Authorship, and Accountability. New England Journal of Medicine 345: 825-827
10. DeMets DL, Califf RM (2002) Lessons Learned From Recent Cardiovascular Clinical Trials: Part I. Circulation 106: 746-751
11. DeMets DL, Califf RM (2002) Part II. Circulation 106: 880-886
12. Deng MC et al., COCPIT Study Group (2000) Effect of receiving a heart transplant. BMJ 321: 540-5
13. Deutsche Forschungsgemeinschaft (1999) Klinische Forschung – Denkschrift. ©2002-3 http://www.dfg.de
14. Dierks C (2003) Legal implications of guidelines. Deutsche Medizinische Wochenschrift 128: 815-819
15. Dissmann W, Ridder M de (2002) The soft science of German cardiology. Lancet 359: 2027-29 Discussion: Interventional cardiology in Germany. Lancet 360: 1694-96
16. Erbel R, Engel HJ, Kübler W et al. (1997) Interventionelle Koronartherapie. Zeitschrift fur Kardiologie 86: 1040–1063
17. Erdmann E, Hoppe UC (2001) Leitlinien zur Therapie der chronischen Herzinsuffizienz. Zeitschrift fur Kardiologie 90: 218–237
18. Fletcher GF (2001) Exercise Standards for Testing and Training. Circulation 104: 1694–1740
19. Freedman B (1987) Equipoise and the ethics of clinical research. New England Journal of Medicine 317: 1415
20. Goodman SN (2002) The mammography dilemma: A crisis for Evidence-Based Medicine? Annals of Internal Medicine 137: 363-65
21. Haynes RB, Devereaux PJ, Guyatt GH (2002) Physicians' and patients' choices in evidence based Practice. Evidence does not make decisions, people do. BMJ 324: 1350
22. Joost HG (2003) Welche Interaktionen gibt es bei gleichzeitiger Gabe von ASS und ACE-Hemmern? Deutsche Medizinische Wochenschrift 128: 398
23. Kuhlmann E (1999) Between informed consent and denying information: patient information under economical constraints. Ethik Med 11: 146-161
24. Kuhn TS (1962) Die Struktur wissenschaftlicher Revolutionen. Suhrkamp, Frankfurt am Main
25. Kunz R, Ollenschläger G et al. (2000) Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis. Deutscher Ärzteverlag, Köln
26. Lauterbach KW, Redaèlli M (2002) Evidence based medicine in der Kardiologie. Cardio News 4: 41-42
27. Lenzer J (2002) Alteplase for stroke: money and optimistic claims buttress the „brain attack" campaign. BMJ 324: 723-726 Saver JL, Kidwell CS, Starkman S (2002) Commentary: Thrombolysis in stroke: it works! BMJ 324: 727-729
28. Messerli F (2002) Life protection: Learnings from new clinical trials in hypertension. The world after ALLHAT. Davos
29. Pearson TA, Mensah GA, Wayne Alexander R et al. (2003) Markers of Inflammation and Cardiovascular Disease. Circulation 107: 499-511 http://circ.ahajournals.org/cgi/reprint/107/3/499.pdf
30. Pendleton D, King J (2002) Values and leadership. BMJ 325: 1352–5
31. Perleth M, Raspe H (1998) Evidenz-basierte Medizin: Möglichkeiten und Grenzen. Münchener Medizinische Wochenschrift 140: 99-102
32. Poole-Wilson P (2003) Priorities in health care and ethical considerations. Davos Lecture, Febr. 14, 2003
33. Praetorius F (1992) Bayes City: The Objectivity of Indications (Bayes-Stadt … Ein Plädoyer für den diagnoseführenden Arzt). Deutsches Ärzteblatt 89: 2113-2120
34. Praetorius F, Sahm S (2001) The Ethical Aspect of Regularisation in Medicine. Ethik Med 13: 221-242
35. Praetorius F (1999) Coronary Angioplasty (PTCA) Registry of the German Community Hospitals: Ethical Consequences of an Internal Quality Control. (Selbstbegrenzung als Modell?) Ethik Med 11: 89-102
36. Praetorius F (2003) Nach der finalen Reform – Traum eines Kardiologen. Deutsches Ärzteblatt 100: 3212
37. Raspe H (2001) „Die Heilkunde wird eine Wissenschaft sein, oder sie wird nicht sein." Zeitschrift für Ärztliche Fortbildung und Qualitatssicherung 95: 495–501
38. Rogler G, Schölmerich J (2000) „Evidence-Biased Medicine" – oder: Die trügerische Sicherheit der

Evidenz. Deutsche Medizinische Wochenschrift 125: 1122-1128

39. Ruskin JN (1989) The Cardiac Arrhythmia Suppression Trial (CAST). New England Journal of Medicine 321: 386-388
40. Schulman KA, Seils DM et al. (2002) A National Survey of Provisions in Clinical-Trial Agreements between Medical Schools and Industry Sponsors. New England Journal of Medicine 347: 1335-1341
41. Senges J, Gitt AK (2002) Reicht die zur Verfügung stehende Datenmenge und Datenqualität aus, um medizinische Entscheidungen abzusichern? Evidence-based Medicine. 68. Jahrestagung der Deutschen Gesellschaft für Kardiologie
42. Smith R (2002) Making progress with competing interests. BMJ 325: 1375-1376
43. Sox HC, Hickam DH, Marton KL et al. (1990) Using the Patient's History to Estimate the Probability of Coronary Artery Disease: A Comparison of Primary Care and Referral Practices. American Journal of Medicine 89: 7-14
44. The ASSENT-3 Investigators (2001) Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin. Lancet 358: 605-13
45. Topol EJ, Califf RM, Van de Werf F et al. (1997) Perspectives on large-scale cardiovascular clinical trials for the new millennium. Circulation 95: 1072-1082
46. Trappe HJ, Löllgen H (2000) Leitlinien zur Ergometrie. Zeitschrift fur Kardiologie 89: 821-837
47. Vogd W (2002) An Increase in Professionalization or Erosion of Medical Autonomy? The Implications of Evidence Based Medicine and the New Functional Elites in Medicine Seen from the Perspective of Systems and Interaction Theory. (Professionalisierungsschub oder Auflösung ärztlicher Autonomie?). Zeitschrift für Soziologie 31: 294-315
48. Von Schacky C (2002) Koronare Herzkrankheit. Deutsche Medizinische Wochenschrift 127: 2429-2431
49. Yusuf S, Cairns JA et al. (2002) Evidence-based Cardiology. BMJ Books, London

7 Are Particular Patients Disadvantaged by EBM? Focus on Frail Elderly Patients

Fabrizio Fabris · Carla Scarafiotti · Barbara Maero

The applicability of Evidence-based Medicine to frail elderly people in community and institutional settings is indeed problematic. Of all those patients with more complex clinical conditions, the outcomes of randomized clinical trials (RCTs) represent a portion, often limited, of the results, because of insufficient relevant evidence. Most elderly patients have multiple impairments and chronic diseases that generally exclude them from controlled trials, or lead to puzzling results.

Firstly, the randomised trial model is adequate for the study of simple interventions such as pharmacological treatments, and not for complex strategies. These complex strategies are often required in the care of elderly persons, whose results are highly variable and dependent on patient-carer interactions.

Even with simple treatments there is a well-known, yet pivotal distinction, between the effect of a medical intervention under experimental conditions (efficacy), and the result that the same intervention currently produces under conditions of normal activity (effectiveness). As pointed out by Frith, effectiveness is a comparative concept, involving the balance of desired effects versus side effects, the seriousness of the condition to be treated, the consequences of no-treatment, and the availability of other treatments (Frith 1999). Remarkably, side effects and pharmacological interactions play a powerfully negative role on the well-being of frail elderly patients.

Those deemed eligible to participate in a RCT are subjected to scrupulous and selective investigations. These investigations are seeking to determine if they are potentially at high risk for the event, are not affected by concomitant illnesses, likely to be compliant with the proposed treatment, and if they accept the proposed treatment favourably. Doctors who follow defined protocols often prescribe experimental treatments; the participants in the study receive special attention and suggestions during the course of the study. Extreme ages, and patients who follow other therapies, are frequently excluded. In actual practice, however, all patients affected by that specific pathology are usually treated, even those who had been excluded from the study for various reasons.

Moreover, in clinical practice, the drug is sometimes prescribed in different doses, with different frequencies of administration, or in combination with other drugs.

Considering these realities of clinical practice, RCTs are essential if one is to know the profile of effectiveness and toxicity of a therapy. They allow the physician to be sure of what does not work; it is not likely that a substance, not active under ideal conditions, could become active in the complexity of practical medicine.

Perhaps the rationale of EBM is, above all, to avoid the subjective choice of the physician in favour of ineffective or obsolete treatments. In specific fields such as palliative, neonatal and psychiatric medicine, it is ethically questionable or difficult to define which patients are to be treated and which are not. Some people maintain that the drawback of these branches of medicine could be overcome, if EBM were not exclusively interested in the purely pharmacological therapies and acute illnesses. In general, comorbidity makes phase III experimentation, which is aimed at evaluating only one therapy, for only one illness, during a given time extension, very difficult. For dementia,

for instance, we do not have specific outcomes, as fractures are for osteoporosis. The elderly are definitely among patients disadvantaged by EBM. In the aged there is not a clear cut-off between normality and illness – usually elderly patients are in a „grey area". This group shows the greatest variability among subjects, and it is more difficult to place elderly people into only one category.

The cost-effectiveness relationship could also serve as something that disadvantages elderly populations, despite the declared wish not to discriminate among patients. The difference is not due to age itself, on the contrary, it is based on the known shorter life expectancy and reduced probability of recovery. According to a study conducted in New Zealand, the cost-effectiveness of treatment for hypertension was deemed to be adequate in people aged 60 to 75 years. The treatment of younger subjects, however, was not considered very cost-effective because of their low absolute risk of CVD, while in subjects older than 75 years of age there are increasing costs associated with non-CVD mortality (Jackson 1998).

It is therefore confirmed that the elderly are not especially suitable for RCT, due to the type of pathology (more often chronic than acute), the increasing impact of the concomitant therapies, and of psychological and environmental interventions. Risk factors are different from those identified in the young-adult and, perhaps more significantly, their importance changes. In older age, both atherosclerosis and its clinical consequences are mainly age-related.

The increase of atherosclerosis in old people is indisputable, but has not been defined how this disease can be attributable to risk factors that are modifiable through specific interventions.

An example is given by our study, performed on a sample of 457 subjects, to investigate the prevalence of extracranial carotid artery atherosclerosis and its relationship to risk factors at different ages (Fabris et al. 1994). In accordance with data from the literature, our results indicate that there is a high prevalence of carotid atherosclerosis in healthy, free-living subjects (up to 60 % in subjects aged 65 to 74 years and over 75 % in the oldest age group). This finding is particularly relevant because of the very low prevalence of symptomatic patients in our series. In all age groups, men are more prone to atherosclerosis than women, but this difference is markedly less in the group aged 65 to 74 years, and nearly disappears in the oldest age group. This may be due to postmenopausal hormonal modifications in women, which increase their susceptibility to the atherosclerotic process, and to the incidence of premature death among men suffering from extensive severe atherosclerosis. Not only the prevalence, but also the extension of the disease (expressed as the number of carotid artery lesions) and the severity of vascular narrowing appear to increase with advancing age. The number of carotid plaques rises progressively with age, averaging two occurrences in subjects older than 74, 20 % of whom had more than three plaques. The severity of stenosis also increased: the mean percentage of narrowing at the site of maximum stenosis was 9 % in those aged 45 to 64 years, 17 % in subjects aged 65 to 74 years, and 27 % in those aged over 74 years. However, most of the subjects aged 75 or older showed minimal or mild stenosis, whereas less than 5 % had severe vascular narrowing. It can certainly be stated that in aged subjects the progression of atherosclerosis is more likely to appear as an extensive, multisegmental vascular involvement rather than a severely narrowing process. On the other hand, age ($p<.01$), with total cholesterol ($p<.05$) and diabetes ($p<.05$), was found to be independently related to the severity of vascular narrowing.

The prevalence of risk factors increases progressively up to the agegroup of 65 to 74 years and then, with the exception of hypertension, declines in the oldest age group. This phenomenon in cross-sectional study design is most likely explained by the premature death, from cardiovascular disease, of high-risk subjects. The physiological trend toward a reduction in total and LDL cholesterol in later life, or the lower number of smokers among elderly women, may also contribute to the decline. Of all risk factors investigated, age has by far the strongest independent association with carotid atherosclerosis. Although hypertension, cigarette smoking, and diabetes are considered important risk factors for atherosclerosis, we failed to confirm this relationship in elderly patients.

Among the elderly, those risk factors that have a purely atherogenic effect seem to have little bearing on atherosclerotic disease.

These findings have been further confirmed by our study on the prevalence of peripheral arterial disease (C W Doppler examination) and its relation to risk factors in 418 patients living in nursing home. The prevalence of this condition was 29 %, although very few patients were symptomatic. Among traditional risk factors, age greater than 95 years ($p<.05$), elevated systolic blood pressure values ($p<.0001$), and a previous history of myocardial infarction or angina ($p<.05$) were positively and independently associated with peripheral arterial disease (Bo et al. 1996).

We have also evaluated the risk factors for hospital and post-discharge mortality in a cohort of 987 patients, aged 70 years and older, admitted to a geriatric care ward (Ponzetto et al. 2003).

Our specific objectives were to identify functional measures that predict 5-year mortality in elderly persons after hospitalization, and to define the prognostic importance of clinical parameters and medical diagnoses.

In-hospital mortality in the sample was 14.9 %. The variables independently associated with in-hospital mortality in multivariate analysis were: functional impairment (ADL), dependence related to medical conditions (DMI), cerebrovascular disease, cancer, low albumin, high creatinine, and high fibrinogen..

During the 5 years of follow-up, 553 patients (67.7%) had died, with a median survival of 36 months. The variables independently associated with mortality in multivariate analysis were age over 75 years, male sex, ADL dependency, cognitive impairment on SPMSQ, medical dependency (DMI), presence of cancer, haemoglobin ≤ 11 g/dl and Charlson's Index ≥2.

In fact, epidemiological data on subjects aged 75 years and over have shown a lower impact, even on survival of the traditional risk factors such as hypertension, hypercholesterolemia, diabetes, obesity, and cigarette smoking, while "new" domains are becoming increasingly important, namely falls, side effects of drugs, malnutrition, physical or mental disability, social isolation, and low physical activity (Fried et al. 1998; Scott et al. 1997; Burke et al. 2001). Clearly, EBM principles are more difficult to apply to these topics.

An additional problem is found in the low compliance of treatments, especially in the case of subjects with serious mental disorders.

At least in some conditions, such as patient's evaluation and management, different methodologies are required (judgement of the caregiver, for instance, or impact on relatives' stress). This does not make it impossible to carry out a trial. EBM should allow clinical judgement in a broader way than a computer.

In geriatrics it is necessary to underline a few aspects:

- EBM has to include, besides RCTs, observational studies and the clinical experience of specialists;
- Emphasis is needed on what it is better not to do, rather than on what is rational to do;
- An extension of the EBM methods to non-pharmacological interventions is recommended,
- and the use of ad hoc criteria of evaluation for caregivers and relatives should be included.

Some problems in the evaluation of drugs are solved by clinical observation rather than by clinical trials. In many cases, it could hardly be ethical and practicable to carry out a RCT.

There is a need to underline the particular sensitivity the elderly show to side effects, especially in the chronic pathologies for which there is a threshold effect, leading to the development of a reaction when the reserve is exhausted. This is clearly described for Parkinson's disease: the illness appears when the reserve of the dopaminergic system is reduced below 20-30% of the basal level of a young subject. In an old person such a limit is often reached, so a drug with negative action on the system, for example flunarizine with its anticalmodulinic action, easily shows its negative action.

Withdrawal from the drug restores the baseline condition (Daniel and Mauro 1995). Few clinical cases, carefully monitored, are enough for verifying this phenomenon.

Another problem is the duration of treatment; a trial seldom lasts so long as to enable the evaluation, in the long or very long term, of the effectiveness and/or tolerability of a drug.

Effectiveness in real practice can be evaluated in observational studies (cohort-studies, case-control studies), which are phase IV investigations. These studies, carried out on large populations, can point out rare side-effects, usually difficult to find out on a more reduced sample. The results are not usually substantially different from those of the RCTs, but they integrate them. The role of the general practitioner becomes central.

People with personal interests in the pharmaceutical field have sometimes tried to extend to clinical medicine the result of a discipline meant to be applied with most of the methodological-scientific rigour (Celemajer 2001). EBM's outcomes, when they corroborate the thesis of the sponsors, which are often correct and legitimate, make them stronger and less questionable. This can also happen when the certification is not absolute, or when it allows some far-fetched extension to certain types of patients, for instance the elderly. Hypertension is an example: at first the trial does not examine a frail elderly population, in particular, but later the results are extended to this population because it represents the majority of the possible recipients of pharmacological care.

References

1. Bo M, Zanocchi M, Gallo R, Bergoglio I, Poli L, Fabris F (1996) Prevalence and risk factors of peripheral arterial disease among older patients living in nursing homes. Journal of the American Geriatrics Society 44: 738-9
2. Burke GL, Arnold AM, Bild DE et al. (2001) CHS Collaborative Research Group. Factors associated with healthy aging: the cardiovascular health study. Journal of the American Geriatrics Society 49: 254-62
3. Celermajer DS (2001) Evidence-based Medicine: how good is the evidence? Medical Journal of Australia 174: 293-5
4. Daniel JR, Mauro VF (1995) Extrapyramidal symptoms associated with calcium-channel blockers. Annals of Pharmacotherapy 29: 73-5
5. Fabris F, Zanocchi M, Bo M, Fonte G, Poli L, Bergoglio I, Ferrario E, Pernigotti L (1994) Carotid plaque, aging, and risk factors. A study of 457 subjects. Stroke 25: 1133-40
6. Fried LP, Kronmal RA, Newman AB et al. (1998) Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. JAMA 279: 585-92
7. Frith L (1999) Evidence-based Medicine and general practice. General Practice and Ethics: uncertainty and responsibility. Routledge, London
8. Jackson R (1998) Which elderly patients should be considered for anti-hypertensive treatment? An evidence-based approach. Journal of Human Hypertension 12: 607-13
9. Ponzetto M, Maero B, Maina P, Rosato R, Ciccone G, Merletti F, Rubenstein LZ, Fabris F (2003) Risk factors for early and late mortality in hospitalized older patients: the continuing importance of functional status. Journals of Gerontology. Series A, Biological Sciences and Medical Sciences 58: 1049-1054
10. Scott WK, Macera CA, Cornman CB, Sharpe PA (1997) Functional health status as a predictor of mortality in men and women over 65. Journal of Clinical Epidemiology 50: 291-96

8 Evidence-based Medicine in Mental Health: Towards Better and Fairer Treatment?

Andreas Reis · Christian Lenk · Cristiane Reis-Streussnig · Nikola Biller-Andorno

Introduction

Discussions on the provision of mental health care over the last decades have focused on the rights of mentally ill patients and the provision of increasingly comprehensive services that reflect the variety of co-existing etiological and therapeutic models. When psychodynamic and psychoanalytical psychotherapies were included into the health care coverage of many countries in the 1960s, this step was welcomed as progress towards equal rights for patients with somatic or diseases (see Duehrssen and Johrswieck 1965).

The debate among patients, mental health care providers, insurers, and policy makers has changed considerably since then. Today, two main concerns shape the discussion: How to maintain and improve the quality of psychiatric and psychotherapeutic treatment measure and how to rationalize the provision of services in a way that leads to a fair allocation of resources by the means of identifying cost-effective, efficient therapeutic options.

Evidence-based Medicine (EBM) plays a role in regards to both the quality of clinical practice as well as the allocation of resources. Supporters and critics of EBM will have to ask, however, which conceptual presumptions underlie EBM, what its methodological limitations are in determining what treatments "work", and which therapeutic options are more "efficient" than others. This paper will discuss these questions with respect to psychiatric and in particular psychotherapeutic care. For this purpose, we will first describe recent attempts of taking up EBM in the area of mental health; we will then summarize and comment on the criticism brought forward against these developments; and finally we will sketch out the conditions under which EBM can be employed for the evaluation of psychotherapeutic and psychiatric approaches in an ethically acceptable way.

The development of EBM strategies in psychiatry and psychotherapy

Even though a large part of psychiatry and psychotherapy, particularly in the psychoanalytic field, were sceptical of the EBM approach at first (and some remaining so), a growing number of researchers and organizations have responded to the demand to prove the effectiveness of specific types of therapy.

As the American Psychiatric Association (APA) recognized that treatment and reimbursement decisions were occurring without scientific and clinical input, it set up the Steering Committee on Practice Guidelines. This committee is tasked to methodically review the data in the literature and produce a consensus document. Moreover, recognizing the growing need for practice-relevant evidence in the field of psychiatry, the APA in 1993 established the Practice Research Network (PRN). Its aim is to complement traditional research methods by generating information across a broad range of patients, treatments, and treatment settings.

In 1998, the APA took a further step and founded the independent not-for-profit American Psychiatric Institute for Research and Education (APIRE). Its mission is to gather data on specific clinical issues and decisions facing psychiatrists, in order to address gaps in the current research base. Moreover, it systematically assesses the effectiveness of different treatments and combinations of treatments for specific psychiatric disorders and patient groups. Finally, APIRE analyzes the impact of changes in the organization, delivery, and financing of care on access, quality, and outcomes of psychiatric treatment (Gray 2002). In the field of psychotherapy, Sifneos (1992) demonstrated the effectiveness of short-term anxiety provocative psychotherapy (STAPP). His pioneering work triggered many other studies in short-term psychotherapy, with manuals and outcome studies.

Further, Grawe, Bernauer, and Donati (1994) performed a meta-analysis of controlled treatment studies in psychotherapy. In their review, they analyzed 897 controlled studies published up to 1983, using a special assessment manual. Only three major methods of psychotherapy, i.e. behaviour therapy, psychoanalytical therapy, and client-centered therapy were shown to be effective. However, in conclusion, the authors judged the evidence for psychoanalytical therapy as "not impressive" ans demanded further investigation with particular respect to long-term psychoanalysis. Another group of treatment methods (e.g. bioenergetics, music therapy) showed a "certain amount" of effectiveness data with equivocal results, and were thus not included in the list of established treatments. Further, the review revealed that for another group of treatments, there were only a few methodologically acceptable studies with results that questioned their effectiveness more than they confirmed it (e.g. transactional analysis). Finally, the analysis showed that a large group of treatments lacked any confirmation of effectiveness (e.g., Jungian analytic psychotherapy). The publication was heavily criticized, especially by the psychoanalytic community, who felt they were being victimized.

Another effort to resolve the controversy over what treatments could be practiced with reasonable assurance of scientific evidence of efficacy was *A Guide To Treatments That Work* by Nathan and Gorman in 1998, commissioned by the Board of Directors of the Division of Clinical Psychology of the American Psychological Association. The editors defined six types of studies, varying in methodological rigor, and ordered them in a way that made it clear to authors to which degree they should rely on their findings (Beutler 2000). The authors then had to inspect the literature and "present the most rigorous scientifically based evidence for the efficacy of treatments that is available". Although the book offered a set of standards by which to evaluate research, it left the exact classification of evidence to the authors, and thus may have introduced some author bias.

A landmark effort to apply a previously defined standard to the body of empirical research was certainly the work of the Task Force on the Identification and Dissemination of Empirically Validated Treatments. The task force was commissioned by the Society for a Scientific Clinical Psychology, a section of the Division of Clinical Psychology. In 1995, they published a first report, Task Force on Promotion and Dissemination of Psychological Procedures, issuing subsequent addendum reports in 1996 and 1998 (Chambless et al. 1996; 1998). In 1998, the task force became a standing committee of the American Society of Clinical Psychology (Beutler 2000). Initially following the FDA criteria, the task force developed a set of criteria by which to rate an entire body of research, rather than individual studies. Thus, a treatment was considered to be empirically valid, if it had been found to be superior to a placebo or no-treatment control group in two independent studies. In the following years, responding to critics, the criteria were refined and somewhat broadened. The criteria initiated by this task force remain one of the most concerned with reliable classification, replicability, and objectivity (Beutler 2000).

Another group that aims to establish a solid evidence base for psychoanalysis is the Psychoanalytic Research Consortium (PRC). This is a group of psychoanalytic clinicians who complement their clinical activities with research efforts. The PRC is mainly interested in exploring the relationships between the processes of psychoanalytic work and the benefit to the patient, particularly the long-term benefit. "Analytic process scales" for investigating recorded sessions were developed. According to Waldron (1997), two steps have to be taken to demonstrate the efficacy of psychoanalysis

through process-outcome studies. First, they developed a precise and reliable rating system based on a manual. Second, they investigated whether their "Analytic Process Scale" (APS) allows for valid estimates of the relationships between specific dimensions of the therapeutic process and the outcome of treatment.

In 2002, the International Psychoanalytic Association (IPA)´s Committee on Psychoanalytic Research published its second *Open Door Review of Outcome Studies in Psychoanalysis*, covering many of the studies of the outcome of psychoanalytic treatment carried out in Europe and North America over the past decades (Fonagy et al. 2002). Randomized outcome studies are reported, as well as process studies. Gray (2002) comments on the results: "The team could find no definitive studies which show psychoanalysis to be unequivocally effective relative to an active placebo or an alternative method of treatment. Most reported studies had major limitations, which might lead critics of psychoanalysis to discount their results. Others had limitations so grave that even a sympathetic reviewer might be inclined to discount the findings." Nevertheless, the authors of the review overall concluded that the results have sometimes "yielded uncomfortable information, but sometimes they have confirmed and strengthened our belief in the appropriateness of our techniques" (Fonagy et al. 2002). They expressed their conviction that the psychoanalytic approach will be acknowledged as a valid and viable alternative for the treatment of mental disorders, in particular as information about the cost of mental illness and psychological distress is increasingly recognized.

Criticism of current strategies applying EBM to mental health

In the area of mental health, as in other areas of health care, attempts have been made to accommodate the requirements of EBM, and to respond positively to the challenges posed by it. At the same time, the increasing importance of EBM criteria in the evaluation and reimbursement of various treatment options for the mentally ill has caused a wave of intense criticism. Several major arguments have been brought forward to justify the rejection of the EBM approach in this field.

General concerns

First, EBM in general has been criticized on the grounds that the supposed methodological universality of randomized controlled trials (RCTs) does not hold true, and that there may be alternative paradigms which are relevant in scientific research. What are the reasons why the so-called "gold standard" in scientific research (i.e. RCTs) may be partly invalid for specific areas of research? As these criticisms apply to the application of EBM as a whole, they will only be briefly discussed in this section. A more in-depth analysis will be devoted to the criticisms specifically raised from a psychiatric and psychotherapeutic perspective.

Seen from a theoretical point of view, EBM has been characterized as a descendant of clinical epidemiology, which enables "rational clinical decision-making", and serves therefore as a kind of "auxiliary science" for the medical doctor (Perleth 1999). Although classical epidemiology is connected with social medicine and other more discursive medical disciplines, modern clinical epidemiology is strongly focused on the proper methodology of pharmacological studies. The rise of EBM as a powerful tool for the evaluation of the outcomes of pharmacological treatments seems to support the dissemination of the same criteria for the evaluation of other therapeutical interventions, despite the question whether these criteria are adequate for medicine on the whole.

Werner Bohleber (2000), for instance, criticizes the new position of power of Evidence-based Medicine for the definition of effectiveness, and the resulting decrease in the influence of psychoanalysts *themselves* to define an effective therapy:

"Required are verifiable proofs for the effectiveness of single therapeutic interventions and en-

tire therapeutic methods. The criteria for the examination are set in advance by an Evidence-based Medicine. EBM has formed a hierarchy of criteria for the evaluation of the object that is to be examined, which put the experimental, comparative, therapeutic study with randomized assignment of patients in the first place. This evaluation system has originally been developed for studies of effectiveness in the case of pharmacological treatments. Its basis is the natural-scientific, positivist model, which tries to measure effectiveness with clearly operationalizable effects that can be objectively determined." [Translation by the authors]

The focus on outcomes and empirical data is viewed as favouring the more natural-scientific fields of medicine. Many are concerned that this positivist-reductionist approach may lead in the end to a systematic bias against, for example, the evaluation of the outcomes of discussion therapy.

Still greater difficulties arise when several RCTs are pooled for systematic reviews and meta-analyses. While variations in procedure and patients may be adequately controlled in a single trial, the task of scrutinizing various protocols becomes much more delicate. Gross differences in methodology will lead to exclusion from the meta-analysis, but subtler procedural nuances are more difficult to detect and are a potential source of inaccuracy (Black 1998).

Another important methodological problem, particularly for the aggregation of trials, is the publication bias. This systematic bias arises from the reluctance of authors to report trials with negative results, and the difficulty of getting negative results accepted for publication (Gilbody and Song 2000). Yet, trials with a negative result may be just as relevant as a "positive" trial in forming an accurate picture (Grahame-Smith 1995).

Specific criticisms

Arguments that have been generated more specifically from a psychiatric and psychotherapeutic perspective raise the question whether the "hierarchy of evidence" of EBM, with RCTs at its top, is unquestionably transferable to psychotherapy with its specific demands for valid studies. Many authors, mainly from psychodynamic schools of psychotherapy, have argued that by their very nature, psychiatric patients and psychotherapy do not lend themselves to the methods of EBM.

First, ICD and DSM-defined diagnoses are perceived by many as being too exclusive and not sufficient in taking into account the complexity of psychiatric patients. By questioning the validity of ICD and DSM-defined categorical diagnoses, some authors even disapprove entirely of psychotherapy research relying on these categories.

Henry (1998) argues that research efforts are particularly flawed for personality disorders. According to him, most researchers involved in the study of personality disorders agree that the current categorical disorders are highly arbitrary. The comorbidity of Axis II disorders is so high, that it calls into question the fundamental nosology that research is required to follow. In fact, the average patient who qualifies for an Axis II diagnosis qualifies for an average of three distinct personality disorders (Fyer et al. 1988). This makes it virtually impossible to adjust for these "variables" in clinical studies. Horowitz (1994) recommends that grouping the subjects by Axis I diagnoses "opens an abyss of heterogeneity by an illusory promise of homogeneity". He calls for shifting away from "designs that foster horse races between schools of psychotherapy rather than new discovery of integrations between previously isolated camps."

In addition, many authors state that "scientific" psychotherapy research underestimates interpersonal and contextual factors, which have a thorough influence on the therapeutic outcome. Henningsen and Rudolf (2000), for instance, argue that randomized controlled trials are not adequately suited to find out whether a specific psychotherapy is effective. In their opinion, the crucial point is whether studies make a clear distinction between the tested variable and possible "confounding variables":

- "The clearer the experimental variable, i.e. the treatment effect, can be distinguished from surrounding, disruptive factors, like the form of application, the expectations of the patients etc., the more applicable RCTs are (for example, for the examination of drug effects). The more heterogeneous the population and the dysfunction that is to be treated, the more complex the possible interventions, the more relevant moral questions become (...), the more artificial is the distinction between specific and non-specific treatment effects (...) and the more important is the real, naturalistic context for the efficiency of health care delivery, and the more difficult it is to generalize the outcomes which result from experimental, context-free settings under very special conditions (...)" [Translation by the authors]

As Henningsen and Rudolf (2000) underline, RCTs are natural-scientific experiments, which are designed to find highly specific relationships between supposed causes and describable effects. Such experiments are more than adequate to verify the special effectiveness of a single substance in pharmacology. Can similar experiments take place in the case of a therapeutic relationship between psychotherapist and patient? In the case of the pharmacological study, it is clear that the administered substance is the supposed active agent. But how can one be sure that the cause of the patient's changed behaviour is the special kind of psychotherapy, and not the psychotherapist or other life circumstances? In contrast to the pharmacological study, such studies cannot be controlled by placebo, because this would require giving mentally ill patients a kind of pseudo-therapy, which would be unacceptable for ethical reasons. The comparison with a pharmacological therapy or a standard psychotherapy is possible, but it seems rather doubtful that all patients in a study would receive the same treatment. For instance discussion therapy essentially includes the individual dealing with the patient's problems.

As for the publication bias, which is a criticism of EBM in general, Gupta points out that psychotherapy might be especially disadvantaged: Publication bias tends to favour pharmaceuticals where the duration of treatment is often short, in comparison with what is needed for psychotherapy. Studies of short duration make it easier for participants to complete the trial, thus leading to a lower dropout rate and increasing the statistical power (Gupta 2003).

Margison raises still another point: He argues that it is not the psychotherapeutic technique, but rather the psychotherapist himself that matters most in terms of success rates (Margison 2000). In an extensive review of the outcome literature, Lambert (1992) concluded that at best 15 % of the variance in outcomes was attributable to the differential "technique", while the quality of the therapeutic dyad or contextual factors accounted for at least 30 % of the variance across different therapeutic approaches.

In fact, the therapeutic alliance, which is highly dependent on personal factors, has emerged as the most consistent predictor of outcome across many studies in different models of psychotherapy (Henry 1994). These factors thus need to be studied in their own right, rather than being "controlled" as error variance. This view is shared not only by psychoanalytic therapists but also by some behaviourists. Smith (1995) stated that it was precisely in recognizing and valuing the personal qualities of the therapist, that the American Academy of Psychotherapists had focused on the personal development of the therapist, and the use of the therapist's self as primary, with technique being of secondary importance.

A number of criticisms have been directed against the previously mentioned Empirically Validated Treatment (EVT) movement, because of its emphasis on standardized treatment manuals: "manualized theory is not the therapy" (Beutler 2000). Although it is commonly accepted by some authors that treatment can be advanced by the use of manuals, two main arguments against them have been brought forward. First, their development from single theories, and second the focus of conventional treatment manuals on particular symptoms or diagnostic groups.

Therapy manuals that are presently on the EVT list are almost exclusively behavioural and cognitive-behavioural, reflecting the divide between academic clinical psychology [tends to be dominated by cognitive and behavioural research] and professional practice, [where interpersonal and

psychodynamic approaches dominate] (Henry 1998). Yet most clinicians adopt an eclectic or multi-theory perspective, and adhering to a manual requires practitioners to constrain the application of their usual interventions (Beutler 2000).

The treatment manuals tend to target narrowly circumscribed disorders such as chronic headache, female orgasmic dysfunction, enuresis, specific phobias, erectile dysfunction, and irritable bowel syndrome. Henry (1998) argues that this spectrum of diseases is highly unrepresentative of a clinician's caseload in everyday practice.

Concerns have been expressed that an EBM approach could have serious effects on the breadth and scope of training in different psychotherapeutic methods. On the one hand, the financial incentive to provide evidence-based treatment is likely to lead to a decreasing number of students interested in studying and specializing in "unproven" treatment methods. On the other hand, the above-mentioned focus on the technique rather than on the psychotherapist himself might result in the neglecting of the development of the personal qualities of future therapists (Smith 1995).

Another criticism deals with outcome criteria. In 2001, Westen and Morrison published a widely debated study on the efficacy of manualized psychotherapies. In their work, they performed a multidimensional meta-analysis of 34 studies, that bear on the clinical utility and external validity of empirically supported therapies, published between 1990 and 1998 in top peer-reviewed journals.

They looked at 17 trials of panic disorder, five trials of generalized anxiety disorder (GAD) and 12 trials of depression including a total of 2, 414 subjects. Other trials were considered, but were excluded because they "did not meet minimal criteria for randomized controlled trials."

The results suggest that a substantial proportion of patients with panic disorder improve and remain improved; that treatments for depression and GAD produce impressive short-term effects, but that existing data do not support long-term effects and that screening procedures used in many studies raise questions about generalizability, particularly in light of a systematic relation across studies between exclusion rates and outcome.

Westen and Morrison stress the importance of reporting, in both clinical trials and meta-analyses, a broader range of outcome indices that provide a more comprehensive, multidimensional portrait of treatment effects and their generalizability. These include exclusion rates, percent improved, percent recovered, percent that remained improved or recovered at follow-up, percent seeking additional treatment at follow-up, and data on both complete and intent-to-treat samples. Aikins (2001), while cautious in the interpretation of the results of meta-analyses, also sees a need for psychotherapy studies to expand their inclusion criteria thus analyzing the influence of comorbid psychopathology (Lott 2003).

Westen and Morrison (2001) essentially argue that researchers and policy-makers may have moved too quickly to deem therapies as successful or superior to other forms of untested treatment, when patients still have clinical or subclinical symptoms at the end of the study, and when long-term prognosis is not investigated thoroughly enough. "The requisites of doing good science from a controlled clinical trial point of view – brief, manualized treatments that are as close to identical across subjects as possible – has made it virtually impossible for anything but a small range of treatments to be tested".

EBM and health policy

Besides the methodological and conceptual criticism concerning the application of EBM criteria for the evaluation of treatment, serious consequences for the future availability of a broad spectrum of treatment options for the mentally ill are dreaded. Psychotherapeutic medicine is facing an ever-increasing pressure from third-party payers to be based on scientific evidence. This phenomenon is not specific to psychotherapy, but part of intense efforts throughout all medical fields to contain the constantly rising costs in the health care sector. For example, Eddy suggested that healthcare funds

be required to cover only those interventions that can be expected to produce their intended effects (Eddy 1996).

As in other disciplines, a selection process among competing treatment options will take place. A distinctivness of psychotherapy, however, is its extremely varied offer of treatment methods. Vigorously applying evidence-based purchasing criteria could mean narrowing the individual patient's choice down to a few reimbursed treatment forms, other therapies not being available just because they do not represent the "best buy".

Limiting the range of therapeutic approaches might not be in the best interest of many of the patients, however. As Henry (1998) points out, most clinicians feel that the increasingly dominant influence of health care managers on treatment decisions is overly intrusive and restrictive. Managed care agents "are becoming our de facto supervisors – telling us what can and what can' t be done". This has potentially detrimental effects on the therapist-patient dyad, the one variable that is shown to be the most important predictor of outcome: Therapists that are forced by managed-care dictates to perform certain types of therapy within certain time limits may often feel less invested in the treatment, and as a result, their client. On the other hand, the patients, left with little choice of treatment options, could engage hesitantly in forms of therapy they deem not suitable for them. Thus, the outcome of such therapies could be seriously weakened.

Moreover, funding only a few types of psychotherapy may have detrimental effects on the development of other psychotherapeutic approaches. In 1997, the British Department of Health announced the financing of one specific type of therapeutic communities as part of providing specialist services for patients with a severe personality disorder. Kiseley (1999) points out that such favouring one particular treatment approach may lead to a number of unfortunate consequences. The funded treatment will have a competitive advantage over therapeutic approaches that do not receive such financial support. Thus, the development of alternative, possibly even more cost effective provisions of care for mentally ill patients, may be hampered.

There is a systematic problem underlying these considerations: Evidence-based Medicine may introduce a systematic bias, resulting in the allocation of resources to those treatments for which there are funds available to show effectiveness (such as new pharmaceutical agents). In areas where rigorous evidence is not available or not attainable for lack of funding, such as psychotherapy, it may only be a short step from the notion that a therapy is "without substantial evidence" to being thought to be "without substantial value" (Kerridge 1998).

Yet there are serious doubts that applying evidence-based purchasing criteria will indeed yield the desired cost savings by health insurance companies. On the one hand, given that many clinical trials are conducted by drug companies, implementing EBM-supported interventions may even lead to increased costs, as pointed out by Sackett and colleagues (Sackett et al. 1996). Practitioners might be obliged to always prescribe the latest, most expensive drug, even if it only has a small but statistically significant clinical benefit.

