Proceedings

of the

Royal College of Physicians of Edinburgh

Supplement No. 9

Clinical effectiveness, clinical guidelines and clinical standards 'from history to the future'

Papers from a symposium held on 11 May 2000 and a conference held on 3 November 2000

2001 Volume 31

CONTENTS .....

Supplement number 9

Publication number 71
ISSN 0953-0932

Clinical effectiveness, clinical guidelines and clinical standards 'from history to the future'

Papers from a symposium held on 11 May 2000 and a conference held on 3 November 2000 at the Royal College of Physicians of Edinburgh

Editorial and Production Team:

G. McAlister	Head of Publications
A. Borthwick	Publications Officer
A. Wood	Assistant Publications Officer
E. Clifford	Publications Assistant

Papers from the symposium on the scientific basis of clinical effectiveness held on 11 May 2000 at the Royal College of Physicians of Edinburgh

Organised by the Royal Society of Edinburgh and
the Royal College of Physicians of Edinburgh with
support from the Clinical Resource and Audit Group (CRAG)

The Royal Society of Edinburgh

Papers from the conference on clinical effectiveness, clinical guidelines and clinical standards 'from history to the future' held on 3 November 2000 at the Royal College of Physicians of Edinburgh

Organised by the Royal College of Physicians of Edinburgh in collaboration with the Scottish Intercollegiate Guidelines Network (SIGN) with support from the Clinical Resource and Audit Group (CRAG)

Acknowledgements

The Royal College of Physicians of Edinburgh would like to thank:

Sheikh Al Hammadi

CRAG

Merck

Novo Nordisk

Organon Laboratories

Solvay

3M Healthcare

for their contribution as sponsors of the conference held on 3 November 2000

and also:

CRAG

The Royal Society of Edinburgh

for their support of the symposium held on 11 May 2000

ORGANISING COMMITTEE FOR SYMPOSIUM HELD ON 11 MAY 2000

Professor Gordon Lowe	Assessor, the Royal College of Physicians of Edinburgh
Professor John Beck	Programme Convener, the Royal Society of Edinburgh
Dr Jill Pell	Consultant in Public Health Medicine, Greater Glasgow Health Board
Professor Jeremy Grimshaw	Dean of the Faculty of Medicine, University of Sheffield
Professor Iain Bouchier	Formerly Chief Scientist, Scotland

ORGANISING COMMITTEE FOR CONFERENCE HELD ON 3 NOVEMBER 2000

Professor Gordon Lowe	Assessor, the Royal College of Physicians of Edinburgh
Professor Jeremy Grimshaw	Programme Director, Health Services Research Unit, University of Aberdeen
Professor James Petrie	President, the Royal College of Physicians of Edinburgh
Ms Juliet Miller	Director of SIGN

PREFACE

Biomedical information continues to accumulate rapidly, and it is not surprising that there is now a demand that the fruits of this research should be reflected in improved clinical care for patients. The notion that medical practice should be based on evidence of the clinical effectiveness of different treatment options can be seen as a response to this demand, and Evidence Based Medicine (EBM) is now challenging medical practice based on patho-physiological theory and personal clinical experience. The development of EBM is not, however, either simple or straightforward; there are problems of patchy availability of reliable information, even for common diseases; differences of opinion regarding the methodology for bringing information together into evidence-based recommendations; the place of clinical effectiveness, irrespective of cost; the necessity for cost-effectiveness in a world of limited resources; the place of equity in determining care; and the mechanisms for producing valid guidelines and standards for clinical practice. Furthermore, the problems do not cease with the production of high quality guidelines or the definition of clinical standards. It is necessary for guidelines to be easily implemented; evidence of implementation (Evidence Based Implementation) should be demonstrated and resources for education and audit to underpin this needs to be allocated.

The RCPE was founded to promote the highest standards of clinical practice and, given the importance of promoting and implementing clinical effectiveness in achieving this end, the College hardly needs to explain two meetings on this subject in one year. Indeed, Fellows of the College such as Charles Lind, Alexander Hamilton and Thomas Balfour were important in producing reliable evidence for medical practice based on sound methods in the eighteenth and nineteenth centuries, long before our current notions of EBM.

The College and our then President, Jim Petrie, recognised that EBM had reached a stage where progress could best be facilitated by bringing acknowledged experts in this field together. The College therefore welcomed an approach from the Royal Society of Edinburgh proposing a joint conference on the scientific basis of clinical effectiveness. Our second conference, on the clinical application of EBM, coincided with publication by the College of the book *"To Improve the Evidence of Medicine". The 18th century origins of a critical approach*.[1] This allowed us to set the conference's deliberations in the context of an historical continuum whilst highlighting current (and future) developments in clinical effectiveness, clinical guidelines and clinical standards which concern all health care professionals.

The College's ultimate interest in standards of medical practice focuses on the care given to individual patients by individual doctors. 'Progress' with EBM needs to take this into account and requires that guidelines, based on group data, are translated into the best clinical standards for an individual patient. This may be the most difficult step of all, requiring that we produce doctors with the necessary knowledge of guidelines and the wisdom and compassion to apply them to individuals. We hope readers will benefit from the wealth of information in this supplement, and we are grateful to Anne Walker for producing a masterly overview[2] which we recommend to readers as their starting point.

NF

REFERENCES

1 Tröhler U. *"To Improve the Evidence of Medicine". The 18th century British origins of a critical approach*. Edinburgh: Royal College of Physicians of Edinburgh; 2000.

2 Walker A. Clinical effectiveness 2000: an overview. This supplement, pp. 2-6.

CLINICAL EFFECTIVENESS 2000: AN OVERVIEW*

A. Walker, Senior Research Fellow, Health Services Research Unit, University of Aberdeen

SCIENTIFIC BASIS OF CLINICAL EFFECTIVENESS

Clinical effectiveness at the sharp end

Evidence-based practice implies that decisions about treatment options will depend largely on evidence of their clinical effectiveness. Developing and using this evidence is not always straightforward, however. Professor Charles Warlow from the Department of Clinical Neurosciences at the Western General Hospital in Edinburgh argued that while scientific evidence of clinical effectiveness is growing, there are numerous gaps. For some health care problems, such as stroke, a substantial body of good quality evidence exists about some treatment options. This is less often the case for rare conditions, e.g. multiple sclerosis, or for 'unfashionable' conditions, such as incontinence or 'medically unexplained symptoms'. Even when a large body of evidence exists, it is often skewed towards particular forms of intervention. For example, whilst much is known about drug and surgical interventions for people who have had a stroke, much less is known about the benefits and costs of physiotherapy, occupational therapy, speech therapy, psychological interventions or nursing procedures. Further research is needed to fill these gaps at a time when it is becoming increasingly difficult to evaluate clinical interventions. The major challenge for the future is to find ways to overcome some of the obstacles to clinical research, e.g. by addressing the way in which research is funded; by improving public understanding of concepts such as randomisation; by reducing the bureaucracy surrounding ethics committees; by increasing public confidence in doctors and clinical research; and by making careers in research more attractive.

Making use of effectiveness information within the complexity and diversity of real practice is the major challenge for physicians working at the 'sharp end'. Around 90% of British health care is provided in primary care settings, and yet this is often the most difficult environment in which to practice evidence-based medicine. The dimensions of primary care that need to be incorporated into evidence-based practice were clearly outlined by Dr Jim Grant, a GP from Auchterarder. The average GP is involved in 5–6,000 face-to-face consultations each year, covering a diverse range of health care problems. The physician must work quickly to achieve a diagnosis, but may have limited access to local specialists or diagnostic tests. Often, the classic 'medical model' of diagnosis and treatment is inadequate in primary care, and biopsychosocial models that take a holistic view of the patient are more appropriate. The challenge here is to create a science base (and implementation strategy) that supports decision-making without losing the 'whole person approach' that is the cornerstone of high quality general practice.

From clinical effectiveness to evidence-based prioritisation

Concern about rising health care costs has been a major impetus behind the rise of evidence-based practice. Information about the relative effectiveness of different interventions can provide a foundation for making rational and equitable decisions about the treatments that will be offered within limited resources. Perhaps surprisingly, evidence of clinical effectiveness is rarely used in the prioritisation of health care services in the UK as Dr Harry Burns, Director of Public Health for the Greater Glasgow Health Board, pointed out. In his view, a key reason for this is that health care in the UK is a government service, funded from taxation. This makes the provision of health care subject to political control in a more overt way than many other health care systems, and means that priorities are often set for the NHS out of political necessity. The classic recent example of this has been the pressure placed on managers to reduce waiting list numbers.

For Professor Alan Maynard of the York Health Policy Group, economics-based medicine (or cost-and-benefit health care) should be more important than evidence-based medicine in determining how limited health budgets are used. Information about clinical effectiveness is a useful but incomplete way of prioritising care. He suggested that, if the NHS budget is to be used to maximise improvements in health for the population, then economic efficiency or cost effectiveness is a better guiding principle. The reasoning behind this is that cost effective procedures or treatments are always clinically effective, but the reverse is not always the case. For example, interferon ß is a clinically effective treatment for some groups of patients, but its price means that the benefits gained by using the drug do not outweigh the costs, so it is not cost effective. However, while cost effectiveness data might be better than clinical effectiveness data, in his view it is still insufficient, because it assumes that efficiency should be the basic principle underlying health-related decisions, rather than equity. The principle of equity aims to reduce inequality in health outcomes across social groups, or to ensure that everyone has an equal chance of a long and healthy life. So, for example, although interferon ß might not be a cost-effective treatment, a policy maker might feel that it should be funded because the people who will benefit are young and deserve a 'fair innings'. There are many other cases in which efficiency and equity might be traded off against each other in this way.

One of the barriers to the use of effectiveness information in health care prioritisation is the absence of a single, trustworthy source of evidence. The primary goal

* *The symposium* 'Scientific Basis of Clinical Effectiveness' *was held at the Royal College of Physicians of Edinburgh on 11 May 2000 and the conference on* Clinical Effectiveness, Clinical Guidelines and Clinical Standards 'From History to the Future'*was held at the College on 3 November 2000*

of the new Health Technology Board for Scotland (HTBS) is to fill this gap, by providing advice on the clinical and cost effectiveness of health technologies to the NHS in Scotland. It has been established as a Special Health Board, and has a similar remit and responsibilities to the National Institute for Clinical Excellence (NICE) in England and Wales. The key difference is that the HTBS agenda will be more sensitive to the information needs of the clinical community in Scotland, rather than being driven by the Department of Health. Dr Angus Mackay, Chairman of the HTBS, told the audience that the Board will be basing its advice on high quality appraisals from elsewhere, but will also conduct (or commission) reviews of its own. The HTBS will be an advice-giving body and not a 'rationing' organisation. Decisions about the availability of treatments will remain in the hands of Health Boards, Trust Boards and Area Drug and Therapeutics Committees.

Clinical effectiveness and clinical practice

The growing acceptance that medical practice should be based on evidence of clinical (or cost) effectiveness creates challenges for doctors as well as policy makers. Not least among these are the problems of assimilating the increasing amount of effectiveness data, and identifying sub-optimal performance. The problem of data overload is not unique to medicine, and Professor Jim McDonald of the Centre for Electrical Power Engineering at the University of Strathclyde introduced the concept of intelligent systems as a potential solution. Using case studies in engineering and medicine, he described the application of the 'knowledge analysis and design structuring' (KADS) technique to capture expert knowledge from individuals and create intelligent decision support systems.[1, 2] Systems such as this have already demonstrated their value as a way of introducing a knowledge management culture into organisations. The issue of clinical sub-optimal performance is equally difficult. Theoretically, it should be straightforward to identify under-performing physicians or surgeons in relation to appropriate outcome measures, after taking case mix into account. Drawing on experience in the Public Inquiry into Paediatric Heart Surgery at Bristol Royal Infirmary, Professor Gordon Murray from the Department of Community Health Sciences at the University of Edinburgh demonstrated that the situation is more complex than this. Even when the technical difficulties inherent in assessing clinical performance can be resolved, profoundly difficult value judgements remain about the level of performance that is acceptable.

Increasingly, medical consultations are based on information: evidence-based information about the effectiveness of different treatments, and a wide range of publicly available health information. Professor Ian Watt of the Department of Health Studies in the University of York, argued that physicians need to respond to these changes by adopting more participatory styles in their individual consultations, facilitating informed decision-making. Often this means that doctors will need to act as interpreters of information, rather than 'experts', using critical appraisal skills to present complex information to patients accurately and appropriately.[3] These thoughts were echoed by Professor Peter Brunt from the Grampian University Hospitals Trust, who urged the medical profession to listen more and become more responsive and communicative. The challenge to doctors, he said, is to maintain those qualities of 'professionalism' that have stood the test of time while being responsive to the needs of the twenty-first century. The limited extent to which these goals have been achieved was clearly demonstrated in the final paper of the symposium. Professor Angela Coulter, Chief Executive of the Picker Institute Europe, presented data, collected in a range of surveys conducted by the Picker Institute, of the quality of health care across Europe and the US.[4] Nearly two-thirds of patients surveyed in British hospitals did not feel involved in decisions about their care, and a third reported that doctors tended to talk in front of them as though they were not there.