On the other hand, within psychotherapy the general use of guidelines would require either that a very large number of practitioners be available to represent the various methods identified, or that a practitioner be both trained in several different theories and models. They must also be willing to apply them all according to a different manual when the suitable patient is identified (Beutler 2000). Providing these kinds of services might lead to the "economic paradox" of EBM (Norman 2003), with increasing costs rather than savings for managed care companies.

Ethical reflections on the use of EBM in psychiatry and psychotherapeutic medicine

Despite some warranted conceptual and methodological criticism, the question today is not whether third-party payers will or will not insist on the "proof" of evidence-based therapies. The questi-

on is rather which type of evidence will be regarded as sufficient to justify inclusion in the package of remunerated services. This process seems legitimate insofar as this serves to sort out some therapies that are identified as verifiably ineffective or as having seriously adverse effects. However, when resources are being allocated on an EBM basis, it has to be clear that a lack of available evidence or of certain levels of evidence does not a priori mean a lack of therapeutic effectiveness but can be due to a variety of obstacles, some of them removable (like lack of funding for studies in a certain area), some of them persistent (like methodological difficulties of doing randomized controlled trials with certain populations). Taking these obstacles into account in an appropriate manner will be an important challenge that proponents of EBM and policy makers have to tackle together if EBM is supposed to lead to better and fairer treatment in mental health.

Consequences of an exclusive reliance on EBM criteria for the quality of psychiatric/psychotherapeutic care

As we have seen above, allocating resources according to the availability of "high level" evidence can imply possibly unjustified restrictions for the provision of psychiatric and psychotherapeutic care, at the meso- as well as at the micro-level. At the meso-level, in addition to the aspects discussed above, EBM could strongly influence the institutional setting of care. For instance, given its high cost, inpatient psychotherapy could face increasing pressure from third-party payers who favour less costly outpatient treatments. Without studies that clearly demonstrate the benefits of inpatient psychotherapy, insurance companies will at least try to cut down on the length of hospital stay of psychotherapy patients if not abandon reimbursement altogether for some diagnostic entities. This has potentially serious implications for hospital departments specialized in psychotherapy and could lead to a restructuring of institutional settings, at least in countries where inpatient psychotherapy plays a major role, for example in Germany.

At the micro-level, faced with mounting pressure from insurers, doctors that are bound to adhere to cost-effective treatment options could increasingly favor "EBM-proof", pharmaceutical treatments or short-term behavioural interventions, over more extensive psychotherapy for the individual patient.

Such a development would narrow down the range of available treatment options, limiting the patient's freedom of choice and the likelihood that the treatment offered matches well with individual expectations and preferences. Psychiatry had responded to the complex needs of psychiatric patients by elaborating a broad spectrum of techniques, from which a highly individualized treatment offer could be put together, which could meet the particular requirements of each case. An exclusive reliance on EBM criteria that does not take the special features of mental health research into account, would risk abolishing past achievements of the field. This would mean further disadvantages to a group of patients that is already suffering from stigma and discrimination.

Consequences for equity and fairness in mental health

Cross-national surveys show that common mental disorders are about twice as frequent among the poor as among the rich (Patel et al. 1999). In the US, children from the poorest families were found to be at increased risk of disorders at the ratio of 2:1 for behavioural disorders and 3:1 for comorbid conditions (Costello et al. 1996). Moreover, some studies show that the course of mental illness is determined by the socioeconomic status of the individual (Kessler et al. 1994; Saraceno and Barbui 1997). This finding could be in part due to service-related aspects and access to treatment, as poverty and associated factors such as lack of insurance coverage can create a big treatment gap (World Health Organization 2001).

As there are already fewer treatment options available to the poor, the consequences on equitable access to services of applying EBM criteria to the provision of psychotherapy have to be scrutinized. Shifting resources away from "unproven" treatments towards pharmaceuticals could indeed have particularly adverse effects on the treatment options available for marginalized groups, thereby aggravating the existing inequalities in psychotherapeutic care. If EBM is to be used as a cost-cutting tool, further narrowing down the catalogue of treatments reimbursed by health insurances, this will disproportionally affect patients who cannot afford to pay out of their own pocket. This is especially true for long-term interventions, that are frequently in psychotherapy. A development towards a future situation where a great spectrum of treatment methods is still available and affordable for the affluent, with few options left for the poor, could be an unintended and undesirable effect of applying EBM in an uncritical manner to mental health.

Broadening the evidence base

The field of psychotherapy, with its highly heterogeneous patient collective and importance of interpersonal and other contextual factors, calls for the acknowledgment of a stronger role of "clinical experience" than in other clinical disciplines. Qualitative and observational studies should be given better recognition within the evidence base.

Many regard the quantitative research paradigm as reductionist (Miles et al. 2003). Eakin and Mykhalovskiy (2003) note a growing importance of qualitative research in exploring clinical decision-making, clinician-patient interaction and patient experiences of illness. While this has led to calls to incorporate these studies into the evidence base, it has been stressed that efforts to do so must not neglect the important differences between quantitative and qualitative studies. Most attempts at systematic review of qualitative research have simply borrowed and sought to impose a template designed for the evaluation of quantitative work (Barbour and Barbour 2003; Miles et al. 2003).

Further, research methods and designs for quantitative research can be refined to include the so-called "confounding variables" of the therapist-patient relationship, personality variables and patient preferences (Henningsen and Rudolf 2000). There is also a great need to expand the inclusion criteria of psychotherapy studies, in order to analyze the influence of comorbidity. As the study by Westen and Morrison (2001) suggests, a further necessity in psychotherapy research is the initiation of a greater number of prospective studies on long-term psychotherapy.

Conclusion

Many mental health professionals have resisted the temptation to just continue their work and wait for the EBM "nightmare" to pass (cf. Yalom 2002). Instead, psychiatry and psychotherapy have embraced up the challenge of proving the evidence-base of their therapeutic methods. Various research groups and professional societies and associations have developed databases, guidelines, and manuals adhering to EBM criteria.

At the same time, the fierce debate about the usefulness of the EBM approach to mental health continues. Various methodological criticisms, from a more general as well as from a specifically psychotherapeutic viewpoint have been raised against "RCTism". A further concern is that Evidence-based Medicine will be hijacked by purchasers and managers to cut the costs of care, thereby narrowing down the spectrum of available psychotherapeutic treatment options.

EBM has certainly had a stimulating effect on psychotherapeutic research and may have been useful in weeding out some clearly ineffective or even dangerous therapeutic practices. It is, however, by no means certain that the use of EBM will indeed lead to better and fairer treatment in

mental health. If EBM is to serve as an appropriate and defensible tool to inform decision-makers and practitioners, it will have to take the specificities of mental health into account, such as highly heterogeneous patient collectives and the importance of interpersonal and contextual factors. This includes acknowledging a stronger role for "clinical experience" and observational studies than in other clinical disciplines.

References

1. Aikins DE, Hazlett-Stevens H, Craske MG (2001) Issues of measurement and mechanism in meta-analyses: comment on Westen and Morrison. Journal of Consulting and Clinical Psychology 69: 904-907
2. Amelang M (2002) [Psychologische Diagnostik und Intervention] Psychological diagnostics and intervention. P 481 ff. Springer, Berlin
3. Barbour RS, Barbour M (2003) Evaluating and synthesizing qualitative research: the need to develop a distinctive approach. Journal of Evaluation in Clinical Practice 9: 179-186
4. Berghmans R, Berg M, Van den Burg M, Ter Meulen R (2004) Ethical issues of cost-effectiveness analysis and guideline setting in mental health care. Journal of Medical Ethics, 30:146-150
5. Beutler LE (2000) David and Goliath. When empirical and clinical standards of practice meet. American Journal of Psychology 55: 997-1007
6. Black D (1998) The limitation of evidence. Journal of the Royal College of Physicians of London 32: 23-26
7. Bohleber W (2000) Symptom – Konflikt – Struktur. Rückkehr einer alten Debatte: Psychoanalyse als Behandlungsmethode im Spannungsfeld zwischen Störungsspezifität und krankem Individuum. Eröffnungsrede zur Herbsttagung der Deutschen Psychoanalytischen Vereinigung vom 21.-24.11.2000 in Bad Homburg. www.dpv-psa.de/html/events/archiv/hg2001/presseinfo.htm
8. Chambless DL, Baker MJ, Baucom DH, Beutler LE, Calhoun KS, Crits-Christoph P, Daiuto A, DeRubies R, Detweiler J, Haaga DAF, Johnson SB, McCurrt S, Mueser KT, Pope KS, Sanderson WC, Shoham V, Stickle T, Williwmas DA, Woody SR (1998) Update on empirically validated therapies. Clinical Psychologist 51: 3-16
9. Chambless DL, Sanderson WC, Shoham V, Johnson B, Pope KS, Crits-Christoph P, Baker M, Johnson B, Woody SR, Sue S, Beutler LE, Williams DA, McCurry S (1996) An update on empirically validated therapies. Clinical Psychologist 49: 5-14
10. Costello EJ, Angold A, Burns BJ, Stangl D, Tweed D, Erkanli A, Worthman CM (1996) The Great Smoky Mountains Study of Youth: goals, design, methods, and the prevalence of DSM-III-R disorders. Archives of General Psychiatry 53: 1129-1136
11. Duehrssen A, Johrswieck E (1965) Eine empirisch-statische Untersuchung zur Leistungsfaehigkeit psychoanalytischer Behandlung. Nervenarzt 36: 166-169
12. Eakin JM, Mykhalovskiy E (2003) Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences. Journal of Evaluation in Clinical Practice 9: 187-194
13. Eddy DK (1996) Benefit language; criteria that will improve quality while reducing costs. JAMA 275: 650-7
14. Fonagy P, Clarkin J, Gerber A, Kächele H, Krause R, Jones EE, Perron R, Allison E, (2002) An Open Door Review of Outcome Studies in Psychoanalysis. International Psychoanalytical Association, London
15. Fyer MR, Frances AJ, Sullivan T, Hurt SW, Clarkin J (1988) Comorbidity of borderline personality disorder. Archives of General Psychiatry 45: 348-352
16. Gilbody SM, Song F (2000) Publication bias and the integrity of psychiatry research. Psychological Medicine 30: 253-258
17. Grahame-Smith D (1995) Evidence-based Medicine: Socratic dissent. British Medical Journal 310: 1126-7
18. Grawe K, Bernauer F, Donati R (1994) Psychotherapie im Wandel – Von der Konfession zur Profession. Hogrefe, Göttingen
19. Gray SH (2002) Evidence-based psychotherapeutics. Journal of the American Academy of Psychoanalysis 30: 3-16
20. Gupta M (2003) A critical appraisal of Evidence-based Medicine: some ethical considerations. Journal of Evaluation in Clinical Practice 9: 111-121
21. Henningsen P, Rudolf G (2000) Zur Bedeutung der Evidence-based Medicine für die Psychotherapeutische Medizin. Psychotherapie, Psychosomatik, Medizinische Psychologie 50: 366-375

22. Henry W (1998) Science, politics, and the politics of science: the use and misuse of empirically validated treatment research. Psychotherapy Research 8: 126-140
23. Horowitz MJ (1994) Psychotherapy integration: Implications for research standards. Psychotherapy and Rehabilitation Research Bulletin 3: 8-9
24. Kerridge I (1998) Ethics and evidence based medicine. BMJ 316: 1151-3
25. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS (1994) Lifetime and 12-month prevalence of DSM-II-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Archives of General Psychiatry 51: 8-19
26. Kisely S (1999) Psychotherapy for severe personality disorder: exploring the limits of evidence based purchasing. BMJ 318: 1410-12
27. Lambert MJ (1992) Psychotherapy outcome research: Implications for integrative and eclectic theories. In: Norcross JC, Goldfried MR (Eds.) Handbook of psychotherapy integration. Basic Books, New York, P 94-129
28. Lott DA (2002) Are Studies Misguiding the Choice of First-Line Treatments? Psychiatric Times XIX: http://www.psychiatrictimes.com/p020301b.html
29. Margison F et al. (2000) Measurement and psychotherapy. British Journal of Psychiatry 177: 123-130
30. Miles A, Grey JE, Polychronis A, Price N, Melchiorri C (2003) Current thinking in the evidence-based health care debate. Journal of Evaluation in Clinical Practice 9: 95-100
31. Miller RB (2001) Scientific vs. clinical-based knowledge in psychology: a concealed moral conflict. American Journal of Psychotherapy 55: 344-56
32. Nathan PE, Gorman JM (Eds.) (1998) A guide to treatments that work. Oxford University Press, New York
33. Norman G (2003) The paradox of Evidence-based Medicine. Commentary on Gupta (2003), A critical appraisal of Evidence-based Medicine: some ethical considerations. Journal of Evaluation in Clinical Practice 9: 111-121
34. Patel V, Araya R, de Lima M, Ludermir A, Todd C (1999) Women, poverty and common mental disorders in four restructuring societies. Social Science and Medicine 49: 1461-1471
35. Perleth M (1999) Evidenz-basierte Medizin: eine Einführung. In: Perleth M, Antes G (Eds.) Evidenzbasierte Medizin. Wissenschaft im Praxisalltag. Medien & Medizin Verlag, München, P 13-18
36. Psychoanalytic Research Consortium. http://www.psychoanalyticresearch.org
37. Sackett DL et al. (1996) Evidence-based Medicine; what it is and what it isn't. BMJ 312: 71-72
38. Sackett DL, Straus SE, Richardson WS (2000) Evidence-based Medicine: How to Practice and Teach EBM. Churchill Livinstone, Edinburgh
39. Sifneos PE (1992) Short-term anxiety-provoking psychotherapy. Basic Books, New York
40. Smith EWL (1995) A passionate, rational response to the "manualization" of psychotherapy. Psychotherapy Bulletin 30: 36-40
41. Saraceno B, Barbui C (1997) Poverty and mental illness. Canadian Journal of Psychiatry 42: 285-290
42. Strauß B, Kächele H (1998) The writing on the wall – comments on the current discussion about empirically validated treatments in Germany. Psychotherapy Research 8: 158-170
43. Task Force on Promotion and Dissemination of Psychological Procedures (1995) Training in and dissemination of empirically validated psychological treatments: Report and recommendations. Clinical Psychologist 48: 3-23
44. Waldron S (1997) How can we study the efficacy of psychoanalysis? Psychoanalytic Quarterly 66: 283-322
45. Westen D, Morrison, K (2001) A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. Journal of Consulting and Clinical Psychology 69: 875-899
46. Widlöcher D (1998) Quality control, condensed analysis and ethics. International Journal of Psycho-Analysis 79: 1-11
47. World Health Organization (2001) The World Health Report 2001. Mental Health: new understanding, new hope. World Health Organization, Geneva
48. Yalom S (2002) The Gift of Therapy. Harper Collins, New York
49. Zepf S, Mengele U, Hartmann S (2003) The state of outpatient psychotherapy in Germany. Psychotherapie, Psychosomatik, Medizinische Psychologie 53: 152-162

9 Current Epistemological Problems in Evidence-based Medicine[1,2]

Richard E. Ashcroft

A quick sketch of Evidence-based Medicine

Evidence-based Medicine (EBM) is an important movement within medicine and health services, which has had considerable success over the past 30 years, in promoting critical scientific and practical awareness of the status of different claims to therapeutic knowledge. Its exponents can generally be characterised as having a strong ethical sense of the importance of avoiding unnecessary harms to patients, and improving health care in the interests of the general good. At the same time, critics of this movement have drawn attention to some alleged weaknesses of the principles and practice of EBM, many of which concern its epistemological credentials.

Epistemology, or the theory of knowledge, is the branch of philosophy concerning the definition of "knowledge", and the establishment of criteria for evaluating claims that something is known, either by individuals or by the community in general. This paper takes epistemological issues as its primary focus, rather than ethical or policy issues raised by EBM, in the belief that many of the latter issues turn, or have been made to turn, on questions of methodology in the evaluation and testing of treatments, in outcome measurement, and in evidence synthesis. When thinking about epistemological issues, it is important to note that raising foundational questions is not identical to raising a sceptical challenge. Philosophical scepticism is a method in epistemology, but it tends to undermine knowledge claims as such, rather than asking, as I do here, what particular methods of inquiry or appraisal do and do not achieve, and how they do so. As I shall show, there are many open questions in the foundations of EBM. I think the challenge here is to solve them, rather than to treat them as fatal objections to the very idea of EBM.

The Evidence-based Medicine movement is normally traced back to a series of lectures given in 1971 by the epidemiologist Archie Cochrane entitled "Effectiveness and Efficiency: Random Reflections on Health Services"(Cochrane 1972). Cochrane argued that too much medical care was using interventions of dubious or unknown safety and efficacy, causing harm at both individual and population levels, through iatrogenic injury, waste of resources, and failure to take up more effective treatments. He argued that treatments should be evaluated systematically, using unbiased methods of evaluation (such as the randomised controlled trial), and that individual practitioners, and the medical profession as a whole, should continuously review and appraise their own state of knowledge. This approach had a strong ethical imperative behind it, rooted in concern to do no harm; to do one's best for one's patients, and to do so justly by eliminating waste.

---

1 This chapter was first published as an article in the Journal of Medical Ethics, April, 2004, published by BMJ Publishing Group and is reprinted with permission by BMJ Publishing Group.

[2]Acknowledgement: The author thanks Ruud ter Meulen and Reidar Lie for their comments on a draft of this paper.

Since the programmatic outline of Evidence-based Medicine in Cochrane's lectures, various elements have been added, including cost-effectiveness analysis, a deepened focus on the types of outcome measures used in evaluation, an expanded range of "research synthesis" tools (notably meta-analysis of existing data sets), and a heightened attention to "patient-relevant" measures and to patient involvement in evaluation. Yet the essence of EBM is arguably the same as it was in 1972, viz., the use of Randomised Controlled Trials (RCTs) to produce (ideally) unbiased evaluations of treatments (and diagnostic tests, health service delivery systems, and so on). Since Cochrane's lectures, there has been a great deal of discussion of the so-called "hierarchy of evidence", a qualitative ranking of different types of evidential support for judgements of the clinical superiority of particular interventions over their comparators, which rests on the notion that it is possible to rank methods of inquiry by their susceptibility to bias. Alongside this discussion has been a discussion of how to combine different sorts of evidence, and how to compare different sorts of evidence (Sackett et al. 1991; Greenhalgh 2000; Egger et al. 2001; Stevens et al. 2001; Goodman 2002).

In the remainder of this paper, I shall review the main "live" issues in discussions of the foundations of EBM. I shall not give a detailed discussion of the criticisms of EBM, nor of the consequent ethical issues surrounding EBM, as these are discussed in detail elsewhere in this issue.

What is knowledge?

The standard account of knowledge in analytic philosophy is this:

- "Knowledge is justified true belief".

That is, for an individual X to know something (a proposition p), they must believe that p is treue, p must in fact be true, and they must have a valid justification for believing that p. For example, for a doctor to know a diagnosis, they must believe that the diagnosis is correct, it must actually be correct, and they must have a good reason for believing that it is correct. What counts as a good reason here is hotly debated. Much of the argument pro and contra Evidence Based Medicine turns on whether "clinical experience" or "diagnostic skill" are sufficient, as reasons for belief, that a patient has the particular condition or that a treatment will prove beneficial, or whether some further reason (such as good quality experimental evidence) is required. In other words: is "clinical intuition" ever self-justifying as a ground for a claim to know something?

This approach to defining knowledge was first proposed in Plato's Theaetetus (Plato 1987). In this dialogue Socrates sets up, then undermines this definition, by pointing out that it involves a vicious regress how do we know our justification of our belief? Since this time, a number of different sceptical challenges to particular knowledge claims, or the idea that we know anything at all, has been proposed, as has a number of different attempts to vary the classical definition in order to evade these challenges.

What is clinical knowledge?

The classical definition of knowledge gives a definition of what it is to know a proposition. As a number of philosophers have pointed out, most famously Gilbert Ryle, there are other kinds of knowledge, such as "know-how", which cannot be reduced to propositional knowledge (Ryle 2000). If one considers the kinds of knowledge which a clinical pathologist might have, for example he or she might know how to take a bacterial culture, which stains to use when looking for certain cellular structures under a microscope, what Staphylococcus Aureus looks like, when an infection is likely to be S. Aureus, when to perform the relevant diagnostic tests, and so on. Hence, clinical knowledge in-

cludes a range of know-how, scientific knowledge, knowledge of rules of practice and capacities for recognition and judgement. Some of these forms of knowledge are particularly resistant to formal analysis, although there has been quite a lot of work both in European and analytical philosophical traditions to try to clarify matters (Wittgenstein 1953; Heidegger 1962; Dreyfus 1992).

Many of the standing criticisms of EBM have turned on the role of these other non-propositional forms of knowledge, and their possible inscrutability to objective evaluation. However, it should be obvious that EBM is not designed to be a comprehensive account of medical knowledge, but only an account of that part of medical knowledge that is propositional. Secondly, while some knowledge that clinicians have could be characterised as capacities to make certain sorts of judgement reliably (such as the capacity to make and use a differential diagnosis), that a particular clinician, or a set of clinicians trained in a particular way, possesses these capacities is a proposition, which can be evaluated for its truth (Zagzebski 1996; Zarkovich and Upshur 2002). For example, the assertion that a particular doctor knows by clinical skill, experience, and judgement what is best to do for his patients looks epistemologically problematic. What is this faculty of knowledge, to which he lays claim? Singular knowledge claims, such as "this patient has this illness, and this treatment will be most beneficial under these circumstances" are very difficult to evaluate, precisely because they involve a faculty of judgement (the application of general rules to particular situations) (Kant 2000). However, the assertion that a doctor (or doctors in general) possesses the capacity to make such claims reliably can be evaluated both analytically and empirically. Analytically, while singular knowledge claims are defeasible because they can be false, the claim that one possesses a capacity to make such claims is no more than the claim that one can make such judgements at or above a certain threshold of reliability. One can then analyse whether this claim of capacity is true, and what makes it true (what its justification is). The claim that one possesses such a capacity is a propositional assertion. Empirically, there are various ways of evaluating whether someone in fact does possess this capacity, by comparing outcomes with other practitioners (using techniques of audit, epidemiology, and outcomes research), and there are ways of evaluating interventions designed to improve such capacities. While it is perhaps more difficult to evaluate claims of skill in medicine and surgery than it is to evaluate the outcomes of particular drug treatments, clinical judgement is nonetheless within the scope of Evidence-based Medicine's analytical techniques.

The content of clinical therapeutic knowledge

Concentrating from now on upon the part of clinical knowledge that is propositional, what sorts of propositions form the domain of clinical knowledge? Since EBM is concerned mainly with therapeutics, I will concentrate here on clinical therapeutic knowledge. This question is too broad to answer, but we can usefully distinguish between propositions about obtaining facts in particular situations and propositions about general truths. Evidence-based Medicine is concerned almost exclusively with the latter. EBM aims at the production and evaluation of law-like generalisations about diagnostic tests, treatments, and other health-care interventions. A typical statement in EBM might be that for condition C, the best evidence we currently have supports the use of treatment T as the most effective treatment for C.

Involved in this assertion are a number of epistemologically interesting claims:

- *E1 T is effective for the treatment of* C – an unqualified statement about T's effectiveness.
- *E2 T is more effective for the treatment of C than other treatments we know of* – of all the treatments for C we have, T is in fact the best, independently of whether we really know that this is the case.

- *E3 T is more effective for the treatment of C than other treatments we know of on the evidence we have at present* – of all the treatments for C we have, the evidence we have indicates that T is the best.

E1 raises an interesting metaphysical question: what is effectiveness? Elsewhere I have proposed an analysis of "effectivenesses" as properties of treatments, defined relative to specific therapeutic ends (Ashcroft 2002). These properties are best understood as causal powers or dispositions (Cartwright 1983). This analysis then prompts two questions, not widely discussed in philosophy of medicine:

- E1A What makes T effective in the treatment of C?
- E1B What therapeutic ends can properly define effectivenesses?

E1A is a question about how the "clinical" property T's-effectiveness-in-treating-C relates to the physical structure of T. E1B is a question about what sorts of ends can properly be understood as being caused by T's "clinical" properties. For example, can treatments properly be understood to cause alterations in patient's quality of life? What sort of mechanism is involved in causing an alteration in someone's quality of life? While this question does take us into very deep metaphysical waters, the clinical point is a simple one – treatments are alleged to bring about all sorts of effects (patient satisfaction, raised CD4 counts, improved 5 year survival rates), and indeed many different sorts of endpoints are used in clinical trials. Are all of them really measurable and comparable in the way physical endpoints are? We will return to this point when we consider the type of knowledge that clinical trial designs can deliver. In this context, it is useful to recall Austin Bradford Hill's famous criteria for identifying causal relationships in clinical epidemiology and clinical trials, and his requirement that there be a "biologically plausible" mechanism connecting putative cause and putative effect (Hill 1991). While often referred to, this point is sometimes more honoured in the breach than in the observance. The point here is that there are serious questions in the metaphysics of medicine, and the foundations of clinical sciences that we have hardly begun to pose and deserve further thought. As an example, consider the way, in which "cost effectiveness" is attributed to treatments as if they were properties of the treatment, when, at best, they are properties of treatments in the context of a particular clinical and economic system.

Clinical therapeutic knowledge as question-relative knowledge (Jardine 2000)

This takes us to a consideration of the meaning of statement E2. It is relatively unusual to gain categorical knowledge in medical science. We can rarely, if ever, say that T is the treatment for C. Even when we can, T is generally compared to a reference class of other possible treatments for T, and may indeed have been formally evaluated through comparative trials against other members of this reference class (including the use of placebo and doing nothing). Two points are important here:

- E2A T's effectiveness is judged superior to the other treatments in the reference class against a specific endpoint.
- E2B T's reference class is defined both by the endpoint and by the set of options available at the time of assertion.

In E2B "availability" is just as problematic in EBM as it was in the debate about the choice of control group in trials in the 2000 revision of the Declaration of Helsinki – what is meant by available? In maximal terms, however, availability here must mean something like "theoretically possible, given the total state of medical knowledge now".

This approach to interpreting E2 can be taken in two ways. The first, and simpler, way is this: E2 is a statement to the effect that treatment T is the most effective treatment for condition C of which we have good reason to be aware. The technical difficulty here is that statements about the effectiveness of T turn out to be statements about our knowledge of T, rather than statements about T directly. The second, more complex, way of interpreting E2 is this: the effectiveness of T is essentially relative to our background knowledge of T and its reference class of alternative treatments. This is to handle our knowledge of T and its properties in a way akin to nineteenth century Idealism, according to which all knowledge is relational, propositions are fictional statements, and we can have true knowledge only of the total system of beliefs and their relationships (Bradley 1930; Vaihinger 2000). While neither of these alternatives is all that attractive from the point of view of common sense, the core of both is the following idea. All our statements about the clinical effectiveness of a treatment are provisional, and asserted in the light of existing evidence. There is a theoretical limit, according to which, all our statements of effectiveness would become categorical statements, when all the evidence is in and all reference classes for comparison become absolute reference classes (all conceivable alternative treatments). Under these conditions, statements about clinical effectiveness would become true or false assertions about the treatments themselves directly (rather than reports on our state of knowledge). Much the same processes would also be gone through to refine our disease concepts and our aetiological knowledge as well, of course. This approach to grasping the nature of effectiveness is a fairly standard strategy in pragmatist and realist philosophy of science, which presupposes that our current scientific "knowledge" is fallible (and may indeed be mostly false), but that "at the end of inquiry" we will have a true representation of the world in all its fine structures (Jardine 1986; Psillos 1999). While this theory has many difficulties, the challenge it presents, of determining what theoretical structure the body of clinical knowledge has and how to determine the truth of clinical propositions, remains pertinent.

Clinical therapeutic knowledge, the ethics of belief and probability kinematics (Morton 2002; Jeffrey 1983)

Statement E3 focuses our attention not on the theoretical structure of clinical knowledge, but on the structure of individuals or communities beliefs about clinical effectiveness at a given time. In other words, it addresses how individuals or communities should maintain their stock of beliefs about what works in medical treatment. There are two different dimensions to this: how should beliefs be updated in the light of new evidence; and what sort of evidence should be sought.

The updating problem is very interesting, and has ramifications for philosophy of science more generally. When should new evidence be sought? This has two elements: when, as a practitioner, should I seek to update my own knowledge base? The rational individual will not update his or her knowledge continuously, since rational (human) individuals are finite beings, and information has costs in time and other resources. If this is the case, what updating heuristic ought the individual adopt? Kenneth Goodman (in philosophy) and others (in the theory of meta-analysis and elsewhere) have thrown some light on this, but it is a question that is as important as I suspect it may be intractable to theoretical analysis (Op. cit. N.6.). Second, how often should the scientific community update its knowledge base and synthesise what "it" knows? Again, this may be an intractable question, but practical steps can be taken in terms of evidence synthesis through such groups as the Cochrane Collaboration. There remain a wide range of "technical" problems in the theory of evidential support for theories, but many of these appear to have no practical consequences (Maher 1993; Kaplan 1996). This appearance may be deceptive. For example, while clinical epistemology appears to concentrate on "proof" or "refutation" of singular propositions, this is a questionable assumption, for statements of therapeutic efficacy should probably be understood as law-like statements, rather than singular statements. As such, they are theoretical statements, albeit at a low level of abstraction,

and so squarely in the domain of problems such as how a theory is to be tested, what counts as a fair test, when evidence can be said to corroborate a theory, and so on. Much of the intellectual difficulty facing clinical therapeutics, I argue, is that the theoretical structure of medicine as an autonomous science (as opposed to a collection of knowledge drawn from diverse more basic sciences) is generally opaque to investigators, and there is the apparent possibility of testing propositional claims one at a time, rather than recognising the role they play in a web of theoretical and empirical commitments.

Evidence, theory and Evidence-based Medicine

The theoretical opacity of medicine leads us to the question of what sort of evidence should be sought for testing propositions of clinical therapeutic effectiveness. Consider the following problem: when should we regard a clinical therapeutic proposition as proven? Given the provisional nature of any such statement when framed as a statement of type E3, the answer may well be "never". However, statements of type E3 are always proposed relative to a reference class of treatments. Does this solve the problem? No, not really, because it is always possible to require further test or up the evidential ante by requesting a more robust or reliable experimental design or a more extensive data set for meta-analysis. Thus, there is a question of epistemological "good behaviour" to solve, which is – when is proof sufficient? At any given time, we could stop – and, as we shall see, there is a broad consensus that a certain type of clinical evidence is generally taken to be sufficient. But frequently parties may dispute whether this is the proper way to resolve the cognitive and therapeutic dispute.

The so-called "gold standard" of clinical evidence is the properly controlled and appropriately "powered" randomised controlled clinical trial (RCT), with appropriate "blinding". (I dislike the phrase "gold standard", as it has a confusing and highly misleading economic meaning irrelevant to this context). The role of the RCT in EBM has been controversial, for a number of reasons. Firstly, many critics hold that RCT evidence may sometimes be unattainable for methodological or ethical reasons, and secondly, they hold that the so-called "hierarchy of evidence" downgrades other sorts of clinical evidence, and provides no way of integrating them into an overall assessment of the evidence for the effectiveness of treatments. Thirdly, the RCT is methodologically wedded to a particular theory of statistical inference, which many statisticians and doctors dispute. And fourthly, the RCT is almost purely a "methodological" solution to clinical epistemology, in that it is "blind" to mechanisms of explanation and causation.

The first two objections have been discussed in many places, and I will not go over these again. It is true that some adherents of the EBM approach have been over-enthusiastic about what can be tested with RCTs, the supposed "meaninglessness" or poor quality of other sorts of evidence, and the ethical superiority of the RCT over other sorts of design and approach to treatment under clinical uncertainty or equipoise (Worrall 2002a; 2002b). Indeed, there are significant questions about the possibility of ascertaining whether or not equipoise obtains and what follows from it epistemologically and ethically (Ashcroft 1999; Veatch 2002). Nonetheless, few critics will deny that in general the RCT does give reliable and robust evidence, and that it has its place in clinical research. Criticisms of the RCT as a methodology can be found, however, which turn on a linked series of problems concerning the theories of inference used to frame and interpret RCTs. Firstly, RCTs produce statements about the truth of E3-type statements within a given confidence interval (usually 95%). The methodology of RCTs is essentially comparative, so that while RCTs sometimes permit us to make estimates of the magnitude of effectiveness, which is not their main purpose. Much of the criticism of RCT methodology depends on this apparent dependency upon "classical" theories of statistical inference (which provides the foundation for talk of "confidence intervals" in the first place) (Edwards et al. 1997; Howie 2002). Critics of this methodology argue that the RCT requires us to collect unnecessarily large sets of data, binds us to excessively large control groups, and re-

quires us to continue with trials for too long, both when there is evidence of danger to patients and when there is evidence of superior effectiveness. They base these arguments on Bayesian theories of inference, which permit regularly "updating" of degrees of belief in the truth of our E3 statement. Indeed, more than this, they hold that statistical information is always and only about rational subjective degrees of belief, rather than measures of objective probability; and indeed that the notion of objective probability as tied to the RCT is meaningless. This controversy will arguably never be settled, and indeed Donald Gillies argues that any reasonable theory of probability must allow both for objective chances (as in physics), and subjective degrees of belief (as in psychology), and must live with the grammatical problems involved in trying to speak of both using the same basic language (Gillies 2000). However, allowing for this diversity of interpretations causes us to ask what the nature of a statement of clinical evidence actually is: a measurement of an objective probability, a statement of rational personal (subjective) probability, or a statement of rational collective (intersubjective) probability? In addition, there is a classical problem of statistical measurement which is unresolved in EBM, which is whether statistical experiments generate true causal knowledge or merely measurements of correlations.

The question of causation is important in a number of ways. First, part of the supposed superiority of placebo-controlled trials in purely scientific terms, is that it permits judgements about the causal efficacy of interventions in bringing about "their" effects and to some extent allows estimation of the size of those effects (Ellenberg and Temple 2000; Carpenter et al. 2002). This justification may not apply if RCTs do not test causal hypotheses, but merely establish correlations or contribute to the "probability kinematics" of degrees of belief. Second, the beauty of the RCT as a methodology is that it seems to operate at a level of scientific theory autonomous from the basic sciences. Apparently, we need to know little or nothing of pathogenesis or drug action in order for a RCT to be designed and implemented and (perhaps) interpreted successfully. Indeed, our theories at this more basic level could simply be wrong. So long as the results of the RCT give an answer to our E3 question (which treatment does better in this population?), measured by using a suitably well-defined and credible endpoint, then questions of mechanism and cause seem to drop out of the analysis. In this regard, RCTs are an admirably "pragmatist" methodology, in the metaphysical and epistemological sense of the term. Much of the appeal of RCTs to methodologists is the way that they can be used to test hypotheses about the effects of interventions in a very wide range of contexts, from clinical pharmacology to social welfare, even when our theories of how interventions bring about their effects may be murky or merely speculative (as in social policy, perhaps) (Davies et al. 2000; Oakley 1999). Most methodologists would challenge designs that have no prima facie theory to support them, but there is no strict methodological requirement for this. Bradford Hill's famous "biological plausibility" requirement, as a necessary condition for an intervention to be testable ethically by RCT, is best understood as a way of screening off obviously implausible treatments from test. Here though, as so often, what counts as "plausible" is contestable: classical examples include the plausibility or otherwise of psychoanalysis for neurosis, or the di Bella treatment for cancer (Grünbaum 1984; Italian Study Group for the Di Bella Multitherapy Trails 1999). The relationship between the theoretical structure of clinical science and the theoretical structures of the more "basic" sciences is as complex, and EBM may prove to be a major contribution to the establishment of clinical medicine as an autonomous scientific discipline. But, as noted above, it's own theoretical structure may be quite opaque. In any event, as Nancy Cartwright and others have suggested is the case in physics, we may be better off expecting medicine to produce a patchwork of phenomenal laws of relatively low generality, rather than a complete and consistent system of universal, metaphysically founded laws (Cartwright 1983).

Concluding remarks

In this paper I have tried to present a range of epistemological issues concerning Evidence-based Medicine and Randomised Controlled Trials. Many of these issues are highly technical; my purpose in drawing them to the reader's attention is to stimulate philosophical debate and research into these problems. But what the practice of medicine? My personal view is that most of these problems are quite generic problems in the philosophy of science: "foundational" questions, so to speak. As a patient, I would still prefer to be treated in the light of the best clinical evidence, and I would still prefer to be "randomised" in a well-designed experiment where genuine uncertainty prevailed about the status of possible treatments for my illness (Ashcroft 2000). Part of my rationale for this is simply to ensure that I benefit from the best treatment, given our state of knowledge at the time. However, as a philosopher, I would also mark certain scepticism about the idea that many of these "foundational" questions admit of metaphysical solutions (Dupré 1995). EBM is the best available bet, and by small methodological and analytical improvements we will make progress in the scientific basis of health services. However, the philosophical challenges to the foundations of EBM are important: methodological modesty is the order of the day.