CLINICAL EFFECTIVENESS, CLINICAL GUIDELINES AND CLINICAL STANDARDS 'FROM HISTORY TO THE FUTURE'

Historical perspectives

The modern evidence-based medicine movement has only been in existence for around 15 years, however, but physicians have always based their practice on evidence in one form or another. Professor Ulrich Tröhler, Professor of Medical History at the University of Freiburg, described three traditions of evidence that are used in medicine, which he calls 'systemic-pathophysiological', 'individual-clinical', and 'socio-analytic'.[5] Systemic evidence is derived from the physician's understanding of pathophysiology. Medical systems ranging from those of the natural philosophers of antiquity through to modern molecular biology provide this evidence. A systemic approach provides certain and objective knowledge and 'precludes' the need for empirical evidence of therapeutic effectiveness because treatment failures can always be explained within the system. Individual clinical evidence is derived from the physician's knowledge of the individual patient, rather than from a prescribed system. This evidence is subjective, but nonetheless 'certain'. Empirical evidence of therapeutic effectiveness for a particular patient has value within this approach, but the physician's knowledge of their patient can always be used to explain the failure of a treatment. In both of these approaches, decisions are arrived at through a process of 'medical judgement', based on the physician's systemic knowledge or their clinical skills. The socio-analytic approach differs from each of the others because it generates 'probable' rather than 'certain' evidence. Empirical evidence of therapeutic effectiveness, derived from the analysis of aggregated data obtained from groups of patients according to a set of external rules, is the cornerstone of this approach.

Each of these traditions has contributed to the success of medicine, but the existence of three types of evidence, with varying levels of certainty and differing implications for the role of medical judgement, creates the potential for tensions between the approaches. At different points in history, one or other of the approaches has been in favour, while another has been marginalised. For example, the socio-analytic approach emerged in the eighteenth century in response to the introduction of new therapies to challenge traditional ones. It became marginalised in the nineteenth century, however, as developments in experimental natural sciences and the introduction of new instruments for observing patients strengthened the pathophysiological and individual-clinical traditions. For Professor Tröhler, true innovation in the search for therapeutic evidence would require integration of all three approaches.

The renaissance of the socio-analytic approach is generally thought to have occurred in the middle of the twentieth century. At this time, the statistician Ronald Fisher[6] described the range of methods that can be used in experiments to ensure that comparisons made between different treatments are fair. The recent celebration of 50 years of randomised controlled trials[7] is based on the widely held view that the MRC Streptomycin trial, reported in 1948,[8] was the first study to address the problem of selection bias by allocating patients to different treatment groups at random. In fact, the first known controlled trial was James Lind's study of citrus fruits as a treatment for scurvy in 12 sailors in 1754.[9] Until now, though, it seemed that empirical studies of treatment effectiveness were completely quiescent between the eighteenth and twentieth centuries.

Iain Milne, the College librarian, and Sir Iain Chalmers, Director of the UK Cochrane Centre, are generating a collection of study reports that challenge this view.[10] They have found many examples of fair treatment comparisons in the early twentieth century; two studies, in the nineteenth century, both of which were conducted by Edinburgh graduates, have recently been uncovered. Thomas Graham Balfour, President of the Statistical Society and Honorary Physician to the Queen, evaluated the effects of homoeopathic belladonna as prophylaxis for scarlet fever in 1854. His study carefully selected children at risk of scarlet fever and alternately allocated them into one of two groups. His results showed no beneficial effects of belladonna, but the report demonstrates the sophistication of his thinking by acknowledging the possibility that his small study may have missed important effects (or type II error, in modern parlance). Alexander Lesassier Hamilton, a former President of the College and a man who led a full and colourful life,[11] reported the use of rotation to generate two comparable groups of sick soldiers during the Peninsular War in his MD thesis in 1816. His trial of bloodletting demonstrated that it was associated with significant harm to the soldiers. The discovery of these studies suggests that the socio-analytic approach was not entirely quiescent for 200 years and, intriguingly, that the flame appears to have been kept alive by Scottish graduates.

The socio-analytic approach is currently in ascendancy in the form of the international evidence-based medicine (EBM or evidence-based practice) movement, which tries to link clinical practice and health policy to systematic scientific evidence of therapeutic effectiveness. Concerns about rising health care costs, evidence of significant geographical variation in the delivery of services and evidence that a proportion of tests and treatments in routine use are ineffective or harmful, created the circumstances for the growth of EBM. Professor Steven Woolf, Professor in the Department of Family Practice at Virginia Commonwealth University, described in his talk the current state of the science.

Work within the discipline of EBM is occurring on several levels. Critical appraisals of individual studies are routinely published in journals such as *Evidence-Based Medicine, ACP Journal Club* and *Bandolier*. Systematic reviews and meta-analyses of evidence are published in most journals and produced by a range of organisations, including the Cochrane Collaboration, the NHS Centre for Reviews and Dissemination and the AHRQ Evidence-Based Practice Centers in the US. Evidence-based practice guidelines, developed according to explicit methods, are being produced in many countries (see below). Finally, tools for implementing the practice changes recommended by guidelines are being developed within systems that deliver clinical and public health services. Professor Woolf argued that these tools need to focus on increasing knowledge, promoting acceptance, enhancing ability and creating reinforcement. The major challenge now facing EBM is the need to individualise choices about treatment options in terms of costs and benefits. Often the best solution to this problem may be for patient preferences to take precedence.

Clinical guidelines: where are we now?

The EBM practice movement is now international, with programmes of guideline development established in many countries. The different approaches taken to guideline development in Europe were clearly apparent in presentations from Juliet Miller, the Director of the Scottish Intercollegiate Guidelines Network (SIGN); Professor Peter Littlejohns, the Clinical Director of the National Institute for Clinical Excellence (NICE); and Professor Günter Ollenschläger, Director of the Agency for Quality in Medicine in Cologne.

In Scotland, the emphasis has been on the methodology of guideline development and the need for health care, professional involvement and 'ownership'. Scotland is at the forefront of this, largely through the activities and achievements of SIGN. Since its establishment in 1993, SIGN has published 50 guidelines, with another 14 in development or under review.[12] Health care professionals nominate topics for guideline development. Each guideline is developed by a multidisciplinary group, based on a systematic review of the scientific evidence, and each contains recommendations for best practice which are explicitly linked to the supporting evidence and graded according to the strength of that evidence. The reputation of SIGN comes not only from the guidelines it develops, but also from the contributions it has made to the process of developing them and its commitment to improving the methodology. For example, SIGN has recently introduced a revised system for grading the strength of evidence on which practice recommendations are based, and better recording systems to ensure that the process underlying these recommendations is as transparent as possible.

The National Institute of Clinical Excellence (NICE) in England and Wales is a relative newcomer to guideline development. The National Institute of Clinical Excellence was created by legislation,[13] rather than by professional groups, in response to public concerns about inappropriate variations in the care provided by the NHS. It has been in physical existence since April 1999 and has two key roles.[14] The first is to provide guidance to the NHS (outwith Scotland) on clinical effectiveness and cost effectiveness of health technologies and methods of clinical audit. The second is to set, deliver and monitor standards of care. Like SIGN guidelines, NICE guidelines will be based on systematic reviews of research evidence. Unlike SIGN, however, NICE reviews will be externally commissioned and paid for by the NHS National Research and Development (R&D) programme. Each guideline is likely to take around one to two years to develop from initiation to publication, and will include recommendations based

on clinical effectiveness, cost effectiveness and effect on the NHS. In addition to guidelines about best practice, NICE will appraise the evidence relating to new health technologies as they emerge for the Department of Health. Currently around two of these appraisals are conducted per month, but this is likely to double as NICE becomes established. The National Institute of Clinical Excellence will also develop national clinical standards (outwith Scotland) and provide support for clinical audit to assess performance against the standards.

In Germany, the situation is different again. There, a wide range of national, federal and local organisations produce guidelines in diverse areas and of varying quality. To assure and promote the quality of German guidelines, the German Guideline Clearing house (GGC) was established in 1999 as a partnership between professional organisations, service providers and purchasers of health services.[15] The German Guideline Clearing house does not produce guidelines, instead it identifies and evaluates existing guidelines. A multidisciplinary group of experts peer review guidelines that meet minimum quality criteria and prepare a clearing report. This includes suggestions and recommendations for practice, supported by explicit evidence drawn from the included guidelines. The GGC finished its first project – the National Hypertension Guidelines Clearing Project – in spring 2000. Thirty-four guidelines were identified, 11 of which met the minimum quality criteria. The final clearing report has been used as a tool for continuing medical education on hypertension, and recommendations about guideline development have also been fed back to the participating organisations.

The problem of proliferating guidelines that have been developed using differing methods is not unique to Germany, but shared throughout Europe. The Council of Europe has made an important contribution to addressing this problem by nominating an Expert Group to prepare an international guideline for producing clinical practice guidelines. Professor Marjukka Mäkelä, Head of Research and Development at the National Research and Development Centre for Welfare and Health (STAKES) in Helsinki, described the work of the Expert Group. It is chaired by Professor Jim Petrie and includes members from nine European countries. Across the Group, agreement on the key functions of guidelines and the best methods for developing them were relatively easily agreed. Legal issues and the relationship between clinical guidelines and clinical standards were more difficult. Ultimately, the Group will recommend that the main aim of guidelines is to support and promote optimal health care. The best guidelines are those produced using methods similar to those developed by SIGN. From a legal perspective, however, guidelines are not binding rules. The Group felt that professionals need to be aware of existing guidelines, and may be expected to justify deviation from them, but compliance with a guideline is not necessarily equivalent to good practice, and deviation from them does not necessarily constitute negligence. The relationship between guidelines and clinical standards was felt to be complex, and likely to vary between different cultures and health care systems. The Group will recommend that guidelines may include suggestions for clinical indicators of process and outcome. However, each nation or health care system will need to decide the level of these indicators that is acceptable for them.

Clinical guidelines: where are we going?

Agreement about the best methods for developing clinical guidelines may be emerging, but that is only the first step in promoting good health care. So, which issues now need to be addressed? Three major areas were identified. First, methodological advances in the systematic reviews underpinning guidelines and the incorporation of economic perspectives within them. Second, the thorny issue of how to implement guidelines in practice. Third, the equally difficult issue of the relationship between evidence-based practice and clinical standards.

Systematic reviews of research evidence have become the 'gold standard' foundation for clinical guidelines. Professor Adrian Grant, Director of the Health Services Research Unit in Aberdeen, described the work of the Cochrane Collaboration in developing the methods for producing systematic reviews and outlined some new challenges in this area. As systematic reviews become more widely used, their strengths and limitations are better understood. In essence, the usefulness of a review depends on the quality, quantity and completeness of the contributing studies. In many areas, systematic reviews are of limited use because there are few good quality studies, because the contributing studies cannot be combined, or because they cannot address the questions that users ask. Debate is ongoing about the best ways of conducting systematic reviews; for example, when (and when not) to combine data, and the extent to which results of trials can be applied in routine clinical care. New methods are also developing to include measures of economic efficiency or resource use within or alongside systematic reviews.

The use of economic information in clinical guidelines is also growing. As Professor Martin Eccles, Professor of Clinical Effectiveness at the Centre for Health Services Research in Newcastle upon Tyne, pointed out in his presentation, health care is not free. The desire to include information about costs in guidelines reflects the recognition that health care must always be paid for; it is only when and how it is paid for that varies. There are a number of ways in which information about the costs and benefits of treatments could be presented within guidelines, all of which are relatively crude and require value judgements to be made. One of the most comprehensible is a profile approach, in which the benefits, costs and consequences of treatment alternatives are listed with the explicit acknowledgement that some of these may be uncertain or unknown. For any particular patient, a profile approach may suggest an obvious recommendation, or it may require 'trade-offs' between different treatment options. However, much more work is needed to develop methods for including cost effectiveness information within guidelines, and to debate the validity and utility of the information.

Guideline development is only one step towards better health care. The most rigorous, evidence-based guideline is useless if its recommendations are not implemented into everyday clinical practice. Speaking on behalf of the Cochrane Effective Practice and Organisation of Care review group (EPOC),[16] Professor Jeremy Grimshaw of the Health Services Research Unit in Aberdeen, argued that 'evidence-based practice should be complemented by evidence-based implementation'. In other words, the choice of strategies to implement clinical guidelines should be based on the best evidence of effectiveness of the

different interventions. The Cochrane Effective Practice and Organisation of Care review group is an international collaboration which undertakes systematic reviews of interventions to improve the quality of care. It has established a specialist register of over 2,000 primary studies, and to date it has produced 18 reviews, with a further 20 protocols published in the Cochrane Library. An overview of these reviews has recently been published as part of the Effective Health Care Bulletin series.[17] The key message from this overview is that passive dissemination of clinical guidelines is generally ineffective in changing practice. Widely used interventions such as audit and feedback or the use of opinion leaders can be effective in some circumstances, although often they have little or no impact. The most effective interventions are reminders, educational outreach visits, interactive educational workshops and multi-faceted interventions.

The final presentation of the day was the inaugural Al Hammadi Lecture, given by Dr David Steel, Chief Executive of the Clinical Standards Board for Scotland (CSBS). Scotland has a long tradition of excellence and innovation in health care. The aim of the CSBS (and the Health Technology Board for Scotland) is to build on existing achievements in the development of clinical audit, clinical guidelines and clinical outcome indicators, in order to address the current challenges that face the health services. These challenges include: coordinating the activities of all the different organisations concerned with clinical effectiveness, and developing a strategic direction; involving the public and patients in decisions about health care standards and quality; and implementing systems to assure the quality of health care. A crucial element in the quality assurance process is the development of clinical standards. Clinical standards identify the key things to get right in order to improve care. As far as possible CSBS standards will be evidence-based, and will make use of existing standards and guidelines. They will focus on the care and treatment provided by health care professionals in terms of patient outcomes and following 'journeys of care'. Generic and condition-specific standards will be developed, with some being categorised as essential, while others are seen as desirable. The Clinical Standards Board for Scotland will use external peer review mechanisms to assess performance against the standards. This will involve both written evidence and review visits by multi-disciplinary teams. The aim of the reviews is to be constructive and to support good practice, not to 'catch people out'; however, if poor performance is identified it will be reported. Progress so far is good, with 19 pilot review visits successfully undertaken. A considerable amount of work is still needed, however. In particular, effective ways of involving the public in the process of quality assurance need to be established, and the culture of the NHS needs to change in order to ensure quality assurance within routine practice.