References

1. Ashcroft RE (2002) What is clinical effectiveness? Studies in History and Philosophy of Biology and Biomedical Sciences 33C: 219-233
2. Ashcroft RE (1999) Equipoise, knowledge and ethics in clinical research and practice. Bioethics 13: 314-326
3. Ashcroft RE (2000) Giving medicine a fair trial. British Medical Journal 320: 1686
4. Bradley FH (1930) Appearance and Reality. Oxford University Press, Oxford
5. Carpenter WJ, Appelbaum PS, Levine RJ (2002) The Declaration of Helsinki and clinical trials: A focus on placebo-controlled trials in schizophrenia. American Journal of Psychiatry 160: 356-362
6. Cartwright N (1989) Nature's Capacities and their Measurement. Oxford University Press, Oxford
7. Cartwright N (1983) How the Laws of Physics Lie. Oxford University Press, Oxford
8. Cochrane AL (1972) Effectiveness and Efficiency: Random Reflections on Health Services. Nuffield Provincial Hospitals Trust, London
9. Davies HTO, Nutley SM, Smith PC (Eds.) (2000) What Works? Evidence-based Policy and Practice in Public Services. Policy Press, Bristol
10. Dreyfus HL (1992) What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press, Cambridge Mass
11. Dupré J (1995) Review of Kitcher The Advancement of Science: Science without Legend, Objectivity without Illusions. Philosophical Review 104: 147-151
12. Edwards SJL, Lilford RJ, Braunholtz D, Jackson J (1997) Why 'underpowered' trials are not necessarily unethical. Lancet 350: 804-807
13. Egger M, Davey Smith G, Altman DG (2001) Systematic Reviews in Healthcare: Meta-analysis in Context. BMJ Books, London
14. Ellenberg SS, Temple R (2000) Placebo-controlled trials and active-control trials in the evaluation of new treatments, part 2: practical issues and specific cases. Annals of Internal Medicine 133: 464-470
15. Gillies D (2000) Philosophical Theories of Probability. Routledge, London
16. Goodman KW (2002) Ethics and Evidence-based Medicine: Fallibility and Responsibility in Clinical Science. Cambridge University Press, Cambridge
17. Greenhalgh T (2000) How to Read a Paper: The Basics of Evidence-based Medicine. BMJ Books, London
18. Grünbaum A (1984) The Foundations of Psychoanalysis: A Philosophical Critique. University of California Press, Berkeley
19. Heidegger M (1962) Being and Time. (Trans. MacQuarrie J, Robinson E) Blackwell, Oxford
20. Hill AB (1991) Principles of Medical Statistics. In: Sackett D, Haynes RB, Guyatt GH, Tugwell P. (Eds.) Clinical Epidemiology: A Basic Science for Clinical Medicine. Little, Brown, Boston, P 284

21. Howie D (2002) Interpreting Probability: Controversies and Developments in the Early Twentieth Century. Cambridge University Press, Cambridge
22. Italian Study Group for the Di Bella Multitherapy Trials (1999) Evaluation of an unconventional cancer treatment (the Di Bella multitherapy): Results of phase II trials in Italy. British Medical Journal 318: 224-228
23. Jardine N (1986) The Fortunes of Inquiry. Oxford University Press, Oxford
24. Jardine N (2000) The Scenes of Inquiry: On the Reality of Questions in the Sciences. Oxford University Press, Oxford
25. Jeffrey R (1983) The Logic of Decision. University of Chicago Press, Chicago
26. Kant I (2000) Critique of the Power of Judgment. (Trans. Matthews E) Cambridge University Press, Cambridge
27. Kaplan M (1996) Decision Theory as Philosophy. Cambridge University Press, Cambridge
28. Maher P (1993) Betting on Theories. Cambridge University Press, Cambridge
29. Morton A (2002) The Importance of Being Understood: Folk Psychology as Ethics. Routledge, London
30. Oakley A (1999) Experiments in Knowing: Gender and Method in the Social Sciences. Polity Press, Cambridge
31. Plato (1987) Theaetetus. (Trans. Waterfield RAH) Penguin, London
32. Psillos S (1999) Scientific Realism: How Science Tracks the Truth. Routledge, London
33. Ryle G (2000) The Concept of Mind. Penguin, London
34. Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) Clinical Epidemiology: A Basic Science for Clinical Medicine. Little, Brown, Boston
35. Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R (2001) The Advanced Handbook of Methods in Evidence-Based Healthcare. SAGE, London
36. Vaihinger H (2000) The Philosophy of 'As If'. Routledge, London
37. Veatch RM (2002) Indifference of subjects: An alternative to clinical equipoise in randomized clinical trials. In: Paul EF, Miller FD, Paul J (Eds.) Bioethics. Cambridge University Press, Cambridge, P 295-323
38. Wittgenstein L (1953) Philosophical Investigations. (Trans. Anscombe GEM) Blackwell, Oxford
39. Worrall J (2002a) What evidence in Evidence-based Medicine? Philosophy of Science 69: S316-S330
40. Worrall J (2002b) What Evidence in Evidence-based Medicine. London School of Economics, London
41. Zagzebski LT (1996) Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge. Cambridge University Press, Cambridge
42. Zarkovich E, Upshur REG (2002) The Virtues of Evidence. Theoretical Medicine 23: 403-412

10 Coordinating the Norms and Values of Medical Research, Medical Practice and Patient Worlds. The Ethics of Evidence-based Medicine in 'Boundary Fields of Medicine'[1]

Rein Vos · Dick Willems · Rob Houtepen

Introduction

Evidence based medicine (EBM) is rightly at the core of current medicine. This holds not only from a scientific or clinical point of view, but also from an ethical point of view: if patients and society put trust in medical professional competency, and delegate all kinds of responsibilities to the medical profession on behalf of that competency, medical professionals had better make sure that their competency expresses the state of the art of medical science. What goes for the ethics of clinical trials also goes, broadly speaking, for the ethics of medicine as a whole: anything that is scientifically doubtful is, ceteris paribus, ethically unacceptable. This particularly applies to so-called boundary fields of medicine, that is, areas of medicine where medical research is weak and diverse, lacking financial incentives, and where the evidence regarding the aetiology and treatment of disease is much less clear than in laboratory and hospital-based medicine. Examples of such 'boundary fields' are physiotherapy, psychotherapy, medical psychology, and occupational health. In these fields, complex syndromes such as repetitive strain injury syndrome (RSI), whiplash, chronic low back pain, and chronic fatigue syndrome (CFS) are investigated. There is an urgent need for well-designed effect studies, preferably randomised clinical trials (RCT), to distinguish effective from ineffective therapies.

It appears that the primary ethical problem in this context is the lack of attention to the 'boundary fields'. The problem may be diagnosed as a lack of evidence per se, or a lack of evidence that is appropriate to these specific fields and patients. Especially when allocation decisions are linked to the availability of evidence, matters of justice are at stake. This is potentially more so because many doubt the seriousness or even the reality of some of the problems in these fields. As documented and supported by the medical historian Shorter, syndromes such as RSI, whiplash and CFS have been explained by primary and secondary gain from illness, by the need of people with troubled lives to be supplied with a legitimate diagnosis and by popular media crazes for 'mysterious diseases', especially if they affect young adults (Shorter 1992). Because of their complexity, their suspect reputation and the need for multiple coordinated interventions, health problems at the intersection of body, mind and society may not get the research and therapeutic effort they deserve. Although we agree that this issue deserves more attention as a matter of potential injustice, we want to argue that this specific form of complexity also calls for other models of analysis and evaluation. Our contention

[1]This chapter is an updated version of an article published in the Journal of Medical Ethics in April, 2004, published by BMJ Publishing Group and is reprinted here with permission by BMJ Publishing Group.

will be that, in order to do justice to the interplay of heterogeneous factors that are so typical to the 'boundary fields' we discuss, ethical models other than justice are required.

Firstly, we will analyse the standard way of reasoning why 'boundary' fields in medicine are such difficult areas for extending the EBM methodology and the ethical problems related to that. This way of reasoning will be called the 'intrusion model' of EMB, that is, EBM should intrude medical fields, which at present still lack an EBM-practice. We will qualify this analysis by using arguments from studies on therapies for chronic whiplash, chronic pain, chronic fatigue patients, and the repetitive strain injury (RSI) syndrome. Secondly, we will propose another way of reasoning this problem ethically, claiming that guidelines and instructions following from EBM have to coordinate different normative 'logics': the logic of clinical trials and other scientific methodologies, the logic of medical practice, and the logic of patient worlds, while each normative logic is beset with a set of normative issues. This will be called the 'coordination model' of EBM: diagnostic, therapeutic and other health care procedures have to coordinate different worlds of norms and values, those of scientists, doctors and other health care workers, and of patients. Thirdly, we will introduce the political philosophy of Laurent Thévenot and Luc Boltanski, who developed a theory on the origin of social conflicts and the solution thereof, as a clash respectively, as the coordination of different contexts of justification, as a way of dealing with different worlds of 'norms'. Thus, it can be shown that the analysed models are not contrary, but complementary models of EBM, and that they represent two complementary strategies of dealing with normative problems related to EBM. Finally, we conclude that the standard ethics of justice approach to EBM ought to be combined with a more procedural 'ethics of practice' approach, emphasising the communicative and deliberative aspects of the practice of health care.

The standard way of reasoning about evidence and ethics in boundary fields: the intrusion model

Discussions on boundary fields such as physiotherapy, medical psychology, occupational health and nursing are very often framed by the notion that such fields lack an evidence-based practice, but that attempts should be made to improve such practices. In many cases this does indeed seem to be the case. The Dutch Health Council stated in its 1999 report 'The Effectiveness of Physical Therapy', based on a systematic review covering 169 randomised clinical trials (RCTs), that in contrast with the widespread use in Dutch health care, there was little or no evidence on the effectiveness of electrotherapy, laser therapy or ultrasound therapy (Health Council 1999). The Health Council advised the professional organization of physiotherapists to revise their guidelines on the use of physiotherapy in the treatment of musculoskeletal disorders. Many other forms of physiotherapy can be shown to be effective, e.g. exercise therapy as an additional treatment to medication in Parkinson's disease (De Goede et al. 2001). Another example is cognitive behaviour therapy (CBT), which has been proven effective in randomised trials as a treatment of chronic fatigue syndrome (CFS), whereas somatic and pharmacological treatment have not. However, professional medical groups have largely neglected these results (Whiting et al. 2001; Prins et al. 2001). These examples show that boundary fields of medicine are not resistant to EBM methodology per se.

In many cases the problem is much more difficult, two types of problems are illustrative of this fact. The first is when medical versus psychological approaches conflict, the second is when medically unexplainable disorders are at stake. The case of chronic neck pain due to whiplash is illustrative of the first type. Chronic neck pain in the case of whiplash is frequently associated with psychological distress. Wallis et al. questioned whether these affective disturbances were the cause or effect of chronic neck pain (Wallis et al. 1997), and designed a clinical trial to test this hypothesis. If a medical intervention, i.e. radio frequency neurotomy, should improve the psychological distress, this would be evidence for the 'medical model', namely that the affective disturbances are epiphenomenal of the organic disorder. If improvement did not occur, then the psychological model

would hold in the sense that psychological factors could sustain pain complaints, even if the organic disorder had been 'cured'. Wallis et al found that psychological distress also improved in a subgroup of patients with fewer complaints of neck pain, thus confirming the medical model. This finding evoked fierce debate, initiated through a letter to the editor in the renowned journal Pain by Kendall (1998). Referring to the biopsychosocial model, these authors claimed that many 'feedback loops' between medical, psychological and social factors occur. EBM would have it that science should resolve this conflict and should collect the necessary evidence to decide which 'model' is correct, since evidence-based decisions have important consequences for diagnostics and treatment in health care of whiplash. The ensuing discussion shows, however, that this may not be very promising when theoretical models are at stake that cross the boundaries between different disciplines. Rather than a Popperian cycle of trial and error, such disputes are reminiscent of Kuhn's reflections on the incommensurability of paradigms.

The second type of problem is exemplified by the case of Chronic Repetitive Strain Injury Syndrome (RSI). RSI has evolved in a short period of time as a generally recognized serious condition (Health Council 2000). In many countries it is considered one of the important occupational risks for workers and employees in the administrative and business sector, potentially even the most important occupational risk. Many programs and preventive measures have been developed or applied. However, medical science has not found a decisive causal scheme of explanation for the syndrome. The number of categories of patients at risk from RSI is still expanding, but experts doubt or at least question the value of RSI-checklists. Most medical treatments and preventive strategies for RSI have not yet been proven scientifically and even the definition and categorisation of RSI differs widely between countries (Arksey 1998). In such cases it is very difficult to decide on what the 'evidence' should be. There is not even international consensus as yet on the very concept of RSI. Epidemiological classifications differ in different countries and figures on incidence and prevalence of RSI-like complaints may vary enormously. In different countries within the western world, people with similar complaints might end up with profoundly different diagnosis and treatment trajectories: preventive measures of all kinds and sorts; treatment by neurologists, orthopaedic surgeons, pain specialists or physicians in rehabilitation medicine; various kinds of physical therapy, stress psychologists and other psychotherapeutic approaches.

Indeed, both types of problem are complex, typical of boundary fields in medicine, and show how difficult it is to expand the EBM-methodology in boundary fields of medicine. In these fields it is much more difficult to define and collect the 'right evidence', due to: long period of time of the course of the disease, the availability of control groups, the interplay of different interventions at the same time, the lack of commonly agreed upon inclusion criteria or effect measures and many other factors showing the complexity of the application of EBM-methodology in these areas. However, attempts to expand the EBM-methodology into the daily practice of professionals and patients are framed in the same way as in standard areas of EBM: scientific evidence needs to be collected and translated into guidelines, protocols, instructions and procedures, which should then 'intrude' into the world of professionals and patients. The term 'intrusion' expresses that criteria, norms and values from the medical scientific world enter the worlds of professionals and patients. Typical of the intrusion model, is the assumption that the criteria from the original field need not to be challenged or adapted fundamentally in order to apply to the adopted field. This model risks encountering defensive reactions in terms of the autonomy of the 'adopted' field.

Based on the intrusion model, the role of ethics would be to assess the consequences of the spread of EBM principles and procedures and to raise criticism if such consequences propose potential threats to a just distribution of health care resources, patient autonomy and other regular ethical focus points. Preferably, this should include that implicit and unintended consequences are made explicit. A number of different recognised ethical frameworks might be employed to perform such an ethical assessment: consequential, deontological, utilitarian, communitarian, etc. These ethical problems have to be taken seriously and discussed. Our claim is that these ethical issues pose impor-

tant problems for the further expansion of EBM-methodology. However, boundary fields of medicine also exhibit another set of ethical problems related to the expansion of EBM-methodology.

The alternative way of reasoning about evidence and ethics in boundary fields: the coordination model

The challenge of collaboration and coordination

The assumption of EBM is that scientific medical knowledge is primarily associated with established outcomes from patient related research, and that guidelines are the best vehicle for introducing these scientific insights into medical practice. Undoubtedly, scientific evidence and guidelines are useful, but they are merely instrumental to the delivery of high quality patient care. The practice of boundary fields in medicine as discussed above shows, however, another assumption: the one-to-one relationship of the medical scientist versus the clinician (or general practitioner) does not fit the complex structure of boundary fields in medicine. Here, different groups of professionals have to collaborate and different perspectives have to be adapted. The literature on RSI, chronic pain and chronic fatigue shows the importance of an optimal collaboration and communication between different professional groups, such as physicians, psychologists, psychotherapists, and physiotherapists. Further, it is shown that patient participation has important effects on health outcome. Baker et al examined whether physical therapists seek to involve patients in therapeutic goal setting and, if so, what methods they use (Baker et al. 2001). The therapists stated that they believed it to be important to include patients in goal-setting activities, and that outcomes improved if patients participated. Further, patients indicated that participation was important to them. Another example, is a recent program in the Netherlands to include patient organisations in the assessment and evaluation of research proposals, one of which focussed on the evaluation of a rehabilitation therapy program in the patients' own neighbourhood (Vink 2001).

These insights regarding professional collaboration and patient participation parallel the growing awareness in EBM literature that EBM has a far wider meaning than in the early 1980's when it first came into use. The Dutch Health Council reports that EBM "is currently understood to incorporate clinical epidemiological data, meaningful deliberations of professionals' pathophysiological knowledge and clinical experience, together with patient preferences" (Health Council 1999 P 13).

If one focuses on the optimisation of patient care, then one has to distinguish between the scientific aspect, the professional knowledge and competence and social developments. This is important, particularly because health care practices have become increasingly interconnected, and have to deal with patients' increased understanding of medical issues and their desire to be involved in determining what constitutes good health care. As the Dutch Health Council notes: "Generally speaking, professional knowledge and competence are characterized by the skilful application of scientific knowledge to concrete situations or put another way: being able to translate from the generic to the specific. In the case of medical professionals, this translation process effectively boils down to integrating epidemiological information (whether or not it is incorporated in guidelines), patient-specific data (including expressed preferences) and a host of organizational preconditions" (Health Council 1999 P 14). Thus, the Council speaks of the 'learning professional' and states that it is vital "that 'learning professionals' also systematically establish and evaluate their own practice data, so as to build up a reservoir of practical knowledge which complements the external knowledge from patient-related epidemiological research" (Health Council 1999 P 14). Thus, we have to think about professionals as individuals with competence and governance, yet also as members of organizational networks. The literature on 'continuing professional education' (CPD) reflects this new way of thinking: not only scientific education and training are important, but also education on management, cooperation and social skills (Health Council 1999). This shows that norms and values from medical

science, clinical practice and patients have to be interconnected, which we think is useful to denote as the coordination model of EBM. Besides norms and values, different forms and sources of evidence also have to be communicated.

Ego-documents as a source of 'evidence'

From an epistemic point of view, 'boundary fields in medicine' show that evidence comes from different sources, runs through different contexts, and meets different perspectives. From an ethical point of view, a different set of ethical issues emerges related to how the patient's experiences and values can be met. This demand to meet perspectives, norms and values of patients and patient groups surpasses the discussion on the integration of patient perspectives, patient satisfaction and quality of life considerations in medical research. Boundary fields in medicine show this through an important source: ego-documents. Patients suffering from chronic diseases, both curable and incurable, explained or unexplained, are generally the authors of ego-documents (Hawkins 1993; Oderwald 1994; Frank 1995), of which an abundance is nowadays offered on the Internet. In general, ego-documents can function as a window on the life of patients. It is of special interest that the medical perspective is usually only one among several, and we can see patients struggle with the way to integrate medical information and advice into their self images, prospects and plans. For people with a relatively long history of several diagnoses, treatment-plans, surgery and recovery periods and encounters with many health care professionals, medical information and options will enter into a cluster of experiences and perspectives entirely different from patients with acute or singular complaints. From ego-documents we can learn that, especially in the field of 'unexplained' or 'controversial' diseases such as chronic pain and fatigue syndromes, many patients clamour for recognition of their status as experts due to experience.

The coordination of different worlds of norms

How should we consider the two analysed models of EBM in boundary fields of medicine? What is their relationship? What kinds of strategies are available for medical ethics to deal with these? The political philosophy developed by Thévenot and Boltanski might help us to clarify the analysed models of EBM in boundary fields of medicine.

In their *De la Justification*, Laurent Thévenot and Luc Boltanski developed a political theory, which analyses societal conflicts as the clash of two or more different contexts of justification, and they claim that all sorts of 'technical objects' may help to enable compromises for such conflicts, that is, these objects are able to coordinate these normative conflicts (Thévenot and Boltanski 1991). Their central case study is the buggy. Buggies have been designed to enable the transport of small children much more intensively than was possible with the classical baby carriage. Buggies are foldable and thus easily transportable by train, car, and airplane. The disadvantage, however, is the potential danger for the baby where the buggy has not completely unfolded. Safety and transportability, two norms from completely different worlds – one of childcare and one of consumer comforts – had to be coordinated. The finalized buggy did accomplish this task, through the mechanism by which the unfolding buggy is fixed, as demonstrated by its 'click'. Such a click is a frequently available and socially acceptable way to show that the coast is clear, as in the case of safety belts in cars and airplanes. The 'click' shows that one normative world – safety for the child – is ensured, and that another normative world can be initiated – transport of the child.

The philosophy of Thévenot and Boltanski has been applied to the field of science and technology studies in order to better understand why and how medical technologies are able to delegate tasks and responsibilities of physicians towards patients (Willems 2001). In many cases of medical

therapy, patients are being involved and made responsible for their own treatment. In the case of asthma, for example, the patients use the peak flow instrument to measure his or her breathing capacity, score the results of these measurements in daily books, and perform all kinds of diagnostic and therapeutic measures, mostly framed under the heading of 'self-management'. In such cases, medical scientific views may conflict with the perspectives of patients. For example, one of the interviews given by Willems explores the case of a 14-year old boy suffering from asthma who is active as a player in his school basketball team: "What I find so troublesome is not that I become breathless during the game; mostly this I can handle, I hang back in the field, I do not run continuously back and forth, and this will do. I am much more troubled by the wheezing after the game. In the dressing-room I see my playfellows look at me, then I have no mind to take out my spray apparatus; yes, I feel ashamed for doing so" (Willems 2001). In this particular case the spray apparatus does not succeed to adapt the norm of good asthma care with the norms of normal behaviour with friends of the 14-year old boy. Later on, the new turbuhaler, which is less tedious and much faster in spraying the antiasthmatic drug – by pulling the turbuhaler and 'clicking' the apparatus immediately enables the boy to inhale the drug – in such a way that it can be done without his playfellows noticing. Along the same line, using the reasoning of Thévenot and Boltanski, it can be concluded that there are two ways of looking at the 'effectiveness' of asthma technologies, one of looking at the consequences or effects of a particular technology (which intrudes the life worlds of physicians and patients), the other at how norms of science, practice and patients are connected (which are compromised and coordinated).

Our claim is that this is a very elegant model for discussing the ethical problems related to EBM, and interpreting the two sets of ethical problems related to boundary fields in medicine as analysed above. What is shown as 'technical artefacts', that is apparatus, can be extended to all kinds of technologies, including the 'soft' technologies of EBM, i.e. evidence-based guidelines, protocols, instructions and procedures. The important point here is that these two ways of evoking and dealing with ethical problems are complementary rather than opposite. The turbuhaler might be successful in terms of coordinating the specific norms of physicians and asthma patients, yet it might still have (unintended) consequences for physicians and patients in other respects which have not been dealt with, and which have to be explicated and valued (in terms of potential threats for autonomy and justice). In the same way, evidence-based guidelines and protocols might be successful in coordinating the different logics of science, practice and patient's worlds, yet might have (unintended) consequences in terms of autonomy or justice. Both models of EBM and related ethical problems have to be considered and evaluated.

Thus, the examples described in this paper show that boundary fields of medicine have an intriguing dynamic in making use of evidence, which also provides a deeper understanding of the standard areas of hospital-based medicine: boundary fields are arenas where different professionals have to deal with 'evidence' in the context of the diagnosis and treatment of complex problems and complex patients. These boundary fields exhibit the pattern of dealing with evidence in the context of cooperation and delegation of responsibilities between physicians and other health care workers. This is of relevance by itself, but more so, because it reflects the increasingly important practice in medicine of integral care for complex, chronically diseased and ill people. However these fields apparently appear to be 'peripheral', these fields are and should be central to the medical understanding of patient care and the ethical analysis of it.

This is why we think it is important not to reject EBM a priori as altogether irrelevant to boundary fields, but to present a different model to include EBM in these fields instead: the coordination model. This model opens a window to important ethical issues that relate to the communication and interaction between scientists, health care workers, and patients. In long term and complex disease and illness trajectories, the role of the professional is not limited to confronting the patient with the best evidence at a specific time in a specific condition and with regard to specific options. Professionals have to discuss long term plans under conditions of uncertainty and in view of diverse

knowledge sources, both from other professionals and from the patient. The ethical issue is also not limited to whether the patient is given proper information and sufficient choice, and whether the patient's autonomy is respected at discrete moments regarding singular decisions. In the 'boundary' fields, ethical issues typically have to do with broader matters of recognition, shared decision-making and a more symmetrical and deliberative relationship between professionals and patients, in which all parties are prepared to learn from the others (Widdershoven and Verheggen 1999). Hence, we need to organize health care practices in such a way that diversity does not result in mutual disregard, but in continuous learning (Vos et al. 2002).

Ethics of justice and ethics of practice

Our aim to complement the intrusion model with the coordination model implies a similar multiperspective view in terms of a general ethical framework. The core question regarding EBM from the standard 'Ethics of Justice' perspective is whether each patient, health problem and field of health care receives the attention and the means it deserves (Hope 1995). This may be argued in deontological terms of rights, needs or fairness: is there no discrimination inherent in implementing EBM against certain groups or is there any other breach of ethical principle? It may also be argued in utilitarian terms: do the overall results of evidence-based health care comply with the social benefits that we expect from health care? The prime focus of discussions concerning 'boundary fields' will be whether older patients, patients with chronic or multiple diseases, preventive medicine, health promotion, mental health care and the like will receive sufficient priority. This is the standard discussion on 'border fields', and remains relevant at all times and in all domains of health care.

The core question regarding EBM from an 'Ethics of Practice' perspective is what kind of evidence it takes to have a flourishing practice. The original Macintyrian version of an ethics of practice stressed that 'external goods' (such as cost-effectiveness) ought to be balanced by 'internal goods': shared understandings by participants on what it takes to be a good participant in that particular practice (MacIntyre 1981). In the case of health care, such a line of reasoning may be put to rather conservative use in claiming that doctors alone should decide what good medicine is. But from our analysis of boundary fields, we must conclude that 'internal goods' or 'shared understandings' are a problematic compass in domains where doctors share the 'ownership' of the field with other professionals and with patients who have much expertise from experience. Although we need not rule out the possibility of shared understandings between these diverse participants, the conditions of the 'boundary fields' of medicine call for a more 'procedural' definition of what it takes to make these practices flourish. An emphasis on shared learning experiences may be the core of an 'Ethics of Practice', which does justice to the diversity and the dynamics of domains of health care where clear definitions, demarcations of diseases and treatment trajectories are rather the exception than the rule. Thus, the political philosophy of Thévenot and Boltanski can be used to reconstruct a concept of 'Ethics of Practice' that is usually associated with conservative virtue ethics.

That the 'Ethics of Justice' perspective and the 'Ethics of Practice' perspective are complementary can now be specified in terms of the need for more substantive ethics to be complemented by more procedural ones, and vice versa. An ethics of justice is substantive in that it uses substantive values and principles as direct criteria for the evaluation of the moral quality of a state of affairs that is or might be the outcome of certain actions. The Macintyrian version of the 'Ethics of Practice' aims to offer an alternative substantive mode. An 'Ethics of Practice' that calls for open learning experiences at institutional level, and for participants who take responsibility in generating such experiences, lacks *prima facie* ethical substance. It aims at quality criteria for communicative processes rather than states of affairs. But some kind of substantive normative criterion will always be required to assess the desirability of the present institutional framework under which communicative processes in health care take place, and to assess the consequences of participant's actions for

this framework. On the other hand, outcome criteria, however tightly argued, are of limited value if they don't do justice to what drives and occupies the participants in health care practice in the first place. Which goals and actions are legitimate for participants in the practice of health care cannot and should not be deduced directly from substantive normative criteria. In a 'well ordered' practice, participants deserve sufficient latitude to mould more general criteria and work out the particular criteria suited to their practice.

Although this argument has special relevance for the so-called 'boundary' fields of medicine, we believe it holds for the ethics of EBM in general. Health care is an arena for diverse and dynamic knowledge claims. The fate of such claims ought to be decided by a whole range of institutional devices linking a plurality of relevant participants. The concept of EBM contributes to the pressure within this system, that all participants are held accountable for their contribution. Institutional devices to make systematic periodic assessments of participants' claims and accounts available, in a transparent way, are indispensable. This does not imply one paradigmatic set of evaluation criteria, but it may imply that all claims and accounts in the practice of health care ought to be assessed in terms of how they might profit from the systematic use and production of evidence. Evidence would not be the impartial judge of practice, but an important incentive to modify accounts or practices. How the use of evidence contributes to a just distribution of health care goods, and how it contributes to reproduction of conditions for shared learning experiences, ought to both be essential elements of such institutional deliberations.

Concluding remarks

We have attempted to make a contribution to the emancipation of boundary fields of medicine, which are often regarded with some suspicion on the part of medicine, especially EBM. We believe such emancipation is not served well with a primarily defensive position, shielding these fields from intrusion by the hard core of medicine. Although we sympathise with pleas for recognition of the fact that the same scientific standard need not and cannot apply in all branches of medicine, we do not want to pursue this particular ethical line of argument, in terms of utility or justice, concerning 'weak' and 'peripheral' disciplines and fields. This has partly to do with knowledge politics. Rather than defending realms from EBM, it might be wise to invest in reforming and expanding the EBM movement. EBM is about "the best available evidence", and this does not rule out methodologies other than the RCT (Ter Meulen and Dickenson 2002). A priori arguments about the complex and idiosyncratic nature of certain fields are doubtful in general, and they may shield conservative 'art versus science' types of attitudes (Vos et al. 2002). Particularly, we want to criticise the idea that boundary fields of medicine are marginal and weak, and that by implementation of EBM these areas will become full members of the medical community. Whether learning and deliberation stimulated by such an approach leads to defensible outcomes, should continue to be laid along the yardstick of an 'Ethics of Justice' and autonomy.

References

1. Arksey H (1998) RSI and the experts. The construction of medical knowledge. UCL Press, London
2. Arksey H (1998) RSI and the experts. The construction of medical knowledge. UCL Press, London
3. Baker SM, Marshak HH, Rice GT, Zimmerman GJ (2001) Patient participation in physical therapy goal setting. Physical Therapy 81: 1118-1126
4. Frank AW (1995) The wounded storyteller: body, illness and ethics. University of Chicago Press, Chicago
5. Goede CJ de, Keus SH, Kwakkel G, Wagenaar RC (2001) The effects of physical therapy in Parkinson's disease: a research synthesis. Archives of Physical Medicine and Rehabilitation 82: 509-515

6. Health Council (1999) De effectiviteit van fysische therapie – electrotherapy, lasertherapie, ultrageluidbehandeling [The effectiveness of physical therapy – electrotherapy, laser therapy, ultra sound treatment]. Dutch Government Press, The Hague
7. Health Council (2000) RSI. Dutch Government Press, The Hague
8. Hawkins AH (1993) Reconstructing illness: studies in pathography. Purdue University Press, West Lafayette
9. Hope T (1995) Evidence based medicine and ethics. Journal of Medical Ethics 21: 259-260
10. Kendall N, Main CJ, Linton SJ, Vlaeyen JW, Nicholas MK (1998) Reply to Wallis et al. Pain 73: 15-22 [letter, comment in Pain 78: 223-225]
11. MacIntyre A (1981) After Virtue. University of Notre Dame Press, Notre Dame
12. Meulen RHJ ter, Dickenson DL (2002) Into the hidden world behind evidence based medicine. Health Care Analysis 10: 231-241
13. Oderwald A (1994) Lijden tussen de regels: overzichtsstudie naar egodocumenten van chronisch zieken [Suffering between the lines: a review of egodocuments of chronically ill patients]. Nationale Commissie Zieken, Zoetermeer
14. Prins JB, Bleijenberg G, Bazelmans E et al. (2001) Cognitive behaviour therapy for chronic fatigue syndrome: a multicenter randomised controlled trial. The Lancet 357: 841-847
15. Shorter E (1992) From paralysis to fatigue. A History of psychosomatic illness in the modern era. Free Press, New York
16. Thévenot L, Boltanski L (1991) De la justification. Les economies de la grandeur. Gallimard, Paris
17. Vink A (2003) Patiënten als referenten van onderzoek. Mediator 14: 10-11
18. Vos R, Houtepen R, Horstman K (2002) Evidence based medicine and power shifts in health care systems. Health Care Analysis 10: 319-328
19. Wallis BJ, Lord SM, Bogduk N (1997) Resolution of psychological distress of whiplash patients following treatment by radio frequency neurotomy: a randomised, double-blind, placebo-controlled trial. Pain 73: 15-22
20. Whiting P, Bagnall A, Sowden A, Cornell J, Mulrow C, Ramirez G (2001) Interventions for the treatment and management of chronic fatigue syndrome: a systematic review. JAMA 286: 1360-1368
21. Widdershoven GA, Verheggen FW (1999) Improving informed consent by implementing shared decision-making in health care. IRB 21: 1-5
22. Willems D (2001) Dokters and patiënten in kleine technologie. In: Berg M, Mol A (Eds.) Ingebouwde normen – medische technieken doorgelicht. Van der Wees Uitgeverij, Utrecht

11 Research Ethics and Evidence-based Medicine[1]

Reidar K. Lie

Introduction

It is increasingly becoming a standard requirement to have to document effectiveness before new interventions are adopted by a health care system. The methods of Evidence-based Medicine, such as meta-analyses or other systematic evaluations of evidence, are often used to document effectiveness or the lack of it. There is an understandable reluctance to fund new experimental interventions when the evidence is not yet complete. Some of the most widely known controversies about the application of Evidence-based Medicine, such as mammography screening for breast cancer, have been concerning disputes about what the evidence says about effectiveness. Similarly, in discussions about the use of Evidence-based Medicine for the issue of allocation of scarce resources, the focus has often been on the problematic assumptions made when utilizing methods, such as Quality Adjusted Life Years.

In this paper I will argue that there is an additional issue that has not been addressed at all in the literature. In my discussion, I will assume that there is no disagreement about the established effectiveness of the intervention. I assume further that there is agreement that the established effective intervention is too expensive for general implementation, i.e. it is not cost-effective. I recognize that both of these claims may more often than not be controversial, but for the sake of argument I will assume that there is no such controversy. The problem I want to address is how we should go about establishing the effectiveness of a *cheaper* intervention that can be widely implemented. According to Evidence-based Medicine we should conduct a randomized clinical trial with the proposed new, cheaper intervention in one of the treatment arms. I will demonstrate that this requirement raises an interesting and difficult ethical issue.

The perhaps most famous recent example of an intervention that was considered too expensive to implement is treatment of mother and baby to prevent peri-natal transmission of HIV. An effective, but very expensive intervention was established, which was unaffordable for developing countries. A proposed shorter, and cheaper, intervention was subsequently tested in randomized clinical trials in various countries, most of which included placebo in the control group. These trials were widely criticized because they did not use the established, proven treatment as a control. Those who defended the trials argued that the design was necessary to answer questions of relevance to developing countries. What is interesting about this discussion is that it has only been regarded as a problem for developing countries: in the developed world it has simply been taken for granted that one should use the established, effective treatment in the control group. In this paper I will argue that this is not true: the problem of the choice of control groups is a general one, arising in all situations of resource *constraints*, not just in situations of extreme lack of resources, such as in developing

[1] This chapter was first published as an article in the Journal of Medical Ethics in April, 2004, published by BMJ Publishing Group and is reprinted here with permission by BMJ Publishing Group.

countries. It follows from this, since no country has unlimited resources for health care, that it is a general problem facing all of us. I will defend this claim by examining a recent case from the UK involving treatment of multiple sclerosis.

National Institute for Clinical Excellence and treatment with beta interferon for multiple sclerosis in the UK

The National Institute for Clinical Excellence (NICE) in the UK was established with the aim of advising health professionals about treatments which work and which are cost-effective. Although it has an advisory function only, its recommendations are quite influential. One important guiding idea behind the establishment of the Institute is the realization that a government cannot afford to provide all expensive new treatments to everyone. The process by which particular new treatments are adopted have until now depended on accidental factors, such as the ability of patient groups and others to pressure politicians to accept funding for their disease. This can very easily lead to inequities, weaker patient groups do not receive the same consideration as patient groups with more powerful allies. Similarly, if reimbursement decisions are delegated to local authorities, one can expect great variations in reimbursement practices. The evaluation process by NICE is supposed to ensure that only cost-effective interventions are adopted by the National Health Service, and to ensure an equitable adoption of new interventions.

When NICE evaluates a new intervention, it will request relevant information from the manufacturer/sponsor of the technology or drug, and commission technology assessments from appropriate academic centres. The decision to evaluate a particular technology ultimately rests with the Ministry of Health. The commissioned reports are then reviewed by a committee set up by NICE, comments are solicited from interested parties, reviewed again by the committee, and the final recommendation is then submitted (Dillon et al. 1999).

Beta interferon was licensed in the UK for treatment of relapsing MS in 1995. In order for a treatment to be reimbursed, a neurologist would have to certify that the drug would be clinically useful. In spite of this, many health authorities did not fund treatment with beta interferon on grounds of unproven benefit and low cost-effectiveness, leading to a charge of rationing by postcode. A High Court also ruled in favor of a patient who had obtained the necessary evaluation by a neurologist arguing that, "a blanket ban was the very antithesis of a national policy, whose aim was to target the drug at patients who could most benefit from treatment". This decision was similar to another High Court Ruling with regard to Viagra on a denial of reimbursement on the grounds of cost-effectiveness. The Department of Health refused funding of the licensed drug for cost reasons. In 1999 a High Court ruled that this was against the law because it "deterred doctors from exercising their duty to use their clinical judgment". The Department of Health's Policy was also deemed unlawful under European law because it "contravened the so-called transparency directive on medicines, which lays down the principle that any decision to blacklist a medicine from a member state's national health service must state reasons 'based upon objective and verifiable criteria'". Subsequent draft guidelines then stated that Viagra could be prescribed for certain conditions, which presumably would fulfil the requirement of denial of treatment based on "objective and verifiable criteria". This constitutes the background both to the establishment of NICE and the particular appraisal of beta interferon treatment (Dillon et al. 1999).

NICE issued its preliminary appraisal on beta interferon treatment for MS in October 2001, based on an evaluation process as described above. NICE did not recommend treatment with beta interferon or glatiramer acetate for patients with multiple sclerosis. There was then a short appeal process, where an appeal board considered objections, but upheld the preliminary appraisal in January 2002 (National Institute for Clinical Excellence 2002a; 2002b; 2002c).

In the final appraisal, NICE did not recommend treatment of multiple sclerosis with beta in-

terferon or glatiramer acetate on the balance of the clinical effects and cost effectiveness. NICE recognized that there is scientific evidence that this treatment modality does reduce relapse frequency and severity, at least during the initial years of therapy, and is perceived to be of great value to people with MS, as the following quotation from the appraisal shows:

- "The Committee considered in detail evidence taken directly from patients and two advocacy organizations. The patient organizations and the patients who attended the Committee meeting spoke of the patients' experience of this distressing disease and the impact of beta interferons and glatiramer on relapses and disease progression. This dialogue provided important insight into the effect of relapses on patients' daily lives and the value that they place on the potential reduction in severity of relapses with the use of these drugs."

The Committee nevertheless concluded that the cost of the product was too high to justify its adoption, when one has to decide to use limited resources for competing interventions. Based on extensive economic modelling, the committee concluded that the cost per quality of life year gained would be between GBP 248.000 and 810.000, for five years of treatment, and between GBP 40.000 – 90.000 for twenty years of treatment.

The NICE appraisal process had several notable features:

- There was an initial cost-effectiveness analysis. These showed great variations in cost-effectiveness estimates, and the modelling was criticized. NICE then commissioned a more thorough cost-effectiveness analysis.
- It is clear that the estimate of effects in relation to costs has a decisive influence on the recommendation that this intervention is not cost-effective. The benefit estimates are done in terms of Quality Adjusted Life Years. The committee argues strongly that, although one may disagree with a number of things with regard to this methodology, it is the best summary measure of the benefits of an intervention, and those who criticize the way benefits are assessed should point to specific alternative evaluations.
- It is clear that there are major uncertainties in the data. There is only clinical trial data for about 2 years of treatment, and beyond 5 years there is no data at all. One therefore has to make estimates concerning treatment effects. Again, one might disagree with the specific choices that the committee has made, but there is no question that they have made a reasonable evaluation of the available evidence.
- It is also clear that whatever reasonable figures one does use for the effects and costs of this treatment, it is going to be a very expensive one compared to benefits, in comparison with other funded treatments within the NHS.
- The process has been very public, with the involvement of industry and patient advocacy organizations at every stage of the process. Key documents have been made available to all interested parties throughout the process.

In spite of this, complaints have been made from both sides about lack of transparency. Patient advocacy organizations have argued that they should have been present during deliberations. Others have complained that NICE is not able to make appropriate cost-benefit calculations because it does not have access to key data that belong to the pharmaceutical companies, and that may be detrimental to their desire to have these drugs funded by the NHS (Cookson et al. 2001; Smith 2000; Sculpher et al. 2002).

This, then, is an example of a recommendation not to adopt a treatment of proven benefit based on cost-effectiveness data, and in spite of input and pressure from patient organizations to adopt the treatment. We may disagree with the specific conclusion, but in principle at least this is an example of the type of decision that my argument is based on: a decision not to implement a treatment, not

because it is of doubtful effectiveness, and not because one might disagree with some of the value judgments made, but because it is found to be too expensive given the limited resources available.

On February 4, 2002, however, the Secretary of Health announced that it would nevertheless fund treatment with these two drugs, as part of a collaborative arrangement with the producers of the drugs to estimate long-term effects. This development after NICE's decision is not important to the problem discussed here, but the ensuing discussion did provide some additional information about the basis for the original decision. Under the proposed scheme, all eligible patients as certified by a neurologist will receive reimbursement of the treatment, initially paid by the NHS. They will have to agree to be part of a monitoring program of the drugs' effectiveness. The results of the monitoring will be fed into a statistical model, which is the same as the one used by NICE in its appraisal, to calculate whether the results are better or worse than those used by NICE in its appraisal. If the results indicate that the cost-effectiveness falls below a sum of 36.000 GBP per QALY, industry will pay a proportional higher cost of the treatment. According to the proposal (Department of Health 2002):

- "If actual benefit is equal to or greater than expected benefit (within a tolerance margin – see below) then the NHS will continue to make payment at the price agreed at the outset of the scheme. However if actual benefit after t years was below the tolerance margin, the price for the period up to the next review point will be reduced to the extent needed to restore cost effectiveness to the cost per QALY "threshold" determining entry into the scheme …., i.e. £36,000. We envisage that the formal monitoring process for assessing cost effectiveness and pricing adjustments will continue for up to 10 years. At the end of this period payments to companies will continue at the level implied by the final review point."