CONCLUSIONS

Evidence-based medicine and clinical effectiveness have a long history, but have come to dominate medical thinking over the last 20 years. This has been an international movement, but one in which Scotland, and indeed the College, have played a leading and influential part. These two meetings serve as a festschrift to Professor Jim Petrie, President of the College and Chairman of SIGN, in developing and promoting evidence-based guidelines both at home and abroad. These contributions were acknowledged by presentations to Professor Petrie on behalf of the German Medical Association, in recognition of his support during the development of the German Guideline Clearing house, and on behalf of Mohammed Al Hammadi as a tribute to the contributions made to raising standards of health care in Saudi Arabia.

Thanks to the work of organisations such as the Cochrane Collaboration, and the many international guideline development agencies, the science of evidence-based medicine has become well established. Many methodological challenges remain, but the foundation is firm and the basic principles are widely accepted. The same cannot yet be said for the art of evidence-based practice. There are numerous barriers to the use of clinical and cost effectiveness information in both medical practice and policy making. Much more work is still needed to understand how best to develop health care systems and to support health care professionals in a way that enables evidence-based medicine to become a reality.

ACKNOWLEDGMENTS

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Health Department. The opinions expressed in the paper are those of the author and may not be shared by the funding body.

REFERENCES

1 Schreiber AT, Wielinga BJ, Breuker JA, editors. *KADS: a principled approach to knowledge-based system development*. London: Academic Press; 1993.
2 Schreiber G, Akkermans H, Anjwierden A *et al*. *Knowledge engineering and management: the common KADS methodology*. Massachusetts: MIT Press; 1999. (See also http://www.commonkads.uva.nl)
3 Towle A, Godolphin W. Framework for teaching and learning informed shared decision making. *BMJ* 1999; **319:**766-9.
4 For more information see: http://www.picker.org
5 Tröhler U. *The History of Clinical Effectiveness*. This supplement, pp. 42-5.
6 Fisher RA. The arrangement of field experiments. *Journal of the Ministry of Agriculture* 1926; **33:**503-13.
7 Smith R, editor. Fifty years of randomised controlled trials. *BMJ* 1998; **317:** 1167-262.
8 Medical Research Council. Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment for pulmonary tuberculosis. *BMJ* 1948; **ii:**769-82.
9 Porter R. *The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present*. London: HarperCollins; 1997.
10 Find the Controlled Trials from History website at: http://www.rcpe.ac.uk/controlled_trials
11 Rosner L. *The most beautiful man in existence: the scandalous life of Alexander Lesassier*. Philadelphia: University of Pennsylvania Press; 1999.
12 Http://www.sign.ac.uk
13 Department of Health. *The New NHS: modern, dependable*. London: Department of Health; 1997.
14 Http://www.nice.org.uk
15 Http://www.azq.de
16 Http://www.abdn.ac.uk/hsru/epoc
17 NHS Centre for Reviews and Dissemination. Getting evidence into practice. *Eff Health Care* 1999; **5:**1-16. (see http://www.york.ac.uk/inst/crd/ehc51.pdf)

THE SCIENTIFIC BASIS OF CLINICAL EFFECTIVENESS: SETTING THE SCENE

C. Warlow, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh

INTRODUCTION

Usually people think of clinical effectiveness in terms of therapeutic interventions, or treatments, if you will. Not just drugs and surgery, but also physical interventions like physiotherapy, or psychological ones like cognitive behaviour therapy, or systems of care like early supported discharge from hospital to the community. The very necessary preceding diagnostic process is important to get right, but is another issue. For all therapeutic interventions, we expect and indeed should demand the best possible evidence of effectiveness. We can then divide them into those that definitely improve outcome and so should be used, if we can afford to; those that don't change outcome at all and should be abandoned; those that are plain dangerous and should be banned; and, finally, those we are not sure about, but which are promising, and so should be subject to further research.

RANDOMISED TRIALS AND META-ANALYSIS

Nowadays, it is generally agreed that the least biased and most reliable evidence of effectiveness usually comes from randomised controlled trials where practicable, and from combining trials in meta-analyses large enough to provide precise estimates of treatment effect overall, and perhaps in a few sub-groups as well. Comparing the outcome of two groups of patients – one given the conventional treatment and the other given the new treatment, who are *not* selected at random – is much more likely to end up as a comparison of two groups who are quite different at baseline, particularly with regard to factors such as age and disease severity which are likely to influence their eventual outcome. No amount of statistical wizardry can correct this bias reliably, and it is all too easy to get the wrong answer from non-randomised comparisons which may lead to untold harm to thousands of patients.

In general, randomised trials and meta-analyses give us an idea of the size of any treatment effect in terms of avoiding death, disability and poor quality of life. For example, a relative reduction in risk of a poor outcome of say 50% (from ten per cent dead to five per cent dead), translates into an absolute risk reduction of ten per cent minus five per cent – or five per cent. These days, the best place to start looking for this sort of information is in the Cochrane Library, which is updated every four months and is available over the internet or on CD-ROM.[1] It presently contains about 1,000 Cochrane reviews and meta-analyses across the whole of medicine, nearly as many protocols of reviews in progress, almost 3,000 non-Cochrane reviews, and about 300,000 references to randomised trials.

FROM EVIDENCE TO GUIDELINES AND PRACTICE

From our knowledge of the frequency of disease, and the expected outcome without treatment, we can then work out what sort of impact the intervention might have on the public health if it was implemented. From the example above, the absolute reduction of five deaths per 100 treated translates into 20 patients having to be treated to postpone one death in the sort of patient entered into the trials (100 divided by the absolute risk reduction of five). The health economists can then tell us how much it might cost us to 'buy' the intervention, or rather what we have to give up to get it. Finally, those who pay for health services have to decide if the intervention is at some level worth it, whether they be government, insurance companies or individual patients with cash in their pockets.

For common disorders like stroke we now have lots of this sort of information. For rarer disorders such as multiple sclerosis we have much less. But even for some very common disorders which are rather nebulous and poorly understood, like patients with 'medically unexplained symptoms', we have almost no information at all – chronic daily headache, atypical non-cardiac chest pain, irritable colon and such like. Patients with these disorders consume vast health care resources; about a third of all new general neurology out-patients have medically unexplained symptoms, for example.[2] But even for something as common and obvious as stroke with a clear onset and reasonably predictable outcome, our knowledge is very incomplete, particularly for non-drug and non-surgical interventions such as physiotherapy, occupational therapy, speech therapy, psychological interventions and nursing procedures. All these interventions cost money and all of them should have some effect on outcome, good or not so good, large or small, it is difficult to say. I am sure there are similar problems in other areas of medicine. Nonetheless, for the things that we *do* have information about, we can get some idea of the shopping list of what works and what doesn't, formulate guidelines, and from that decide what we can deliver and what we are prepared to afford.

BARRIERS TO PROPER EVALUATION OF THERAPEUTIC INTERVENTIONS

All this gathering of data and putting them together requires tremendous energy, time and resources. Sadly, however, the obstacles to the proper evaluation of treatment seem to be increasing – or, at least, they appear to be rather high.

Ethical barriers

Ethical hurdles are high and getting higher. Well-meaning ethicists and lawyers are certainly inhibiting progress by stopping research in suddenly unconscious patients in Scotland by requiring consent from a relative or carer. This makes clinical trials of promising very early treatments of stroke, head injury and cardiac arrest impossible to do at all – we will no longer be able to discover what treatments work, putting us back into the Victorian age of cupping and leeches. Restricting access to individual patient records without their explicit consent is a bombshell for

observational epidemiology and, indeed, for audit of patient outcomes.[3] In the meantime, ethicists seem to almost wilfully avoid confronting the huge double standard of unfettered medical and surgical opportunism and experimentation which is the norm in routine medical practice.[4] The bureaucracy surrounding application to ethics committees also seems to be getting worse: recently, after gaining Scottish Multicentre Ethical Committee approval for an innocuous observational epidemiological research project, we had to then get approval from 15 local ethical committees – the usual procedure. This took four months and nearly 6,000 pages of A4 weighing literally an arm and a leg, but at the end of the day only three modifications to the protocol were suggested – hardly encouraging for the young researcher who had to endure all this before even starting the research.[5]

Misunderstanding of randomisation

Randomisation is mistrusted, I suspect, because it is understood neither by the general public nor by journalists and the media. It is important to emphasise time and time again that patients who are randomised to the normal accepted treatment in a trial of a new, exciting treatment may actually be protected from an unexpected consequence of that new treatment, and there are many examples.[4] Moreover, joining a randomised trial is not just an altruistic act for future patients, but also a selfish act, because patients in trials, even in the control group taking the conventional treatment, tend to do better than expected. This is hardly surprising, given all the advantages of being in a trial compared with being treated in routine higgledy-piggledy clinical practice: trial treatment is peer reviewed by colleagues, ethics committees and those providing funding. Inclusion in a trial is not just decided on the whim of an individual clinician; medical records are kept for years, not lost or destroyed in a few years; the patients get written and detailed information, which is rare in routine practice; follow-up is regular and tenacious and any important ancillary treatments more likely to be given; and the risks and benefits of the new treatment are prospectively monitored by a data monitoring committee set up to protect the treated patients and the controls from harm.

Mistrust of doctors

These days doctors seem to be more mistrusted, if not under actual attack, and we have to put our own house in order after the recent series of medical scandals ranging from fraud through incompetence to outright murder. I believe we would also be more convincing if we – and those providing research funding – concentrated on the common disorders which are serious threats to public health, not just the rare and fascinating.

Paying for research

A perennial problem is getting research funded. Nowadays, the refereeing of grant applications seems capricious at best and biased at worst – after all, the referees are probably competing for the same pot of money. Under these circumstances researchers may be tempted to fall into the arms of industry where funding appears easier to get. But clearly industry will mostly be interested in drug interventions, and usually patented and expensive drugs at that – it is not their job to evaluate surgical interventions. In acute stroke, as the number of trials has increased over the decades, so the proportion sponsored by industry has risen to about 50%. Half of those are even authored by industry employees, giving rise to the possibility for conflict of interest, spin and bias.[6] But still these trials go on while the government looks the other way in the naïve expectation that someone else will pay for research, i.e. industry. So it will, on its own terms, but the NHS ends up not only paying a vast drugs bill but also not knowing whether promising non-drug interventions and old non-patented drugs are effective. Curiously and encouragingly, though, and again in *acute* stroke, the recent advances have all come from trials and meta-analyses which were *not* sponsored by industry – aspirin,[7] thrombolysis[8] and stroke units.[9]

Lack of time

Finally, there are increasing time pressures on NHS clinicians and university clinical academics. How can they evaluate treatments in randomised trials when they are increasingly involved with management without being allowed to drop clinical sessions to do the job properly; with (rightly) increased demands to train junior doctors, but at the same time relatively fewer junior doctors who are spending less and less time in service work as part of their training; with continuing medical education counted in terms of time spent and not either in terms of quality or appropriateness; with revalidation and the enormous amount of time that will take, both for the revalidated and the revalidators; and with non-doctor tasks like taking blood in out-patient clinics and filing chaotic patient records? There are also extraordinary amounts of red tape that are strangling our research efforts, and clinical care for that matter. For example, the research assessment exercise cripples medical research for several months every few years.

But red tape is not just a problem for medicine; it is just as bad, if not worse, for teachers, police, social workers and even those in the arts. In a letter to the *Scotsman* in 1999, the director of a small Scottish theatre company wrote: *'We are swimming in a bureaucratic sludge, awash with consultancies, incomprehensible accountancy language, over-management, juvenile debate on aims and strategies, pseudo business practice and the dreaded mission.'* Sound familiar?

It is hardly surprising that young bright clinicians are avoiding research careers, and we now have to have the Savill report to suggest ways of tempting them back from the NHS, private practice, abroad or even financial institutions in the city.[10] Someone, somewhere, has got to address these problems.

AUDIT

So that is all about research, but what about audit of patient outcomes, specifically sorting out how well we are putting research findings into routine practice – the reality of clinical effectiveness? Clearly it would be nice to know, and you might imagine essential to know, which hospitals, or services within hospitals, or even individual doctors, are producing the best outcomes. For how else can the consumer be empowered to choose, supposing a choice is comprehensible (which it so often isn't, e.g. in deciding on the best telephone or gas supply deal) and available (which of course in real life it seldom is, particularly in medical emergencies)? Here we must not lose our heads. Scientific rules to do with bias and sample size must be

applied just as rigorously in auditing patient outcomes as in randomised trials. But there are real problems, and let me mention three.

Firstly, the government wants us to evaluate the outcomes of care given by different hospitals; someone has advised them that this would be a good thing. In response to this, Nic Weir and others have looked at the risks of death six months after stroke in five Scottish hospitals.[11] There were very different death rates which clearly didn't happen by chance. But, of course, these differences could all be due to bias, specifically differences in case mix, because different severities of stroke are admitted to different hospitals. Routinely collected statistics like age and sex cannot possibly correct for stroke severity. One has to collect, at great effort and expense, baseline variables like conscious level from the records of a large number of patients (not forgetting the ethicist's impractical demand that we seek their consent first), develop a statistical model to predict outcome and then apply it to every patient. It then becomes apparent that four of the hospitals have about the same outcomes and that one seems worse. But it is difficult to be sure because the confidence intervals around the estimates are wide, even with a large number of patients. Are we seriously expecting this sort of effort and expense to audit disease outcomes from all hospitals, or even just of some diseases in some hospitals?

Secondly, what of the apparently easy target of comparing the results of individual surgeons? In fact this is far from easy, despite the screeching demands of the tabloid press, amongst others. For example, the operative risks of all the surgeons in the European Carotid Surgery Trial by the number of cases they operated on are shown in Figure 1.[12] Only one surgeon is outside the 99th percentile, but once his or her results are adjusted for case mix the discrepancy vanishes. One needs to evaluate a surprisingly large number of carotid endarterectomies to get a precise estimate of risk, but each individual surgeon doesn't do all that many, perhaps one or two a week in a busy place. If it is difficult for a common operation like carotid endarterectomy, what hope have we for evaluating the outcomes of the rare operations that are done, for example, by paediatric cardiac surgeons?

Finally, are we to allow clinicians and hospitals to audit their own outcomes? Might that not lead to biased, if not fraudulent, reporting with so much at stake from the point of view of funding and patient referrals, and even a call from the General Medical Council? You think that wouldn't happen? We have looked at reports of carotid endarterectomy in the literature and divided them into those by individual surgeons, those by groups of surgeons, those where a physician was an author and those where a physician saw the patient after surgery. It will be no surprise that the risk of death is about the same.[13] But strokes were different, not very common in the reports by single surgeons but far more common when a physician was involved. Are we to believe that physicians cause strokes, or that somehow the operator doesn't quite notice his or her disasters?

Unfocused demands for audit using biased methods with inadequate sample sizes, and without the resources to do it properly – remembering the pressures on clinicians' time that I have already mentioned – are doomed to failure and ridicule, will waste money and are completely unacceptable. Given the difficulties, it is hardly surprising there is so much awful audit around – perhaps a definition of most audit is 'bad research which doesn't have to be approved by an ethical committee but which does have to be repeated in five years'.

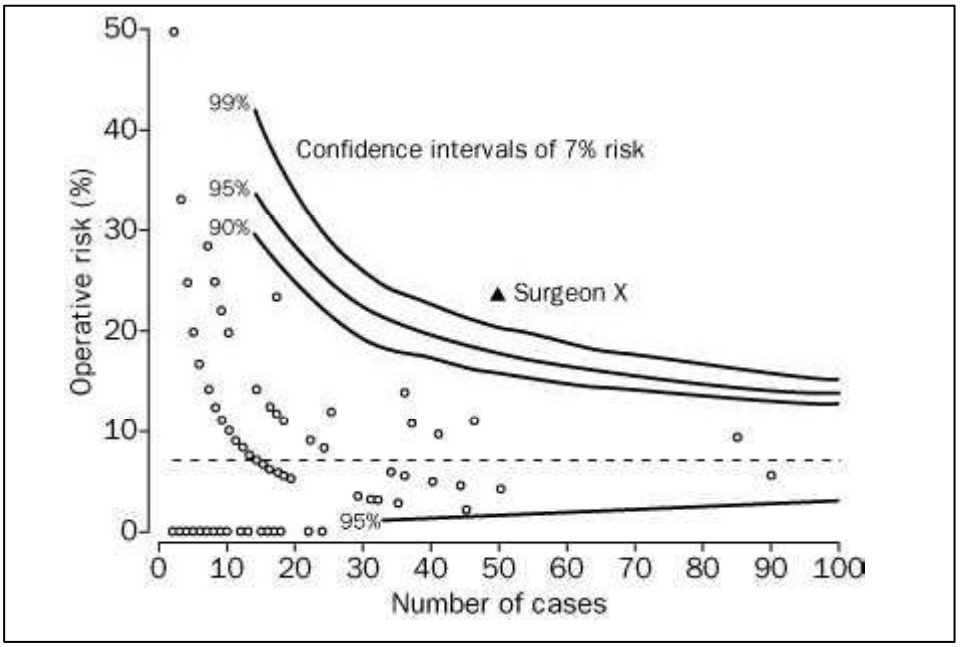

FIGURE 1

Operative risks of individual surgeons in the European Carotid Surgery Trial according to the number of patients they operated on. The dotted line represents the overall 7% operative risk of stroke and death. Reprinted with permission from Elsevier Science. Rothwell PM, Warlow CP on behalf of the European Carotid Surgery Trialists' Collaborative Group. Interpretation of operative risks of individual surgeons. *Lancet* 1999; **346:**1325.

CONCLUSIONS

I hope I have given you some idea of how clinical effectiveness – at least of therapeutic interventions – should be evaluated, where the best evidence comes from, what the problems are in collecting that evidence and how even more problematic it is to assess the effectiveness of routine clinical practice. Mostly I have taken my examples from stroke, which is relatively straightforward because it is common and quite easily diagnosed. But if stroke is difficult, and it is, spare a thought for those researchers and clinicians who have to grapple with disorders such as migraine, depression, osteoarthritis, premenstrual tension and all those patients with medically unexplained symptoms which take up so much time in specialist clinics, and even more time, I suspect, in primary care.

REFERENCES

1 Cochrane Library. Issue 2. Oxford: Update Software; 2000.

2 Carson AJ, Ringbauer B, Stone J *et al.* Do medically unexplained symptoms matter? A prospective cohort study of 300 new referrals to neurology outpatient clinics. *J Neurol Neurosurg Psychiatry* 2000; **68:**207-10.

3 Al-Shahi R, Warlow CP. Using patient-identifiable data for observational research and audit. *BMJ* 2000; **321:**1031-2.

4 Chalmers I. Minimising harm and maximising benefit during innovation in health care: controlled or uncontrolled experimentation? *BIRTH* 1986; **13:**155-64.

5 Al-Shahi R, Warlow CP. Ethical review of a multicentre study in Scotland: a weighty problem. *J R Coll Physicians Lond* 1999; **33:**549-52.

6 Dorman P, Counsell C, Sandercock P. Reports of randomised trials in acute stroke 1955-1995: what proportion were commercially sponsored? *Stroke* 1999; **30:**1995-8.

7 Chen ZM, Sandercock PAG, Pan HC *et al*. Indications for early aspirin use in acute ischaemic stroke: a combined analysis of 40,000 randomised patients from the Chinese Acute Stroke Trial and the International Stroke Trial. *Stroke* 2000; **31:**1240-9.

8 National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995; **333:**1581–7.

9 Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke (Cochrane Review). In: The Cochrane Library. Issue 2. Oxford: Update Software; 2001.

10 *Training in Academic Medicine*. Recommendations from the Academic Medicine Committee of the Royal College of Physicians. London: Royal College of Physicians; 2000.

11 Weir N and Dennis MS on behalf of The Scottish Stroke Outcomes Study Group. Towards a national system for monitoring the quality of hospital-based stroke services. *Stroke* 2001; **32:**1415-21.

12 Rothwell PM, Warlow CP on behalf of the European Carotid Surgery Trialists' Collaborative Group. Interpretation of operative risks of individual surgeons. *Lancet* 1999; **346:**1325.

13 Rothwell PM, Warlow CP. Is self-audit reliable? *Lancet* 1995; **346:**1623.

SHARP END EXPERIENCE

J.A. Grant, Head Clinician, Perth and Kinross Health Care Cooperative, St. Margaret's Health Centre, Auchterarder

If an accountable and comprehensive framework of clinical effectiveness is to be developed in this country, then primary care must be fully involved and committed. If it is not, then the overall standards of health care in the United Kingdom (UK) will not improve. More than 90% of all medical care in this country occurs in the community. As the time patients spend in acute hospital beds gets shorter, this percentage will inevitably grow rather than diminish.

Primary care in the UK is not a corporate body of equal members doing equal jobs in equal surroundings; it is a diverse collection of health care delivery systems which differ widely in scope and effectiveness. Within the UK approximately 45,000 independent medical contractors and their health care teams deliver medical care. Their circumstances vary from the traditional single handed practitioner to large multi-partner practices working within huge out of hours cooperatives. Some modern purpose-built primary health care premises can now provide in-house diagnostic facilities including ultrasound and endoscopy, while many rural area practices with community hospitals still deal with significant amounts of in-patient care, including maternity services. Within this spectrum of primary care models there is almost the same variation to be found in medical opinion and practice.

The latest NHS reorganisation has attempted to bring general practitioners (GPs) together in local health care cooperatives (LHCCs) on the basis principally of their geographical location rather than their types of practice. Local health care cooperatives have the potential to improve the clinical effectiveness of primary care provided they are supported, resourced and given time to effect change.

I would like to consider some aspects of clinical effectiveness and primary care within the following framework:

1. The dimensions of primary care
2. Practical difficulties in diagnosis and treatment
3. To refer or not to refer?
4. The whole patient

1. THE DIMENSIONS OF PRIMARY CARE

Approximately 70–80% of the members of a practice consult their doctor once per year; 20% will do so with an illness that is classified by ICD9 as being of a serious nature. Ten per cent of consultations will occur at home, 50% will be at the beginning or end of life. The average GP in the UK will have approximately three consultations and make something less than one visit per person, on his or her list each year.[1]

Against this background of constant demand, an urban practitioner will have to develop skills and coping strategies to deal with the problems of social deprivation and drug abuse, while the GP in a rural area may require to maintain skills in intrapartum obstetrics and community hospital intermediate care. Though the core skills of the generalist are similar, the context of each practice can significantly alter the dimensions of the primary care services the practitioner has to provide.

The contextual framework of general practice will always be determined to a major degree by geography, demography and social circumstance. However, in the future the dimensions of primary care will be defined to some extent by the outcome of the current debate as to what should or should not be done within the community as opposed to the acute hospital. There is little doubt that the issue of clinical effectiveness will be central to determining the future breadth and depth of these dimensions.

2 PRACTICAL DIFFICULTIES IN DIAGNOSIS AND TREATMENT

As the doctor of first contact, the GP faces special challenges and risks in his or her clinical practice. He or she is aware that in a typical surgery a small number of patients may have an acute problem, many will have chronic problems, while many others will remain completely undifferentiated in terms of diagnosis even after they have been seen and assessed. Unfortunately, patients do not come in to the surgery with an appropriate label. The practitioner must work quickly as the average consultation, at seven minutes, is not long. The use of time to find out how a situation resolves or develops, provided there is no clinical risk, is an essential skill.

Patient presentations are often oblique and fragmented: the sore back in the young woman who has major sexual problems; the painful neck secondary to muscular tension, which is masking an underlying depressive illness; the persistent cough, among dozens of others, that is the first sign of bronchial carcinoma. All must be dealt with and all must be diagnosed and managed. The range of investigations to help the diagnostic process in primary care is by no means comprehensive. The lab or the CT scanner is not down the corridor!

There is now increasing pressure being put upon GPs and their teams by patients and regulatory bodies to have more accountability and to demonstrate evidence-based medicine is being practised. Such pressures can make for more and more defensive practice with all the ramifications that such practice inevitably brings.

The diagnostic process is, in subtle ways, different in primary care. The GP cannot simply use the medical model paradigm as taught by medical schools. He/she has to learn to function with a far more complex bio psycho-social model where the influences of work, family and the whole psychological and social situation of the patient are often far more relevant than the knowledge his or her cholesterol is 5·4 or 7·2!

Let me give you an example of what I mean. Standards are set by government for primary care in terms of recording a patient's blood pressure, weight, cholesterol, smoking habits etc. – so-called process measures by which governments can assess, or so they say, 'the quality' of primary care. All GPs know that these measures are deeply flawed and probably have little to do with quality in terms of what is really important, namely the doctor/patient relationship and what goes on within the consultation.

For several years I had a patient who I had tried to persuade to refrain from smoking and attend the surgery for inadequate blood pressure control and hypercholesterolaemia. One day I met up with him socially and asked him why he had defaulted on so many occasions. He looked at me and said: 'My son has a serious heroin habit and is probably not going to survive. I would give my right arm if he would start smoking tobacco rather than take heroin and you want me to come into the surgery and talk about my cholesterol.' In today's society, this unfortunately is the lot of the many, not just the few. General practice has the very sobering habit of daily confronting you with the harsh realities of life. Our protocols, standards and outcomes are worthy products of the rigorous academic process but can pale into insignificance against the reused needle or the stale breath of the 19-year-old unemployed drug addict.

3. THE REFERRAL PROCESS

The referral process is one of the most complex aspects of GP behaviour and has been the subject of much research. Published referral figures range from less than 0·5 to more than 15 referrals per 100 consultations and from 0·6% to 25% of patients on a practice list a year,[2] a huge range and diversity. Unfortunately, most studies are limited by the quantitative basis on which they were conducted, thus failing to provide a complete understanding of the reasons behind these variations. Coulter *et al*.[3] conducted a questionnaire-based study on 127 GPs referrals to various hospital specialists. In total, they looked at over 18,000 referrals and pre-formulated seven reasons for referrals, including specific investigation, diagnostic uncertainty, treatment, operation or management advice, patient pressure etc. Though one of the largest published studies, like nearly all published studies it does not address the question of why the variations or the complex reasons why GPs make particular referrals. For instance, 'I referred this patient because the last one I had with the symptoms died unexpectedly' or 'I referred this patient because I was tired and didn't want to argue.' These are some of the actualities of general practice which make standard scientific methods difficult to apply and interpret. As with so much of general practice, the reality is very much more than the simple scientific method can expose.