The figure of 36.000 was arrived at in the following way:

- "A retrospective analysis of appraisal determinations in its first year of operation, as summarized by Sir Michael Rawlins at NICE's annual public meeting, suggests that positive recommendations were in general associated with a cost per QALY of £30,000 or less; higher cost per QALY figures were accepted only if there were special factors accepted as relevant by the Appraisal Committee and not covered by the formal modelling. A number of "special factors" which might be considered to be relevant to the cost effectiveness of treatments for MS have been put to us in discussion. The FAD has specifically referred to two unquantified factors: i) the impact of treatment on the severity (independent of the frequency) of relapses, and ii) possible cost offsets from the avoidance of severe levels of disability requiring intervention by the Personal Social Services. In the light of all these considerations the threshold will be set, for the purpose of this scheme only, at £36,000."

Patients immediately claimed victory, and saw this as a reversal of NICE's recommendation, although the Department of Health pointed out that there was no contradiction between this policy and the appraisal by NICE. NICE had urged the government to consider ways of making the treatments more cost-effective, by for example negotiating price reductions, and thereby achieving an acceptable cost-effective level. The proposed scheme would only mean that the NHS would pay the full amount for the treatment if it was found to be cost-effective according to the set threshold. If the monitoring process should find that the benefits are less than that, industry will have to pay for part of the expenses of the drugs. The initial cost to the NHS in this scheme is estimated to be GBP 7.000 – 10.000 per patient, with a total estimated annual cost of 50 million pounds.

Testing new potentially cost-effective interventions in clinical trials

The fact that a government body, after a public process, whatever its flaws, has rejected a recognized clinically effective treatment on cost-effectiveness grounds raises some interesting issues of research

ethics which have not been explored fully. If an intervention is rejected on these grounds, it raises the question about what the appropriate control group should be in a future clinical trial with the aim of identifying an effective, but cheaper intervention for the condition.

Fundamental to research ethics is the concept of clinical equipoise. When testing a new, promising therapy in a randomized clinical trial, there should be no evidence that one of the therapies offered in the trial is more effective than the other. Associated with this idea is the claim, most famously connected with the debate about the revision of the Declaration of Helsinki, that the control group in a clinical trial is entitled to the 'best proven' intervention. If there is clinical equipoise, if the control group receives the 'best proven' intervention, and the experimental group a new, promising treatment, it is felt that nobody who enters a clinical trial is disadvantaged. They will either receive an intervention they would ordinarily receive, if they were randomized to the control group, or an intervention that, according to the best available evidence, is indistinguishable from the 'best proven' therapy in terms of effectiveness. Hence, it would not ordinarily be appropriate to have a placebo group, or no intervention group, in a clinical trial where proven interventions are known to exist.

During the past few years there has been an extensive discussion about the choice of a control group in the context of research in resource poor settings. The question is whether it is permissible to use as a control group a treatment that is known to be less effective than the 'best proven' therapy, in order to obtain knowledge that will be useful for the country in which the trial takes place. Thus, in the peri-natal HIV transmission trials, it was argued that it was permissible to have a placebo group to establish whether a short course treatment would be better than a placebo, even though a known, effective treatment was already available in resource rich countries. This position has been highly controversial, as witnessed by the revision process of the Helsinki Declaration. The World Medical Association affirmed that the control group in a clinical trial should receive the 'best proven' therapy, irrespective of where the trial takes place.

The discussion about the permissibility of choosing an intervention that is less effective than the 'best proven' therapy, or best current therapy, has usually been seen as only relevant to resource poor settings. However, by reflecting on the situation after NICE rejected beta interferon for treatment of multiple sclerosis, it can easily be shown that the issue is one that applies in all cases of resource constraints, that is, is applicable to all countries.

Let us assume that the relevant government authorities in the UK decided that, all things considered, it should not reimburse beta interferon treatment for multiple sclerosis, neither as ordinary care nor in the context of the research project described above. Let us also assume, given the wide consultation process before the decision, that there was general consensus in the country that beta interferon treatment should not be covered, in light of all the other alternative uses of funds within the NHS. Given the unsatisfactory current treatment options for MS, it would be important to develop new, but cheaper interventions. The issue is how one should design a trial for a promising, cheaper intervention in the future.

In such a trial there would be two choices for the control group. One could either provide the 'best proven' treatment, i.e. beta interferon treatment, or a placebo. In light of the discussion about the revision of the Declaration of Helsinki, one might want to decide that placebo use would be unjustified. After all, there is a 'best proven' therapy, but a very expensive one, that one would seem to be obligated to provide to the participants in the trial. From a methodological point of view, there is something to be said in favor of this strategy. We have identified a treatment that is known to be effective against MS, and we would therefore presumably want to know whether the new, promising treatment is at least as good as the established, effective, but too expensive treatment. If that were really what we are interested in, an equivalence trial would be unproblematic. However, we might very well expect that the new, cheaper treatment is not going to be as effective as the expensive, proven treatment, but still expect it to be better than what is currently offered within the NHS. If it is also the case, and this is a crucial point, that the course of untreated multiple sclerosis is highly variable so that we do not know what the relapse rate is in a particular group of patients, an equivalence trial

would not provide us with useful results. Let me defend this claim by using results from a realistic, but hypothetical trial, the results of which are provided in *Table 11-1*.

Table 11-1
Results from a hypothetical trial

Placebo/No treatment	New	Beta inteferon
10	11	12
9	11	12

Let us assume that from the data available we expect untreated MS patients of a certain age to expect 10 QALYs if they are left untreated. Let us also assume that if we treat them with beta interferon we can increase their quality of life so that they now can expect to improve their prospects to 12 QALYS, but at a cost of GBP 160.000, giving us a cost per QALY of GBP 80.000, way over the accepted limit of GBP 36.000. A new, promising, cheaper treatment costing GBP 50.000 is tested against beta interferon in an equivalence trial, showing an expected QALY of 11, which is in-between the QALYs gained by beta-interferon and no treatment, giving a cost per QALY of 50.000, still above the accepted limit of GBP 36.000. Based on this equivalence trial, we would therefore reject the new, promising treatment as not cost-effective.

However, it turns out, unknown to us because we did not include a placebo group in the hypothetical trial, that the expected, untreated QALY in the trial population is actually 9, not 10. The cost-benefit of beta-interferon would therefore be 53.000 in this patient group, still above the accepted limit. The cost-benefit of the new treatment would, however, be 25.000, which falls within the accepted limit, and the new treatment should be adopted.

The important point is that we could not have known this if we had not included a placebo group in the hypothetical trial. If we had done an equivalence trial we would have wrongly rejected the new treatment as not cost-effective, whereas, if we had included a placebo group, we would have seen that the new treatment is indeed better than the current acceptable treatment, which in this case is no treatment.

It is evident that the structure of this hypothetical trial is exactly the same as the structure of the peri-natal HIV transmission trials, showing that problem of a choice of control group in the presence of established effective interventions is the same in all settings of resource constraints, not just in settings of extreme scarcity of resources.

In this case, since there has been a decision not to introduce the intervention into the National Health System, one might want to argue that one is not denying anybody anything they have a claim to, in spite of the requirement of the Declaration of Helsinki. In the UK, however, the health authorities decided to introduce beta interferon treatment for MS in the context of a research project in an attempt to gather data about the long term effects of beta interferon treatment, as well as limit the costs to the NHS, should the treatment turn out to be quite cost-ineffective. As some commentators have noted, the value of the research project is quite doubtful, as it is going to be difficult to establish reliable effectiveness data in the absence of reliable comparative data about the course of MS without treatment. Quite apart from that problem, there is an additional problem if the scheme should result in a judgment that beta interferon treatment does indeed fall outside the accepted limit of cost-effectiveness. One would then also have to search for a cheaper, but still effective treatment, and the issue of the choice of a control group would return. But in this scenario, all eligible patients would then be receiving the expensive treatment, and one would presumably have to take some of them off this treatment in order to establish the cost-effectiveness of the new, promising treatment. This is clearly morally more problematic than not providing something to the control group that they would not have received anyway from the national system.

Concluding remarks

The case of beta interferon treatment for multiple sclerosis shows that we face the problem of the choice of an appropriate control group, whenever there are resource constraints, and whenever equivalence trials cannot be carried out for scientific reasons. The debate about what is an appropriate control group is therefore not only relevant to developing countries, but is going to be increasingly important for all countries as there is increasing pressure to prioritize among expensive, new interventions. This problem raises the issue of the relativity of ethical standards in research. If we assume that all countries face exactly the same resource constraints, the design of the trials will be the same everywhere. However, if we assume that resource constraints will vary significantly, even in relatively resource rich settings, treatment access will vary in different countries, necessitating different standards of care. Although this might be considered to be ethical relativism by some, it should more appropriately be seen as adapting universal ethical principles to different local circumstances. What the beta interferon case shows, is that it is likely that local circumstances may vary more dramatically than we have been used to so far, as we are facing increasingly expensive, but effective interventions.

References

1. Cookson R, McDaid D, Maynard A (2001) Wrong sign, nice mess: Is national guidance distorting allocation of resources. British Medical Journal 323: 743-745
2. Department of Health (2002) Cost effective provision of disease modifying therapies for people with multiple sclerosis. Department of Health, London
3. Dillon A, Gibbs TG, Riley T, Sheldon TA (1999) The National Institute of Clinical Excellence and coverage for Relenza by the NHS. Informing judgment: Case studies of health policy and research in six countries. Milbank Memorial Fund, New York
4. National Institute for Clinical Excellence (2002a) Appraisal of the use of beta interferons and glatiramer acetate in the treatment of multiple sclerosis. National Institute of Clinical Excellence, London
5. National Institute for Clinical Excellence (2002b) Beta interferon and glatiramer acetate for the treatment of multiple sclerosis. National institute for clinical excellence, London
6. National Institute for Clinical Excellence (2002c) NICE appraisal of beta interferon and glatiramer for MS. Process to 2002. National institute for clinical excellence, London
7. Sculpher M, Drummond M, O'Brien B (2002) Editorial. Effectiveness, efficiency and NICE. A nice start but evidence costs money. British Medical Journal 322: 943-944
8. Smith R (2000) The failings of NICE. Time to start to work on version 2. British Medical Journal 321: 1363-1364

12 Clinical Evaluative Research: Which Patients Benefit, How and When? A Contribution to a European Discussion

Heiner Raspe

Introduction

Clinical medicine is a pragmatic science, quite different from "arts" or "applied" sciences. One of its central characteristics is the continuous scientific evaluation of actual consequences of clinical judgements, decisions, and actions. This is achieved mainly by clinical-evaluative patient-centred research with its most powerful tool, the randomised controlled trial (RCT). Its results are the main substrate of evidence-based clinical prevention, diagnosis, prognosis, therapy and rehabilitation. These interventional (and artificial) trials establishing "efficacy" have to be supplemented by non-interventional ("effectiveness") studies such as outcome observations, register studies, and post marketing studies.

Recently the European Parliament and the Council issued a Directive "on the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use" (2001/20/EG). It has to be adopted and transformed into national law of the member states at the latest by the 1st of May 2004.

Article 3 section 2 (a) states:

- "A clinical trial may be undertaken only if, in particular: the foreseeable risks and inconveniences have been weighed against the anticipated benefit for the individual trial subject and other present or future patients."

Minors and incapacitated adults "should (normally) be included in clinical trials only when there are grounds for expecting that the administering of the medicinal product would be of direct benefit to the patient, thereby outweighing the risks" (Consideration 3). Later article 4 permits clinical trials on minors but not incapacitated adults which provide "some direct benefit for the group of patients" to which the study subject belongs. However, Article 2 Definitions do not provide any definition of "benefit" and "direct benefit".

Similar statements can be found in the World Medical Association Declaration of Helsinki (2002; No. 16), the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine issued by the Council of Europe in April 1997 (Article 16 (ii)) and its Draft Additional Protocol on Biomedical Research (Article 6 (1)).

Trials, on the other hand, have to produce "credible" results, should have "relevance" and scientific merit, and must not include "obsolete or repetitive tests".

Thus the directive seems to require two discrete qualities: scientific originality and validity and a positive balance between the trial subjects' benefits and risks.

In the following, these two qualities are studied more closely with a main focus on the potential of trials to directly benefit their participants.

Medical-scientific aims and functions of clinical trials

A clinical trial has more than one aim and function

It first has the potential to test a specific biologic hypothesis, that for instance drug A (e.g. pravastatin) is able to influence certain biologic processes (e.g. total serum cholesterol or LDL cholesterol); second, it tests the more general hypothesis that the observed effects were – at least to a considerable part – caused by the intervention and not by chance, co-interventions, bias or confounding. It thus establishes a causal nexus between an intervention and an outcome. This is achieved mainly by the use of a comparative intervention (most often placebo or standard treatment). Efficacy is then estimated by subtracting (absolute risk difference) or relating (risk ratio) the effects of the two treatments to each other. As the main goal of medicine is not to produce effects but benefits, the outcomes additionally have to be "clinically relevant". And as any drug or other medicinal product is expected to do "more good than harm", "net-benefit" is what finally counts. The net benefit of an intervention results from the difference between all positive and all negative effects of an intervention.

Clinical trials – third – have the function to assess all relevant inconveniences and risks. It is especially this function in which RCTs are to be supplemented by observational studies. RCTs usually have limits in their disease and co-morbidity spectrum, size and length of follow-up that make them less apt at identifying subgroup-specific, rare and serious adverse reactions.

All these questions have to be answered "validly" and with sufficient precision. Clinical trials are to produce new and credible knowledge. This is their core aim and, if it is achieved, trial results ultimately strengthen the evidence base of clinical medicine. The results should further transcend the group of individuals from which they were derived and bear relevance for the population the sample was originally drawn from. It is this internal and external validity that makes results and reviews of single valid studies the basis of more general recommendations, and guidelines for further present or future doctors and patients.

Concerning their multiple medical and scientific aims and functions, clinical trials are both indispensable and without alternative. Medicinal products must not be marketed and used without tests on limited groups of patients under firmly controlled conditions. It would be unethical to widely apply drugs, for example, only on the grounds of personal impressions and unbalanced clinical experiences. It is no coincidence that mere opinions, even those of respected authorities, "first principles" and pathophysiological considerations are at the bottom of any hierarchy of evidence (e.g. www.cEBM.net/levels_of_evidence.asp).

A paradigmatic RCT: the West of Scotland Coronary Prevention Study (WOSCOPS)

To illustrate design and results of an RCT the WOSCOP-Study was introduced (Shepherd et l. 1995). It tested the hypothesis that "lowering the blood cholesterol level may reduce the risk of coronary heart disease". The researchers randomly assigned more than 6,500 Scottish men with hypercholesterolemia, 45 to 64 years of age, to receive – double-blinded – either a lipid lowering drug (pravastatin 40 mg/d) or placebo. The average follow-up period was 5 years. Outcome measurement included the (clinically relevant) composite endpoint of severe coronary events (nonfatal myocardial infarction (MI), or death from coronary heart disease (CHD) – confirmed by *an independent End-Points Committee). ❯ *Table 12-1* shows the main results (from ❯ *Table 12-2* of the publication).

The absolute risk of experiencing an unfavourable outcome was 6.5 % under pravastatin and 9.0 % under placebo. Pravastatin thus reduced the risk of the placebo group by 28 % ((9.0 minus 6.5) divided by 9.0 and multiplied by 100). The same preventive effect can be expressed in form of a relative risk (here 0.72, derived by dividing 6.5 by 9.0). The absolute risk reduction amounts to 2.5 % (9.0

Table 12-1
Nonfatal myocardial infarction or death from CHD after 5 years of WOSCOPS (definite and suspected cases)

Treatment	MI or death due to CHD	Other Outcomes	All
Pravastatin	215	3087	3302
Placebo	295	2998	3293
All	510	6085	6595

% minus 6.5 %), an absolute difference which can be transformed into a Number Needed to Treat (NNT) to prevent one additional coronary event: if the treatment of 100 patients over 5 years averts 2.5 additional events, then 40 patients need to be treated to avoid 1 event (NNT = 40).

It is obvious that no one is presently able to identify or predict the successfully treated patient(s). The very understandable question: "will I be the one in 40 patient?" is so far and for the foreseeable future unanswerable. Efficacy turns out to be a stochastic quantity. All the effect measures reported are statistically highly significant ($p < 0.001$) and thus very unlikely a result of mere chance. More precisely: a false positive trial is not totally excluded, this risk is however below 1 in 1000. But are the effects of this well conducted pivotal trial clinically relevant (providing "class 1a" evidence according to the levels of evidence hierarchy mentioned above? Does pravastatin produce benefit? The results of the trial can be presented in very different ways (Box 1) – with probably different conclusions.

Box 1: Different ways of expressing treatment effects (benefit?) of pravastatin compared to placebo (adapted from Shepherd et al. 1995)
Pravastatin given over 5 years reduces the risk of severe cardiac events in hypercholesterolaemic men by about 30 %, nearly a third as compared to placebo.
Among 1,000 men over 5 years treated with placebo 90 will experience a potentially fatal cardiac event; it will be 65 under pravastatin treatment.
Of any 1,000 men 910 will be still be alive or without MI after 5 years of treatment with placebo compared to 935 after 5 years of pravastatin treatment.
Of 1,000 men treated with pravastatin over a period of 5 years 65 will have an unfavourable outcome despite the treatment, 910 will have no benefit beyond placebo effect, and 25 will "directly" benefit from the treatment. This effect requires the prescription and intake of about 1.8 million tablets (40 mg) of pravastatin.

The benefit for 2.5 % of the intervention group is far from being certain. All effect measures are so called point estimates or statistics; they provide an estimate of efficacy, not its true quantity that actually implies some uncertainty. It can be quantified by means of a confidence interval. This is a range of values derived from the study, which includes the true value (parameter) with a high degree of probability (e.g. 95 %). For pravastatin, Shepherd et al. (2002) report a relative risk reduction of nonfatal MI or death from CHD of 31 % (based on a multivariate analysis, our unadjusted value was 28 %, cf. Table 12-1) with a 95%-confidence interval ranging from 17 to 43 %. This means that the true reduction could be as low as 17 % (worst case) or as high as 43 % (best case) under a remaining uncertainty of overall 5 %.

Patients as potential beneficiaries of clinical research

Much has been written on the concept, measurement and reporting of inconveniences, burdens and risks of subjects participating in controlled clinical trials, be they a consequence of the medicinal product to be tested itself ("therapeutic", Tri-Council Statement 2003) or the procedures necessary to monitor its effects ("non-therapeutic").

The category of benefit has widely been neglected, as if even the slight possibility of a remote and undefined "benefit" outweighs a wide range of burdens and risks. A more serious consideration requires a qualification and quantification of benefits equally careful as that of risks and burden.

This may additionally require a change in terminology: it is to some degree misleading to speak – in the same breath – of "benefits and risks" as "risk" refers to a probability of a negative state or event whereas "benefit" seems to take a favourable outcome for granted. It seems more appropriate to combine benefit with harm or damage and risk with chance or possible benefit.

The unbalanced situation gives grounds for focusing here mainly on the possible benefits of clinical research in general and "direct benefits" in particular.

None of the Directives, Conventions or Declarations mentioned above includes a definition or classification of "benefit" or "direct benefit".

Nevertheless, all refer to the participants or individual trial "subjects" (Directive 2001/20/EC Article 2 (i)) as first line beneficiaries. Two (or more) groups of subjects are to be distinguished: one or more intervention groups (IG) and one or more control groups (CG). As any benefit was shown to be a stochastic quantity emerging from the comparison of two or more summary statistics relating to groups (see section 3 above), it must fundamentally be conceived as a "group benefit".

The documents quoted address further possible beneficiaries: patients from the same group of persons, "same" in terms of disease status or biologic characteristics. An early profiting group could be the IG and/or CG itself, depending on the study results. If they favour the investigational product, then its use should be continued in the IG and de novo started in the CG. Its use should be discontinued in the IG and not started in the CG if the results showed the investigational product to be inferior to the placebo or standard treatment. In this case the members of the IG obviously had to suffer some "direct" harm or better "risk" since the disadvantage is again expressible only as morbidity or mortality differential, i.e. a relation of summary statistics (percentage, mean etc.) of two or more groups. In other words, indirect benefit may materialize by avoiding direct harm.

The next profiting group can be seen in other already existing patients outside the study groups, regional, national and international. Finally, the benefit can extend over future, yet-to-be born patients.

At last there could be no (presently imaginable) persons as beneficiaries at all. This is the case, when the research aims predominantly at the advancement of knowledge, as it is typical for basic research (e.g. today's stem cell research).

Box 2: Some basic distinctions

Possible beneficiary
Patient/proband him/herself
Members of the intervention and/or control group(s)
Group of similar present or future patients
Science, knowledge
Source of possible benefit
Medicinal product, intervention
Trial results
Concomitant actions, study circumstances
Mode of action
Direct: if produced by the investigational product/procedure

Indirect: if produced by concomitant actions or "negative" results
Timing of possible benefit
Immediate (within hours or days)
Delayed (within weeks or months)

A clinical trial is "directly" beneficial to its participants if the benefit follows – earlier or later – from the intervention tested. In WOSCOPS it was pravastatin that reduced the risk of acute coronary events over a period of five years with first visible effects after 6 to 12 months, and an increase to a relative reduction of 31 % over the next four years.

This primary direct benefit can be distinguished from a secondary still direct advantage as patients can profit – in the IG often additionally, in the CG de novo – from the trial results and their practical consequences. The results may then – thirdly – become directly beneficial for other existing or future patients.

So far we avoided giving a definition of "benefit" itself. The term is hard to define, for understandable reasons:

"Benefit" in any case is more than "effect". An example: in 1990 Riggs et al. published a study that compared high dose fluoride with a placebo in postmenopausal women with osteoporosis. The drug significantly increased bone mineral density in the lumbar spine by 35%; it nevertheless resulted in a disturbing increase of non-vertebral fractures. The drug thus produced a measurable effect but no benefit.

It follows, that the type of outcome is of relevance first. Benefit should be measured in terms of clinically relevant outcomes or "clinical endpoints". Surrogate measures are unsatisfactory and may sometimes be misleading. Two classes of surrogates can be distinguished: 1. "Intermediate Outcomes" which are part of a pathophysiological chain between a disease process and its ultimate clinical consequences (such as bone density as a causal factor of osteoporotic fractures) and 2. "Proxies". Proxies are different but closely related manifestations of the same disease process that leads to the endpoint of interest. Body weight and peripheral oedema can in this sense be seen as proxies for certain types of heart failure. The degree of thoracic kyphosis for osteoporotic deformities of the thoracic spine may give another example.

However, to become a "benefit", even a favourable outcome has to satisfy more criteria: it must be more than trivial at least in terms of size (e.g. relative risk), frequency (e.g. NNT) and duration. To identify an effect as a benefit thus requires several judgments. "Benefit" is a social construct and a judgemental category dependant on historical, cultural and personal circumstances. This does not preclude from categorizing and measuring it, but only within a defined social and psychological context, and this is the problem of any attempt of a general context-free definition.

In Evidence-based Medicine (EBM), the overall usefulness of an intervention or product is often expressed as a "net-benefit". The term again refers to a difference, the difference between all positively and negatively evaluated consequences of an intervention. Obviously, the difficulty in defining "net-benefit" is not to be underestimated.

Benefit from the WOSCOP-Study: ex ante and ex post

In the last section, it became clear that the actual quality and degree of benefit as well as the actual group of beneficiaries, is unable to be finally established before the end of the trial, and only after careful analysis and (peer-reviewed) publication of all results. Clinical trials produce data, but data are not to be confused with evidence, they have to be transformed into it. It is only evidence, not data that shall influence clinical judgments, decisions, and actions.

On the other hand: the researchers and eventually an ethics committee have to weigh "the fore-

seeable risks and inconveniences … against the anticipated benefit for the individual trial subject" ex ante.

We thus have to deal with two perspectives, one in advance, and the other in hindsight. The differences and some of the problems will be considered concerning the WOSCOP-Study mentioned above.

The situation ex ante

Before 1989, when the first patient was enrolled in WOSCOP, what could be the "anticipated benefit" of 40 mg pravastatin daily? Who were the foreseeable beneficiaries (besides the researchers, publishers and the drug company)? What was the estimated size of the benefit? What about side effects and risks that may offset all positive effects?

This question deliberately does not address the "non-therapeutic" benefits that may derive from being a study participant independent of the actual allocation (placebo or pravastatin). Trial subjects are said to be under beneficial scrutiny; they are seen on a regular basis and subjected to unusually close monitoring. Another advantage of participants of trials of innovative drugs may be seen in the possibility of exclusive access to an otherwise still unobtainable medication. This of course depends on the allocation scheme, with a chance of 50 % in a one to one randomised controlled trial.

Our question instead focuses on only those direct benefits for the WOSCOP participants that may arise from the exposure to pravastatin. This implies the impossibility of any direct benefit for the control (placebo) group. We do not exclude the possibility of a placebo-related benefit for this group, compared to no intervention at all; but this benefit could not be attributed to specific ingredients of the dummy, only to the sham intervention as such.

We further reject the idea that the mere inclusion within a placebo-controlled RCT is sufficient to satisfy the requirement of "a possible direct benefit to all study subjects". Its proponents argue that each participant of the study has a priori the same probability of being allocated to the active drug (and its likely benefit). But it neglects the fact that this average probability results from two extreme allocation probabilities, one of 0 for the control group, the other of 1.0 for the group randomized on the active drug. And the subjects of each group are, person by person, at least in principle, identifiable after randomisation.

Back to WOSCOPS: In 1989 when recruitment started, there was virtually no evidence to support the preventive use of pravastatin in hypercholesterolaemic but otherwise healthy male subjects. There was, however, some data of a tertiary preventive effect of statins in subjects having suffered a myocardial infarction (Scandinavian Simvastatin Survival Study Group 1994). The study plan thus followed, at best, a promising idea based mainly on the hypothesis of high cholesterol as a risk factor of CHD, the potential of pravastatin to reduce its concentration and on analogy. In the meantime, it has repeatedly been shown that statins are effective even in persons with "normal" cholesterol levels (Heart Protection Study Collaborative Group 2002; Shepherd et al. 2002).

There was at that time no universally accepted standard drug treatment of ordinary hypercholesterolemia in middle-aged men (besides unpleasant, overall scarcely accepted and hardly successful behavioural interventions). Thus, the criterion of clinical equipoise may be seen as fulfilled. It implies a genuine uncertainty amongst the medical community (not of each single physician) about the comparative therapeutic benefit of each of the two or more study groups (see Tri council policy statement Section 7.1).

It was this background against which WOSCOPS tested the hypothesis (called H0) that there would be no statistically significant difference between the effects of pravastatin as compared to a placebo.

However, the sample size calculation was based on the assumption (The West of Scotland Coronary Prevention Study Group 1992) that pravastatin could provide a 30 % risk reduction in a popu-

lation with a baseline risk of 6.9 fatal and 13.8 non-fatal incident CHD-cases per 1000 a year when compared to a placebo. The study planned with about 6,500 men had a 99 % power to detect such a difference as statistically significant (under $\alpha = 5$ % two-sided). This sample size calculation: 1) expresses the conviction of the group that H0 had a predominant chance to become rejected, and: 2) points to a relative risk reduction (30%) upon which the worldwide medical community may have agreed as a "minimal clinically important difference" (Wells et al. 2001) in the secondary prevention of severe cardiac events. The authors assumed an annual risk of 20.7 combined fatal and nonfatal events per 1000, amounting to an absolute risk of 10.4% over five years. A 30% reduction would result in an absolute difference of 3.1% and a NNT of 33.

The ethics committee of the University of Glasgow, on this basis, approved the study plan. It is not totally excluded that other committees would have been more critical, though in this field a RR of 0.70 and a NNT of 33 can be accepted as relevant, both from a clinical and an epidemiological point of view.

Finally, an evaluation ex ante has to balance the expected benefit against all possible "therapeutic" and "non-therapeutic" risks and inconveniences. In 1989, pravastatin seemed to be a comparatively safe and innocent drug with rare and mostly minor side effects.

This led, and most probably would still lead, to an overall favourable scientific and ethical evaluation of the study plan.

The situation ex post

The results of WOSCOPS were published in 1995 (Shepherd et al. 1995). Only the members of the intervention group could be seen as possible "direct" beneficiaries.

The results broadly confirm the assumptions made nearly a decade before. The crude event rate was 9.0 % for definite and suspected coronary events in the placebo, and 6.5 % in the pravastatin group, resulting in a relative risk of 0.75 (adjusted 0.69) and a NNT = 40. A less impressive description of the results was given in Box 1, paragraph 4), also considering those 65 subjects (out of a thousand) who became ill despite five years of pravastatin treatment, and those 910 who would have stayed healthy without any specific intervention.

It is not always the case that the results of RCTs confirm the assumptions made in advance as perfectly as in WOSCOPS. However, trials sponsored by pharmaceutical companies tend to reject H0 more often than not in favour of the medicinal product (Djulbegovic et al. 2000; Lexchin et al. 2003; Melander et al. 2003). It is not clear how much would be too much; but studies which only duplicate widely known results are of no scientific merit and thus unethical, a wasting of scarce resources including patients' trust and cooperativeness. The directive 2001/20/EG states that "in order to achieve optimum protection of health, obsolete or repetitive test will not be carried out, whether within the Community of in third countries" (consideration 6).

What about direct benefit ex post, if H0 could not be rejected and no significant difference was detected by a sufficiently powered study? At best, the members of the intervention group had no benefit at all; but it is more likely that they experienced some disadvantage or even harm compared with the members of the control group. Statins are now known to produce rare but rather specific side effects (muscle disorders, advancement or induction of malignancies; cf. Shepherd et al. 2002) that are fortunately missed in placebos. The ratio of benefit and harm would become even more unfavourable if a placebo turned out to be superior to the active drug.

Thus, a trial that might have seemed directly beneficial a priori would have changed into a detrimental experiment a posteriori. This result would not speak against the scientific character of the study. Its external and/or internal validity does not depend on the final outcome.

A special case is provided by trials testing the hypothesis that two treatments, one standard or placebo, the other innovative, are equivalent in respect to all positive outcomes. If the hypothesis

can be accepted, then the participants of the intervention group will again have no more direct benefit than those of the control arm. They may, however, undergo less foreseeable inconveniences and risks.

Other clinical questions and research designs

RCTs are in the heart of evaluative clinical research. They typically study the efficacy of clinical interventions, be they preventive, therapeutic, or rehabilitative. Other most relevant clinical problems, however, require other study designs.

Three will be discussed in more detail: testing the validity of diagnostic procedures by controlled cross-sectional studies, studying course and outcome of cohorts of cases with incipient diseases (prospective inception cohort study), and analysing adverse events by means of case-control studies.

It will result in there definitely not being any direct benefit for any participant in any of the three study types.

Validating a new diagnostic test

The first two phases of diagnostic studies estimate sensitivity, specificity and predictive values of a positive and negative test (cf. Sackett and Haynes 2002).

Table 12-2
A typical 2x2-table of an early diagnostic study (hypothetical data)

Test Result	Disorder present	Disorder absent	No of Subjects
Positive	80	20	100
Negative	20	80	100
	100	100	200

It starts with 200 subjects, 100 patients and 100 health controls. Their disease or health status is to be clarified by applying the current "diagnostic gold standard" in advance. Myocardial infarction diagnosed according to WHO criteria may serve as an example. Let us think of cardiac troponins as the target of the new test. Obviously, the test identifies 80 % of all MI patients as diseased and 80 % of all healthy controls as non-diseased and 80 % of test-positives and test-negatives are diseased and healthy respectively (sensitivity, specificity, predictive value of a positive as well as negative test are all equal 80 %). As the disease status of all subjects is known beyond any reasonable doubt at the start of the study, and the new test does not add anything to this knowledge, it is hard to see any direct benefit for any participating subject. This does not exclude a "secondary" direct benefit for further existing or future patients if the new test turns out to be more economic, secure, rapid, predictive or less invasive.

Assessing prognosis

A very similar situation emerges for prognostic studies of patients with defined incipient diseases (e.g. early rheumatoid arthritis). The simplest design prospectively observes the course and outcome of an "inception cohort" over a couple of years. A more sophisticated study stratifies the whole

sample according to predefined prognostic markers. Again, a primary direct benefit of the follow up or stratification itself cannot be expected before the end of the study. After that, and in view of the results, a secondary benefit may emerge as far as prognostic stratification becomes therapeutically relevant (e.g. for a "risk adapted" treatment).

Analytical case-control study

A third example of a study design without any direct benefit for the participants is given by a case-control study enquiring into the causes of uncommon adverse events possibly related to certain drugs. Here, cases (with the event) and controls (without) are to be compared in respect to a previous potentially noxious exposure. How many of the cases (e.g. newborns with phokomelia) and healthy controls have been exposed to a certain drug, e.g. thalidomide. Again, the disease status is to be clarified in advance, and the exposure occurred in the past, thus excluding any immediate direct benefit for the cases and controls. Under certain circumstances, the controls may profit from the results of the study by avoiding any exposure in the future (secondary indirect benefit).

Summary and conclusions

1. In all relevant directives, conventions, and declarations regulating clinical research in order to protect trial subjects, risks and burden within and from clinical trials are extensively and carefully considered. In contrast, the category of benefit, especially "direct benefit", seems widely neglected. This leads to difficulties when "foreseeable risks and inconveniences have (to be, HR) weighed against the anticipated benefit for the individual trial subject and other present and future patients" (Directive 2001/20/EG, Article 3).
2. To approach the problem, I first distinguish between effect and benefit as two different though overlapping consequences of clinical interventions, and define "direct benefit" as any clinically useful consequence originating from direct exposure to the investigational product (e.g. a drug) within the study ("primary direct") or after its completion, then mediated by its results ("secondary direct"). An indirect benefit can be produced by either avoiding a harmful product/intervention or by certain organisational arrangements, such as an exceptionally careful follow-up of all trial participants.
3. In any case, direct benefit turns out to be a stochastic quantity, expressed in complex statistics such as relative risk, NNT, or weighted mean difference, all to be derived from comparisons of summary statistics relating to groups. The idea that a clinical study produces a direct benefit to identifiable or predictable single subjects is not tenable. Benefit is always and in a fundamental sense a group benefit. It is thus prudent to speak of "predictable risks and burdens" and "reasonable likelihood" (World Medical Association Declaration of Helsinki 2002) or "the potential to produce results of direct benefit" (CDBI 2003) or "anticipated benefit" (Directive 2001/20/EG).
4. In assessing benefit, one has to consider further uncertainties: the possibility of a false positive trial (by alpha error), and the uncertainty of any statistic unavoidably contaminated by sampling error. Statistics are estimates to be surrounded by confidence intervals, no more, no less.
5. As clinical studies address open questions and test hypotheses their results are to some degree unpredictable. Would the answers be known in advance, the "studies" would not be acceptable both scientifically and ethically. A priori expectations of benefit may thus be disappointed in hindsight.

6. Some clinical questions require studies and study designs from which direct benefit can never be expected. This is regularly the case with diagnostic, prognostic and analytical case-control studies (e.g. to clarify the cause of adverse events). That does not make them less relevant for evidence-based clinical medicine. They are still necessary and indispensable. It is from my point of view problematic that the Directive 2001/20/EG seems to exclude this type of clinical study on "incapacitated adults not able to give informed legal consent". Article 5 (i) states that there have to be "grounds for expecting that administering the medicinal product to be tested will produce a benefit to the patient outweighing the risks or produce no risk at all." This requirement cannot be satisfied by studies on sensitivity, specificity, predictive values and likelihood ratios of new test (e.g. in the field of dementia). A direct benefit can only be expected from controlled studies that try to answer the question whether „patients who undergo this diagnostic test fare better (in their ultimate health outcome) than similar patients who are not tested?" (phase 4-studies according to Sackett and Haynes 2002).

Clinical evaluative research is – for many good reasons explained earlier – essential for the development of medicine as a pragmatic science and the sake of present and future patients. The tension between clinical-evaluative research and individual clinical care cannot totally be resolved: "Clinical research is distinct from clinical practice, in that the purpose and goals of each, although not mutually exclusive, are quite different... In fact, by participating in good clinical research, an individual may receive the highest quality of patient care and treatment, but that is not the goal of research, and much research does not directly benefit individual participants" (Grady 2002).

It is an essential task of researchers and ethical committees to safeguard the rights and integrity of clinical trial subjects. The weighing procedure required has to carefully consider both risk and possible benefits, be they primary or secondary, direct or indirect. But as the Ethics Advisory Committee of the British Royal College of Paediatrics and Child Health stated: "A research procedure which is not intended directly to benefit the child subject (I would add: and incapacitated adults as well) is not necessarily either unethical or illegal" (2000: 177; Principle 4).

References

1. Council of Europe (1997) Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine
2. Council of Europe, Steering Committee on Bioethics (2003) Draft Additional Protocol to the Convention on Human Rights and Biomedicine, on Biomedical Research.
3. Djulbegovic B, Lacevic D, Cantor A et al. (2000) The uncertainty principle and industry-sponsored research. Lancet 365: 635-638
4. Grady C (2002) Ethical principles in clinical research. In: Gallin JI (Ed.) Principles and practice of clinical research. Academic Press, San Diego, P 15-26
5. Heart Protection Study Collaborative Group (2002) MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial. Lancet 360: 7-22
6. Levitan HL (1981) Patterns of hostility revealed in the fantasies and dreams of women with rheumatoid arthritis. Psychotherapy and Psychosomatics 35: 34-43
7. Melander H, Ahlquist-Rastad J, Meijer G, Beermann B (2003) Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326:1171-1173
8. Riggs LB, Hodgson SB, O'Fallon M et al. (1990) Effect of Fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. New England Journal of Medicine 322: 802-809
9. Royal College of Paediatrics and Child Health (2000) Ethics Advisory Committee: Guidelines for the ethical conduct of medical research involving children. Archives of Disease in Childhood 82: 177-182

10. Sackett D, Haynes RB (2002) Evidence base of clinical diagnosis. The architecture of diagnostic research. BMJ 324: 539-541
11. Scandinavian Simvastatin Survival Group (1994) Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: The Scandinavian Simvastatin Survival Study (4S). Lancet 344: 1383-1389
12. Shepherd J, Cobbe SM, Ford I et al. (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. New England Journal of Medicine 333: 1301-1307
13. Shepherd J, Blauw GJ, Murphy MB et al. (2002) Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet 360: 1623-1630
14. The European Parliament and the Council (2001) Directive on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Official Journal of the European Communities 121: 34-44
15. The West of Scotland Coronary Prevention Study Group (1992) A coronary primary prevention study of Scottish men aged 45-64 years: trial design. Journal of Clinical Epidemiology 45: 849-860
16. Wells G, Beaton D, Shea B et al. (2001) Minimal clinically important differences: Review of methods. Journal of Rheumatology 28: 406-412
17. World Medical Association Declaration of Helsinki (2002) Ethical Principles for Medical Research Involving Human Subjects. WMA, Washington

13 Defining a Proper Background for Discussing Evidence-based Medicine[1]

Alessandro Liberati · Paolo Vineis

Background

The Evidence-based Medicine (EBM) movement has provoked strong restatements from within the clinical world about the essence of the patient-clinician relationship and the balance between scientific approach and personal experience.

Some commentators saw the movement partly as an attempt by clinicians to keep control of decision-making, in the face of governments set on increasing intervention in the previously relatively autonomous professions. Health policies worldwide, however, reveal the growth of mechanisms aimed at establishing parameters for acceptable clinical practice and a range of apparatus for monitoring and enforcing these parameters. On another track, some critics have questioned the movement's sometimes 'exclusive' focus on one particular research design (i.e. the randomised controlled trial) as unnecessarily narrow, and reinforcing the cultural and political values of particular research groups. Also embedded in this phenomenon is a staging of the confrontation between science and progress on the one hand and myth and reaction on the other.