Figures in themselves tell you little, especially when talking about clinical effectiveness. Though primary care is often blamed for an apparently increasing number of referrals to secondary care, such blame is often, though not always, based on ignorance and a failure to understand the reality behind the facts.

We are products of our training, experience, tolerance of uncertainty, sense of autonomy and personal values. We work in widely different situations with different pressures. All these characteristics operate through a final common pathway – a doctor's referral threshold. Effectiveness as a doctor is not inversely or directly proportional to the number of referral letters he or she writes.

4. THE WHOLE PATIENT

Just as the specialist deals with a specific area or disease process within part of a patient, the generalist or primary care physician deals with the whole patient. Thomas Kuhn, in his influential book *The Structure of Scientific Revolutions*, refers to a set of received beliefs in science as a paradigm. Accumulative research that follows the acceptance of a paradigm is called normal science, research normal to science which he describes as 'A strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education.'[4] The biomedical model is such a paradigm, i.e. patients suffer from diseases which can be categorised in the same way as natural phenomena. Many, though hopefully a declining number, still think that people suffering from such a disease can be viewed independently and separately from their social context. Many, understandably, also think that a physician's main task is to diagnose and to treat. This works for certain categories and conditions but in many others, especially in general practice, it encounters anomalies which cannot be ignored.

The old medical model paradigm has never been a good fit with family practice, and for this reason many GPs only partially accept it. Possibly this is one reason why the guideline production process is so firmly rooted in the medical model and why the quantitative scientific method has had difficulties in gaining widespread acceptance in primary care.

SUMMARY

The underpinning paradox of general practice is that its independence and diversity is both its greatest strength and its greatest weakness. It allows the patient the choice of the physician they can best relate to and trust. It can provide the continuity of care in a doctor/patient relationship which is the very centre around which all that is best in primary care is built. It also allows a doctor the scope to develop his or her own clinical practice in a relatively unregulated manner, governed principally by his/her own professionalism and the demands of his/her practice populations.

Such personal interactions and inputs do not readily lend themselves to external regulation, yet most thinking doctors believe that this is essential if our discipline is to move forward. But how is this to be achieved? The range of problems presented is often obscure and fragmented. The processes involved in managing these problems can be extremely complex and depend on many factors. The pressures of inadequate consultation times and ever increasing patient demand and expectation are unremitting. I have tried to illustrate some of this complexity in terms of the problems of primary care diagnosis, treatment and referral.

In a recent qualitative study exploring GPs' perceptions of effective health care, it was clearly identified that patient factors were the main reason for not practising effectively. Other reasons given were lack of time, lack of knowledge, lack of resources and 'human failings'. The authors suggested that the provision of effective health care in general practice requires a broader vision and a more pragmatic approach.[5]

Systems of clinical effectiveness in primary care must be measured against the whole person approach. Primary care presents very special challenges in terms of the methods being used to set and establish more effective clinical care within the NHS. To establish robust systems of clinical effectiveness within primary care, it is important to acknowledge these difficulties and to attempt to address them. Methods must therefore be tested and re-tested against the realities of clinical practice if they are to achieve the necessary widespread acceptability and reliability.

REFERENCES

1 McWhinney IR. *A Textbook of Family Medicine.* Oxford: Oxford University Press; 1989.

2 Wilkin D. Patterns of referral: explaining variations. In: *Hospital referrals*. Roland H, Coulter A. Oxford: Oxford University Press; 1992; 76-91.

3 Coulter A, Noone A, Goldacre M. General practioners' referrals to specialist outpatients clinics. *BMJ* 1989; **299:**304-8.

4 Kuhn T. *The structure of scientific revolutions*. Chicago: University of Chicago; 1967.

5 Tomlin Z, Humphrey C, Rogers S. General Practitioners' perceptions of effective health care. *BMJ* 1999; 318:1532-5.

ECONOMICS-BASED MEDICINE: AN EVOLVING PARADIGM

A. Maynard, Director of York Health Policy Group, Department of Health Studies, The University of York

The last decade has seen the emergence of evidence-based medicine.[1] Whilst the Cochrane Collaboration and health service researchers worldwide have invested heavily in systematic reviews and randomised clinical trials, economists have continued to insist that a myopic focus on benefits alone is an incomplete basis for making decisions about where to invest society's scarce health care resources.[2] They continue to argue that if the goal of society is to focus resources so that population health gain is maximised from a given and limited budget, then choices have to be based on the value of what society gives up when an intervention is funded (the cost) and the value of what society gains (the benefit or health gain). They argue that what is clinically effective may not be cost effective, but what is cost effective is *always* cost effective. Let us explore these issues.

THE UBIQUITOUS NATURE OF RATIONING

There are two certainties in life: death and scarcity. Resources are always and everywhere scarce. The individual, the group and the nation face similar constraints: demands are practically infinite and resources are limited. As a result of scarcity, decision makers have to make difficult choices and prioritise those activities which best meet their needs and are affordable within the available budget. The issue confronted by all decision makers is not *whether* to ration but *how*.

Rationing takes place when such choices are made. Rationing occurs 'when someone is denied (or simply not offered) an intervention which everyone agrees would do them some good and which they would like to have'. In the majority of cases patients who are ill wish to have interventions which doctors tell them may be beneficial. Occasionally they may refuse care when the benefits are marginal and the intervention offered is grossly unpleasant (e.g. 'commando surgery' for cancer patients). With most people confronting health problems which are allegedly amenable to medical interventions over their life, the demand for health care is apparently infinite and certainly exceeds the capacity of society to fund it from NHS and private sources. Thus the question which has to be addressed is, as is set out so nicely in the Yom Kippur (Day of Atonement) prayer book: 'who shall live and who shall die, who shall fulfil his days and shall die before his time?'

The medical profession is appointed by society as the agency whose members have to make these difficult choices. What data do they need to inform their choices? Even if they have the data, what incentives will ensure that their choices comply with what is best for the patient and society?

THE PRINCIPLES OF RATIONING

The dominant ethic in all health care systems is that of the physician who is trained to do his or her best for the individual patient. Thus, if there are two competing interventions, X and Y, available to treat a condition and the former produces five years of good quality life and the latter produces ten years of good quality life (or quality adjusted life years (QALYs)), the doctor will choose Y which gives the greatest benefit to the patient.

However, if X costs £100 and Y costs £10,000, the cost per QALY for X is £20 and the cost per QALY for Y £1,000. Consequently, if the Health Board had a budget of £500,000 for the treatment of patients with this condition, the physician who chooses treatment Y (which is clinically the most effective) would generate less health gain for the community (5,000 QALYs) than if treatment X (which is cost effective) were used (25,000 QALYs).

The choice of rationing criteria is between that which flows from the individual ethic (and the Hippocratic oath) and that which flows from utilitarianism, which prioritises the social ethic. This clash between the individual and the social ethic can confuse the policy debate considerably. Thus, practice guidelines produced by systematic review and expert consensus of trial data about clinical effectiveness may, if they affect practitioners' behaviour, lead to resource allocation and access criteria which are inefficient and fail to maximise population health gains from a limited budget. Such guidelines, and the behaviour they induce, are inefficient and waste society's scarce resources. Furthermore, they are unethical, because inefficiency deprives patients waiting for care interventions which would improve their health.

Archie Cochrane set out clearly the basis for economics-based medicine (the proper kind of EBM!):

> Allocations of funds and facilities are nearly always based on the opinions of senior consultants, but, more and more, requests for additional facilities will have to be based on detailed argument with "hard evidence" as to the gain to be expected from the patients' angle and the cost. Few could possibly object to this.[3]

Unfortunately, 30 years later, many reject this position, often not explicitly but by their behaviour. The laments about 'underfunding' cannot be supported in a world where medical practice variations are considerable, where clinical knowledge is not translated into practice and the economic perspective lost by, for instance, Royal Colleges when formulating practice advice. Cochrane would, without doubt, criticise vigorously this clinical myopia if he were alive today!

But is this narrow economic perspective adequate? Economics-based medicine, as set out here, requires us to identify, measure and value the cost and health benefits of competing medical interventions and rank them in a 'league table' in terms of their cost QALY characteristics. The given budget should then be allocated by working down the league table in relation to which are best value and the epidemiological base until the funding is exhausted. In

this way resources would be targeted at those interventions which give the 'biggest health bang for the health care buck' and population health gains from the limited budget would be maximised.

What if the recipients are mostly old or middle class? Society may wish to discriminate in favour of the young and the disadvantaged. The social consensus may be that resources should be taken away from the old and rich and given to the young and poor, even if such decisions were inefficient. In effect, such decision-making is based on the belief that QALYS received by the poor and young are of higher value than those received by the old and rich. Thus Williams has developed the 'fair innings' argument.[4] He argues that society may wish to discriminate against the old who have had a fair innings, even if there are efficient treatments available for them, so as to redistribute resources and care to the young (a person, for example, with multiple sclerosis) who have not had a fair innings.

The NICE issue with such arguments is: what are the distributional issues which society values; and how much redistribution should take place? It is obvious that society is not interested in efficiency alone; we spend resources on low birth weight babies because we value young life highly and in spite of the fact that such expenditure is often inefficient.

THE PRACTICE OF RATIONING

Not only is it difficult to achieve a consensus about the principles by which society should allocate its scarce health care resources, it is also difficult to alter clinical practice.

For decades, clinical practice has been based on trust. Society allocated resources to the NHS and left the professions to set standards and police practice. The policing of practice has been based on general principles inculcated at medical school and reinforced by the Royal Colleges and the General Medical Council. The emphasis on the analysis of practice has been slight, with most of these agencies having little capacity to monitor individual practice standards. Following the recent clinical 'scandals', there has been great emphasis on investment in clinical governance with many policy makers advocating structures and processes of unknown cost and effectiveness. This rush for a 'quick fix' raises issues identical to those associated with the principles of rationing: what are the costs and benefits of interventions and why is there an absence of economics-based policy making in this area, too?

As a consequence of this, routine activity data collected by the NHS for decades has not been owned and validated by practitioners, let alone used to monitor practice by individuals or the public guardians of good practice. The problems of translating evidence into practice in terms of particular therapeutic interventions are well defined. Sometimes change can be induced by use of financial incentives; one example is the payment of a fee of £6·40 to general practitioners to vaccinate the elderly and other high risk groups against influenza, which proved very successful during the winter of 2000–1, with coverage rates doubling, and more in some cases.

The challenges are to establish economics-based standards of clinical practice and governance and then to monitor and police the behaviour of practitioners. This will have to be led by clinicians, but if there is a failure of leadership there is an obvious risk that such roles will be usurped by others. Consequently, it is important that these challenges are met sooner rather than later.

Furthermore, the application of economics-based medicine requires the development of criteria of expected compliance and the sampling of practice to identify compliance and deviance. Once again, this will be resource intensive and it will be necessary to pilot alternative approaches and develop an evidence base.

RATIONING HEALTH CARE: AN OVERVIEW

Only a knave or a fool would argue that rationing health care is avoidable. Despite this, there are many knaves and fools who, for political and other reasons, assert that easy solutions are within our grasp. However, even if we had the same defence policy as Costa Rica with no armed forces, there would still be a gap between what is needed and what can be afforded: the Americans spend twice as much *per capita* as the UK and have rationing of access to care.

As a consequence, it is necessary to achieve a consensus about the principles of rationing. Such a consensus is the product of social values, one of which might be the efficient use of resources. This economic paradigm is central to the rationing processes as it makes explicit what the goals are, and what the opportunity costs are of goals, other than efficiency. Such explicitness makes decision-makers accountable.

It is said that the comedian Groucho Marx asserted, 'the secret of life is honesty and fair play. If you can fake it that you have made it'! In the processes that determine patient access to care there is apparent honesty and fair play but, after careful scrutiny, it is clear there is also much faking. Medical practitioners assert an evidence base for their interventions which is often absent in terms of clinical effectiveness, let alone cost effectiveness. They dominate the rationing process with incomplete medical knowledge and the assertion that their perspective, clinical effectiveness, should determine who will live and who will die.

Economics-based medicine challenges this paradigm by requiring that explicit choices are made between clinical effectiveness, efficiency and equity as the determinants of who will get access to care. Whichever principles are adopted, and increasingly it seems that the economic perspective is beginning to dominate, there is a need for a collaborative effort to develop systematically the evidence base.

Voltaire argued that 'the role was to amuse the patient whilst nature took its course'. The doctor's role is more than this but the clinical perspective alone is an inadequate basis to ration society's scarce resources.

REFERENCES

1 Maynard A, Chalmers I, editors. *Non-random reflections on Health Services Research. On the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Publishing; 1997.

2 Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. *Lancet* 1997; **349:**126-8.

3 Cochrane AL. *Effectiveness and Efficiency: random reflections on health services*. London: Nuffield provincial Hospitals trust.

4 Williams A. Rationing health care by age: the case for. *BMJ* 1997; **314:**820-2.

THE RELATIONSHIP OF THE HEALTH TECHNOLOGY BOARD TO THE SCIENCE OF CLINICAL EFFECTIVENESS IN SCOTLAND

A.V.P. Mackay, Chairman, Health Technology Board for Scotland, Glasgow

The Health Technology Board for Scotland, previously known as the Scottish Health Technology Assessment Centre, was born in April 2000 after a fairly long, measured and systematic gestation. Its job will be health technology assessment, that being a comprehensive, systematic evaluation of the consequences of the decision to use a health technology. This includes evaluating relative alternative technologies and thoroughly evaluating clinical and cost effectiveness to provide advice about value for money.

It is high time for health technology assessment to be taken seriously; indeed, it could be seen as long overdue. In the early years of expansion of the pharmaceutical industry in the 1950s and 1960s, a new drug could be introduced to the NHS with no fuss and a minimum of scrutiny. Thalidomide changed all that. In 1970, the Committee on the Safety of Medicines was established under Article 4 of the Medicines Act (1968), charged with advising the Secretary of State for Health not only on matters of safety, but also on the quality and efficacy of all new substances or articles to which the Medicines Act applied.

The definition of efficacy is important; it is the benefit observed under strictly enforced clinical trial conditions – in contrast to effectiveness, which is the benefit observed in ordinary, everyday clinical practice.

Quality, safety and efficacy represent a crucial set of three hurdles over which all new agents have to pass before being granted a Product Licence and Marketing Authorisation. Over the past 20 years prescribers have been spoiled for choice – able to choose from a wide range of safe and efficacious agents, several representing real quantum advances in therapeutics, many more representing 'me too' variants on a theme.

The 1990s have seen the drugs bill for the NHS rise to proportions which, even in a reasonably well funded NHS, represent a possibly unacceptable opportunity cost. For example, the recent publication in *Health Bulletin* by Cameron Stark and colleagues[1] shows startling figures for the escalating cost of antipsychotic drug prescription in primary care in Scotland. Over the period from April 1994 to December 1997 the quarterly prescription costs increased from just over £0·5 million to £1·3 million, an increase of over two and a half fold. Ninety per cent of the increase was accounted for by the use of new 'atypical' antipsychotics. The obvious question arises: is the net benefit derived from new atypical antipsychotics two and a half times greater than the older drugs? Almost as striking as the magnitude of this increase in cost was the variability between health boards across Scotland.

Any measure which could make even a small percentage impact on such enormous expenditure could free considerable resources for other uses. We have had blacklists and whitelists of drugs, the creation of area drug and therapeutics committees and the patchy introduction of local drug formularies, but these measures have been taken on a sometimes flimsy evidence base and applied inconsistently.

We have been elegantly 'muddling through' in a typically British way, trying to devise and apply non-Draconian ways of guiding prescribers to achieve best value for money, but without the required evidence on either clinical effectiveness (as distinct from efficacy) or costs. The general public hasn't realised this, and neither has most of the medical profession!

Almost inevitably, certain new, high profile and expensive treatments – beta-interferon and clorazil and the other atypical antipsychotics, and the so-called anti-dementia drugs – were to test the system, and frankly they found it wanting. Their introduction to prescribers has been confusing and the availability to patients inconsistent. In the case of the anti-dementia drugs, the medical press and the Health Board advisory machinery, not being in possession of all of the efficacy data, and certainly not in possession of effectiveness data, came to variable conclusions. There appeared to be *ad hoc* rationing and the catchy headline 'postcode prescribing' appeared for the first time. One of the issues to which the Health Technology Board is a response is the perceived 'right' of a patient to be given any drug which has a UK marketing authorisation and which his or her doctor feels he or she deserves.

This is a problem because it represents an obligation on a publicly-funded NHS to provide agents which may be neither effective in everyday clinical terms nor cost-effective. Of course, these issues are not confined to Scotland or the UK; they are worldwide.

South of the border the Government response was the creation of the National Institute for Clinical Excellence (NICE), an organisation with which the Health Technology Board will interact closely and thoughtfully.

The Scottish political initiative was signalled in the 1997 White Paper with the following aims: 'A Scottish centre should be established to "evaluate and provide advice to the NHS on cost-effectiveness of all innovations in health care, including new drugs"'.[2]

The term 'health technology' covers virtually the full spectrum of health interventions. The remit of the Health Technology Board is therefore very broad, but within it drug treatments will undoubtedly figure large. A health technology is defined as any intervention used to promote health; prevent, diagnose or treat disease; or provide rehabilitation or long-term care. This includes medicines, devices, clinical procedures and health care settings.

A fairly detailed implementation plan based on broad policital intentions was developed over the summer of 1999 and is encapsulated in the Implementation Paper published and circulated in September 1999.[3] The following is a summary of just a few of the main issues dealt with in the Implementation Paper.

Core aims and functions

Essentially, through its advice, the Health Technology Board gives an authoritative lead on judgements about clinical and cost effectiveness of new and existing technologies through which:

- new technologies which are clinically effective and cost effective should be introduced more smoothly into the NHS;
- consistency should improve; and
- existing, less effective technologies should be more readily abandoned.

Key aspects of this are:

- high quality advice from a single source;
- openness; and
- support for drug and therapeutics committees which, after all, will continue to be the main source of advice to prescribers, trusts and health boards for the majority of new agents.

The climate in which this advice will be given is vitally important; it is clinical governance.

Output and targeting

Advice will be the output, supported by an explanation of how and why it was arrived at and structured according to certain categories. Depending upon the type of technology, the target decision-makers will vary but the advice will always be widely disseminated. It must be emphasised that this will be advice on clinical and cost effectiveness and *not* on need, budgetary priority or affordability.

But isn't this really just rationing with a politically acceptable 'spin' and cloaked in the jargon of health economics? Arguably not, although there should be no difficulty with the term rationing, which simply represents the equitable distribution of limited resources. Health technology appraisal will certainly inform the process of rational and equitable distribution of resources, but it will *not* be about what a health care provider *should* afford or *can* afford, nor will it be about assessing population needs or prioritising these against available resources. It will be to give advice on *relative value for money* from health interventions.

Regardless of priorities and 'affordability', any reasonable person would surely not condone the provision of treatments which were either only equivocally beneficial or where the benefit per pound spent, or per hour of medical time taken from some other activity, was less than an alternative. Advising on the degree of real benefit and value for money will be an aid to the evidence-based, equitable distribution of limited resources but will not be 'stumbling into rationing'.[4] The Health Technology Board will therefore put a value on an intervention, it will not decide whether or not the NHS should provide it.

Viagra is a useful example with which to illustrate the point. Although not subject to clinical and cost effectiveness appraisal by NICE (its marketing authorisation predated NICE), I would wager that NICE's conclusion would have been that, on the basis of relative cost effectiveness, Viagra should be made freely available according to its licensed indications. It is impressively clinically effective and indeed arguably cost effective relative to alternative therapies. However, policy-makers would probably still have wanted to restrict its availability on the grounds of their priorities for health improvement of the population.

Many problems await the organisation when it opens for business. Some sources of misunderstanding have already been identified, but the following are some other difficult issues which will have to be tackled early in the life of the Health Technology Board.

Thorny issues

The first is the narrow minded and, dare one say, even bloody minded doctor who wants to exercise a right to prescribe as he or she sees fit – a thorny problem for which the Health Technology Board will not provide a quick fix. However, through its contribution to the growing climate of clinical governance, one anticipates that the Board will make a difference.

The second is the problem of rapidly acquiring useful data on clinical effectiveness and costs of a new drug given the relatively limited data set required for marketing authorisation to which I have already referred. It will not be uncommon for a new drug to be licensed for which information on effectiveness and financial costs is simply too limited for a clear decision on cost effectiveness. An idea which has been discussed in Scotland is that of the post licensing pragmatic trial (Figure 1).

A crucial element in this idea is the intention of the Board to work closely with the existing network of major area drug and therapeutics committees throughout Scotland and to use their local contacts and involvement as a vehicle on which to generate relatively robust and useful data in a short time. If, as the result of a careful assessment, the Board decides that clear advice is impossible due to data limitations, it is intended that in certain instances the organised collection of the required data would be overseen by a topic-specific project team. The team would comprise representatives from the Board's core staff, from the original topic-specific group, the sponsor(s) of the technology, the area drug and therapeutics committee network (in the case of a drug), and the Chief Scientist Office (CSO). Funding would be sought through the CSO and possibly other sources, including the technology sponsor(s). The purpose of the project team would be to oversee the enrolment of clinical, and perhaps economic, researchers who would be responsible for data collection on a national multi-centre pragmatic basis. Data on clinical effectiveness (including unwanted effects) would be recorded by clinicians actually using the technology. Outcome measures would have to be valid, simple and essentially those which the clinician would anyway be recording in case records, otherwise clinicians would be discouraged from enrolling as raters. For many disorders, it would seem sensible to look carefully at the use of patient-derived outcome measures quantified on a simple global impression scale. The development of a methodology for pragmatic multi-centre clinical trials will be the subject of early dialogue between the Board and the CSO, with the prospect of creating an entirely new approach to clinical research, one with a highly practical pay-off.

Health technology assessment has been a worldwide activity for many years but so far seems to have had relatively little real impact on decision-making by either individual

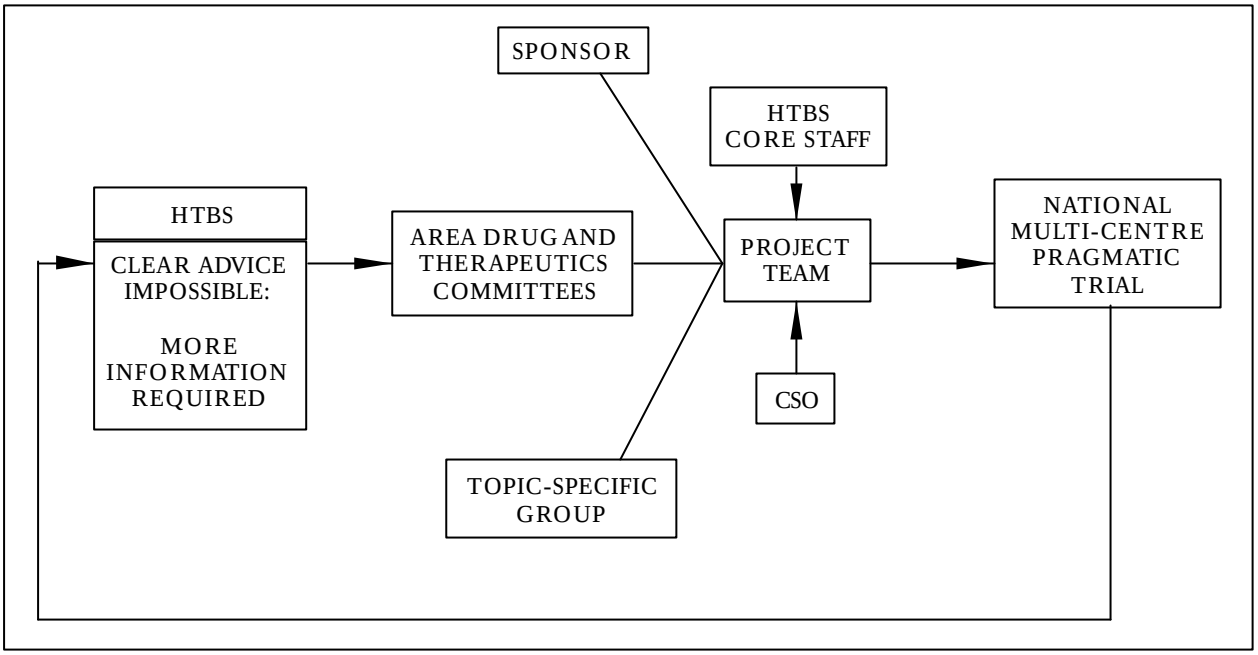

FIGURE 1
Possible model for Scottish multi-centre pragmatic trials.

prescribers or health service budget-holders. The timely creation of the Health Technology Board offers the promise of new methodologies and a useful engagement of the clinical community in the appraisal process. It could be the catalyst to the development of pragmatic clinical research in which Scotland led the international field.

ACKNOWLEDGEMENTS
The help of Doreen Pedlar with the production of this paper is gratefully acknowledged.

REFERENCES

1 Stark C, Jones J, Agnew J *et al*. Antipsychotic drug prescribing trends in primary care in Scotland 1994–1997. *Health Bulletin* 2000; **58 (2):**96-101.
2 White Paper. *Designed to Care: Renewing the National Health Service in Scotland.* December 1997.
3 Scottish Executive. *A Scottish Health Technology Assessment Centre*. Report of the Implementation Working Group. 2000.
4 Smith R. Stumbling into rationing. *BMJ*; **319:**936.

INDUSTRIAL EXPERIENCE AND CASE STUDIES OF INTELLIGENT DECISION SUPPORT SYSTEMS

J.R. McDonald, Rolls-Royce Chair in Power Engineering; S.D.J. McArthur, Lecturer, Centre for Electrical Power Engineering, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow

INTRODUCTION

Industry and commerce require new and advanced support systems to cope with ever increasing data volumes. There is now a climate of mass electronic data storage, leading to organisations becoming 'data rich, but information poor'.

To combat this, intelligent system technology is being used to automate the interpretation of such data sets. The result is that the data is rendered as useful information which can inform decision-making and remove the data analysis burden from individuals. Additionally, the next step can be taken where the decision itself can be automated using intelligent system (IS) technology. Knowledge management underpins this technology.

1. INTRODUCTION TO INTELLIGENT SYSTEMS

Intelligent system technology is a collection of approaches and methods which can be applied to data interpretation. They offer the ability to interpret large volumes of data quickly. The following sections describe three well known IS methods.

A *knowledge based system* (KBS) is based upon codified human knowledge and comprises four key components: a rulebase; an inference engine; a database; and a user interface. These are shown in Figure 1. The rulebase stores the knowledge, typically as rules or heuristics. The database gathers the real-world data used by the rulebase. The inference engine is the 'intelligent processing engine'. It uses the knowledge and database to reason about what has happened. It delivers its conclusions via the user interface.