It is the aim of this paper to discuss the extent to which the current debate addresses the real issues, or is rather confounded by extraneous factors. We will do so by starting from the personal conviction that there is within the current debate a mixture of epistemological confusion about the proper definition of "proof" and "evidence", resistances to cultural and professional changes from within the medical profession, misplaced criticisms from EBM-skeptics and, to some extent, overenthusiasm and reductionism from those who fail to recognise EBM's practical and methodological limitations.

In this paper we will therefore briefly discuss: a) what is EBM and what it is not; b) the difficulties related to the understanding of uncertainty and the appropriate role of pathophysiological reasoning vs. the use of empirical evidence; c) EBM's main limitations and enemies; d) the proper directions that clinical practice and health care policies should pursue to take advantage of the innovations that EBM can bring about. In the Appendix, some examples drawn from the Cochrane Library will be used to illustrate where systematic reviews of available evidence can and cannot be used in order to inform clinical practice in a meaningful way.

What is EBM and what it is not

The same group of people that, years before, started the discipline called "Clinical Epidemiology" (CE) (Evidence-based Medicine Working Group 1992), introduced the term EBM, as we use it no-

[1] This chapter is an extended version of an article first published in the Journal of Medical Ethics in April, 2004, published by BMJ Publishing Group and is reprinted here with permission by BMJ Publishing Group.

wadays, in 1992. CE stemmed essentially from the idea of adapting and expanding epidemiological methods to medical and health care decision-making; CE was in fact defined as "the discipline dealing with the study of the occurrence of medical decisions in relation to its determinants" (Spitzer 1986).

CE has been very successful in illustrating new ways of teaching medicine and training health professionals, and positioned itself around the notion of "critical appraisal skills", as yet another essential ability that – in addition to the interpersonal, diagnostic and prognostic ones – a good doctor should master. An important CE by-product was the documentation that much of the available evidence on diagnosis, prognosis and treatment of diseases was of poor methodological quality and quite often of dubious transferability to everyday clinical practice.

This led to a strong call for improving the scientific basis of clinical practice that was seen as too often dominated by practices of unproven effectiveness. This was the background for the 1992 Journal of American Medical Association (JAMA) article that first used the term "Evidence based Medicine" (Evidence-based Medicine Working Group 1992).

In essence, proponents of EBM said that: "all medical action of diagnosis, prognosis and therapy should rely on solid quantitative evidence based on the best of clinical epidemiological research". Also they stated that: "we should be cautious about actions that are only based on experience or extrapolation from basic science". Indeed, this is not a new concept, as recent research into the history of medicine has documented (Vandenbroucke 1996). Vandenbroucke recently discussed the well rooted historical precedents for the CE and EBM movements in the history of methodological research in medicine quoting, among others, Alexandre Louis who led in 1830 in France an initiative called "Médecine d'Observation" (Vandenbroucke 1996). Finding, not surprisingly, strong resistance from his fellows' environment, Louis stated that "physicians should not rely on speculation and theory about causes of disease, nor on single experiences, but they should make large series of observations, and derive numerical summaries from which real truth about the actual treatment of patients will emerge".

Parallels and differences between now and then are worth noting here. In the early 1800's, proponents of Médecine d'Observation were reacting against a kind of medicine that derived its theories from many things that we would consider "nonsense" by today's scientific standards. Today, EBM acts in the context of a very different environment where modern medical basic science has a solid experimental background. We now know that "Médecine d'Observation" failed shortly after its appearance. A strong reaction from the medical profession together with the absence of contextual conditions account for this unfavourable outcome. Will EBM experience a different outcome as it takes in a more scientifically oriented medical world? In many ways, a similarly strong negative reaction has emerged against EBM today. No doubt one of the reasons of such a negative reaction against EBM has been the fact that it was labelled as a "shift in medical paradigm" (Evidence-based Medicine Working Group 1992; Vandenbroucke 1996). Such a definition would imply that EBM means scientific medicine, and that all medicine practised before it was unscientific. This is not only simplistic but, to any closer scrutiny, profoundly wrong. The difference that needs to be marked between the pre- and EBM era is not that before it people did not use the evidence. Rather, the real failure was the lack of a framework and a set of rules to use the evidence in a systematic and explicit fashion.

Seen in this way, the current fight around EBM and its nature could be advanced by moving the discussion from principles into a more pragmatic perspective where the attention is centred on a "Better use of Evidence in Medicine". This would have the distinct advantage of indicating that it is the way and the rules according to which we use and interpret evidence that needs to be changed.

In contrast with the traditional wisdom of clinical practice, stressing the need for "a better use of evidence in medicine" would indicate that the intuition and unsystematic clinical experience, as well as the pathophysiologic rationale, are an insufficient ground for clinical decision-making. On the contrary, the modern practice of medicine finds its way with formal rules aimed at interpret-

ing the results of clinical research effectively; these rules must complement medical training and common sense of clinicians, whose uncontrolled dominance is no longer ethically and scientifically acceptable.

Struggling for a better use of evidence in medicine also has other important advantages. It challenges the paternalistic and authoritarian nature of much medical practice, and helps the understanding that – even when based on scientific methods – there is a selective and structural imbalance in the nature of the evidence that is available. This is skewed and biased toward therapeutic vs. preventative interventions, and toward simple pharmacological vs. complex behavioural/social care. Acquiring critical appraisal skills – one of the most important tenets of the EBM movement – is the necessary (though not sufficient) immunisation against ignoring that there is a structural imbalance in the research agenda. An imbalance that should be overcome in order to make the sort of evidence that is needed to provide effective and comprehensive health care to all patients fully available(Garattini and Liberati 2000).

The many faces of evidence – proof, causality and uncertainty – and their implications for clinical decisions

Having set this background, it should be clear that some definition of "proof" is also needed to distinguish between scientific medicine and charlatanism. Pathophysiology – i.e. the reference to a mechanism to support the introduction of a new drug – is a criterion that has failed several times in the past: for example, the widespread practice of phlebotomy in eighteen century medicine had some "pathophysiological" basis, but no effectiveness at all. To define what we accept as 'proof' is clearly a problem of transparency of medical practice.

Our thesis is that, unfortunately, the "evidence/lack of evidence balance" is not a black/white one for several reasons:

- For many clinical practices, even if we have well-conducted Randomised Controlled Trials (RCTs), all we can achieve is a "weight-of-evidence" overall evaluation, because we face conflicting results from RCTs (see example 1 in Appendix);
- In other instances, RCTs are not available simply because they have not been conducted, and we only have access to observational investigations;
- Or, the quality of the RCT is poor, so that a meta-analysis is not easily interpretable (examples 2-3 in the Appendix);
- Or, RCTs cannot be easily conducted for practical or ethical reasons (example 4 in the Appendix).

Of course, we also have clear instances in which meta-analyses contribute in an unequivocal way to the adoption or banning of a treatment (examples 5-8 below).

In addition, we need to integrate the scientific evidence with the patient's preferences, with economic constraints, with the health care organization and with ethical obligations. This kind of integration is the object of clinical guidelines, in which ideally evidence is a necessary but insufficient component.

The model we can use comes from a different field, causality, and was suggested by the philosopher John Mackie. Mackie claims that causality cannot be reduced to single necessary and sufficient causes, but rather should be described in terms of elements that he calls INUS (Insufficient Non-redundant component of an Unnecessary Sufficient complex). In his example – "Why did the house burn?" – the causal complex is formed by the association of fire in the fireplace, a strong wind, a defect in the alarm system and the fact that the house is wooden. If we analyse each component, none of them is a single sufficient cause, but only their conjunction gives origin to an overall sufficient complex. However, the complex is not necessary, because the house could burn in many different ways (for example, because I deliberately set it on fire). According to Mackie, although none of the elements is sufficient, at least one is necessary (non-redundant), i.e. in its absence the complex would be ineffective (in the example: eliminating the fire in the fireplace would make the

whole complex ineffective). Let us try to apply this same reasoning to medical decisions. The physician has to integrate several elements into a decisional complex. Let us consider, for example, the prescription of ovariectomy in young women (below the age of 50) with a diagnosis of breast cancer. According to the systematic review in the Cochrane Library, there are 12 RCTs on this topic (example 9 in the Appendix). Most of them show some advantage associated with ovariectomy in that particular category of patients, but none of them reaches statistical significance. However, when a meta-analysis is performed, a statistically significant Odds Ratio of 0.72 is obtained, indicating an 18% reduction of mortality associated with treatment.

Should the oncologist decide to prescribe ovariectomy, based on this statistical evidence, with a rather weak "mechanistic" (biological) basis? Instead of being an exception, the ovariectomy example is the rule: in very few instances are either an RCT (or a meta-analysis) or a biological explanation so strong as to be considered "definitive". In other words, we have to face a large "grey" area that seems to restrict the expectations of both supporters and detractors of the RCT. The practical oncologist might decide that, based on the Cochrane meta-analysis, the weight of evidence is quite strong because there was good a priori evidence on the hormone-sensitivity of breast cancer in young women. Thus, by weighing the empirical evidence coming from trials with a mechanistic background, she can decide to prescribe ovariectomy. But one can reason exactly the opposite: considering the relatively small advantage (18% reduction in mortality), particularly in women who have received chemotherapy, and the important side-effects including reproductive problems, the weight of the empirical evidence could be reduced.

However, it is important to correctly perceive one feature of Mackie's definition of INUS, i.e. that at least one component is necessary (non-redundant); we believe this component is evidence: without evidence there will never be good clinical decisions.

If we accept that evidence is a necessary component, how to weigh the evidence still depends on the definition of effectiveness we adopt. Our point of view is that effectiveness, like disease, is a "fuzzy" concept. Concepts are almost never sharp, i.e. defined on the basis of a single property, but they tend to be fuzzy. In particular, the concept of effectiveness cannot be defined on the basis of a singular property (reducing mortality), but on several properties that are partially overlapping: for some people effectiveness is mainly subjective, for others it is mainly objective, and no single definition is the right one. In summary, we have to face the fact that effectiveness is a "fuzzy" concept. This means that we cannot use the results of clinical trials (or of their synthesis in the form of systematic reviews or meta-analyses) as the only source of information and decisions about care: the work of the physician consists of just integrating different kinds of knowledge, although evidence is a necessary component.

Internal and external enemies of EBM

The difficulties that hamper prudent and systematic use of evidence do not only come from its imperfect and limited nature and from the medical establishment's resistance to change. There are also "internal enemies" (which we will call the "enthusiasts" here) who seem to have limited understanding of EBM's structural limitations, and are dominated by unduly (optimistic) expectations of its sufficiency to guide medical practice. We mention below some of the relevant problems that should be kept in mind before blaming EBM as the sole culprit of its limitations.

First is the bias in the research agenda and the lack of mechanisms to prioritise it with respect to health needs. The increasing commercial influences in health care have produced a structural distortion in the setting of the research agenda, and we see today a systematic bias in research priorities with a lot of (often redundant) data on pharmacological treatments, and a dearth of information on potentially very relevant non-pharmacological interventions. Only recently has this started to attract attention, but this is still far from what would be needed to bring about the necessary changes.

Health services, on the other hand, have not traditionally been interested in investing in research, and with some noticeable recent exceptions (see the UK R&D program as well as part of the NIH research program in the US) this is still the case. Consumers' input into the research agenda is far from systematic, and often patients' charities end up the role of lobbying for a particular disease or health problem rather than for the advocacy of an open and transparent prioritisation (Liberati 1997).

The lack of independence of medical information and the "pollution" caused by the commercial interference in it is another key factor. The imbalance between commercial and independent information is so striking that it may be naive to imagine that EBM alone can maintain its credibility without structural and cultural investments. When relevant information is not properly disseminated and implemented it is as if it does not exist. The recent example of the pharmacological treatment of hypertension is a case in point: here very expensive drugs have been marketed for many years without good evidence of their superiority over the equally effective and much less expensive old diuretics, only thanks to a publicly funded large scale trial (Appel 2002) we now know that millions of dollars have probably been wasted without substantial benefits to patients. Lack of independence and monopoly of scientific information also manifests itself in the increasing medicalisation of common problems, and in the making of "new diseases" as a way to make the health care market bigger and more profitable (Freemantle and Hill 2002). In this scenario, as long as people identify as "evidence-based" procedures and interventions for which studies exist, and as "non-evidence based" areas where studies could exist but have not been carried out because there is no commercial interest in running them, EBM is at high risk of being used as a fashionable and misleading key-word (Charlton 1998).

The "paternalism" inherent in the idea that experts "know it better", and that they are therefore entitled to make decisions on behalf of their patients is a third important enemy. Paternalism has many components, all of which are dangerous and should be recognised. One component comes from the idea that the increasing complexity of modern medicine requires increasing specialisation. More and more medicine is fragmented into sub-specialties, where people have a very deep knowledge of an increasingly narrow spectrum of problems. This technical knowledge leads to an overemphasis of the yield of a particular intervention, where benefits are much too overrated with respect to risks (Sackett 2002). Linked to this, is the inherent conflict of interest that unavoidably links the social and professional prestige of those who are experts in a given field, to the success of the intervention/technology of which they are champions. Like the bias of the research agenda, there are signs of increasing awareness that conflicts of interest are a threat to an equitable and effective practice of medicine, but this is still much less than it should be (Bekelman et al. 2003). And, again, a narrow technical view of EBM could be insufficient, and perhaps even misleading, in this respect.

Lack of awareness of the problems mentioned above is – we believe – a great danger to EBM. Assuming that all relevant "information needs" can be derived from published studies, that all practical skills can be derived from being updated with the medical literature, that methodological rigor is the only dimension that matters – even divorced from clinical and epidemiological relevance – and that health policies should be dictated (rather than more humbly "informed") by evidence of effectiveness alone, are all internal threats that should be seriously considered and challenged.

Have we learned something from EBM?

Having discussed EBM's epistemological, structural and practical limitations, it is also fair to reflect on what has helped us to understand of the major problems and limitations of today's clinical practice and health policies. Given the space constraints of this paper, we will summarise in short statements what we believe are the issues that should inform a health care policy agenda that takes seriously some of the challenges that are ahead of us, if we care for effective and equitable systems of delivering health care. The list is tentative and incomplete and would hopefully be instrumental to stimulate our discussion on EBM's benefit/harm balance thus far.

Clinical Practice

There are no organised mechanisms and efforts to transfer and disseminate information on interventions that work from research to clinical practice; these efforts should become an integral part of the functioning of a good health care system.

Medical practice is fraught with ineffective interventions and long delays before effective care enters clinical practice. Special attention to this should be given a special emphasis in continuing medical education activities.

Doctors and health professionals are not, by themselves, able to critically appraise the results of clinical research; consequently they can be (easily) misguided by unintentionally or intentionally wrong messages. Teaching critical appraisal skills should be an essential part of medical education.

Clinical practice should (and can) be informed by results of systematic reviews of the best available information; knowledge of a given field based on just the few better known studies is dangerous, because it ignores "publication bias" and false negative results, etc. Medical education should stress the idea that knowledge is a "cumulative" rather than a "discrete" process, and appropriate information tools should be made available to all health professionals (The Cochrane Library 2002).

Clinical Research

The quality of medical research is often poor and urgent improvements are needed. Poor quality has to do with both failure to apply appropriate designs and methodologies as well as paying attention to the search for relevant outcomes and interventions that are generalisable outside the research settings (Godwin et al. 2003).

There are not explicit and transparent mechanisms for prioritising research. Health care systems have almost exclusively delegated the responsibility to pharmaceutical companies and the commercial sector in general.

Public and independent support for research is urgently needed (Garattini and Liberati 2000). Conflicts of interest and lack of independence of investigators represent an increasing threat to the credibility of research (Korn 2000; Angell 2000).

Patients' participation can be instrumental in both improving relevance and applicability of clinical research and in facilitating shared decision-making. There is some evidence that, if properly involved at the level of planning and identifying priorities, patients and consumers can provide valuable inputs for research (Liberati 1997). However, this is a process that requires a governance effort in order to avoid increasing fragmentation included in the prioritisation process (Hanley 2000).

Health Care policies

Evidence should "inform" but may often be inadequate to "guide" decision-making at policy level. The sort of evidence that is usually produced by traditional clinical research is too narrow and lacks important elements that are otherwise crucial in policy-making (Maynard 1997).

Resources are often wasted by not acting against the use of ineffective interventions, or by implementing effective interventions with ineffective strategies. A better link between efforts to improve quality at the micro level is needed.

Health care systems should assume more responsibility in knowledge production and they should promote research into areas that are not likely to attract resources due to limited commercial return (Garattini and Liberati 2000).

Conclusions

There is no doubt that EBM does not, and cannot, answer all the epistemological and practical questions surrounding the practice of medicine. On the contrary, it is important that expectations from EBM are appropriate in order to prevent conceptual and practical mistakes. EBM provides methodological tools and a cultural framework. Methodologically, it is useful to understand how we can produce valid and relevant information about the effectiveness of medical care. Culturally, its anti-authoritarian spirit is important to increase the participation of different stakeholders and to increase the opportunity for a multidisciplinary approach to health care problems.

It is clear that, thus far, the potential of EBM has not been fully exploited, and that too narrow views of it have created avoidable confrontations with those that may be concerned that an "EBM-dominated view" can do more harm than good. As efforts by methodologists have chiefly focused on how to design, conduct and interpret studies aimed at assessing efficacy/effectiveness of drugs, EBM these days is mostly "Evidence Based Therapy" with robust tools (i.e. randomised controlled trials) designed especially for assessing the worth of relatively simple interventions. The fact that we currently have limited ability to reliably assess complex interventions, preventative care in general as well as diagnosis or prognosis, should not only be seen as the results of the greater intrinsic complexity of these areas, but also as the consequence of the lower intellectual investments. A reflection, in turn, of the more limited commercial interests is at stake here.

It is our view that – despite the many limitations we have highlighted in this paper – EBM has, at least in some areas of medicine, resulted in better clinical research and greater awareness of health professionals, health administrators and policy makers. A lot remains to be done in order to create a better understanding of the nature of proof, evidence and uncertainty; a more balanced research agenda; more coherent mechanisms to improve quality of care; more substantial cultural efforts to empower patients and consumers. We should, however, be ready to recognise that most of this goes beyond what EBM can do alone and depends, more broadly, on health policy and politics with a capital "P".

Appendix

The Appendix shows summaries from the Cochrane Library (2002). We have chosen nine examples that can be considered typical of a few categories that have been mentioned above. In the first example, the evidence is rather sparse (only 453 patients) and results are conflicting, with a non-statistically significant trial showing protection and one significant trial showing an excess of deaths in the corticosteroid arm. The second example, treatment of giardiasis, is paradigmatic of the lack of good trials, at least in some fields; in this case 34 trials were identified, but only one was methodologically acceptable. The third example is more complex, since the information available was not enough to evaluate the efficacy of treatment. If all missing data (drop-outs) are attributed to disease progression (worst-case scenario), then treatment is associated with a slightly adverse effect. Lack of data of good quality is the main problem in this example. In the fourth case, the subject itself is difficult (information relating to children and adolescents on their cancers), and ethically sensible. One might ask: is it ethical to start a randomised trial that implies that one arm does not receive information or receives information that is considered a priori to be worse than for the other arm? Is the randomised trial an adequate tool for this research subject? What is the best way to ascertain effectiveness? It is not surprising that trials are extremely heterogeneous in this example: one wonders, however, whether heterogeneity could really be overcome, or is not inherent in the subject.

The fifth, sixth and seventh example show how useful meta-analyses can be. In these three cases, individual trials were equivocal, but the overall consideration of their results showed (a) that post-operative radiotherapy causes damage to patients, (b) that aminophylline only causes side-effects

in patients with asthma treated with beta-agonists, and (c) that anticoagulants do more harm than benefit in acute ischaemic stroke. On the opposite side, the eighth example (Warfarin in atrial fibrillation) is paradigmatic of a situation in which a meta-analysis clearly reveals that – more than single trials – the benefits are considerable and the treatment should be transferred into practice. Finally, the ninth example is commented upon in the text above.

1. Corticosteroids in ischaemic stroke. Seven trials involving 453 people were included. Details of trial quality that may relate to bias were not available from most trials. No difference was shown in the odds of death within one year (odds ratio 1.08, 95% confidence interval 0.68 to 1.72). Treatment did not appear to improve functional outcome in survivors. Six trials reported neurological impairment, but pooling the data was impossible because no common scale or time interval was used. The results were inconsistent between individual trials. The only adverse effects reported were small numbers of gastrointestinal bleeds, infections and deterioration of hyperglycaemia across both groups.
2. Treatment of giardiasis. Thirty-four trials were included. Only one trial was without serious methodological flaws. Compared with a placebo, drug treatment was associated with an improved cure rate (odds ratio 11.5, 95% confidence interval 2.3 to 58). Metronidazole treatment longer than three days had a better parasitological cure rate than other long treatment courses (odds ratio 2.4, 95% confidence interval 1.3 to 4.4), but there was significant heterogeneity between the trials. Available evidence has not detected a difference in cure between single dose therapy and longer treatment courses (odds ratio 0.33, 95% confidence interval 0.08 to 1.34). Within the single dose regimens, the available evidence did not demonstrate a difference in parasitological cure rate between tinidazole and other short therapies (odds ratio 3.4, 95% confidence interval 0.95 to 12), but had a higher clinical cure rate (odds ratio 5.3, 95% 2.7 to 10.7).
3. Interferon and multiple sclerosis. Although 1215 patients were included in this review, only 919 (76%) contributed to the results concerning exacerbations and progression of the disease at two years. Specifically, interferon significantly reduced the occurrence of exacerbations (RR=0.80, 95% CI [0.73,0.88], p <0.001) and progression of the disease (RR=0.69, 95% CI [0.55,0.87], p=0.002) two years after randomisation. However, the correct assignment of dropouts was essential to the demonstration of efficacy, most conspicuously concerning the effect of the drug on disease progression. If interferon-treated patients who dropped out were deemed to have progressed (worst case scenario) the significance of these effects was lost (RR=1.31, CI [0.60,2.89], p=0.5). The evolution in magnetic resonance imaging (MRI) technology, in the decade in which these trials were performed, and different reporting of data among trials, made it impossible to perform a quantitative analysis of the MRI results. Both clinical and laboratory side effects reported in the trials were more frequent in treated patients than in controls. No information was available regarding side effects and adverse events after two years of follow-up. The impact of interferon treatment (and its side effects) on the quality of life of patients was not reported in any trial included in this review. Reviewers‘ conclusions: The efficacy of interferon on exacerbations and disease progression in patients with relapsing remitting MS was modest after one and two years of treatment. It was not possible to conduct a quantitative analysis beyond two years. Longer follow-up and more uniform reporting of clinical and MRI outcomes among these trials might have allowed for a more convincing conclusion.
4. Communicating with children and adolescents about their cancer. Six studies met the criteria for inclusion. They were diverse in terms of the interventions evaluated, study designs used, types of people who participated and the outcomes measured. One study of a computer-assisted education programme reported improvements in knowledge and understanding about blood counts and cancer symptoms. Two out of two studies of school reintegration

programs reported improvements in some aspects of psychosocial well-being (one in anxiety and one in depression), social well-being (two in social competence and one in social support) and behavioural problems; and one reported improvements in physical competence. Reviewers' conclusions: Interventions to enhance communication involving children and adolescents with cancer have not been widely or rigorously assessed. The weak evidence that exists suggests that some children and adolescents with cancer may derive some benefit from specific information-giving programs, and from interventions that aim to facilitate their reintegration into school and social activities. More research is needed to investigate the effects of these and other related interventions.

5. Post-operative radiotherapy in non-small cell lung cancer. 2128 patients from 9 trials were included (median follow-up of 3.9 years). The results show a significantly adverse effect of PORT on survival with a hazard ratio of 1.21 or 21% relative increase in the risk of death. This is equivalent to an absolute detriment of 7% at 2 years (95% confidence interval 3 to 11%) reducing overall survival from 55% to 48%. Exploratory subgroup analyses suggested that this detrimental effect was most pronounced for patients with stage I/II, N0-N1 disease, whereas for stage III, N2 patients there was no clear evidence of an adverse effect.
6. Aminophylline in acute asthma. Fifteen studies were included. Overall, the quality of the studies was only moderate; concealment of allocation was assessed as clearly adequate in only seven (45%) of the trials. The doses of aminophylline and other medications and the severity of asthma varied between studies. There was no statistically significant effect of aminophylline on airflow outcomes at any period of time. The aminophylline treated group had higher values of PEFR at 12 (PEFR 8 L/min or 2.3%) and 24 hours (PEFR 22 L/min or 6.4%), but these were not significant (p >0.05). Analyses of two subgroups were performed, by grouping studies according to mean baseline airflow limitation (n=11 studies) and the use of any steroids (n=9 studies). There was no relationship between baseline airflow limitation, nor the use of steroids, on the effect of aminophylline. Aminophylline treated patients reported more palpitations/arrhythmias (OR: 2.9; 95% CI: 1.5 to 5.7) and vomiting (OR: 4.2; 95% CI 2.4 to 7.4), but no difference was found in tremor or hospital admissions. Reviewers' conclusions: In acute asthma, the use of intravenous aminophylline did not result in any additional bronchodilation compared to standard care with beta-agonists. The frequency of adverse effects was higher with aminophylline. No subgroups in which aminophylline might be more effective could be identified. These results should be added to consensus statements and guidelines.
7. Anticoagulants in ischaemic stroke. Twenty-one trials involving 23,427 patients were included. The quality of the trials varied considerably. The anticoagulants tested were standard unfractionated heparin, low-molecular-weight heparins, heparinoids, oral anticoagulants, and thrombin inhibitors. Based on eight trials (22,450 patients), there was no evidence that anticoagulant therapy reduced the odds of death from all causes (odds ratio 1.05, 95% confidence intervals 0.98-1.12). Similarly, based on five trials (21,846 patients), there was no evidence that anticoagulants reduced the odds of being dead or dependent at the end of follow-up (odds ratio 0.99, 95% confidence intervals 0.94-1.05). Although anticoagulant therapy was associated with about 9 fewer recurrent ischaemic strokes per 1000 patients treated, it was also associated with a similar sized 9 per 1000 increase in symptomatic intracranial haemorrhages. Similarly, anticoagulants avoided about 4 pulmonary emboli per 1000, but this benefit was offset by an extra 9 major extracranial haemorrhages per 1000. Sensitivity analyses did not identify a particular type of anticoagulant regimen or patient characteristic associated with net benefit.
8. Warfarin in patients with atrial fibrillation. Fourteen articles were included in this review. Warfarin was more efficacious than a placebo for primary stroke prevention {aggregate odds ratio (OR) of stroke=0.30 [95% Confidence Interval (C.I.) 0.19,0.48]}, with moderate evi-

dence of more major bleeding {OR=1.90 [95% C.I. 0.89,4.04]}. Aspirin was inconclusively more efficacious than a placebo for stroke prevention {OR=0.68 [95% C.I. 0.29,1.57]}, with inconclusive evidence regarding more major bleeds {OR=0.81[95% C.I. 0.37,1.78]}. For primary prevention, assuming a baseline risk of 45 strokes per 1000 patient-years, warfarin could prevent 30 strokes at the expense of only 6 additional major bleeds. Aspirin could prevent 17 strokes, without increasing major haemorrhage. In direct comparison, there was moderate evidence for fewer strokes among patients on warfarin than on aspirin {aggregate OR=0.64[95% C.I. 0.43,0.96]}, with only suggestive evidence for more major haemorrhage {OR=1.58 [95% C.I. 0.76,3.27]}. However, in younger patients, with a mean age of 65 years, the absolute reduction in stroke rate with warfarin compared to aspirin was low (5.5 per 1000 person-years) compared to an older group (15 per 1000 person-years). Low-dose warfarin or low-dose warfarin with aspirin was less efficacious for stroke prevention than adjusted-dose warfarin. Reviewers' conclusions: The evidence strongly supports warfarin in AF for patients at average or greater risk of stroke, although there is clearly is a risk of haemorrhage. Although not definitively supported by the evidence, aspirin may prove to be useful for stroke prevention in sub-groups with a low risk of stroke, with less risk of haemorrhage than with warfarin. Further studies are needed for low molecular weight heparin and aspirin in lower risk patients.

9. Ovariectomy in breast cancer. Among 2102 women aged 50 or under when randomised, most of who would have been premenopausal at diagnosis, 1130 deaths and an additional 153 recurrences were reported. 15-year survival was highly significantly improved among those allocated ovarian ablation (52.4 vs. 46.1%, 6.3 [SD 2.3] fewer deaths per 100 women, logrank 2p=0.001), as was recurrence-free survival (45.0 vs. 39.0%, 2p=0.0007). The numbers of events were too small for any subgroup analyses to be reliable. The benefit was, however, significant both for those with ("node positive") and for those without ("node negative") auxiliary spread when diagnosed. In the trials of ablation plus cytotoxic chemotherapy versus the same chemotherapy alone, the benefit appeared smaller (even for women with oestrogen receptors detected on the primary tumour) than in the trials in the absence of chemotherapy (where the observed survival improvements were about six per 100 node-negative women and 12 per 100 node-positive women). Among 1354 women aged 50 or over when randomised, most of who would have been perimenopausal or postmenopausal, there was only a non-significant improvement in survival and recurrence-free survival.

References

1. Angell M (2000) Is Academic Medicine for sale? New England Journal of Medicine 342: 1516-18
2. Appel L (2002) The verdict from the ALLHAT: thiazide diuretics are the preferred initial therapy for hypertension. JAMA 288: 3039-3041
3. Bekelman JE, Li Y et al. (2003) Scope and impact of financial conflicts of interests in biomedical research. JAMA 289: 454-465
4. Charlton B (1998) The rise and fall of EBM. Quarterly Journal of Medicine 91: 371-374
5. Evidence-based Medicine Working Group (1992) Evidence based Medicine: a new approach to the teaching of medicine. JAMA 268: 2420-2425
6. Freemantle N, Hill S (2002) Medicalisation, limits to medicine or never enough money to go around. BMJ 324: 864-5
7. Garattini S, Liberati A (2000) Bias by omitted research. BMJ 321: 845-6
8. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R (2003) Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 3: 28
9. Hanley B (2000) Involvement works: the 2nd report of the Standing Committee on Consumer Involvement in NHS research. www.conres.co.uk
10. Korn D (2000) Conflicts of interest in biomedical research. JAMA 284: 2234-36

11. Liberati A (1997) Consumer participation in research and health care. BMJ 315: 499
12. Maynard A (1997) Evidence based medicine: an incomplete method to inform treatment choice. Lancet 349: 126-128
13. Sackett DL (2002) The arrogance of preventive medicine. CMAJ 167: 363-4
14. Spitzer WO (1986) Clinical Epidemiology. Journal of Chronic Disease 39: 412-15
15. The Cochrane Library, Issue 2, 2002 http://www.update-software.com/Cochrane/HTM
16. Vandenbroucke IP (1996) Evidence based medicine and "Médecine d'Observation". Journal of Clinical Epidemiology 49: 1335-38

14 Evidence-based Medicine and Equity: The Exclusion of Disadvantaged Groups

Wendy A. Rogers

Introduction

Being healthy is almost universally acknowledged as a significant good, both in itself and for the instrumental value of good health in achieving other important goals in life. At some periods in history, poor health has been seen as a matter of fate or chance, a natural disaster akin to earthquakes or floods. If this were the case, the moral duty on society would be to help those who were afflicted, but there could be no responsibility to influence the distribution of ill health as this would be considered to be beyond human control. In the early twenty first century, we have a different view. We know that the burdens of health are unevenly distributed, both within societies and between societies. Poor health is strongly associated with poverty, with disadvantage, with social exclusion. We also know that access to good health care tends to vary inversely with the need for it, and that this Inverse Care Law is as true today as when it was first described thirty years ago (Watt 2002). This information should change the nature of our moral responsibilities with regard to health care; not only should we treat those who are sick, but we should act to prevent this maldistribution of both ill health and health care, and to ensure that the disadvantaged receive the best rather than the worst of our health care. We might do this in a number of ways, for example through individual clinical interventions, through systemic approaches to health care, or by using policy to direct activities in the health sector and beyond.

Evidence-based Medicine is an approach to health care that seems to offer multi-level assistance in creating and delivering fairer health care. This implicit promise of fairness operates in at least two ways. First, the processes of EBM are committed to objectivity through the use of strictly standardised methods, thereby eliminating opportunities for subjective decisions and possible discrimination. Second, the findings of EBM can be used to ensure fair distribution of effective interventions across the population, at the individual level through the use of evidence-based practice, and at the population level through the use of evidence-informed health policy and resource-allocation decisions. Avoiding discrimination and ensuring fair distribution of effective treatments are potentially powerful tools in addressing the exclusion of disadvantaged groups and achieving greater equity in health care. This paper explores the effects upon health equity of using EBM as the basis for clinical care and health policy decisions, using a focus on disadvantaged groups. „Disadvantaged" is a very general descriptor; here I take the term to include groups subject to social exclusion or deprivation for reasons such as low socio-economic status, ethnicity, age, gender, mental ill health or similar. If EBM improves the health care (and so health) of those with the greatest burden of ill health, then we are some way towards meeting our responsibilities. The paper is divided into three main sections, looking at the way evidence is created, the way EBM is clinically applied, and the policy uses of EBM.

Creating the evidence for EBM

One of the major claims of Evidence-based Medicine is that it provides objective evidence about the effectiveness of interventions. This is achieved through the use of research methods that aim to minimise the risk of bias, such as randomised controlled trials (RCTs). Results from multiple RCTs are pooled using systematic reviews and meta-analysis to give an overall result, which is considered to be the definitive best available evidence about a specific intervention. The evidence for EBM is created from scientific research. To date the majority of this research relates to testing the effectiveness of treatment interventions on research populations in clinical controlled trials. The first question we should ask is whether, and to what extent, disadvantaged groups participate in the production of research evidence. The production of research includes commissioning, research design, and participation in clinical trials.

Research commissioning and design

Historically, research commissioning and design has been the domain of scientists together with funders of research. Funders may be private such as pharmaceutical companies, or public such as government-funded bodies. Almost by definition, the disadvantaged are not well represented amongst these groups. Information in this area is scarce, however, one report of a UK survey investigating consumer involvement in designing, conducting and interpreting RCTs found that one third of specialised trial centres had involved consumers in some part of the process (Hanley et al. 2001). Consumers were defined as "patients and potential patients, carers, organisations representing consumers' interest, members of the public who are targets of health promotion programmes and groups asking for research because they believe they have been exposed to potentially harmful circumstances, products or services." This broad definition potentially includes members of disadvantaged groups. The survey asked for information about the background of consumers who had been involved, but this information was not reported in the published paper. Without this information it is difficult to reach any conclusions about the presence of people from disadvantaged groups. However, even without this information, the type of consumer involvement reported indicates that their role is largely limited to helping the research run smoothly once the really important decisions have been finalised. The commonest form of consumer involvement was drafting or reviewing information for trial participants, followed by promoting recruitment, and membership of the steering committee. There seems little opportunity in this system for substantial input (for example in determining research topics, or interventions to be tested, or outcomes to be measured), from any consumers, let alone from people who's voices are more generally unheard (Rogers 2002).

Participation in research

Discussions about the kinds of populations who are generally included as trial participants are not new. There are two issues here, the first is whether or not potentially important variables such as gender, ethnicity and socio-demographic markers are recorded, and second, what does this information (if available) tell us about the participation of disadvantaged groups. There is some limited information available on the recording of variables. Swanson and Bailar assessed the heterogeneity of subgroups in cancer treatment and prevention trials published in eleven journals between 1990 and 2000. They found that age and gender were reported in over ninety percent of the 261 published trials, but that less than thirty percent reported race or ethnicity (Swanson and Bailar 2002). This lack of reporting precludes the possibility of any sub-group analysis to identify significant differences in racial or ethnic sub-groups. Sub-group analysis presents its own challenges, however it

may be possible for appropriately rigorous analysis of sub-group data with some innovation in the application of accepted trial techniques (Moye and Powell 2001).

The gender imbalance in research is well known and well reported (Dresser 1992; Sherr 2000), yet despite this, there continue to be important gender discrepancies (Ebrahim and Smith 1997; Rochon et al. 1998; Hooper et al. 2001). Gender is easy to define and record as a variable in trial data. Other potential markers for disadvantage are less so, but there is enough recorded information to suggest that people from ethnic minorities and low socio-economic groups are generally under-represented. Much of the information about the research participation rates of ethnic minorities comes from the US, where study after study reports either under-representation or lack of data about ethnicity.

In adult cancer treatment and prevention trials, participation rates for minority populations and the medically underserved are lower than those of white non-Hispanics. The most represented group in these trials are white, middle class, highly educated men (Giuliano et al. 2000). The findings are similar for breast cancer research. A review of trials of selective oestrogen receptor modulators, drugs that may be useful both in the prevention and treatment of breast cancer, found limited ethnic variability amongst the 50,000 women who had participated in the trials, compared with the ethnicity and racial make-up of the population of women who need treatment for breast cancer (Taylor 2001). This is despite the fact that breast cancer rates in women vary with race and ethnicity. In the US white non-Hispanic women have the highest rates, closely followed by Hawaiian, African-American, Japanese and Alaska Native women (Miller et al. 1996). A study of trials run by the National Cancer Institute found that patients enrolled onto clinical trials were significantly less likely to be uninsured and more like to have Medicare health insurance than the patients with cancer in the community, and that geographic areas with higher socio-economic levels had higher levels of clinical trial accruals (Sateren et al. 2002). For children, the situation seems to be different, as minority children with cancer have been found to be proportionately represented (Sateren et al. 2002; Bleyer et al. 1997).

There are a number of recognised barriers to the participation of minorities in cancer research, such as study duration, cost, time, follow-up visits and side effects, as well as cultural characteristics and attitudes, beliefs and knowledge about research (Giuliano et al. 2000). However, there is little evidence to date that these barriers are being addressed in robust or systematic ways.

Other areas of research demonstrate similar patterns. A review that compared the characteristics of patients with heart failure in RCTs with those of patients with heart failure in the community, found that the participants in the trials were markedly different from patients in the community. In particular, trial patients with heart failure were younger, more often male, more likely to have a sub-normal systolic ejection fraction and were most commonly white. This represents only a relatively small segment of the heart failure population, with significant under representation of minorities, women and the elderly. Perhaps most concerning, the authors did not find any marked change in the characteristics of patients in trials over time: RCTs of the 1990s continued to focus on young, white, male patients (Heiat et al. 2002).

These results are not surprising. The difficulties in recruiting people into research from areas with severe economic deprivation have been recognised, and tend to be accepted because of the increased cost of trying to increase recruitment rates in these populations (Watt 2002). If the distribution of ill health was a matter of fate, we might be able to accept this reasoning. But given our present state of knowledge, exclusion on grounds of cost to the researchers seems inequitable, given the likely consequences. What are the implications of the absence of disadvantaged groups in trials, or the lack of identifiable data? First there is the general observation that people in trials often fare better than people who receive treatment outside of a trial. Lack of participation in trials effectively removes this benefit from disadvantaged groups. More importantly, however, the lack of participation means that there is a paucity of research evidence about which interventions are effective in disadvantaged groups. The generalisability of the findings are limited to people who are sufficiently

similar, in relevant respects, to the trial participants. This begs the question as to which respects are relevant; physical, cultural and structural issues may all be relevant in different ways.