Case based reasoning is underpinned by another approach to problem solving. It is based upon reasoning by analogy or example. Fundamentally, when it is given a new problem to deal with it will attempt to match it to the closest example(s) or nearest neighbour(s) it has dealt with in the past. These previous cases are stored in a case base. Cases contain all the information relating to a previous problem *plus* the solution or conclusion. An adaptation procedure is used to adapt the solution to the closest matching known case in order to solve the present problem. Finally, the new case, plus its solution, is entered into the case base. As a result, this approach is seen to exhibit machine learning characteristics. Figure 2 demonstrates this architecture.

Model based reasoning uses detailed functional, physical, mathematical or logical models as a basis for the reasoning. In particular, model based diagnosis (MBD) is useful for data interpretation. Model based diagnosis systems utilise models of correct behaviour to predict how the device being diagnosed should have operated. A comparison is made between the observed (actual behaviour of the system being diagnosed) and the predicted behaviour. A difference in the observed and predicted behaviour is indicative of a device failure.

Each of the above techniques is based upon accurate capture and representation of human knowledge. As a result, a rigorous and controlled knowledge capture process must be undertaken.

2. KNOWLEDGE CAPTURE AND MANAGEMENT

A key aspect of building intelligent decision support systems, based upon the technologies described in Section 1 above, is the elicitation, validation and modelling of tacit knowledge and expertise. Tacit knowledge is that which experts and experienced practitioners have never recorded or documented. It comprises their core expertise, rules of thumb and experiences. This knowledge must be captured and codified.

The KADS (knowledge acquisition and design structuring) methodology can be used to capture the tacit knowledge required for intelligent decision support

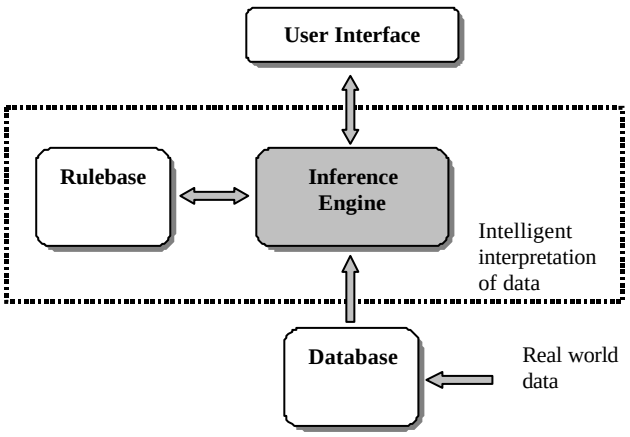

FIGURE 1
Knowledge based system architecture.

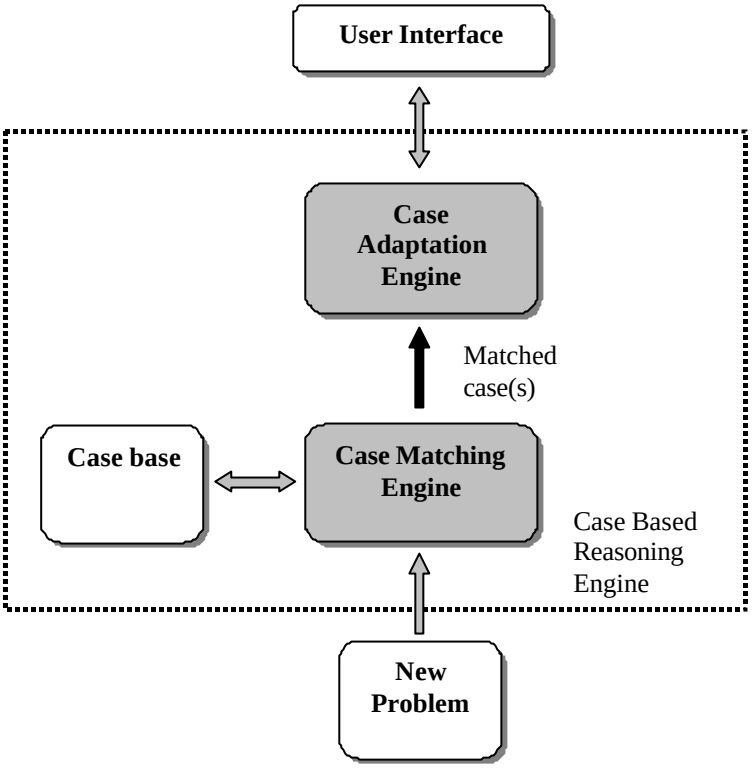

FIGURE 2
Case based reasoning architecture.

systems.[1] Knowledge acquisition and design structuring is a formal and standardised method of capturing knowledge and representing it within models. Research within the case studies described later has also enhanced KADS by adding a more rigid knowledge elicitation process to the activity.

There are five key activities within the knowledge capture process:

- knowledge elicitation and transcription;
- knowledge modelling;
- knowledge validation;
- knowledge implementation; and
- knowledge maintenance.

2.1 *Knowledge elicitation and transcription*
Formal knowledge elicitation meetings are held with experts. These are formally structured, detailed discussions concerning previous case studies, events and experiences. Knowledge transcripts are generated from each meeting to detail the captured expertise. These are validated by the experts and their peers.

2.2 *Knowledge modelling*
The KADS 'model of expertise' is used to represent the knowledge in a graphical format (e.g. as a set of semantic, object based, graphical relationships). This describes exactly how different types of knowledge are used within each stage of the problem solving process. The model of expertise breaks the knowledge into three distinct layers:

- task layer: knowledge of the problem solving tasks and activities undertaken by the expert;
- inference layer: a model of the inferences and decisions made within each of the tasks identified; this layer also represents which domain knowledge feeds into each decision and inference; and
- domain layer: provides semantic networks representing core domain knowledge, cases and rules.

The experts and their peers also validate the knowledge models.

2.3 *Knowledge validation*
The validation of knowledge is critical to ensure accurate reasoning within the IS. As a result it is validated at two points in the knowledge capture process: once the transcripts are generated and once the models have been created.

2.4 *Knowledge implementation*
The validated knowledge models provide the specification for the decision support system to be built. At this stage the type of reasoning to be applied can be determined, e.g. knowledge based, model based, case based, etc.

2.5 *Knowledge maintenance*
Decision support systems, and knowledge, are not static. Therefore, it is important that a process of knowledge maintenance is continuous. New knowledge must always be elicited, validated, modelled and implemented.

The complete knowledge capture cycle is shown in Figure 3. While it is extremely important to use this process in order to build intelligent decision support systems, it also has value in terms of knowledge management. It allows organisations and companies to ensure that valuable expertise is not lost when individuals retire or move on to

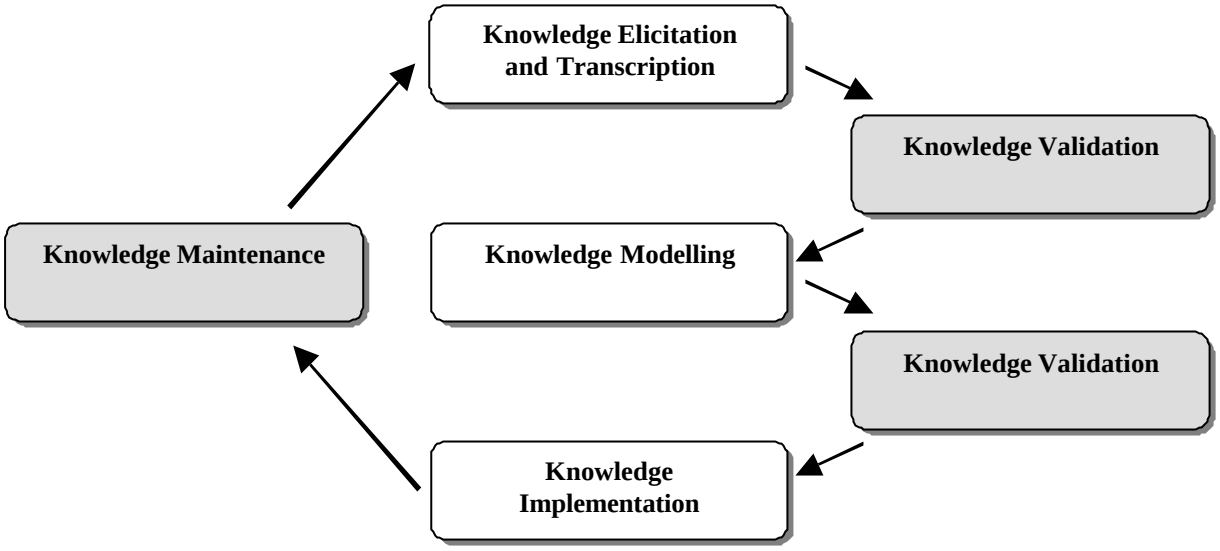

FIGURE 3
Knowledge capture life cycle.

another organisation. Therefore, the knowledge can be used for training and support.

3. CASE STUDIES OF DECISION SUPPORT SYSTEMS

3.1 *ScottishPower protection decision support system*

ScottishPower is responsible for all of the electrical network within the central and southern areas of Scotland. One of its prime activities is ensuring that protection systems operate correctly. Protection systems ensure that when a fault occurs on a section of an electrical network then it is de-energised. This prevents physical harm to humans and expensive damage to the electrical plant.

ScottishPower's problem is the volume of data gathered automatically which has to be interpreted to decide whether there are any problems with protection devices. The supervisory, control and data acquisition (SCADA) system provides a continual stream of alarms indicating what is happening in the electrical network. During storms, or significant faults, the volume generated can be in the order of tens of thousands of alarms per hour. In addition to this, other devices called fault recorders gather very detailed electrical information. Fault locators also provide information in the form of the exact location of a particular fault.

A decision support system was created to aid the engineers working with these data sets.[2,3] Their knowledge was elicited and structured within three intelligent modules:

- a knowledge based alarm processor, called APEX, was used to quickly process the SCADA system alarms and indicated what was happening;
- a knowledge based fault diagnostic expert system was used to identify the nature of the faults occurring; and
- a model based reasoning system used the detailed electrical data to determine if the protection devices were operating correctly.

The output from each intelligent module can be used to automatically generate a fault report for an engineer. The reports generated can quickly alert them to situations which they must resolve, while buffering them from the minute to minute data analysis task. The architecture of the system is shown in Figure 4.

3.2 *European blood and marrow transplant (EBMT) database*

Over the past several years the EBMT database has been developed by combining the individual databases from several research institutes comprising patients' records covering the past decade. The original aim of this was to create a single reference database for all the institutions, thereby sharing the knowledge of symptoms and treatments for various types of cancer. When all the individual databases were combined into one, the main problem was the dimensionality of the final database, for although there are only 3,610 records, each has 126 fields. To extract knowledge from the database, knowledge discovery in databases (KDD) techniques were applied. The preliminary aim of the KDD activity was to extract rules concerning the likelihood of success for different treatment regimes applied to a patient.[4] These could then be applied within a decision support system. The KDD approach can also be combined with the structured knowledge capture approach described previously.

KDD offers techniques such as rule induction and artificial neural networks. Rule induction allows heuristics to be automatically derived from a database, while neural networks allow complicated non-linear cause and effect relationships to be captured and re-used.

The research applied various data reduction, data cleansing, rule induction and neural network techniques to the original EBMT database. As a result, several useful rules and neural networks were developed, which will provide the basis for predicting the outcome in patients depending upon their characteristics.

3.3 *Decision support for nuclear reactor refuelling*

Evaluation of the data produced during the refuelling process in a nuclear power plant is required to ensure proper 'sit-down' of the fuel assembly, thereby allowing the continued and safe operation of the station. During

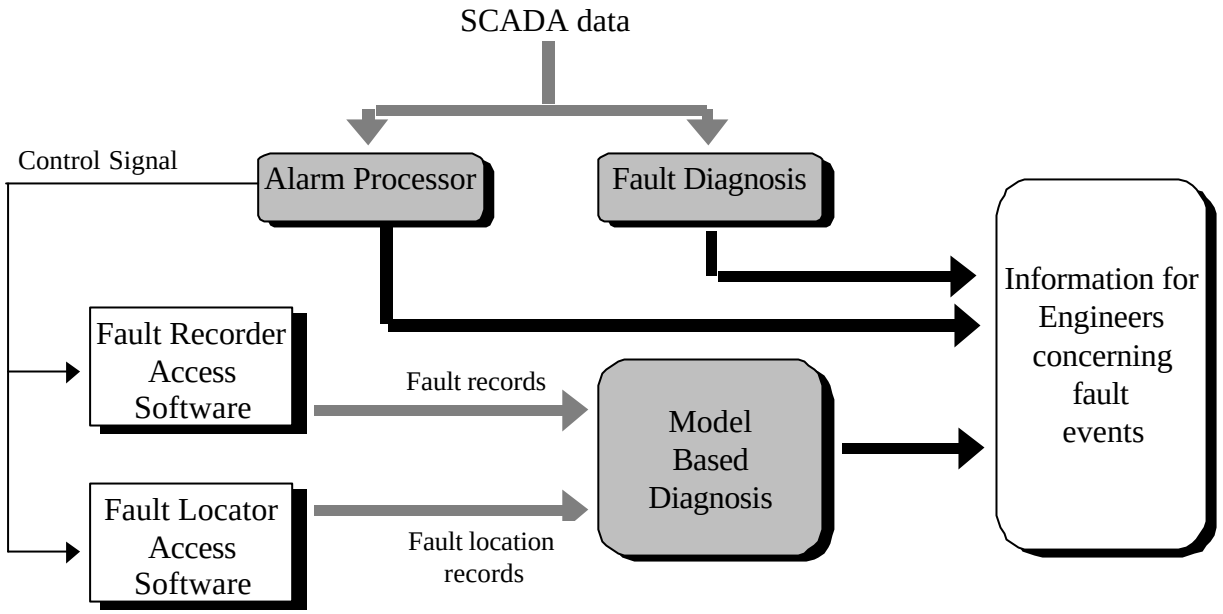

FIGURE 4
ScottishPower decision support system.

the refuelling process, measurements regarding the position of the grab, holding the fuel element, are recorded along with their time. Throughout this process the operator monitors these measurements to ensure that they are within given bounds. Once the process is completed, a senior experienced engineer inspects the measurements to ensure that the fuel assembly has settled properly into the fuel channel. This is time consuming due to the volume of data and requires considerable domain experience. If the refuelling has been successful, the operator can then repeat the refuelling process at another channel. However, if any errors have been encountered the new fuel assembly will have to be repositioned or removed from the fuel channel.