There are two main ways that physical differences may be important in working out whether trial results are applicable to different populations. The first is to do with the presence or absence of co-morbidities. Most RCTs exclude people with more than one disease, as the trial aims to find out specific information about the effect of a single intervention on a single disease state. Indeed, part of the strength of the evidence derived from RCTs lies in strictly controlling as many variables as possible, so that any differences between the intervention and control group may be attributed to the intervention rather than to some other factor. The presence of multiple diseases and their various treatments would weaken this process, hence the exclusion of people with co-morbid conditions. This means that the high quality evidence generated and used in clinical practice and for policy, applies only to people with single conditions. However, co-morbidity, like ill health, tracks disadvantage. The more disadvantaged a person is, the more likely he or she is to have greater severity and complexity of a greater number of conditions (Watt 2002). The people with the single conditions, to whom the evidence applies, are under-represented amongst the disadvantaged. The current system of generating evidence about single conditions leads to the lack of an evidence base for the treatment of people with co-morbidities. Many EBM-derived guides are inapplicable in the care of disadvantaged individuals and communities (Starfield 2001).

Second, there is reason to believe that there are some racial or ethnic and gender variations in responses to drug treatments (Johnson 2002). These variations, due to differences in the metabolism of some drugs, result in variable circulating concentrations of active drug, so that the same doses of particular drugs given to people of different races can have variable effects. The extent and nature of these potentially significant differences have not been fully investigated, but there are several classes of drugs known to be affected, including cardiovascular, psychotropic and central nervous system drugs (Matthews 1995). The consequences of these differences are also not fully understood, but known examples include the increased sensitivity of Asian-Americans to beta blockers (used to control high blood pressure), and the decreased effectiveness of ACE inhibitors in African-Americans (Matthews 1995). However, this information does not find its way onto the labels of products. A review of product labelling for 185 drugs approved by the Center for Drug Evaluation and Research between 1995–1999 found that eight percent (15/185) described differences related to race, but only one product label recommended a change in dosage based on racial differences (Evelyn et al. 2001).

It should be relatively easy to create the evidence about correct doses of various drugs, where relevant, for different ethnic groups and for women. More difficult is the task of unravelling cultural and structural issues to do with accepting health care, but this is a crucial part of generating good evidence. We need to know not only that an intervention works in ideal trial circumstances with a well-defined population, but also that it works in the context of routine care, with heterogeneous populations. A drug may be very effective at controlling pain relief, but if it is culturally unacceptable to admit suffering from pain, the drug will not work for that person because they will not be able to articulate the need for it.

Some of these problems could be overcome by performing new research targeting disadvantaged groups, generating evidence about the efficacy of interventions in the trial situation, and also about effectiveness when interventions are delivered as part of normal care in the community. However, this is unlikely to occur. As already mentioned, the increased cost is one factor, together with the difficulties in recruiting and retaining disadvantaged participants in trials. In addition, research that repeats testing of an existing intervention, albeit with a different population or in a different setting, is far less attractive both to funders and to researchers, than research involving new interventions. Such trials are likely to be small rather than the mega-trials currently in vogue, and this raises its own problems. Small trials are methodologically challenging: the smaller a trial, the larger the treatment effect necessary for the results to be significant, so that it is easy to miss small effects that

may be clinically, but not statistically, significant (Sterne et al. 2001). Trials that show no statistically significant benefit are less likely to be published, making them unattractive to researchers who rely on publications for career advancement and further research funding. The potential lack of publication creates problems in terms of the information entering the evidence base. In theory, systematic reviews should include unpublished results, but in practice, budgetary and temporal constraints can preclude this.

Further reasons why small studies that target disadvantaged populations are unlikely to be performed concern the funding of research. Pharmaceutical companies, who are major funders of research worldwide, are interested in products that will find a market, irrespective of the causes and distribution of ill health within populations. There is no incentive for these companies to fund small studies with disadvantaged groups. Even if the research identifies effective drug treatments, the market will be small in terms of absolute numbers, and will comprise people who may not be able to afford the treatment. Also, such research is unlikely to lead to new patents, and as the existing patent is already in place (and running out), the research is unlikely to add to the existing and expected profits from the drug. On the other hand, pharmaceutical companies are interested in developing newer and more effective versions of existing successful drugs, the so-called "me too" research. This means that comparative trials are an important part of their research and development programmes. Comparing new treatments with existing ones leads to clustering of research around a narrow range of pharmaceutical interventions in research-accessible populations with no co-morbidities. There is no incentive in this system to trial existing drugs in poor populations with multiple co-morbidities.

In summary, disadvantaged groups rarely have a voice in commissioning and designing research, and appear to have limited participation in trials of new interventions. In addition, co-morbidity is more common in disadvantaged groups, so that the evidence that is generated may not be applicable. The impact of race, ethnicity, gender and age on treatment effects are not well documented, nor is the practical effectiveness of interventions delivered in disadvantaged communities. These factors result in a lack of research evidence about effective interventions for this group. As the presence of evidence of effectiveness is increasingly a requirement for the provision of health care, this is a serious matter. Health care funders, both government and private, are accountable for their spending; funding interventions that are of proven effectiveness is seen as part of this. Apart from the intuitive attractiveness of funding interventions that work, no health care provider wants to be accused of wasting resources on interventions that do not work. For disadvantaged groups, this can be a vicious cycle; to receive the newest and best treatments requires evidence, but exclusion from research prevents the generation of evidence and hence access to treatments. The barriers to performing research with disadvantaged groups are considerable, but unless they are overcome, the disadvantaged will remain disenfranchised from the goods of EBM. The power of EBM to mandate treatment cannot function if there is no applicable evidence.

The clinical application of EBM

For the reasons outlined above, there is frequently a lack of evidence about the effectiveness of interventions in disadvantaged groups. However, where there is applicable evidence, we might hope that this can be used to ensure the equitable distribution of evidence-based interventions for all of those who need them. The presence of evidence should ensure access for all, especially in situations where the delivery of care has been variable. In some cases, this has occurred; evidence-based guidelines have been used to improve treatment rates in disadvantaged groups. For example, one study found that the introduction of an evidence-based guideline on haemodialysis led to dramatic improvements in access to treatment for African American patients, with a ninety two percent increase in

the proportion receiving adequate haemodialysis (Owen et al. 2002). The use of evidence-based guidelines helped to ensure equal treatment for all those needing dialysis, irrespective of race.

This is encouraging, however, other areas of health care are not doing so well. Improved treatment of cardiovascular disease has been one of the flagships of EBM, especially the use of thrombolytics ("clot-busting" drugs) in the treatment of acute myocardial infarctions, or heart attacks. Despite widespread acceptance of the efficacy of thrombolytic drugs, it is clear that not all those who would benefit actually receive these drugs. A review of 26,575 Medicare beneficiaries in the US found that, despite meeting all of the eligibility criteria, and after adjusting for differences in clinical and demographic characteristics and clinical presentation, African Americans with heart attacks were significantly less likely than whites to receive treatment (Weissman 2000). The situation is not so different in Europe; a review of 4035 patients with acute myocardial infarctions found that women and the elderly were less likely to receive thrombolytic treatment compared with younger and male patients (European Secondary Prevention Study Group 1996). The tendency to undertreat heart disease in the elderly extends beyond the use of thrombolytics; a Canadian review found that elderly cardiac patients receive consistently fewer prescriptions for proven cardiac therapies, including aspirin and statins, despite the fact that the effectiveness of these therapies is not affected by age (McAlister 1999). Given the higher mortality risks from heart disease for the elderly, this is the opposite of what we might expect.

These studies show us that evidence about efficacy has not succeeded in changing patterns of treatment. In particular, disadvantaged groups are less likely than others to receive at least some efficacious treatments. These studies do not explain why this happens; we are left speculating about various reasons. Under-representation in research has been suggested as a possible reason (Anon 1996; Rochon et al. 1998). Clinicians might be aware of a lack of research evidence for some groups, and fall into the trap of mistaking the absence of evidence of efficacy as evidence of no efficacy. Unless careful, it is quite possible to confuse lack of proof about effectiveness with proof of ineffectiveness. Perhaps the best evidence in the world is not enough to overcome deeply entrenched patterns of discrimination. Either way, the idea that EBM will lead to fairer treatment is challenged.

The examples discussed here concern variable access to proven treatments in eligible patients. There is less information about withholding treatments because patients do not fit the profiles of the research populations. The presence of other illnesses (co-morbidities) may have unknown effects on proven therapies, so that clinicians do not know whether evidence about effective interventions applies to patients with multiple illnesses. There does not seem a way out of this vicious circle, for as we have already seen, the presence of other illnesses, which is more likely in disadvantaged populations, precludes participation in research and hence the generation of applicable evidence.

The policy uses of EBM

EBM is playing an increasingly prominent role in health care policy (Biller-Andorno et al. 2003). The aim of health policy is to make the best overall health choices for the population (Norheim 2003). Health policy makers have to make resource allocation decisions, about which treatments to provide and which to limit or withdraw. One of the ethical imperatives in this process is to produce the greatest good for the greatest number. EBM has been described as a "natural building block" in this political process of rationing (Norheim 2003 P 310), in part because EBM seems to offer clear cut proof about which treatments work and which do not, and in part because its epidemiological methods deal with benefits to aggregated populations in ways that echo the utilitarian reasoning of public health.

One of the major policy roles of EBM is to inform purchasing decisions about health care at regional or national levels. In the UK, the National Institute for Clinical Excellence (NICE) is the body charged with assessing evidence about efficacy as part of making national recommendations about

the introduction of new interventions. This task was described in the following terms: "NICE is crucial to our plans for fair and equal treatment of patients within a truly national health service. Internal markets and postcodes were never an acceptable way to run a health service. NICE guidance will provide a common currency of effectiveness for the NHS, to inform and assist decision-making about treatment and care at all levels, national, local and individual" (Dobson 1999).

As mentioned, part of the rationale for NICE was to end uneven access to various treatments across the UK, known as postcode rationing. This was widely perceived to be unfair, as people living in some areas were able to receive treatments that were unavailable in other areas. Health authorities are now obliged to fund interventions that are approved by NICE, guaranteeing nation-wide access to approved interventions. Since 1999, these have included new drugs for the treatment of various cancers, obesity, Alzheimer's and motor neurone disease, diabetes, attention deficit hyperactivity disorder, flu and others (Raferty 2001). Prima facie, this is good news for disadvantaged groups, as no matter how poor their community or district, their local health authority must provide access to NICE-approved interventions.

There are now concerns however, that the aim of equitable distribution may fail. NICE does not have an explicitly formulated standard of proof, so that once a treatment reaches this benchmark, it is approved. The process is far more opaque, and involves cost-effectiveness analysis as well as responses to lobbying by patient groups and confidential submissions from the pharmaceutical industry. NICE publishes information about the estimated costs of its guidance, using cost per quality adjusted life year (QALY). Costs for approved interventions vary between GBP 3,000 and GBP 43,000 per QALY (Raferty 2001). All approved interventions are deemed to be cost-effective, and have to be made available. The result has been described as "health authorities slavishly funding marginally cost effective drugs approved by NICE and diverting funding away from more cost effective existing services that lack politically powerful advocates … NICE has effectively become an advocacy mechanism by which lobbies of specialists and their supporters in the pharmaceutical industry extract more public money from the NHS" (Cookson et al. 2001 P 744).

The overall effects of this process on patterns of expenditure and service provision are unknown, and to some extent unknowable: we just do not know what else would have been funded if the money had not been spent on NICE-approved treatments. However, as described the process has serious implications for disadvantaged groups. The presence of evidence is used by both clinicians and the pharmaceutical industry as a trigger to seek approval from NICE for certain interventions. This gives those interventions a spurious advantage over other interventions that are discounted, because of lack of evidence and lack of advocacy by powerful groups. Given the predominance of pharmaceutical interventions amongst those reviewed by NICE, we should be concerned about the long-term extent of this pharmaceutical creep. Given the way that disadvantaged groups are underrepresented in research, as recipients of evidence-based care, and amongst the advocates lobbying NICE, we should be very concerned about the impact of evidence-based purchasing on health care for the disadvantaged. There is the very real risk that services to disadvantaged groups that lack an evidence base are being sacrificed to pay for new, marginally cost-effective evidence-based interventions that do not apply to the disadvantaged.

EBM and fair health care

To be fair or just, we might think that a health care system should take into account various factors, such as need, benefit, equity (of access, of opportunities, of outcomes), or personal preferences. It is almost impossible to rank these, so that probably the best we can do is try to take account of all of them to some extent. Evidence-based Medicine addresses only one of these factors: capacity to benefit. A treatment may be effective at preventing something as important as coronary heart disease, or as trivial as premature greying of the hair. There is no logical relationship between proof

of effectiveness and the urgency/importance of the condition for which the intervention is effective. However, once the capacity to benefit is proved, this diverts attention away from more important questions, such as whether or not this is an important health problem, or whether this should be provided given the impact on other aspects of health care. Rather than starting with a set of priorities, and then performing the research to find out how best to achieve the agreed ends, EBM inverts the system. The research is performed, often for largely commercial reasons, and then the presence of this "evidence" is taken as some kind of imperative. Of course, institutions such as NICE have criteria other than proof of effectiveness that enter their considerations, but hard evidence is very seductive, especially when used by lobby groups to argue their cause. Politically it can be very difficult to refuse people access to treatments that are evidence-based, even if there are robust justice-related reasons for doing so.

If we take into account the social determinants of health, this primacy accorded to capacity to benefit seems particularly invidious, as it does little to address inequalities in health. Much ill health is the result of disadvantage; although the exact mechanisms are not well understood, we can assume that they are more to do with the material circumstances of people's lives than individualistic factors. Yet EBM turns our attention away from social and cultural factors that influence health, and focuses on a narrow biomedical model of health and disease that is primarily individualistic. Instead of looking at ways to prevent ill health and ameliorate disadvantage, we are directed towards a system of health care that is very good at delivering highly sophisticated, and often expensive, individual treatments to those who are able to access them. This leaves those with the greatest burden of health disenfranchised, as there is little relevant research, poor access to treatments, and attention is diverted away from activities that might have a much greater impact on health.

The WHO has recently published the World Health Report 2002, which identifies ten major preventable risks that account for forty percent of annual deaths worldwide. The list includes childhood and maternal underweight, unsafe water, and sanitation and hygiene. As we might expect, the greatest burden of health risks is borne by the poor countries, and by the disadvantaged in all societies. A recurrent theme in the report is the need for evidence-based interventions. Possible interventions are listed for each risk; interventions for underweight include micronutrient supplementation and fortification (WHO 2002). This evidence-based advice almost beggars belief: if people are underweight and undernourished, surely they need food, rather than micronutrient supplementation? We are left with the uneasy suspicion that as there are no RCTs to prove that food is an effective method of reducing malnutrition, such a common-sense measure lies outside the evidence-base, and is therefore excluded from consideration.

Conclusion

So far my criticisms of EBM have been of its current processes and applications, that we still need to ask whether EBM is intrinsically inimical to the interests of disadvantaged groups, or whether it is just that its potential is not being realised. This is a difficult question.

Of course there is a place for evidence in health care: we need to know what kinds of interventions improve health outcomes, and which ones do not. There is no necessary reason why appropriate and applicable evidence (broadly interpreted) cannot be gathered about the effectiveness of interventions for disadvantaged groups. This research will most likely be difficult, expensive and have little commercial potential, but these are not morally valid reasons for not performing such research. Once performed, the results of this research should inform health care policies and expenditure, and lead to real improvements for the disadvantaged. Understood like this, EBM is a potentially valuable tool.

However, the current research climate does not inspire optimism about this possible change in the direction of EBM. Medical research reflects the priorities of the rich, with ninety per cent of re-

search funding investigating the diseases of ten percent of the world's population (Global Forum for Health Research 2002). A range of forces shapes this research agenda, but commercial forces have become dominant, so that research into interventions that are unpatentable is less and less likely (Horrobin 2003). EBM has become a tool for commercial ends, narrowing the range of possible interventions and seeking ever more marginal benefits with ever more expensive drugs. The overall effect of EBM occurring within a biomedical framework is to provide evidence about the effectiveness of medical interventions for a narrowly defined range of disease states. This narrow focus is especially worrying when governments adopt policies based upon evidence. A commitment to funding only health interventions that are supported by evidence legitimises governments' continued reluctance to tackle the wider causes of ill health. Treating the effects of poverty and discrimination rather than the causes, allows perpetuation of the social structures which cause ill health.

It is possible that EBM could serve the interests of the disadvantaged, but this will only happen with a commitment to justice in health care at the highest possible levels, funded accordingly. Without such commitment, EBM will continue to foster an individualistic treatment-oriented approach to health care, using a system that largely excludes the vulnerable and disadvantaged, and may in fact increase inequities.

Acknowledgements

I would like to thank members of the Evibase group for their support and helpful discussions, and also participants in the Beijing Symposium on Feminist Approaches to Bioethics, November 2001. A shorter version of this paper appeared in the *Journal of Medical Ethics*. This research was funded by the National Health and Medical Research Council of Australia in the form of the Sidney Sax Fellowship (ID 007129).

References

1. Biller-Andorno N, Lie R, Ter Meulen R (2002) Evidence based medicine as an instrument for rational health policy. Health Care Analysis 10: 261-275
2. Bleyer W, Tejeda H, Murphy S, Brawley O, Smith M, Ungerleider R (1997) Equal participation of minority patients in U.S. national pediatric cancer clinical trials. Journal of Pediatric Hematology Oncology 19: 423-427
3. Cookson R, McDaid D, Maynard A (2001) Wrong SIGN, NICE mess: is national guidance distorting allocation of resources? British Medical Journal 323: 743-5
4. Dobson, F (1999) Secretary of State's speech. http://www.nice.org.uk/Embcat.asp?page=oldsite/back/frank_dobson.htm&d=907
5. Dresser R (1992) Wanted, Single white male. Hastings Center Report 22: 24-29
6. Ebrahim S, Smith GD (1997) Systematic review of randomised controlled trials of multiple risk factor interventions for preventing coronary heart disease. British Medical Journal 314: 1666-1674
7. Evelyn B, Toigo T, Banks D, Pohl D, Gray K, Robins B, Ernat J (2001) Participation of racial/ethnic groups in clinical trials and race-related labelling: a review of new molecular entities approved 1995-1999. Journal of the National Medical Association 93: 18S-24S
8. European Secondary Prevention Study Group (1996) Translation of clinical trials into practice: a European population-based study of the use of thrombolysis for acute myocardial infarction. Lancet 347: 1203-1207
9. Giuliano A, Mokuau N, Hughes C, Tortolera-Luna G, Risendal B, Ho R, Prewitt T, McCaskill-Stevens W (2000) Participation of minorities in cancer research: the influence of structural, cultural and linguistic factors. Annals of Epidemiology 10: S22-S34
10. Global Forum for Health Research (2002) The 10/90 report on health Research. Global Forum for Health Research, Geneva. Cited in: Horton R (2003) Medical journals: evidence of bias against the diseases of poverty. Lancet 361: 712-713
11. Hanley B, Truesdale A, King A, Elbourne D, Chal-

mers I (2001) Involving consumers in designing, conducting, and interpreting randomised controlled trials: questionnaire survey. British Medical Journal 322: 519-523
12. Heiat A, Gross C, Krumholz H (2002) Representation of the elderly, women, and minorities in heart failure clinical trials. Archives of Internal Medicine 162: 1682-1688
13. Hooper L, Summerbell C, Higgins J, Thompson R, Capps N, Smith GD, Ebrahim S (2001) Dietary fat intake and prevention of cardiovascular disease: a systematic review. British Medical Journal 322: 757-763
14. Horrobin D (2003) Are large clinical trials in rapidly lethal diseases usually unethical? Lancet 361: 695-697
15. Johnson J (2002) Influence of race or ethnicity on pharmacokinetics of drugs. Journal of Pharmaceutical Sciences 86: 1328-1333
16. Kohl H (1995) Physical activity and cardiovascular disease: evidence for a dose response. Medicine and Science in Sports and Exercise 33: S472-S483
17. Matthews H (1995) Racial, ethnic and gender differences in response to medicines. Drug Metabolism & Drug Interactions 12: 77-91
18. McAlister F, Taylor L, Teo K, Tsuyuki R, Ackman M, Yim R, Montague T (1999) The treatment and prevention of coronary heart disease in Canada: do older patients receive efficacious therapies? Journal of the American Geriatrics Society 47: 811-818
19. Miller B, Koloel L, Bernstein L, Young J, Swanson G, West D, et al. (1996) Racial/ethnic patterns of cancer in the United States 1988-92. Bethesda, MD: National Cancer Institute, Bethesda. Cited in: Giuliano A et al. (2000) Participation of minorities in cancer research: the influence of structural, cultural and linguistic factors. Annals of Epidemiology 10: S22-S34
20. Moye LA, Powell JH (2001) Evaluation of ethnic minorities and gender effects in clinical trials: opportunities lost and rediscovered. Journal of the National Medical Association 93: 29S-34S
21. Norheim OF (2002) The role of evidence in health policy making: a normative perspective. Health Care Analysis 10: 309-317
22. Owen W, Szczech L, Frankenfield D (2002) Healthcare system interventions for inequality in quality: corrective action through Evidence-based Medicine. Journal of the National Medical Association 94: 83S-91S
23. Raferty J (2001) NICE: faster access to modern treatments? Analysis of guidance on health technologies. British Medical Journal 323: 1300-1303
24. Rochon P, Clark J, Binns M, Patel V, Gurwitz J (1998) Reporting of gender-related information in clinical trials of drug therapy for myocardial infarction. Canadian Medical Association Journal 159: 321-327
25. Rogers WA (2002) Evidence-based Medicine in practice: Limiting or facilitating patient choice? Health Expectations 5: 95-103
26. Sateren W, Trimble E, Brawley O, Breen N, Ford L, McCabe M, Ungerleider R, Christian M (2002) How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials. Journal of Clinical Oncology 20: 2109-2117
27. Sherr L (2000) Women and clinical trials. In: Sherr L, St Lawrence J (Eds.) Women, health and the mind. John Wiley and Sons, Chichester, P 47-58
28. Starfield B (2001) New paradigms for quality in primary care. British Journal of General Practice 51: 303-309
29. Sterne J, Egge M, Smith GD (2001) Investigating and dealing with publication and other biases in meta-analysis. British Medical Journal 323: 101-105
30. Swanson GM, Bailar JC (2002) Selection and description of cancer clinical trials participants – science or happenstance? Cancer 95: 950-959
31. Taylor AL (2001) SERMs, ethnicity, and clinical trials: opportunities and challenges. Annals of the New York Academy of Sciences 949: 292-294
32. Watt G (2002) The inverse care law today. Lancet 360: 252-4
33. Weissman N (2000) Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. New England Journal of Medicine 324: 1094-1100
34. World Health Organisation (2002) The World Health Report 2002. World Health Organisation, Geneva

15 The Role of Formal Outcome Evaluations in Health Policy Making: A Normative Perspective

Ole F. Norheim

Introduction

Limit-setting decisions can be defined as the withholding of potentially beneficial health care through financial or organisational features of the health care system in question (Norheim 1999). The definition is broad enough to encompass the view that, the withholding of treatment perceived to be beneficial should be seen as a question of rationing. Limit-setting decisions involve outcome evaluations, typically considering whether a health care intervention has a proven effect, what the value is of that outcome, and whether the intervention is cost-effective. The aim of this paper is to examine the role of formal outcome evaluations in health policy decisions concerned with setting limits. The perspective is normative, and the normative position articulated is that of a contractualist liberal theory (Rawls 1993: Scanlon 1982; 1998; Daniels 2000; Daniels and Sabin 1997).

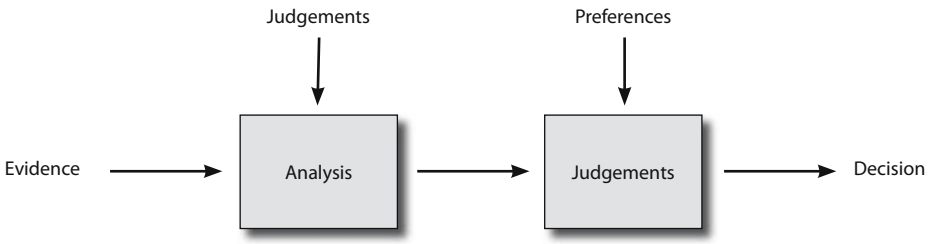
Types of outcome evaluations

David Eddy's article "The anatomy of a decision" identifies two kinds of judgements in outcome evaluations (Eddy 1990). Such judgements are often seen as subjective, but could properly be handled as intersubjective judgements.

The first step involves analysis of evidence for the outcome of a given intervention (the left part of figure 1). Scientific experts analysing outcomes may have different views as to the validity and relevance of the evidence in question. Disagreement in judgements can arise from what we might call different "scientific" preferences over what is considered good evidence. Articulating and deliberating the standards of validity can reduce such disagreement. This is the area where Evidence-based Medicine has provided a systematic and explicit method for assessing the quality of clinical studies, such as randomised trials.

The second step identified by Eddy (*Figure 15-1*) concerns the judgements explicitly involved in assessing the value of the outcomes. In the outcomes movement, and here I include both the practitioners of Evidence-based Medicine and health economists, the value assigned to outcomes are derived solely from individual preferences. Typically, one tries to elicit the preferences of a representative sample of individuals, such as patients, relatives, nurses, physicians or citizens. Formal methods for this type of outcome evaluation include the time trade-off method, standard gamble, visual analog scales, and the person trade-off method (Drummond et al. 1997). There is little consensus as to what method provides the most reliable and valid results, but all methods are used in health economic evaluations, such as in cost-effectiveness studies (Nord et al. 1993). In clinical medicine, the ideal of informed decision-making includes as a minimum the inclusion of patient preferences.

Figure 15-1
Two main steps of a decision process, according to David Eddy (1990)



For the purposes of this article – to assess the role of formal outcome evaluation in health policy – it is also possible to identify a third step not discussed by Eddy in his paper, and that is the judgement involved in assessing limit-setting decisions. In the overall assessment of an intervention, not only the quality of evidence and the value of the outcome are considered, but also whether it's funding will improve efficient service delivery (is the intervention sufficiently cost-effective?) and satisfy the requirements of fairness (does it's funding reduce inequalities in access to health care, does the target population suffer from a condition that is especially severe so that it deserves preferential treatment, etc.?) Whether it is desirable to fund an intervention thus involves normative and political judgements about efficiency and fairness.

Summing up, judgements on health outcomes are involved in at least three steps of the process of evaluating whether a given intervention should be funded and implemented:

1. Assessing the quality of evidence on outcomes involves scientific judgements,
2. Assessing the value of health outcomes involves citizen's preferences,
3. Assessing limit-setting decisions involves citizen's distributional preferences or values.

Each step deserves careful attention. Firstly, the status of judgements, preferences and values in a pluralistic democracy needs to be clarified. The idea of democratic deliberation imposes constraints on how values should be measured, handled and aggregated towards political decisions (Elster 1986; 1998; Gutman and Thompson 1996).

Public preferences and the assumption of value pluralism

According to most positions in normative political theory, the „fact" of value pluralism is the basic assumption for all accounts of how institutions should distribute goods and burdens in society (Rawls 1993). Citizens have different, and sometimes incompatible, conceptions about what constitutes a good life – and what kinds of goods and burdens affect their life prospects. Society's institutions need to acknowledge and respect this fact.

In health care, pluralism concerning citizen's preferences has implications for the role of outcome evaluations in steps 2 and 3 outlined above. For people trained in normative theory this is well known, but the problem of value pluralism is not fully acknowledged in the field of Evidence-based Medicine – to which we now turn.

Assessing the quality of evidence on outcomes

The method of EBM can be seen as: a) a systematic search for all available evidence capable of answering a specifically developed question (e.g. on outcomes of treatment), and b) a systematic assessment of the quality of this evidence. The quality of evidence is typically tested according to three criteria: validity, importance and applicability (Sackett et al. 1997). Validity is related to the assessment of representativity of the study in question, the proper design, elements of analysis, proper blinding and randomization in RCTs, etc. Importance is another way of testing the value of the outcome in question, and a patient centered perspective is often introduced – emphasising patient preferences. Applicability relates to the clinical setting, whether e.g. the intervention considered is applicable to the individual patient in front of the doctor.

The guiding idea behind EBM has always been that if an intervention is proven effective, then it should be provided if the patient wants it (the ideal of Archie Cochrane, see (Norheim 2002)). Further developments in the theory have extended its application to evidence-based health policy (Drummond et al. 1997). Health technology assessment agencies have incorporated the main ideas of systematic search and assessment of evidence into a framework concerning evidence on outcomes, cost-effectiveness and organisational issues that provides a basis for giving guidance about a given technology's impact on the health care system.

It is worth noting that there is a lesson to learn here in terms of epistemology or philosophy of science. In the empirical tradition that we could see EBM in, evidence is not considered the truth. All we can conclude from a systematic assessment of clinical trials etc. is that the evidence, so far, could be seen as a provisional truth, until new evidence falsifies the conclusions given. This provisional truth is what constitutes objective knowledge about the likelihood that a given intervention will have the expected outcome. There is in the scientific community a set of institutions and procedures that ensure that evidence is produced and can be relied upon – at least provisionally, until new evidence challenges our previous beliefs. These institutions and procedures (the educational system, university hospitals, scientific journals, conferences, etc.) are far from faultless, but they are the grounds that traditionally have guaranteed the authority of medical decisions and medical reasons. These institutions:

- Rely on procedures that secure the quality of the documentation (peer review, the hierarchy of medical journals, impact factors, standards of excellence in research);
- Encourage the publication of methods and results, securing that evidence can be scrutinised, challenged and revised;
- Foster a continuous debate about the criteria of validity;
- Ground the legitimacy of medical judgements.

Evidence-based Medicine can be seen as a methodology that makes the scientific judgements of the profession explicit, grounds clinical decisions in good medical reasons and clarifies what the standards of good evidence are. At the institutional level, health technology assessment could be viewed as a further development towards a systematic, transparent, and explicit method of performing outcome evaluations. The paradigmatic example of such an institutional development is the Centre for Clinical Excellence (NICE) in the UK (Smith 1999).

However, there is a remarkable lack of awareness about the value judgements embodied in the assessments these institutions perform. Outcome evaluation is seen as an exercise performed by experts with scientific methods. Although there are attempts at incorporating e.g. patient preferences into health technology assessments, the proper nature of these evaluating exercises are not fully recognised. With the exception of health economic studies, EBM and health technology assessment rarely incorporate formal methods for assessing the value of an outcome described as step two in Eddy's scheme (*Figure 15-1*).

Assessing the value of health outcomes

Health economics is the field where the valuation of health outcomes is most fully incorporated. One example is the QALY measure, which gives the formal concept of outcomes a content: the patient who can derive the most life years weighted for the quality of life in those years has the best outcome (Drummond et al. 1997).

Several methods have been developed to elicit the preferences that express the trade-offs between quality and quantity of life. Common to all these methods, is that a representative sample of people is asked to value a year gained in various health states, with reduced quality of life compared to a year gained in full or normal health. The final map of relative health-state evaluations is transformed to a scale ranging from 0 to 1.0. This can be done *directly* for each condition in question, but an *indirect* approach is more common. First one attempts to elicit the relative values for different health states described in general terms abstracted from particular patients and diagnoses. This allows for a certain level of generality, which permits others to classify the outcomes of particular treatments for particular indications in relation to this „generic" index.

The Rosser and Kind index is an example (Rosser and Kind 1978). One health state in this index – „Severe social disability or slight impairment at work. Able to do housework except for very heavy tasks. Moderate pain." – is assigned the relative value of 0.956. Successful treatment from this health state to „No disability" will give the difference 0.044 (1.0–0.956) as the value of the outcome per year. Now, if hip-arthrosis for the average patient fits the description, and hip replacement improves the health state to „no disability", we can indirectly assign the value 0.044 to this outcome per year. If the specificity of this generic index is good, i.e. if it measures what it is supposed to measure, and if the results are representative, health outcomes can be measured and compared in a relatively simple manner.

The first normative issue of concern is the wide variations in valuations, even when the same method is used for the same health state. Most studies, including Rosser & Kind's, show quite large inter-individual variations between the expressed valuations. It comes as no surprise that people simply have widely different strength of preferences. The technique used to solve this problem is to take the *mean or median* value of the expressed preferences for a certain outcome. Taking the mean or median value as a representation of public preferences is rarely accounted for, but sometimes given a kind of „democratic" justification. In an article by Erik Nord et al. (1993), the procedure of taking the median value is explained and justified for the person trade-off technique:

- „[A]nalysis was carried out using median values because, with the person trade-off technique, mean values may be influenced unduly by individuals responding with very high numbers that do not have real cardinal significance. The median may be interpreted simply as a measure of central tendency but also in terms of a majority view: if X is the median, then a majority are against assigning a lower value than X and also against assigning a higher value than X." (Nord et al. 1993)

A similar argument could probably have been given for taking the mean. Minority views are ruled out by a mechanism similar to majority vote. If value pluralism is to be taken seriously, this seems to be a move that deserves more attention. Is this mechanism, similar to majority vote, sufficient in a pluralistic democratic society when it comes to health, an issue of fundamental importance to citizens?

The second problem relates to the problem of error. As noted by Norman Daniels, what do we do when expressed preferences are clearly irrational, inconsistent, prejudiced, or strike us as clearly objectionable (Daniels 1994)? Nord, cited above, justifies taking the mean by arguing that individuals responding with very high numbers should be given less weight, because high numbers do not have „real cardinal significance". The response, as seen from value pluralism, would be: who is to judge? Moreover, what measure of external validity can we legitimately use if we give primary importance

to preference autonomy? It is, for instance, possible to imagine that the majority's preferences as to improvements from a psychiatric disease, as compared to a physical disease, might show a prejudice against psychiatric diseases in general. Should we be allowed to correct for prejudiced value-judgements if they are extreme? What about strikingly counterintuitive weights?

For a purely preference-based approach to health-state evaluation, questions like these create problems. The possible sources of „error" are numerous, and the methods have few internal mechanisms that can rule out even obvious errors.

These two problems, the democracy problem and the problem of error are parallel to issues much discussed in normative political and economic theory. For example, contractualist liberal theory, in the tradition of John Rawls, takes the fact of value pluralism as a starting point for how we are to evaluate the goods and burdens to be distributed by society (Rawls 1982). In this theory, equal respect for persons requires equal respect for peoples' values. The consequences have direct implications for the democracy problem. With respect to principles, Charles Beitz has compared the contractualist idea of hypothetical agreement, or the requirement of unanimous consent of all affected parties, with the utilitarian aggregative view:

- „[A]lthough aggregative conceptions hold that principles should be acceptable from the perspective of society, this does not imply that they should be acceptable from the perspective of everyone, taken seriatim. In contrast to contractualist conceptions, utilitarian views might therefore be seen as an application of the idea of rule by the majority." (Beitz 1989)

Contractualism requires that principles (and in our context here: evaluations to be used in public policy) be acceptable from the perspective of all, and rejects the rule of the majority in matters of fundamental importance. This contrast between unanimous agreement and majority vote can be, however, misleading. Contractualism does not necessarily demand *actual* unanimous agreement. Another characteristic of the social contract view is its emphasis on *reasonable* agreement (Scanlon 1998). Preferences can be justified, and justifications are always presented to others. An acceptable justification must be able to stand up to public scrutiny. The reasonable is, as Rawls notes in *Political Liberalism*, always public (Rawls 1993).

Amartya Sen has argued in a similar vein about the role of individual and public evaluations. He discusses how weights can be assigned to „functionings" in relation to his capabilities approach (an exercise that parallels the assignment of weight to health states). First, he emphasises that individual evaluation is a judgemental exercise, and „that it can be resolved only through reasoned evaluation" ((Sen 1997), my emphasis). When people hold a preference, they have good (personal) reasons for it: „But in arriving at an „agreed" range for social evaluation (…), there has to be some kind of reasoned „consensus" on weights (even if it is of an informal kind) (Sen 1997). Drawing on lessons from the literature on social choice theory as well as public choice theory, he elaborates about what kind of agreement or consensus he has in mind for public decision-making:

- „It is not so much a question of holding a referendum on the values to be used, but the need to make sure that the weights – or ranges of weights – used remain open to criticism and chastisement, and nevertheless enjoy reasonable public acceptance. Openness to critical scrutiny, combined with – explicit or tacit – public consent, is a central requirement of non-arbitrariness of valuation in a democratic society." (Sen 1997)

What we can learn from political and economic theory is that, the problem of democracy and the problem of error could be approached from a different perspective. First, it is not enough to refine the methods of health status evaluations with better questionnaires, better questions, or more representative samples of respondents. „Tapping" citizen's preferences – as they are out in the world – is insufficient both in terms of validity and acceptance. Second, Sen's argument suggests that eva-

luations should be recognised as a public exercise where evaluations can be examined, discussed, challenged and criticised. Explicit or tacit consent through reasoned evaluation is a prerequisite for legitimacy, and legitimacy for the decisions taken is surely needed, when health policy makers decide about whether to fund an intervention or not. Third, the weights assigned to health states through deliberation derive not only from individual preferences, but should properly be seen as collective valuations grounded in explicit or tacit public consent.

Assessing limit-setting decisions

Assessing limit-setting decisions involves citizen's distributional preferences or values. In the overall assessment of an intervention, two questions must be asked: first, is the intervention sufficiently cost-effective, and second, would its introduction satisfy the requirements of fairness? Whether it is desirable to fund an intervention thus involves normative and political judgements about efficiency and fairness.

Earlier developments in medical ethics have focused on finding general principles for health care rationing. One such account is Norman Daniels' „fair equality of opportunity" approach (Daniels 1985; 1994a; 1994b). The idea is that health is an all-purpose or substantial good, that is a necessary condition for achieving fair equality of opportunities to develop, form and revise a rational life plan. Disease, defined as departure from art-specific normal functioning, restricts the range of opportunities open to an individual. Health care that can maintain or improve health related opportunities should thus be distributed equally.

Another principled approach derives from the ethical framework of consequentialism in general, and utilitarianism in particular. Various variants are found in the literature on health care rationing, but in its general form, the principle of utility maximisation is recommended for resolving the conflicts between competing claims for resources (Broome 1993; Gerard and Mooney 1993; Mooney and Olsen 1991; Wagstaff 1991; Williams 1996; Nord 1994; Cubbon 1991; Nord et al. 1999; Singer et al. 1995). Utility can be defined in terms of e.g. disability adjusted life years or quality adjusted life years (Williams 1985). This approach can be operationalised into the principle of giving highest priority to interventions with the best cost-utility ratio.

Recent debates on health care rationing have increasingly acknowledged the difficulties, observed by Bill New and others, involved in the task of developing criteria for priority setting (New 1996). In empirical studies, strong disagreement about the proper principles for priority setting has been found (Nord et al. 1995; 1999). At a theoretical level, the ethical justification for these different principles is also quite different, and no consensus on acceptable first principles has developed.

Daniels (and others) concluded that it is hard to find principles that can guide us in solving these rationing dilemmas (Daniels and Sabin 1995; 1997; Daniels 1994a; Daniels et al. 1996; Singer et al. 2000; Holm 1998; McKneally et al. 1997). Instead, he argues that decision-makers must turn to fair processes:

- „In pluralist societies we are likely to find reasonable disagreement about principles that should govern priority setting. For example, some will want to give more priority to the worst off, some less; some will be willing to aggregate benefits in ways that others are not. In the absence of consensus on principles, a fair process allows us to agree on what is legitimate and fair." (Daniels 2000)

Although this debate in medical ethics indicates that there is no overarching theory of justice to balance competing claims, it should be noted that there are some reasons for rationing that every theory of resource allocation in health care would accept. For example, most theories of distributive justice in health care focus on a specific set of information. Based on recent literature on principles for priority setting (Ham 1995; Ham and Locock 1998; Ham and Coulter 2000), it is possible to

outline an overlapping consensus on a minimal-information set. It is important to emphasise this, because a focus on deliberative democratic procedures does not exclude the need to assess evidence on outcomes and cost effectiveness, it only clarifies the proper role of formal methods of outcome evaluation.

This minimal information set can form a basis for developing relevant reasons for rationing. Accordingly, the priority of a given condition and its intervention should be assessed in terms of:

- The burden of disease, if untreated;
- The benefit from the intervention;
- Treatment costs,
- The quality of evidence on 1-3.

As we can see, outcome evaluations have a central role in criteria for setting limits. The information can be formulated in terms of characteristics of the condition, and characteristics of the intervention in question. The set of information outlined here also excludes certain characteristics of patients as irrelevant. There is general consensus that individual characteristics such as race, ethnicity, place of origin, religion, sex, social status, sexual orientation and physical or mental disability should be considered irrelevant in devising rationing criteria (McKneally et al. 1997).[1]

By careful examination of the cases in question, the information required can be explored and discussed; evidence can be consulted, thus making it possible to test evidence against arguments and against conflicting systems of norms (Klein 1993). It should be noted, however, that reasonable disagreement is also to be expected over the application of such criteria. How are they to be interpreted? Disagreement over the proper judgements on issues of this kind is precisely the reason for turning to fair procedures.

Accountability for reasonableness is a framework developed by Daniels and Sabin, and could be used as a starting point for the discussions about procedures and institutions aiming at setting limits fairly. Space does not allow a full description of this framework, but its main elements are rather straightforward (Daniels and Sabin 1997; Singer et al. 2000; McKneally et al. 1997). It includes four conditions:

- Publicity condition: decisions regarding coverage for new technologies (and other limit setting decisions) and their rationales must be publicly accessible.
- Relevance condition: these rationales must rest on evidence, reasons, and principles that all fair minded parties (managers, clinicians, patients, and consumers in general) can agree are relevant to deciding how to meet the diverse needs of a covered population under necessary resource constraints.
- Appeals condition: there is a mechanism for challenge and dispute resolution regarding limit setting decisions, including the opportunity for revising decisions in light of further evidence or arguments.
- Enforcement condition: there is either voluntary or public regulation of the process to ensure that the first three conditions are met.

This framework has been applied to assess coverage exclusions by institutions such as managed care organisations in the USA (Daniels and Sabin 1997), public agencies in Canada (Singer et al. 2000), and most recently, decisions made by such institutions as NICE in the UK.[2] By taking the fact of value pluralism seriously, accountability for reasonableness is a framework that satisfies the

---

[1] Of course there are some contentious personal characteristics as well, such as age, responsibility for own disease, etc, but space does not allow discussion of these here (see (Norheim 1996)).

[2] International study on priority setting. Reports forthcoming (2002) in a collection of articles edited by Chris Ham and Glen Robert (Open University Press).

requirement of a pluralistic society's need for mechanisms and procedures that can handle citizen's conflicting distributional preferences or values.

The role of formal outcome evaluations in health policy making

We are now able to tie the various strands of the argument together. Modern liberal democracies must recognise the fact of value pluralism, as so should any public health care system. This implies that any disagreement over outcome evaluations must be brought out into the open, so that the arguments for judgements concerning medical interventions and their outcomes can be critically assessed. Outcome evaluations have been discussed in three steps: 1) analysis of evidence involves scientific judgements; 2) assigning values to outcomes involve citizen's judgements about health states; and 3) limit-setting decisions involve citizen's judgements about, among other things, efficiency and fairness.

All three steps need to handle disagreement in various ways. Evidence-based Medicine can be seen as a methodology that handles scientific disagreement: it makes the scientific norms of the profession explicit, it grounds clinical decisions in proper medical justifications, and the scientific community has institutionalised mechanisms that secure what we may call the trust in the soundness of medical judgements. The norms and judgements about validity and quality of evidence are not always right, but they are open to scrutiny and criticism.

Accountability for reasonableness, the framework developed for assessing limit-setting decisions, is also based on the idea that decisions and their underlying norms and justifications should be public and transparent. The requirements of openness can be seen as a question of legitimacy. Health policy should seek legitimacy by requiring accountability at all levels. Rationing decisions satisfy the requirements of accountability, if all relevant reasons for the decisions are given by those responsible for it to those who are affected by it (Gutman and Thompson 1996).

The theory and practice of assigning value to health outcomes is grounded in an entirely different framework. Most methods are designed for eliciting the individual preferences of a representative group. These preferences are taken as given, and normally not questioned or criticised. Inter-individual variation and disagreement is handled by taking the mean or median values from the whole group. A fundamental objection to this framework is that it is not open to „criticism and chastisement", and hence less likely to arrive at valid and acceptable results. The methods developed to „measure" individual preferences do not incorporate safeguards that secure trust in the results arrived at. Can we, for contested limit-setting decisions, expect the public or political decision-makers to trust non-transparent evaluations that form such an important part of the information necessary for making judgements?

Interestingly, several methods for assigning values to outcomes are now being developed in a way that could satisfy the description of „reasoned evaluation" outlined above. In some formal models of health state evaluations, deliberation has been introduced as an integral part of the process of eliciting preferences. For disability weights used by WHO and others, in the construction of disability adjusted life years, Murray and Lopez initially used the so-called deliberative person trade-off technique originally developed by Erik Nord (Nord 1995; 1999; Murray and Lopez 1996). Paul Dolan introduced the discussion through focus groups in health state evaluations, as have Peter Ubel and others (Dolan et al. 1999; Ubel 1999).

Space does not allow a full description of the requirements of deliberate models for health state evaluations, but four features would be central:

- Assessing health status evaluation should not be seen as empirical research, but as an integral part of public priority setting.[3] The results should be publicly available.

[3] This requires that the process should be institutionalised, but it is beyond the scope of this chapter to explore this point further.

- The ranking and values assigned to health states will be based on some mechanisms transforming individual preferences to public evaluations. The mechanisms used should be stated.
- The reasons given for the rankings should be provided as far as possible. Reasons should be expressed in terms that can be seen as relevant and acceptable to all.
- There exists no external standard of the „correct" evaluation. The results should therefore be open to further scrutiny and revision.

I believe no existing method for assigning values to outcomes satisfies all these features. Of course, further research is needed to develop more detailed requirements of this kind.

Concluding remarks

This chapter has explored the role of outcome evaluation in health policy decisions at three different levels: in the analysis of scientific evidence, in assigning values to outcomes, and in limit-setting decisions. The lessons learned from EBM and accountability for reasonableness was, first, that disagreement in evaluations should be acknowledged, and second, that it could be handled through open deliberation. I have argued that the methods and practice of assigning values to health outcomes have also underestimated the requirements of legitimacy for such evaluations: Explicit or tacit public consent is as central for valuation in a democratic society as is inter-subjective agreement on validity in Evidence-based Medicine. Health state evaluation should therefore be seen as an integral part of public priority setting. What remain, as open questions are, first, whether a deliberative model for health state evaluation could be seen to „produce" valid and legitimate weights, solely on the basis of how they are derived, and if yes, how to further describe the requirements of such a method.

References

1. Beitz C (1989) Political equality. Princeton University Press, Princeton
2. Broome J (1993) QALYs. Journal of Public Economics 50: 149-167
3. Cubbon J (1991) The principle of QALY maximisation as the basis for allocating ealth care. Journal of medical ethics 17: 181-5
4. Daniels N (1985) Just health care: studies in philosophy and health policy. Cambridge University Press, Cambridge
5. Daniels N (1994a) Four unsolved rationing problems. A challenge. Hastings Center Report 24: 27-9
6. Daniels N (1994b) Principles for national health care reform. Hastings Center Report 24: 8-9
7. Daniels N (2000) Accountability for reasonableness. BMJ 321: 1300-01
8. Daniels N, Light D, Caplan RL (1996) Benchmarks of fairness for health care reform. Oxford University Press, New York
9. Daniels N, Sabin J (1997) Limits to Health Care: Fair Procedures, Democratic Deliberation, and the Legitimacy Problem for Insurers. Philosophy & Public Affairs 26: 303-50
10. Daniels N, Sabin JE (1995) The yin and yang of health care system reform. Professional and political strategies for setting limits. Archives of Family Medicine 4: 67-71
11. Dolan P, Cookson R, Ferguson B (1999) Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study. BMJ 318: 916-8
12. Drummond M, Stoddart G, Torrance G (1997) Methods for the economic evaluation of health care programmes. Oxford University Press, Oxford
13. Drummond MF, Richardson WS, O'Brian BJ, Levine M, Heyland D (1997) User's guides to the Medical Literature. XIII How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? The Evidence-based Medicine Working Group. JAMA 277: 1552-7

14. Eddy DM (1990) Clinical decision making: from theory to practice. Anatomy of a decision. JAMA 263: 441-43
15. Elster J (1986) The market and the forum: three varieties of political theory. In: Elster J, Hylland A (Eds.) Foundations of social choice theory. Cambridge University Press, Cambridge, P 103-32
16. Elster J (Ed.) (1998) Deliberative Democracy. Cambridge University Press, Cambridge
17. Gerard K, Mooney G (1993) QALY league tables. Handle with care. Health Economics 2: 59-64
18. Gutman A, Thompson D (1996) Democracy and Disagreement. The Belknap Press of Harvard University Press, Cambridge Mass
19. Ham C (1995) Synthesis: what can we learn from international experience? British Medical Bulletin 51: 819-30
20. Ham C, Coulter A (2000) Where are we now? In: Coulter A, Ham C (Eds.) The global challenge of health care rationing. Open University Press; London
21. Ham C, Locock L (1998) International Approaches To Priority Setting In Health Care: An Annotated Listing Of Official And Semi-Official Publications. With A Selection Of Key Academic References. University of Birmingham, Birmingham
22. Holm S (1998) The second phase of priority setting. Goodbye to the simple solutions: the second phase of priority setting in health care. BMJ 317: 1000-2
23. Klein R (1993) Dimensions of rationing: who should do what? BMJ 307: 309-11
24. McKneally M, Dickens B, Meslin E, Singer P (1997) Bioethics for clinicians: 13. Resource allocation. CMAJ 1157: 163-7
25. Mooney G, Olsen JA (1991) QALYs. Where next? In: McGurie A et al. (Eds.) Providing Health Care. The economics of alternative systems of finance and delivery. Oxford University Press, Oxford
26. Murray C, Lopez A (Eds.) (1996) The Global Burden of Disease. Harvard School of Public Health, Cambridge Mass
27. New B (1996) The rationing agenda in the NHS. Rationing Agenda Group. BMJ 312: 1593-601 H, Singer P (1995) Maximizing health benefits vs egalitarianism: an Australian survey
28. Nord E (1994) The QALY – a measure of social value rather than individual utility? Health Economics 3: 89-93
29. Nord E (1995) The person-trade-off approach to valuing health care programs. Medical Decision Making 15: 201-8
30. Nord E (1999) Cost-Value Analysis in Health Care. Cambridge University Press, Cambridge
31. Nord E, Pinto JL, Richardson J, Menzel P, Ubel P (1999) Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Economics 8: 25-39
32. Nord E, Richardson J, Macarounas-Kirchmann K (1993) Social evaluation of health care versus personal evaluation of health states. Evidence on the validity of four health-state scaling instruments using Norwegian and Australian surveys. International Journal of Technology Assessment in Health Care 9: 463-478
33. Nord E, Richardson J, Street A, Kuhse H, Singer P (1995) Maximizing health benefits vs egalitarianism: an Australian survey of health issues. Social Science and Medicine 41: 1429-37
34. Norheim O (2002) The role of evidence in health policy making: a normative perspective. Health Care Analysis 10: 309-317
35. Norheim OF (1996) Limiting access to health care: a contractualist approach to fair rationing. University of Oslo, Oslo
36. Norheim OF (1999) Healthcare rationing-are additional criteria needed for assessing evidence based clinical practice guidelines? BMJ 319: 1426-1429
37. Rawls J (1982) Social unity and primary goods. In: Sen AK, Williams B (Eds.) Utilitarianism and beyond. Cambridge University Press, Cambridge, P 159-185
38. Rawls J (1993) Political Liberalism. Columbia University Press, New York
39. Rosser R, Kind P (1978) A scale of valuations of states of illness. Is there a social consensus? International Journal of Epidemiology 7: 347-358
40. Sackett DL, Richardson WS, Rosenberg W, Haynes RB (1997) Evidence-based Medicine. How to Practice and Teach EBM. Churchill Livingstone, New York
41. Scanlon TM (1982) Contractualism and utilitarianism. In: Sen AK, Williams B (Eds.) Utilitarianism and beyond. Cambridge University Press, Cambridge, P 103-128
42. Scanlon TM (1998) What We Owe to Each Other. The Belknap Press of Harvard University Press, Cambridge Mass
43. Sen A (1997) On economic inequality. Clarendon Press, Oxford
44. Singer P, Martin D, Giacomini M, Purdy L (2000) Priority setting for new technologies in medicine: qualitative case study. BMJ 321: 1316-18

45. Singer P, McKie J, Kuhse H, Richardson J (1995) Double jeopardy and the use of QALYs in health care allocation. Journal of Medical Ethics 21: 144-150
46. Smith R (1999) NICE: a panacea for the NHS? BMJ 318: 823-4
47. Ubel P (1999) Pricing life: Cost-effectiveness in health care rationing. MIT Press, Cambridge Mass
48. Wagstaff A (1991) QALYs and the equity-efficiency trade-off. Journal of Health Economics 10: 21-41
49. Williams A (1985) Economics of coronary artery by-pass grafting. BMJ 291: 326-329
50. Williams A (1996) QALYS and ethics: a health economist's perspective. Social Science and Medicine 43: 1795-804

16 The Usefulness of Formal Outcome Evaluations in Health Policy Making: Looking for the Baby in the Bathwater

Erik Nord

Introduction

There is increasing interest among health care fund holders in implementing „Evidence Based Health Policy". Behind the slogan lies a belief in economic rationality in health care: Resources should go where they do most good. In order to be able to make decisions to this effect, policy makers need evidence about the health benefits of different interventions, the values placed on these benefits by individuals and society as a whole, and the relationship between values and intervention costs.

Health economists have developed various ways to value different kinds of health benefits on a common numerical scale, in order that different health programs may be compared to each other in terms of their „value for money" Such common value indicators include the Quality Adjusted Life Year (QALY), the Disability Adjusted Life year (DALY), the Healthy Year Equivalent (HYE) and the Saved Young Life Equivalent (SAVE). Many non-economists (and even some economists) are sceptical of these indicators. One reason is that some of the indicators are constructed in ways that are inconsistent with societal concerns for fairness in resource allocation (Smith 1987; Hadorn 1992; Arnesen and Nord 1999). Perhaps more profoundly, it is felt that assessments of value and worthiness of funding in health care are essentially subjective and judgemental. The priority ratings of different programs can, according to these writers, not be calculated in a precise, objective way, but need to be decided in a fair, democratic process of deliberation between interested parties (Daniels 1998; Nygaard 2000; Biller-Andorno et al. 2001).

I agree with the view of priority setting as a complex deliberative process, however, there is not necessarily a conflict between quantification and deliberation. The interesting question is whether numerical indicators of value, and the relation of these to costs, in some contexts may serve as pieces of information that may facilitate deliberations and decision-making.

When trying to answer this question, we should distinguish between what looks plausible and is well intended on the one hand, and how the world really works on the other hand. It is one thing to observe a widespread interest in summary value indicators among both researchers and decision makers. For instance, the QALY is widely used in economic evaluations of medical technologies published in academic journals; analyses in terms of QALYs are encouraged in guidelines for pharmacoeconomic evaluations issued by drug administrations in countries like Australia, Canada, Norway and the UK; and the WHO has set up league tables of the cost-effectiveness of various large scale health programs in developing countries in terms of DALYs. However, it is a different thing to know whether the presence of information in terms of QALYs or DALYs in some situations has actually led to more satisfactory decision processes and/or better decision outcomes. I am not aware of studies that shed proper light on this issue.

We are therefore basically left to speculate and wait and see what experiences the future brings.

To stimulate discussion, I shall draw up a scenario in which at least some people would think that numerical value indicators might be useful as a means of summarizing complex information. I shall then bypass much criticized concepts like QALYs and DALYs, and instead present an indicator that is fairly easy to understand and at the same time encapsulates concerns for both efficiency and fairness in resource allocation. I shall leave it to the readers to judge whether they find the scenario likely and the indicator promising. Other scenarios could be constructed in a similar way.

A possible scenario

A regional health authority gets an increase in its budget for the next year of 5 million Norwegian kroner, which allows it to expand activity in some areas of care. Deliberations in various professional and political fora, lead to the identification of a number of diagnostic groups for whom the community would like to do more than what is presently being done. Professionals are asked to specify possible evidence based programs for each of these groups, within a cost frame of 1 million Norwegian kroner. The results are summarized in *Table 16-1*, which forms the basis for deliberations at the level of the board of directors of the health authority as to how the available five million kroner should be spent.

It is clear from Table 16-1 that the task of making a decision is complex, inasmuch as the programs are numerous and differ widely in terms of the nature of the problems they address, the severity of these problems, the expected health gains from the various interventions, and the number of people that may be treated given a program restriction of a million kroner. Each member of the board is bound to have a hard time in making up his/her mind regarding the relative value of the various programs, and the board as a whole is likely to have a hard time in reconciling different individual judgements.

To make things easier for the board, the regional health administration prepares a table that *summarizes* the information in Table 16-1, and assigns a *tentative value score* to each program by means of some standard formal procedure. One may imagine that if the board members (a) were used to such scores, (b) understood how they were determined and (c) knew from experience that they were usually roughly consistent with their own values, they would perceive them as helpful, for instance in ruling out some poor candidates from more lengthy deliberations.

Let us take a closer look at a formal valuation tool that the administration could use to arrive at its tentative values, and also see what the summary table could look like.

A formal valuation procedure

Let us assume that the administration has in its tool kit *Table 16-2*. It shows a scale of severity of illness. By severity is meant the burden of the illness to the individual concerned, in terms of loss of well-being, quality of life, or loss of utility as economists would say. To illustrate what degree of severity the various levels of the scale are meant to cover, an example of a corresponding health state is provided at each level. All the examples refer to mobility problems. But the scale itself refers to loss of well-being, whatever the nature of the health problem. So for instance, a problem with pain that reduces well-being as much as the mobility impairment described at level 4, would also be assigned to level 4 on this scale. Similarly, a hearing problem that reduces well-being as much as the mobility impairment described at level 3, would be assigned to level 3.

The scale is a modified version of a scale constructed by Sintonen (1981). The examples of mobility problems were chosen with a view to making each step up on the scale appear equally significant in terms of individual utility gains. With a few exceptions, subjects involved in a pilot study

Table 16-1
Candidate health programs that each cost one million Norwegian kroner

A Surgery for persons with hallux valgus (displacement of the great toe toward the other toes). NOK 4.000 per patient

Age/sex:	**60-70 years, most often females.**	**250**
Without intervention:	Little to moderate pain that becomes worse with movement. Difficulty with outdoor activities and climbing stairs.	
With intervention:	Freedom of movement and without symptoms for life.	

B Treatment program for alcoholics: residence at a treatment center for three months every other year for 8-10 years. NOK 250.000 per patient.

Age/sex:	**40-50 years, mostly males.**	**4**
Without intervention:	Chronic alcoholism, physically reduced, receives disability insurance, divorced. Death around 60 years.	
With intervention:	Abstinence for long periods of time, maintain job and family, moderate psychological problems 20-30% of the time, relapse after a couple of years following the first treatments, permanent abstinence following 4-5 stays at the treatment center. Death around 70 years.	

C Psychotherapy two times a week for three years for persons with chronic anxiety. NOK 50.000 per patient.

Age/sex:	**20-50 years, primarily females.**	**20**
Without intervention:	Must be accompanied by another person to all activities outside the home. The anxiety "can be lived with" in the home. Unable to work.	
With intervention:	Symptoms are considerably reduced. Can go out alone and tackle most social situations. In part able to work.	

Table 16-1 (Continued)
Candidate health programs that each cost one million Norwegian kroner

D Physical treatment and discussion therapy for one year for whiplash injury. NOK 20.000 per patient.

Age/sex:	**20-50 years, mostly females**	**50**
Without intervention:	Daily headaches. Concentration difficulties. Must significantly restrict reading. On part-time sick leave. Complaints considerably reduced when in the company of family and friends.	
With intervention:	Without headaches 50% of the time. Can read a little more and has more social vitality. Continues to be on part-time sick leave.	

E Surgery for persons with crooked partition between the nostrils. NOK 3.000 per patient.

Age/sex:	**20-30 years, both sexes**	**333**
Without intervention:	Chronically congested in the right nostril. Sinus infection one month of the year. Moderate headaches and concentration difficulties 10-20% of the time. On sick leave 6-8 weeks a year.	
With intervention:	The vast majority become completely healthy.	

F Bypass surgery for older persons with unstable angina pectoris. NOK 100.000 per patient.

Age/sex:	**65-70 years, mostly males**	**10**
Without intervention:	Chronic chest pains that restrict all outdoor activities and demanding chores in the home. Awake during the night due to chest pains. Side effects of medicine such as headaches, nausea and fatigue some of the time.	
With intervention:	50% are without symptoms while the other 50% have considerably less pain and are able to partake in most activities without difficulty.	

G Meniscus surgery. NOK 2.500 per patient.

Age/sex:	**From 15 years, primarily males.**	**400**
Without intervention:	Periodic pain and failure of the knee joint. Cannot take part in sports, jogging, trekking in the mountains, etc..	
With intervention:	Complete function of the knee is restored for most.	

Table 16-1 (Continued)
Candidate health programs that each cost one million Norwegian kroner

H Employment of nurses to follow-up patients with poorly regulated insulin-dependent diabetes once a month for three years. NOK 15.000 per patient.

Age/sex:	15-25 years, both sexes.	67
Without intervention:	Weekly episodes of bllod sugar being too high or too low. Acute hospitalization lasting many days a couple of times a year. Significantly reduced vision from 40 years of age, 25% chance of leg amputation from 50 years of age, death at 55-65 years.	
With intervention:	Fluctuations in blood sugar levels significantly reduced. Acute hospitalization reduced by 75%. Onset of complications and death delayed by 5-10 years.	

I Consultations of 30 minutes every 14 days "for the rest of life" for persons with chronic depression (in addition to low doses of antidepressives). NOK 100.000 per patient.

Age/sex:	All ages, both sexes.	10
Without intervention:	Unable to work. Periods with serious melancholy. Hospitalized several weeks every year. One in five with chronic depression commit suicide within 10 years.	
With intervention:	Continues to suffer from depression but seldom experiences melancholy that requires hospitalization. Still unable to work.	

J Total hip replacement for patients with hip joint arthrosis. NOK 80.000 per patient.

Age/sex:	65-75 years, mostly females.	12
Without intervention:	Pain when walking and at night. Cannot climb stairs. Uses crutches. Difficulty getting dressed. Restricts movement to reduce pain.	
With intervention:	Pain eliminated for 95% of cases. Patients are able to walk almost normally and get dressed without significant difficulty. The effect lasts for life in 75% of the cases. 25% of the patients experience symptoms after 5 years and the surgical procedure must be repeated.	

Table 16-2
A scale of severity

1. Healthy	
2. Slight problem	e.g. can move about anywhere, but has difficulties with walking more than 2 kms
3. Moderate problem	e.g. can move about without difficulty at home, but has difficulties with stairs and outdoors.
4. Considerable problem	e.g. moves about without difficulty at home. Needs assistance with stairs and outdoors.
5. Severe problem	e.g. can sit. Needs help to move about – both at home and outdoors.
6. Very severe problem	e.g. to some degree bedridden. Can sit in a chair part of the day if helped up by others.
7. Completely disabled	e.g. permanently bedridden.
8. Dying	

said that they perceived the examples as having fairly equal intervals in this sense (Nord 1993). (The states described at each level were also mapped into two different quality of life instruments of the kind used to assign values to health states on the zero-one utility scale used in QALY-calculations. One of the instruments gives utilities based on magnitude estimation and category rating, while the other gives utilities based on standard gamble and time trade-off (see Nord 1993 for details). Both mappings supported the impression that the 8-point severity scale has fairly equal intervals in terms of individual utility.)

Let us now assume that the administration also has in its tool kit ❯ *Table 16-3*. It uses the severity scale of Table 16-2 to indicate at a numerical level the weight that society places on: (a) the severity of the initial condition and (b) the size of the health gain (the reduction in severity) in its appreciation of health outcomes. The table does this by using the concept of „equivalent numbers of people treated". Consider for instance the lower right-hand cell. It describes a program that prevents 10 people from dying (bottom row), and instead leaves them in full health (right-hand column). Each of the other cells describes a program that society would consider roughly equally valuable and worthy of funding. For instance, the table suggests that society considers a program that could prevent 10 fatalities to be as valuable as a program that could take 50 people from level 5 to full health, or 1100 people from level 3 to level 2.

The validity of such numbers, of course, needs to be documented, and I return to this below. At this point, suffice it to say, numbers of this kind can most likely never be more than very rough estimates of societal preferences, since very few people can be expected to have precise and stable views on the relative values of different health programs in terms of equivalence numbers. But let us, for the sake of the argument, for a moment assume that the numbers in Table 16-3 are roughly valid in terms of their order of magnitude.

Now consider ❯ *Table 16-4*. This is the summary table provided by the administration to the members of the regional board together with the program descriptions. The information in columns A – C is taken directly from the descriptions. Columns D and E indicate at what level of severity the various patient groups typically are located before and after intervention. These placements are the result of „informed judgements", based on comparison of the information on functional status in the program descriptions, with the illustrative examples of health problems provided with the various severity levels of Table 16-2 (I return to problems associated with this below). Column F gives the relevant equivalence numbers from Table 16-3.

Table 16-3
Equivalence numbers for health improvements. Severity level after intervention

Severity level without intervention	7	6	5	4	3	2	1
1. Full health							
2. Slight problem							100000
3. Moderate						1100	1000
4. Considerable					140	130	125
5. Severe				80	60	55	50
6. Very severe			60	50	40	35	30
7. Completely disabled		40	25	20	19	18	17
8. Dead			10	10	10	10	10

Table 16-4
Summaries of programs and tentative value scores

A	B	C	D	E	F	G
Program	**Cost per person (NOK)**	**Number of patients 1 mill / B**	**Severity before (table 2)**	**Severity after (table 2)**	**Equivalence number (table 3)**	**Value for money (C / F)x10**
Diabetes	15.000	67	4 5	3	60-140	5-11
Nose	3.000	333	3/4	1	125-1000	3-25
Whip lash	20.000	50	4/5	3/4	80-140	4-6
Depression	100.000	10	6/7	5/6	40-60	2-3
Hallux	4.000	250	3	1	1000	2.5
Hip	80.000	12	5	1/2	50-55	2.5
Anxiety	50.000	20	4	2/3	130-140	1.5
Angina	100.000	10	4	1/2	125-130	0.8
Alcoholism	250.000	4	5/6	3/4	40-80	0.5-1
Meniscus	2.500	400	2	1	100.000	0.04

The point is now to compare columns C and F. If the number of patients in a given program is high compared to the equivalence number, there is an indication that the program yields good value for money, given what we roughly know about societal concerns of severity and treatment effect. The ratio for each program is shown in column G (multiplied by ten to avoid unnecessary decimals), and the programs are ranked according to this ratio. Columns B – E allow readers to see why the different programs come out the way they do. Tables 16-2 and 16-3 open the „black box of valuations“ even further.

Discussion

As noted above, one could imagine that a rank ordering of the ten programs, according to the ratios in column 11, might facilitate the deliberations of the regional board (without in any way „dictating" answers). Whether this is true or not, remains to be tested. The reader may start by asking him/herself whether he/she thinks she would have found Table 16-4 useful if he/she had been a board member facing the candidate programs of Table 16-1, and had been used to seeing information in terms of the severity scale of Table 16-2 and the equivalence numbers of Table 16-3.

The answer is not necessarily positive. Two methodological questions presumably spring to many readers' minds. One is whether equivalence numbers of the kind shown in Table 3 really can be established in a reliable and valid way. The other is whether the conditions described in the various programs can be reliably mapped onto the severity scale of Table 16-2. I address each of these questions in turn.

The empirical basis for equivalence numbers

A number of countries have developed guidelines for priority setting in health care. The guidelines are based partly on ethical reflection in academics and policy makers (e.g. Daniels 1985; Menzel 1990; The Norwegian Commission on Priorities in Health Care 1987), and partly on measurements of values, attitudes and preferences in samples of the general population (e.g. Charny et al. 1989; The Oregon Health Services Commission 1990; Campbell and Gillett 1993; Olsen 1994; Nord et al. 1995; Ubel et al. 1996; Pinto Prades 1997; Dolan and Cookson 1998).

A review of existing materials of the above kinds in industrialised countries like Australia, England, Holland, New Zealand, Norway, Spain, Sweden and the US (Nord 1996) suggests that ethicists' and policy makers' reflections and results from public preference measurements, converge on the following points:

A. To be willing to fund a medical intervention, society demands that it has some beneficial effect. (Severity alone is not enough.)
B. Given some beneficial effect, society's appreciation (valuation) of medical interventions increases strongly with increasing severity of the patient's condition.
C. Life saving or life extending procedures are particularly highly valued, and even significantly higher than interventions for patients with severe chronic conditions.
D. When the minimum requirement of effectiveness is satisfied (point A above), society worries less about differences in the size of the health benefits provided by treatment programs for different patient groups, the underlying attitude being that people are entitled to realising their potential for health, whether that be large or moderate, given the state of art in different areas of medicine.
E. As a special case of point D, society in most cases does not wish to discriminate between people with different potentials for health in decisions about life saving or life extension. For instance, society regards the prevention of premature death in people with chronic disease as equally worthy of funding as the prevention of premature death in otherwise healthy people. (Life extending interventions for people in vegetative states, or states of very low subjectively perceived quality of life, is a different matter.)

Points A-E constitute a set of guidelines for resource allocation in a national health service rooted in careful ethical reflection. Table 16-3 is essentially a representation of these guidelines in a more precise form. The procedure for arriving at the more precise form consists of questions to samples of the population, regarding how they would prioritise between programs that affect different numbers of people, with different degrees of severity of illness and different capacities to benefit. Such "per-

son trade-off" studies have been conducted in Australia, England, France, Holland, Norway, Spain, Sweden and the US (with less problematic versions of the person trade-off technique than the one used by the WHO). Table 16-3 is a synthesis of the results. For a review, see Nord (1999). Central tendencies are similar across countries, one aspect of the verbal guidelines particularly also shows up strongly in the population preference measurements, namely the strong concerns for severity expressed in points B and C.

I emphasize that the method of synthesis was informal, the ambition being only to indicate somewhat roughly what seems to be a widespread societal structure of concern. All numbers thus simply express *the order of magnitude* by which the general public, in the countries included in this research, emphasize initial severity and the size of health improvements when asked to prioritize between programs.

The reader should note the consistency between the numbers in the table and the verbal guidelines presented above. The concerns for severity and life saving expressed in guidelines B and C come through in the upper diagonal of the table: The increase in the equivalence number, from going one step up on the scale, is higher (and much more so) the higher the start point. The concerns for effectiveness and realisation of potential expressed in points A and D come through in each horizontal line: A movement from any given start point has a lower equivalence number the better the end point, but the decrease is small. For instance, a movement from level 7 to level 4 has only a slightly higher equivalence number than a movement all the way from level 7 to level 1. The concern for non-discrimination in matters of life saving or life extension is expressed in the bottom line of the table, according to which, the avoidance of death has the same equivalence number no matter what the resulting state is, although the bottom left hand cell is left void, to indicate that at some level of severity the value of life extension will be questioned.

Mapping of conditions onto the common severity scale

The levels of the severity scale in Table 16-2 are illustrated with examples of reduced mobility. This is because so much of the existing societal preference data for resource allocation in health care pertains to this particular dimension. To apply the equivalence numbers of Table 16-3 to other kinds of health problems, such as those occurring in the ten programs of Table 16-1, one needs to know where these other problems belong on the severity scale of Table 16-2. This requires an assessment of the burden of those other problems, compared to the effects on quality of life of the various mobility problems indicated in Table 16-2. Studies of quality of life in patients with disabilities and chronic illnesses, combined with the experience and expertise of health professionals, may facilitate such assessments. There is evidence that disagreement between different assessors rarely exceeds one level on the scale (Nord 1994). There is, nonetheless, a margin of error in the mapping procedure that needs to be accounted for in a table like Table 16-4, by giving estimates in terms of intervals rather than exact numbers. As may be seen from the table, this was done for most of the programs in question.

Concluding remarks

Priority setting in health care is a complex deliberative process. An interesting question is whether numerical indicators of value, and the relation of these to costs, in some contexts may serve as pieces of information that may *facilitate* deliberations and decision-making. Since there is little direct evidence on this issue, I have drawn up a scenario that hopefully may help readers see both possibilities and problems. Essentially, I am saying that the tools in Tables 16-2 and 16-3 could, in theory, be applied to the information in Table 16-1, to obtain the summary Table 16-4. Other scenarios could

be constructed in a similar way. For instance, national drug administrations are routinely deciding whether or not new pharmaceuticals should be included in public reimbursement schemes. It would theoretically be possible to summarize extensive information on different products, by the same tools, to obtain summary tables like Table 16-4. The important questions are: Does all this sound relevant to real world decision contexts? Does it look feasible? Does it look helpful?

My own response is that numerical estimates of the relative value a society places on different health outcomes can be theoretically meaningful and easy to understand. There is also some evidence that they can be roughly correct. On the other hand, it remains to be seen whether decision makers will subjectively perceive such numerical estimates as reliable. It is fair for researchers to try to increase the precisions of such estimates, to encourage their tentative use in deliberations and decisions about resource allocation, and to examine whether their use leads to greater satisfaction with decision processes and/or to better decisions.

References

1. Arnesen T, Nord E (1999) The value of DALY life. BMJ 319: 1423-5
2. Biller-Andorno N, Lie R, ter Meulen R (2001) Paper for workshop in Göttingen 2001
3. Campbell A, Gillett G (1993) Justice and the right to health care. In: National Advisory Committee on Core Health and Disability Support Services. Ethical Issues in Defining Core Services. Wellington
4. Charny MC, Lewis PA, Farrow SC (1989) Choosing who shall not be treated in the NHS. Social Science and Medicine 28: 1331-8
5. Culyer AJ, Lavers RJ, Williams A (1971) Social indicators: Health. Social Trends 2: 31-42
6. Daniels N (1985) Just Health Care. Harvard University Press, Cambridge Mass
7. Daniels N (1998) Distributive justice and the use of summary measures of population health status. In: Institute of Medicine. Summarizing population health: Directions for the development and application of population metrics. National Academy Press, Washington DC
8. Dolan P (1998) The measurement of individual utility and social welfare. Journal of Health Economics 17: 39-52
9. Dolan P, Cookson R (1998) Measuring preferences over the distribution of health benefits. Mimeo. University of York,York
10. Hadorn D (1991) Setting health care priorities in Oregon. JAMA 265: 2218-25
11. Menzel P (1990) Strong Medicine. Oxford University Press, New York
12. Murray C, Lopez A (1996) The Global Burden of Disease. Harvard University Press, Cambridge Mass
13. Nord E (1993) The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy 24: 227-238
14. Nord E (1994) Workshops on a value table for prioritising in health care. Working paper 1/1994. Norwegian Institute of Public Health, Oslo
15. Nord E (1995) The person trade-off approach to valuing health care programs. Medical Decision Making 15: 201-208
16. Nord E (1996) Health status index models for use in resource allocation decisions. A critical review in the light of observed preferences for social choice. International Journal of Technology Assessment in Health Care 12: 31-44
17. Nord E, Richardson J, Street A, Singer P, Kuhse H (1995) Maximising health benefits versus egalitarianism: An Australian survey of health issues. Social Science & Medicine 41: 1429-1437
18. Nord E, Pinto JL, Richardson J, Menzel P, Ubel P (1999) Incorporating societal concerns for fairness in numerical valuations of health programs. Health Economics 8: 25-39
19. Norwegian Commission for Prioritizing in Health Care (1987) Retningslinjer for prioritering innen helsevesenet. (Guidelines for prioritizing in health care.) Universitetsforlaget, Oslo
20. Nygaard E (2000) Is it feasible or desirable to measure burdens of disease as a single number? Reproductive Health Matters 8: 117-25
21. Olsen JA (1994) Persons vs years: Two ways of eliciting implicit weights. Health Economics 3: 39-46
22. Oregon Health Services Commission (1991) Prioritization of health services. A report to the Governor and Legislature. Oregon Health Services Commission, Oregon

23. Pinto Prades JL (1997) Is the person trade-off a valid method for allocating health care resources? Health Economics 6: 71-81
24. Sintonen H (1981) An approach to measuring and valuing health states. Social Science and Medicine 15: 55-65
25. Smith A (1987) Qualms about QALYs. The Lancet 329: 1134-36
26. Torrance GW (1986) Measurement of health state utilities for economic appraisal. Journal of Health Economics 5: 1-30
27. Ubel PA, Kamlet M, Scanlon D, Loewenstein G (1996) Individual utilities are inconsistent with rationing choices: A partial explanation of why Oregon's cost-effectiveness list failed. Medical Decision Making 16: 108-119
28. Williams A (1988) Ethics and efficiency in the provision of health care. In: Bell JM, Mendus S (Eds.) Philosophy and Medical Welfare. Cambridge University Press, Cambridge
29. Williams A (1997) Intergenerational equity: An exploration of the 'fair innings' argument. Health Economics 6: 117-132

17 Evidence-based Medicine and Managed Care

Marion Danis

Introduction

Evidence-based medical practice is essential for the best delivery of managed care and managed care organizations (MCOs), and their clinicians are as well positioned as anyone to practice state-of-the-art Evidence-based Medicine (EBM). Furthermore, MCOs are uniquely well situated to advance the evidence base on which medicine is practiced. In this chapter, I will review the reported use of evidence in managed care, and offer an analysis of the ethical challenges that face managed care organizations and clinicians as they perform evidence-based practice. I will argue that it is their responsibility to deliver the highest quality evidence-based medical care, which requires familiarity with the evidence, and appreciation of the strengths and limitations of evidence-based practice. Further, I will suggest that there is a tight link between decisions about the validity and relevance of medical evidence, and decisions about whether to offer or limit effective treatments, thus the ethics of these two decision-making processes must be considered in tandem. While health care organizations should make decisions on a population-based perspective, providers in these organizations should make individual decisions based on an evidence-based patient-focused perspective. A strategy for fostering these two perspectives simultaneously is offered, and an example from the literature that is consistent with this strategy is cited. Finally, I will also suggest that managed care organizations have some responsibility to promote progress in EBM, given their unique capacity to manage the care of large populations and collect outcome data.

Defining managed care

Managed care is medical care in which financing and delivery of services are linked, so that payers exercise some economic control over the way that services are delivered (Sultz and Young 1999). A variety of arrangements are often used to deliver managed care:

- Provider panels: a selected list of physicians and other providers that care for plan members.
- Limited choice: enrollees must use the selected list of providers or pay an additional out of pocket cost.
- Gatekeeping: enrollees must obtain referrals from a case manager for specialty or inpatient services.
- Risk sharing: providers carry some of the managed care plan's financial risk through capitation and withholding of reimbursement.
- Quality management: a managed care plan monitors provider practice patterns and medical outcomes to identify deviations from defined standards of quality and efficiency.

In thinking about managed care, a distinction should be made between the techniques of managed care and the organizations that perform them (Kongstvedt 2001). The techniques, such as those mentioned above, should be aimed at promotion of wellness, early detection of disease, patient education and self-care, financial incentives for providers, and utilization management.

The variety of managed care organizations are often difficult to understand clearly. Health main-

tenance organizations (HMOs) provide the simplest arrangement, involving a panel of salaried providers or an otherwise restricted network of providers, which leads to a close alignment of financing and delivery of care. Preferred Provider Organizations (PPOs), which are networks of providers who have agreed to accept managed care contractual agreements, use a strategy of imposing lower out-of-pocket payment rates for care provided to enrolled patients in the network, and higher out-of-pocket payment rates for care provided outside the network. There are also mixed arrangements, such as point-of-service (POS) programs, that function like PPOs except that to receive the highest level of benefit the enrollee must have a referral from a primary care provider who is part of the network. Managed care strategies are also used by employers, insurers, union management, trust funds, and the Medicare and Medicaid programs (Kongstvedt 2001). As Kongstvedt suggests, the arrangements are becoming increasingly difficult to characterize or statistically profile.

The role of EBM in managed care

Managed care attempts to offer the best quality medical care that is possible at the lowest cost (Zelman and Berenson 1998). In order to aim for such outcomes, organizations attempt to set standards of care that reduce practice variation by aiming for appropriate utilization of diagnostic and therapeutic procedures. Quality improvement programs in managed care organizations are generally run by quality assurance committees, and their activities may be numerous, including: focused studies of identified clinical problems areas, identification of "centers of excellence" for high risk and high technology services, rigorous credentialing of affiliated practitioners, medical audits, provider peer review, and member satisfaction surveys, and of greatest interest to this discussion, the development and implementation of evidence-based practice guidelines.

Opportunities for practicing EBM in managed care

Review of the literature points to both challenges and opportunities for practicing EBM in managed care. In pursuing the types of activities mentioned above, managed care organizations have the potential to provide evidence-based care in a manner that exceeds the capacity of the individual clinician. The high degree of practice organization and the large number of patients, make it feasible to offer evidence-based practices in a standardized fashion, through the systematic implementation of treatment protocols with the help of multidisciplinary teams when necessary.

The need to monitor the quality of care they provide has pressured MCOs, as well as other types of health care delivery organizations, to adopt evidence-based practice. In the United States, the National Committee on Quality Assurance (NCQA) has been measuring quality of care and releasing State of Managed Care Quality reports annually since 1997. Over the subsequent five-year period, the rates of childhood and adolescent immunization, cervical cancer screening, hypertension control, use of beta-blockers after heart attack, and control of diabetes, asthma, and hyperlipidemia have been monitored.

This process of measuring and reporting quality of care has been associated with a significant improvement in quality, as judged by the measures used. Over the course of three years of reporting by NCQA, managed care organizations have had documented improvement in the rates mentioned above. Health plans that publicly report their results have significantly higher rates of effectiveness in achieving these quality-of-care goals than health plans that do not report to the public. For example, publicly reported plans have rates of adolescent immunization of 46%, compared to rates of 25% for non-publicly reported plans. While NCQA first began reporting results for private, commercial health plans, in 2003 it began reporting on the results of publicly funded plans (Medicaid and Medicare) as well. The NCQA advocates the measurement and reporting of quality information on the

grounds that it saves lives and money. In keeping with this justification, the organization calculates estimates of the number of lives in the U.S. population saved through measures such as beta-blocker treatment, breast and cervical cancer screening, and diabetes and cholesterol control (NC QA 2002). Some large employers, who are among the largest purchasers of health plans on behalf of their employees, have reportedly begun to pay more to managed care plans that show a commitment to accountability through measurement and reporting of quality indicators than they pay to plans that are not committed (NCQA 2002).

The private sector is not alone in fostering the measurement of the quality of care. The U.S. federal government, through the Centers for Medicare and Medicaid Services (formerly the Health Care Financing Administration), and the Agency for Healthcare Research and Quality have begun an initiative to develop national standards for assessing the quality of hospital patient experiences (CMS 2003).

Multiple influences have served to pressure clinicians to adhere to evidence-based practices. There is evidence that compensation incentives for clinicians are associated with comprehensiveness of care management. Market pressures, as indicated by the percentage of a clinician's patients who are in health maintenance and preferred provider organizations, are associated with use of managed care practices. Thus, compensation incentives and managed care market pressures have increased the use of evidence based care management practices (Shortell et al. 2001).

The literature indicates that the use of Evidence-based Medicine does lead to cost-effective care in the managed care setting (Anonymous 2000). One particularly pertinent question regarding EBM is, how managed care and other traditional fee-for-service care compare, when quality measures are applied to both. In a review by Robinson, managed care offers care that is less expensive but equal in quality. Unfortunately, managed care yields less patient satisfaction (Robinson 2000).

Along with the possibility of practicing EBM, large managed care organizations offer the opportunity to generate new medical evidence by providing large clinical and administrative databases (Hornberger and Wrone 1997). Substantial pressures on the managed care industry have actually forced the pace of the generation of new medical evidence for several reasons. First, it has done so, because sufficient evidence does not always exist to guide practitioners in a cost-conscious, high quality practice environment. Secondly, there is often a need to find a credible means of arbitrating between parties who must pay for care, and those who are marketing new technology or pharmaceutical products (Sheingold 1998). These pressures have actually fostered the development of evidence-based methods for guiding assessments whether or not to adopt new technology (Ramsey et al. 1998).

As a result of pressures to generate evidence, technology assessment has been a growing industry (Sheingold 1998). The Agency for Healthcare Research and Quality has contracts with several evidence-based practice centers, and also has an internal assessment capacity to conduct technology assessments. AHRQ purchases assessments for the Medicare program so that coverage decisions for Medicare are based on these assessments.

Several examples of the advancement of Evidence-based Medicine in managed care are noteworthy. One of the most impressive efforts has been the development of disease management strategies for chronic illnesses using evidence-based guidelines. The chronic care model was reported after demonstrating that the majority of patients with hypertension, diabetes, tobacco addiction, hyperlipidemia, congestive health failure, chronic atrial fibrillation, asthma, and depression were inadequately treated (Bodenheimer et al. 2002). The chronic disease model involves 6 components: self management support, clinical information systems, delivery system redesign, decision support, health care organization, and community resources. The intent of this disease management model is to have well –informed, activated patients interact with prepared proactive practice teams, in order to have higher quality chronic disease management.

In an example illustrating the model, the authors describe a diabetic patient who attends a diabetes self management class, and subsequently comes to clinic for a scheduled diabetes-manage-

ment visit with a record of the results of her home glucose monitoring. A medical assistant takes this record and scans it into the electronic medical record, reviews the results with the patient, and after being prompted by a reminder pop-up message in the computer, refers the patient for an eye exam, a test for urine microalbumin and prints out a graph of the results of the patient's glycosylated hemoglobin level (Hgb A1C) results and low-density lipoprotein (LDL) cholesterol from the past two years, for the patient and physician to review. While the results of glucose monitoring are normal, the LDL-cholesterol level is elevated, so the physician refers the patients to the pharmacist who adjusts the patient's lipid lowering medicine according to a practice guideline-based protocol, and to the nutritionist who instructs on ways to adopt a low fat diet. These strategies take advantage of the organizational structure of managed care, in a manner that would be difficult to reproduce outside of such highly coordinated care systems (Bodenheimer et al. 2002). A random survey of 47 healthcare systems and managed care organizations in 2000, found that 89% had or were working to develop disease management programs (Whellan et al. 2002).

Like programs to promote better management of chronic disease, evidence based strategies have been developed for health promotion. For example, the U.S. Public Health Service has recommended a clinical guideline for treating tobacco dependency as a means of reducing smoking (Cawood and Morrow 2001). The Public Health Service reviewed 6000 studies to formulate its guidelines. The recommended approach involves four directives: 1) Ask all patients about their tobacco use and document it regularly; 2) Advise all patients who smoke to quit; 3) Assess patients' willingness to quit and initiate cessation treatments for those who are ready to do so; and 4) Arrange follow-up to prevent relapse.

As of 1998, 75% of managed care organizations in the U.S. either fully or partially covered one or more smoking cessation interventions (Cawood 2001). Full coverage usually involved self-help materials and smoking cessation classes. Thirty eight percent of health plans offered a combination of nicotine replacement therapies along with smoking cessation classes. The American Association of Health Plans (AAHP) sponsors an award program that showcases health plans that have developed innovative smoking cessation programs that are based on the Public Health Service guidelines. The award-winning program reduced smoking rates among its enrollees from 21% in 1994 to 16.7% in 1999.

As mentioned earlier, the need to decide whether or not to adopt new interventions has prompted the adoption of several strategies. A paradigmatic example is the adoption of evidence based formulary decisions by the Academy of Managed Care (Fullerton et al. 2001; Gricar et al. 2002). In 2000, the Academy of Managed Care Pharmacy (AMCP) published the Format for Formulary Submissions, in order to address the growing need of managed care pharmacies to ensure that utilization of medicines, biopharmaceutical and vaccines was appropriate, and to make it possible to determine whether new products would bring added clinical and economic value to covered populations. The format is designed to provide a high standard of objectivity for information for pharmacy, and therapeutic committees, when they are making formulary decisions. The format is that of an unsolicited letter from a health system to a manufacturer, for all possible clinical and economic information necessary to assess the overall clinical utility of a product. Manufacturers are asked to submit all possible published and unpublished studies, and information regarding both FDA-approved indications and off-label uses of the product (Gricar et al. 2002). The valuable features of this format are that it establishes a uniform approach for presentation of evidence, which involves demanding standards and includes data that would otherwise be unavailable in the published literature.

Managed care plans have also looked to evidence-based methods to make decisions about whether or not to adopt new technologies (Sheingold 1998). The Task Force on Technology Assessment of Medical Devices has published guidelines to help health-plans, providers and manufacturers resolve conflicts about decisions to pay for use of new devices (Ramsey et al. 1998).

Challenges

While there are numerous opportunities for practicing and advancing EBM in managed care, a number of challenges warrant caution. Since many of these cautionary notes are not unique to the managed care setting, only those concerns that have been particularly discussed in the managed care literature will be addressed.

As managed care emphasizes strategies for population-based approaches for care, a particular concern about EBM in the managed care setting is the limited data that exist for some specific populations. This is particularly true for the young and the old.

The leading causes of morbidity and mortality in children are infrequent in number, which makes it difficult to conduct RCTs of interventions to address these causes (Wehr 2001). For example, neonatal sepsis from group B streptococcal infection, the leading cause of death in newborns, occurs at a rate of 1.8 per 1000 live births. Trials involving thousands of newborns would be required, to test strategies to prevent group B streptococcal sepsis.

Even for conditions that occur in both children and adults, data are unavailable for the paediatric population. As of 1999, 80% of drugs approved for adult indications were not approved for paediatric use, or had not been labelled for various ages.

For geriatric patients, the applicability of clinical trials is also not evident. Many studies do not include the elderly. In a recent review of guidelines for management of hypertension, it was not known whether the recommended target blood pressure for individuals under age 65 should be applied to the elderly (August 2003). For the elderly and the chronically ill, the effect of co-morbid conditions on the applicability of treatment guidelines is not known.

A second challenge for managed care, along with other organizations, is the extent to which inferences can be made from evidence regarding one product to another. For example, many studies of older medications may show important medically beneficial outcomes. When new analogues, designed to achieve the same outcome but with better safety profiles, become available, to what extent can the benefits demonstrated in large trials with older drugs be extrapolated to the newer agent? At the same time, many older medications became the standard of practice without being well studied. How should these less expensive older medications be judged in comparison to newer expensive medications that have been well studied?

A third challenge for EBM is that the benefits of many interventions are less than they could be, because evidence-based guidelines are underused (Holloway et al. 2000). One major reason is that physicians often lack sufficient knowledge and skills to practice EBM. Community based physicians have greater knowledge deficits than physicians who are in teaching settings, but both have substantial deficits (Beasley and Wooley 2002).

Yet another challenge is that the public does not necessarily understand the rationale for EBM. The public often has great faith in medical interventions and is enthusiastic about new technology. Those who wish to market new technologies often foster this faith. Rather than seeing EBM as a potentially useful strategy for assessing the value of treatments and a fair basis for arbitrating coverage decisions, the public may view complicated analyses of the cost effectiveness of interventions with scepticism and frustration (Mechanic 2002).

Another challenge for MCOs is the discrepancy between adopting evidence for the practice of clinical care and adopting evidence related to management techniques. Health care managers and policy makers encourage clinicians to use evidence-based practice, but don't necessarily apply the same expectations to themselves. However, there is evidence that the underuse of effective interventions, and overuse of ineffective interventions, is widespread in health care management (Walshe and Rundall 2001).

Normative suggestions

Managed care organizations are in a rather unique position regarding EBM. As the literature mentioned earlier indicates, these organizations are in a powerful position to foster the development of evidence, and as such should take on this responsibility. In meeting this challenge, several concerns should be bourn in mind. First is the need to generate data where it is unavailable, but that data is needed for populations that cannot be well studied in RCTs. Setting the agenda of what questions will be addressed should occur in a manner, that avoids undue influence by interested parties who are the manufacturers of innovations and attend to the needs of populations such as paediatric, geriatric and populations with multiple medical conditions. Second is the need to carefully consider the development of policies for caring for patients when data does not exist.

MCOs can play a role in development, evaluation and incorporation of clinical trials into their clinical practice strategies, as some have suggested (Vakil 2001), or can simply maintain databases to record the outcomes of their practices. There are certainly examples of managed care practices that have organized into teams to conduct randomized trials. The Department of Medicine at the Indiana University School of Medicine and the Regenstrief Institute for Health Care, for example, has organized its adult primary care practice into four parallel practices which permits utilization and comparison of alternative health interventions (Tierney et al. 2001), and has also conducted trials involving practices in the community (Overage et al. 2001).

Many reports published in the literature involve efficacy studies that have tested the use of interventions under ideal circumstances. To what extent the results of such studies are applicable in non-idealized practice settings, cannot be fully determined without further effectiveness studies. Managed care organizations would do well to publish evidence from practices about the effectiveness of these interventions.

In considering how managed care organizations should go about building the evidence base, an important concern that should be taken into account is the need to incorporate patient preferences into assessments of health care interventions. Given the large patient populations in their practices, MCOs have the unique potential to readily collect data on patient utilities, for improvements in quality of life associated with interventional therapy. In doing so, it will be more readily possible to incorporate patient preferences into practice guidelines (Brown et al. 2000).

In the absence of requisite evidence, purchasers of managed care and policy designers should understand the limits of evidence-based coverage standards (Wehr 2001).

When evidence is lacking, managed care contracts and statutory rules should require that procedures for making decisions on medical necessity (or coverage decisions however they are defined) include consideration of the quality of the evidence that is available. Regarding paediatric care, coverage decisions should take into account the paediatric epidemiology and issues of the need to take patient growth and development into account. In the absence of evidence, clinical guidelines and expert opinion should be used instead to resolve coverage decisions. Professionals who are assigned to decide coverage decisions should be required to have clinical expertise in managing the condition for which an intervention is recommended (Wehr 2001).

MCOs should also make significant educational efforts to promote understanding of EBM. This includes educating clinical trainees and patients. There are published models of managed care organizations, and medical schools are collaborating to teach Evidence-based Medicine (Matson et al. 2000). MCOs should also educate consumers, so they have the capacity to understand evidence-based decision-making (Mechanic 2002). This would allow enrollees to be more knowledgeable, active, and responsible participants in clinical decisions with their physicians.

Thus far, the discussion has focused on ethical issues related to MCOs as large organizations. As organizations with responsibility for the health care of large populations, it is reasonable to argue that MCOs should take a population-based perspective in evaluating and utilizing medical evidence. MCOs have a fiduciary responsibility to the entire population of patients for whom they provide

care. Ibrahim et al have noted that MCOs should take a public health perspective, and should focus on the relevance of five principles pertinent to this emphasis: a community perspective, a clinical epidemiology perspective evidence-based practice, an emphasis on outcomes, and an emphasis on prevention (Ibrahim et al. 2001).

Arguing that the organization as a whole should take a population perspective is insufficient. The ethical analysis remains incomplete without examining the interactions of the organization and individual provider. While much has been written about the way the individual clinician should utilize evidence-based guidelines at the bedside, the interesting question to address here is the intertwined ethics of the organization and the individual provider under managed care. I would suggest that the interaction is worth noting because it may serve as a paradigm of organization-provider interactions in general.

Before discussing a strategy for negotiating differences in the population and individual perspectives, it seems appropriate to make clear that EBM per se is not the major ethical difficulty faced in resolving treatment decisions in managed care. While EBM concerns itself with *identifying effective care*, an equally or more pressing problem, is that of rationing – the question of *denying effective care* due to limited resources. It is the combination of the two that are at the heart of the ethical dilemmas faced by organizations that manage the health care of large populations.

Hence, before outlining a defensible strategy for making practice decisions in managed care, I will refer to some assumptions that ought to underlie rationing. I adopt here the assumptions attributed to Alan Williams and published by Alan Maynard (1999).

1. The role of health care is to improve health and reduce inequalities in health. Thus, it is taken as axiomatic that European health care policy makers are seeking the twin goals of efficiency (the production of health status improvements at least cost) (benefit) and equity ("fairness").
2. It is not appropriate to determine access to care by willingness and ability to pay. This is replaced by an allocation or rationing mechanism based on „need".
3. Need can either be a demand or a supply concept: patients may demand care for conditions for which there is no cure or for which benefits are illusory; doctors may supply care to meet a need when there is a cost-effective intervention. Usually the latter technocratic definition of need is used. (i.e. judgment about ability to benefit).
4. How, if at all, should the benefit (need) principle, which achieves efficiency, be weighted by equity goals?
5. The role of judging the needs of competing patients should be given a disinterested group of independent and neutral experts. Doctors are the social proxy for this group.
6. To ensure that doctors are independent and avoid the temptation to undermine their neutrality by supplier-induced demand, it is necessary to create mechanisms to manage and monitor their performance.

It is useful to combine these assumptions with a practical strategy for the clinician, who must act to balance the dual roles of steward for an organization's resources and advocate for the patient. While it is not possible to adequately address the ethical quandary posed by these conflicting perspectives and give due attention to the large literature on this topic, I will make use of several published analyses for tackling the gap between population and patient perspectives in clinical decision-making. Perhaps one of the best ethical strategies for striking the proper balance is, the ethic of evidence-based patient choice as described by Tony Hope (1997) and Michael Parker (2001). They suggest that the physician can respect the interests of the individual patient by engaging in deliberation with her. In doing so, the clinician hears of the patients needs, articulates the population perspective to help her appreciate it, and then negotiates a strategy for meeting her needs.

Based on the assumptions and strategies reviewed here, it is possible to derive a set of recom-

mendations for using evidence to make treatment decisions in managed care, at either the population or individual level:

1. Organizations should take a population based perspective in assessing evidence about the effectiveness of clinical interventions and in making coverage decisions.
2. They should use available standardized strategies for evaluating evidence, such as the Format for Formulary Submissions (NCQA), and apply such strategies to their setting.
3. Where evidence is unavailable or not sufficient to make coverage decisions, organizations should rely on specialized experts who can appreciate the consequences of coverage decisions rather than simply forgoing coverage.
4. Evidence about organizational standards should be taken as seriously as evidence about clinical practice. For example, managed care purchasers should subcontract with high volume providers, given the evidence that outcomes are functions of volume (Wehr 2001).
5. To the greatest extent possible, MCOs should develop the capacity to maintain databases regarding patient care in order to expand the evidence base.
6. Individual clinicians practicing in managed care organizations should adopt the strategy of evidence-based patient choice:
    a. They should offer treatments according to the guidelines developed by their organizations. They should make efforts to explain the evidentiary basis for these recommendations to patients
    b. When a patient requests a treatment strategy that diverges from the recommendations, the clinician should have the leeway to negotiate an alternative plan of care. An evidence-based assessment of individual risk should be combined with patient preferences to guide the decision to provide treatment.

This set of recommendations builds on, but differs from, the strategies proposed by Hope and Parker, by explicitly suggesting that the guidelines should include permissible strategies for the individual provider to negotiate satisfactory patient-centered care, rather than assuming this process will be outside of evidence-based guidelines. This suggestion does not imply that guidelines can ever be designed to fully address the need for flexibility in the doctor patient relationship. However, it is based on the concern that, as practice becomes increasingly evidence based and emphasizes measurement of the quality of care, we must explicitly include options for flexibility where the evidence warrants it. Otherwise patient-centered care will be at odds with evidence-based care. An example of guided strategies for addressing patient preferences is, the use of tiered pharmacy benefits in which prescription of brand name medications is covered, but involves a higher co-payment.

It is useful to illustrate the combined processes of evaluating evidence and deciding about the allocation of an intervention, to see how it plays out at the population and individual level. As an example, consider the published recommendation of the National Institutes of Health Consensus Development Conference for the use of mammography in women ages 40–49 (NIH Panel 1997). While the evidence about mammography may have changed since the time of this publication, the strategy for making evidence-based decisions recommended in this statement, still demonstrates a strategy for straddling the population and individual perspectives. The committee initially reached a consensus, and subsequently two members of the panel could not agree to the consensus and wrote a minority report. Taking into account the evidence available at the time about the risks and benefits of mammography, the panel thought that the data did not support a recommendation for universal mammography screening for all women in their forties. A minority of the panel believed that the risks had been overemphasized and supported a recommendation for screening all women in this age group. Hence the entire panel agreed that women in this age group should be provided with information on these issues upon which to base their decisions. Panellists agreed that for women in there forties, who choose to have mammography, the costs should be covered by health maintenance organizations.

Consider how this interpretation of evidence might be translated by an MCO into an evidence-based policy about the allocation of mammography. As an organization, it might not recommend or routinely offer mammography to its female enrollees in this age group, but if a patient wished to have it, in discussions with her physician, it would be made available to her. Alternatively, the organization might have guidelines for handling the decisions at the individual level, by having the clinician assess a patient's risk level using a risk stratification model for breast cancer (Armstrong et al. 2000) and then cover mammography for women at high risk, but ask women at low risk to pay a co-payment for mammography if they wish to have it done.

Conclusion

The ethical issues that arise in the practice of EBM in managed care are not unique. Like other large health care organizations, MCOs must grapple with the combined ethical difficulties involved in identifying and rationing effective care. MCOs should develop guidelines that simultaneously steer effective care for their population as a whole and include permissible strategies for the individual provider to negotiate satisfactory patient-centered care.

On the other hand, MCOs are in a uniquely strong position to generate evidence. Hence, MCOs have a responsibility to realize that the nature of the evidence they generate can have a profound impact on medical practice. They should seek to generate evidence that promotes population health, attempts to fill the gaps in needed data, particularly for understudied segments of the population, incorporates patient utilities, and avoids undue influence from those who are marketing innovations, or from those who wish to cut costs without concern for good patient outcomes.

Acknowledgement

I wish to thank Samia Hurst, MD and Reidar Lie, MD, PhD for their inspiration and suggestions. The views expressed here are those of the author and do not reflect policies of the National Institutes of Health or the Department of Health and Human Services.

References

1. Anonymous (2000) Evidence-based Medicine helps providers offer cost-effective care. Capitation Management Report 7: 88-90
2. Armstrong K, Eisen A et al. (2000) Assessing the risk of breast cancer. New England Journal of Medicine 342: 564-570
3. August P (2003) Initial treatment of hypertension. New England Journal of Medicine 348: 610-617
4. Beasley B, Wooley D (2002) Evidence-based Medicine knowledge, attitudes, and skills of community faculty. Journal of General Internal Medicine 17: 632-639
5. Bodenheimer T, Wagner E et al. (2002) Improving primary care for patients with chronic illness. JAMA 288: 1775-1779
6. Brown G, Brown M et al. (2000) Health care in the 21st century: Evidence-based Medicine, patient preference-based quality, and cost effectiveness. Quality Managed Health Care 9: 23-31
7. Cawood J, Morrow T (2001) Putting to use a clinical guideline for treating tobacco dependency. Managed Care Interface 14: 79-83
8. CMS (2003) Centers for Medicare and Medicaid Services
9. Fullerton D, Atherly D et al. (2001) Showing outcomes and proving value brings success. Managed Care Interface 14: 63-65
10. Gricar J, Langley P et al. (2002) Academy of Managed Care Pharmacy's Format for Formulary Submission, AMCP
11. Holloway R, Benesch C et al. (2000) Stroke prevention: narrowing the evidence practice gap. Neurology 54: 1899-1906

12. Hope T (1997) Evidence-Based Patient Choice. The King's Fund, London
13. Hornberger J, Wrone E (1997) When to base clinical policies on observational versus randomized trial data. Annals of Internal Medicine 127: 697-703
14. Ibrahim M, Savitz L et al. (2001) Population-based health principles in medical and public health practice. Journal of Public Health Managed Practice 7: 75-81
15. Kongstvedt P (2001) Essentials of Managed Health Care. Aspen Publishers, Gaithersburg
16. Matson C, Morrison R et al. (2000) A medical school-managed care partnership to teach Evidence-based Medicine. Academic Medicine 75: 526-527
17. Maynard A (1999) Rationing health care: an exploration. Health Policy 49: 5-11
18. Mechanic D (2002) Socio-cultural implications of changing organizational technologies in the provision of care. Social Science and Medicine 54: 459-467
19. NCQA (2002) NCQA: Measuring the quality of America's health. NCQA
20. Overage J, Perkins S et al. (2001) Controlled trial of direct physician order entry: effects on physicians' time utilization in ambulatory primary care internal medicine practices. Journal of the Medical Informatics Association 8: 361-371
21. National Institutes of Health Consensus Development Panel (1997) National Institutes of Health Consensus Development Conference statement: Breast cancer screening for women ages 40-49, January 21-23, 1997. Journal of the National Cancer Institute 49: 1015-1025
22. Parker M (2001) The ethics of evidence-based patient choice. Health Expectations 4: 87-91
23. Ramsey S, Luce B et al. (1998) The limited state of technology assessment for medical devices: Facing the issues. American Journal of Managed Care 4: SP188-SP199
24. Robinson R (2000) Managed care in the United States: a dilemma for evidence-based policy? Health Economics 9: 1-7
25. Sheingold S (1998) Technology assessment, coverage decisions, and conflict: The role of guidelines. The American Journal of Managed Care 4: SP117-SP125
26. Shortell S, Zazzali J et al. (2001) Implementing Evidence-based Medicine: the role of market pressures, compensation incentives. and culture in physician organizations. Medical Care 39: I62-78
27. Shye D, Porath A et al. (2000) Adapting a national guideline for local use: a comparative study in a US and an Israeli health maintenance organization. Journal of Health Services Research and Policy 5: 148-155
28. Sultz H, Young K (1999) Health Care USA: Understanding its Organization and Delivery. Aspen Publishers, Gaithersburg
29. Tierney W, Dexter P et al. (2001) The effects of discussions about advance directives on patients' satisfaction with primary care. Journal of General Internal Medicine 16: 32-40
30. Vakil N (2001) Evidence-based health care: making health policy and management decisions. Managed Care 10: 22-24
31. Walshe K, Rundall T (2001) Evidence-based management: from theory to practice in health care. Milbank Quarterly 79: 429-457
32. Wehr E (2001) Evidence-based health coverage for children: proceed with caution. Ambulatory Pediatrics 1: 23-27
33. Whellan DJ, Cohen EJ, Matchar DB, Califf RM (2002) Disease management in healthcare organizations: results of n-depth interviews with disease management decision makers. American Journal of Managed Care 8: 633-641
34. Zelman W, Berenson R (1998) The Managed Care Blues & How to Cure Them. Georgetown University Press, Washington DC

18 Recommendations from the Evibase Project

Rob Houtepen · Nikola Biller-Andorno · Donna Dickenson · Paolo Vineis · Reidar Lie · Ruud ter Meulen

Evidence-based criteria are increasingly expected to guide clinical decision-making and the allocation and rationing process in global health care systems. Health policies worldwide reveal the growth of mechanisms aimed at establishing evidence-based parameters for acceptable clinical practice, and a range of apparatus for monitoring and enforcing these parameters. In this context, the Evidence-based Medicine (EBM) movement has provoked strong restatements from within the clinical world about the essence of the patient-clinician relationship, and the balance between scientific approach and personal experience. Some of the chapters in this volume, particularly in the section on clinical practice, reflect this critical attitude. On another track, some critics have questioned the movement's, sometimes exclusive, focus on one particular research design (i.e. the randomised controlled trial) as unnecessarily narrow and reinforcing the cultural and political values of particular research groups. These concerns are put forward in the section of this volume on the ethical issues of research methodologies related to EBM.

There are high expectations among health policy-makers about the role EBM could play in the context of allocation decisions in health care: when these decisions are based on the outcomes of (economic) evaluations, they may become more rational and transparent. However, there is concern, that this rationalisation of health policy and of the allocation decisions that may be part of it, will go at the expense of widely shared social values like equal access to care and solidarity. Moreover, it is believed that, when these policy decisions are taken on the basis of specific research methodologies, like the RCT, weaker groups in our societies, like minority groups, the elderly and the chronically ill, may be seriously disadvantaged. These critical voices can be heard in the section of this volume on EBM and health policy. In this section, one can also find doubts about the feasibility of a purely economic approach to allocation problems, and a plea for a better link with the democratic process. EBM is not a value free enterprise, and should be enshrined in a moral and political perspective on the goals of medicine and the fair distribution of resources. In the first section of the volume there is a noticeable similar political critique: EBM is not moving around in a social and political vacuum, but is inherently influenced by the socio-political context.

On the basis of these concerns, the participants of the EviBase Project (see Introduction) made a number of recommendations at the final conference of the project in Maastricht in March 2003. At this conference, there were, in addition to the core group members, a number of participants from outside the project, representing various organisations, including the WHO/Geneva, The Dutch Health Council, The Royal Dutch Physicians Organisation, the National Institutes of Health (U.S.) as well as a number of others (see below).

The editors of the volume decided to publish the recommendations together, in the same way as the conference participants had written them down, and refined by the members of the project group. The recommendations are drawn into three sections: research, allocation of resources and quality of care

A. Concerning research priorities

1. The Evidence-based Medicine (EBM)-movement has been criticised due to the narrow range of accepted methodologies, especially the predominance of Randomised Controlled Trials (RCT). The EBM-movement has responded to these criticisms by attempting to broaden the range of accepted evidence, with evidence from non-RCT-methodologies. These efforts are promising and need to be pursued further. They will, however, have to be conducted within the general spirit of critical attitude and careful procedures that is characteristic of the EBM-movement.
2. The whole methodology of medical research puts much emphasis on the randomised trial, which has been developed for the investigation of simple treatments (typically drugs). However, it is more and more evident that many complex interventions (prevention, personal 'care' as opposed to 'cure') are as, or even more, effective as drugs. However, demonstrating the effectiveness of such complex interventions with RCT is difficult and also conceptually questionable. Therefore, we believe that a methodological investment into more flexible research tools is necessary within the EBM movement.
3. Vulnerable social groups (such as the aged and the poor), and 'weak' parts of medicine and health care (such as health promotion and care for chronically ill people), are still underrepresented in the current evidence base, despite efforts to remedy this deficiency. In order to increase the practical relevance of EBM, research concerning these groups, and these fields, deserves extra priority. Public research bodies have a special responsibility to this end.
4. In view of their relevance, social and cultural determinants of health and the effectiveness of health care are still an under researched subject. Ideally, any evidence base, concerning a specific subject in medicine and health care, should include references to the available evidence on the social and cultural determinants of that specific subject.
5. In order to increase the practical relevance of evidence, stakeholders in the practice of health care (such as patients and different health care providers) should be enabled to participate in the decision making process concerning research priorities. This should include the selection of issues, target groups and interventions to be tested.
6. The call for more evidence-based therapies does not mean that more treatment is always better. In some (stages of) illnesses, it might be better to have no treatment at all. In this respect, by supplying evidence on the outcome (or poor outcome) of medical treatment, EBM could act as an ally against over-treatment or medicalization.

B. Concerning allocation Issues

7. There is agreement that the concept of justice is one of the cornerstones of an ethical framework concerning EBM. The quality of care provided to vulnerable groups (the poor, the elderly) is an important ethical touchstone, to assess the consequences of EBM with regard to the concept of justice. This recommendation supports the second recommendation, that EBM should include vulnerable social groups in its research as well as the 'weaker' parts of medicine particularly, care for chronically ill and elderly patients, even when it is difficult or not possible to conduct RCTs in these areas.
8. So far, the methodological instruments to assess the efficacy and utility of treatment and care are mainly geared towards comparison within certain sub domains of medicine. To improve the relevance of EBM for allocation issues, it is desirable that more and better instruments are developed for comparison between sub domains.
9. To strengthen the legitimacy of Evidence Based allocation procedures, it is vital that maximum transparency is safeguarded concerning the procedures for the use of evidence. Such transpar-

ency might be furthered by the existence of a plurality of institutions involved in the process of assessing evidence for the use allocation proposals. In general, it is undesirable for a single institution to have a monopoly on this process.

C. Concerning the implementation of evidence to improve quality of care

10. Next to the concept of justice, respect for the autonomy of patients belongs to the ethical framework for the assessment of the consequences of EBM. To uphold and preserve the respect for autonomy, the process of deliberation between patient and physician on treatment and care options is vital. EBM should always be regarded as a tool to enhance this process, rather than limiting the options to a single rational one, against all other irrational options.
11. EBM must be treated as an instrument of empowering doctors and other health care professionals, rather than restricting them. The whole process of translating evidence for practical use by health care professionals must be shaped in such a way, that professionals experience benefits for the practical problems and choices they face.
12. It is important that all future physicians (and in appropriate ways also other health care professionals) are trained to understand the basic principles and procedures of EBM, and the potential practical use of EBM. It remains debatable, however, whether a practicing doctor should have methodological training to the extent that he or she can independently assess the value of research reports. The basic goal for education in this regard, should be to promote a thorough critical attitude toward all evidence claims, including those from the EBM movement.
13. EBM is so important, that it should not be restricted to doctors. One consequence is that EBM principles and procedures, with all the considerations and provisions mentioned earlier, should be extended as much as possible to include other fields and professions in health promotion and health care. Another consequence, is to strive for an integrated approach in the production and implementation of evidence, i.e. to start with health care problems instead of doctor's problems when and where relevant. This should include the involvement of other health care professionals in the process of producing evidence and practice guidelines.

The following persons attended the final conference of the EviBase Project where the recommendations were discussed:

Richard Ashcroft	Imperial College, London
Nikola Biller-Andorno	University of Goettingen
Marion Davis	National Institute of Health, Washington DC
Donna Dickenson	Centre for Global Ethics, University of Birmingham
Michaele Ghisleni	University of Torino
Don Hill	Ethox, Oxford
Rob Houtepen,	University of Maastricht
Christian Lenk	University of Goettingen
Reidar Lie	University of Bergen
Hans Maarse	University of Maastricht
Dick Meerman	Royal Dutch Physician's organisation
Ruud ter Meulen	Institute for Bioethics, University of Maastricht
Josy Moust	University of Maastricht
Nico de Neeling	Health Council, The Hague
Herman Nys	University of Louvain/University of Maastricht
Tessa Tan-Torres	WHO Geneva
Rein Vos	University of Maastricht
Trudy van der Weijden	University of Maastricht

Though the participants took part in the discussion, the recommendations as they are formulated above do not necessarily reflect their individual opinion or the opinion of the organization they were coming from.

Notes on the Contributors

Richard Ashcroft, M.A., Ph.D., is Leverhulme Senior Lecturer in Medical Ethics at the Imperial College, London.

Nikola Biller-Andorno, M.D., Ph.D., is Professor and Director of the Institute of Medical Ethics at the Charité – Joint Medical Faculty of Humboldt and Free University, Berlin, Germany.

Marion Danis, M.D., is Head of the Section on Ethics and Health Policy of the Department of Clinical Bioethics at National Institutes of Health. Please note that the author's contributions to this book do not necessarily reflect the views of the National Institutes of Health.

Donna Dickenson, B.A., M.Sc, Ph.D., is Professor at the Centre for the Study of Global Ethics at the University of Birmingham.

Fabrizio Fabris was Full Professor of Gerontology and Geriatrics at the University of Torino, Director of the Postgraduate School of Geriatrics and Director of the Turin section of the Italian Atherosclerosis Study Group. Professor Fabrizio Fabris prematurely died on August 21st, 2003 at the age of 68 years.

Klasien Horstman, Ph.D., is Professor of Philosophy at the University of Eindhoven and Associate Professor at the University of Maastricht.

Rob Houtepen, Ph.D., is Assistant Professor at the Department of Health Care Ethics and Philosophy at the University of Maastricht.

Christian Lenk, Ph.D., is Assistant Professor for Medical Ethics at the Department of Ethics and History of Medicine at the University of Goettingen.

Heike Leonhardt-Huober is Medical Doctor and collaborator of the Institute of Clinical Economics in Ulm.

Alessandro Liberati, M.D., is Professor of Biostatistics and Clinical Epidemiology at the University of Modena and Reggio Emilia and Director of the Italian Cochrane Centre, Italy.

Reidar Lie, M.D., Ph.D., is Senior Research Fellow at the Department of Clinical Bioethics at the National Institutes of Health of Bethesda, Maryland, USA and Professor at the Department of Philosophy of the University of Bergen, Norway.

Barbara Maero is working as a geriatrician at the University Hospital of Torino.

Ruud ter Meulen, Ph.D., is Professor of Philosophy and Director of the Institute of Bioethics at the University of Maastricht.

Erik Nord, Ph.D., is Senior Researcher at the Norwegian Institute of Public Health.

Ole Frithjof Norheim, M.D., Ph.D., is Professor in Medical Ethics at the Department of Public Health and Primary Care of the University of Bergen, and Consultant at the Department of Internal Medicine of the Haukeland University Hospital.

Franz Porzsolt, M.D., Ph.D., is Associate Professor at the University of Ulm and the Ludwig-Maximilians-University of Munich and works at the Institute of Clinical Economics.

Heiner Raspe, M.D., Ph.D., is Professor for Social Medicine at the University Clinics Schleswig-Holstein, Campus Lübeck.

Andreas Reis, M.D., is Associate Professional Officer at the Department of Ethics, Trade, Human Rights and Health Law at the World Health Organization in Geneva, Switzerland. Please note that the author's contribution to this book does not necessarily reflect the views of the World Health Organization.

Wendy Anne Rogers, B.M., B.S., BAHons, Ph.D, MRCGP, FRACGP, is Associate Professor of the Flinders University of Adelaide, Australia.

Carla Scarafiotti M.D. is Medical Assistant-Geriatrician and is working at the Geriatric Department of the Molinette University Hospital of Torino.

Christien Sepers M.A. was researcher at the Institute for Bioethics in Maastricht and is now working for the Province of North-Holland in Haarlem.

Professor Mariachiara Tallacchini, Ph.D., teaches Science, Technology & Law at the Law Faculty of the Catholic University of Piacenza, Italy.

Paolo Vineis, M.D., is Associate Professor of Medical Statistics at the University of Torino, Italy.

Rein Vos, PhD, is Professor of Theory of Health Sciences at the University of Maastricht.

Dick Willems, MD, PhD is Professor of Medical Ethics at the Academic Medical Centre/University of Amsterdam.

Index of Names

Index