The knowledge from the experienced senior engineers was captured through the process described in Section 2. The resulting knowledge and case base was used for two purposes. Firstly, it fulfilled a knowledge management function by allowing the expertise to be available over an intranet to less experienced personnel. Secondly, the knowledge base was combined with neural networks and

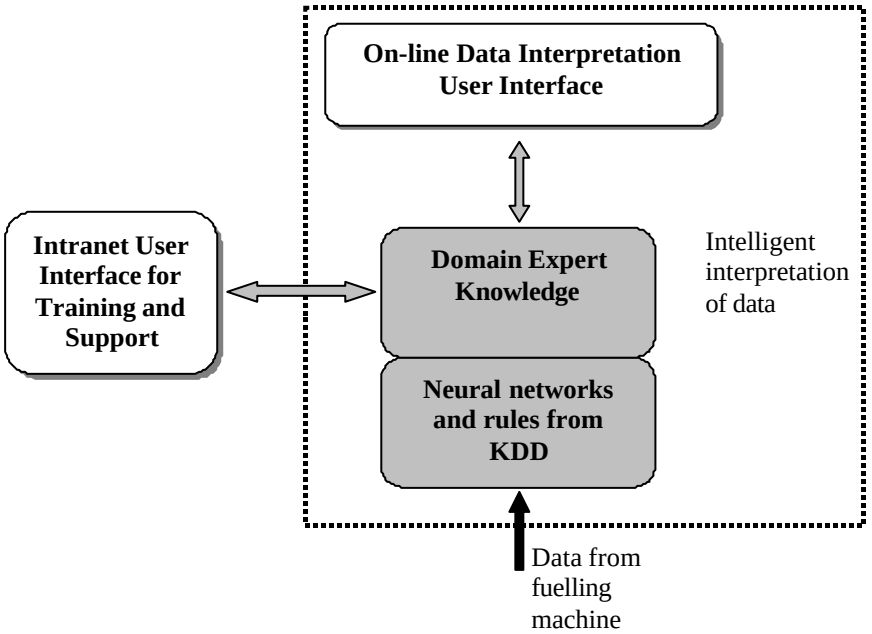

FIGURE 5
Nuclear reactor refuelling decision support system.

rules, generated through an extensive KDD exercise, to provide automated analysis of the data for the operators.[5] Figure 5 shows the architecture of this system.

CONCLUSIONS

This paper has provided an insight into the various intelligent system technologies which are able to provide decision support when data overload becomes a problem. Case study examples have been provided to demonstrate how they may be used. The capture of knowledge, through a structured methodology, has been shown to be key to the success of such systems and helps to create a knowledge management culture within organisations.

REFERENCES

1 Schreiber G *et al*. *Knowledge Engineering and Management: The Common KADS Methodology.* Boston: MIT Press; 2000.

2 Bell SC, McArthur SDJ, McDonald JR *et al*. Model based analysis of protection system performance. *IEE Proceedings – Generation, Transmission and Distribution* 1998; **145(5):**547-52.

3 Bell SC, McArthur SDJ, McDonald JR *et al*. On-line protection performance analysis using intelligent systems. *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications* (Special issue on AI methods in Power System Protection) 1997; **5(4):**235-43.

4 Steele JA, McArthur SDJ, McDonald JR *et al*. Knowledge discovery in medical databases: what factors influence a successful bone marrow transplant? *IEE Colloquium on Knowledge Discovery and Data Mining* 1998; digest no: **98/434:**3/1-3/8.

5 Steele JA, McArthur SDJ, McDonald JR *et al*. Towards an Estimation Aid for Nuclear Power Plant Refuelling Operations. *Proceedings of the 13th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE*. New Orleans, Louisiana, USA: June 2000; 56-66.

REDUCING ERRORS IN MEDICAL PRACTICE: BARRIERS TO MONITORING CLINICAL PERFORMANCE

G.D. Murray, Head of Public Health Sciences, Department of Community Health Sciences, University of Edinburgh

Reducing errors in medical practice, or in general avoiding sub-optimal performance, ought to be straightforward. One draws up a league table of clinicians using an appropriate outcome measure, one identifies any individuals who are under-performing and one offers them suitable support and additional training. Clearly, differences in case mix can invalidate a direct comparison of outcomes, but statisticians can adjust performance measures for case mix to ensure that one is comparing like with like.

If only life were so simple! There are numerous practical and technical difficulties in constructing case mix adjusted league tables. These difficulties are discussed using case studies including the Public Inquiry into Paediatric Cardiac Surgery at the Bristol Royal Infirmary. The issues concerned include the level at which performance is assessed (should one be assessing individual clinicians, or teams, or units, or hospitals, or health boards?), the selection of clinical areas to be assessed (e.g. would one assess the entire workload of a surgical unit, or concentrate on indicator procedures?), the quality of the data upon which the assessment is based, the choice of suitable outcome measures, the sensitivity of the statistical procedures to detect differences in performance and the robustness of the statistical models used in the case mix adjustment.

Even if these technical difficulties could be resolved, there are other profoundly difficult value judgements to be made. In particular, even if one had a perfect performance measure, one would still find that 50% of clinicians were 'below average'. How far below average can performance fall and still remain acceptable?

BACKGROUND

Since at least the 1850s there has been pressure to measure surgical performance in terms of clinical outcomes. The fact that such an activity depends crucially on careful statistical reasoning is perhaps reflected in the fact that in 1858 Florence Nightingale, one of the first proponents of surgical audit, was elected a Fellow of the Statistical Society of London (now the Royal Statistical Society). In the intervening years the debate on surgical audit has cycled through appraising performance in terms of structure, process and outcome. Each approach has advantages and disadvantages, but there is no doubt that in the early 2000s the spotlight is firmly focussed on assessing outcomes. The momentum that has built up behind 'clinical governance' goes far beyond surgery (e.g. the Shipman case), but the following discussion will focus largely on surgery as a (relatively) simple special case.

OUTCOME AUDIT MADE SIMPLE

At least conceptually, assessment of performance via clinical outcomes is straightforward. One chooses the level of assessment and the scope of the assessment, which should lead to an appropriate outcome measure. One then needs a mechanism to collect reliable data and must accumulate a sufficient number of cases for analysis. Given the accumulated data, one needs a statistical approach which can adjust for the effects of case mix and, finally, given the case mix adjusted measure of performance, one must refer to an appropriate norm. However, each step along this process raises difficult questions and often quite arbitrary decisions require to be made.

LEVEL OF ASSESSMENT

It seems obvious that one would want to measure the performance of individual clinicians, but generally clinicians do not work as individuals. For example, should a poor operative outcome be ascribed to the junior surgeon who undertook the procedure, or to the responsible consultant who should perhaps have been more directly involved in supervising the operation? What if in fact the post-operative death was largely the result of poor anaesthetic management, or a lack of adequate intensive care facilities in the recovery phase? What if the patient died because they were basically unfit for surgery as a direct result of their tumour not being diagnosed until it had reached a very advanced stage? Is this not a reflection of a breakdown in the 'pathway to care' rather than an indication of surgical incompetence? There are no simple answers to such questions, but they show that it is very difficult to isolate the performance of an individual clinician and suggest that one should be looking at the 'big picture,' from diagnosis and referral through to management, rather than focussing on one step along the pathway.

SCOPE OF ASSESSMENT

There is a basic conflict in choosing the scope of an assessment. If one assesses the entire workload of a clinician then this will be representative and will maximise the available information. However, the workload will be very heterogeneous. Can one mix major resections for colorectal cancer with minor surgery for ingrowing toenails? The relevant outcome measures would be very different and there will be no consistent prognostic factors to be used in case mix adjustment. The alternative is to focus on a limited number of indicator conditions or procedures. This has the complementary advantage of standardising the technical difficulty of the procedure, but unless a very common condition is chosen there might be too little information to gain any worthwhile indication of performance.

Stark *et al.*[1] give an example from paediatric cardiac surgery. Taking one year's data from five surgical units, there was a total of 501 operations from six indicator procedures, with ten post-operative deaths. One does not need to perform a complex power calculation to see that with such data it would be *impossible* to differentiate between the five surgical units, never mind between individual surgeons, on the basis of post-operative mortality rates.

OUTCOME MEASURE

The use of indicator procedures does allow one to choose an outcome measure which is appropriate to the context. For example, for major paediatric cardiac surgery, mortality might well be an appropriate outcome measure (always provided that the sample size is adequate), whereas for laparoscopic cholecystectomy it might be more appropriate to look at time to return to work or some measure of patient satisfaction or freedom from symptoms.

Another point which needs to be considered is the time scale. In the context of surgery for early stage colorectal cancer, one might want to assess ten year survival rates to get a true indication of the quality of the surgery. However, this is of limited value in audit if a surgeon needs to wait ten years plus to get an indication of their performance.

RELIABLE DATA

If the results of an audit are to be reliable then it is self-evident that the underlying data must be reliable. Indeed, with the increasing likelihood of such data being used in the context of litigation, it is crucial that data are not only reliable but can also be *shown* to be reliable. This is in many ways analogous to the data quality requirements for clinical trials which lead to drug registration. There are stringent procedures defined under 'Good Clinical Practice' which must be followed in order for the data to be robust. This has major resource implications, but there is little point in gathering poor data.

The quality of routine data sources was a major issue in the Bristol Royal Infirmary Inquiry, and part of the commissioned research was an assessment of data quality in the UK Cardiac Surgical Register. This was one of the key data resources available to the Inquiry, but there were major problems with the data quality.[2] Even with the very objective outcome measure of 30 day mortality there were major differences from centre to centre in how this was interpreted and reported. First, there were generally no systematic procedures in place to assess outcome at 30 days for children discharged before 30 days. Second, there was ambiguity over whether the '30 days' was from the first procedure, the 'definitive' procedure or the final procedure. Third, some units only reported deaths which were a direct result of the surgical procedure, rather than reporting all cause mortality. Fourth, some units reported deaths after 30 days if they were still prior to discharge. With such an easy definition raising major inconsistencies, it is hardly surprising that coding diagnoses and operative procedures were even more inconsistent.

CASE MIX ADJUSTMENT

There are many published scoring systems which can be used for case mix adjustment, such as APACHE[3] for intensive care and POSSUM[4] for general surgery. However, this area raises yet more difficult issues.[5] In POSSUM, for example, one of the covariates which is used in the case mix adjustment is blood loss. However, one could argue that blood loss reflects technical competence, and adjusting for this factor excuses incompetence rather than highlighting it.

ASSESSING ACCEPTABLE PERFORMANCE

Even if one could address all of the difficulties raised earlier, and devise a perfect case mix adjusted measure of performance, one is still faced with the very difficult judgement of what constitutes acceptable performance. Just as with every other profession, there is true variation in competence from clinician to clinician. Indeed, by definition, 50% of clinicians will be below average. Moreover, if we give these 50% of under-performing clinicians early retirement, then 50% of those who remain will still be below average! This fundamental truth needs to be recognised, and it can be helpful to think about the precise aim of any audit exercise. Are we following the traditional model of trying to inform individual clinicians of their performance so that they can fine-tune their work and gradually improve – 'closing the audit loop'? If so, then I believe that there are almost insuperable difficulties in basing such an audit on outcome. Such methods are instead more appropriate for identifying gross outliers, where performance is far outside the range of 'acceptable' individual to individual variation.

In their summary of the statistical evidence presented to the Bristol Royal Infirmary, Spiegelhalter *et al.*[6] concluded that for open heart surgery in children aged under one year, the mortality rate at Bristol was roughly double that at other comparable units in England. They also concluded that this constituted robust evidence that Bristol was an 'outlier' in terms of outcome. However, if the observed excess in mortality had been 50% rather than 100%, the authors commented that the many flaws in the data were such that this could not have been interpreted as robust evidence of excess mortality at Bristol. In other words, this enormously complex and expensive exercise, based primarily on data collected over five years (1991–5), lacked the sensitivity to detect major variation in performance which would have been of enormous clinical relevance.

My own opinion is that if we want to use audit to raise the performance of all clinicians rather than as a rather blunt tool for identifying gross outliers, then we should focus much more on process and audit compliance with agreed guidelines.

REFERENCES

1. Stark J, Gallivan S, Lovegrove J *et al*. Mortality rates after surgery for congenital heart defects in children and surgeons' performance. *Lancet* 2000; **355:**1004-7.
2. Lawrence AE, Murray GD. The UK Cardiac Surgical Register: Assessment of data quality issues for the Bristol Royal Infirmary Inquiry. Available from: URL http://www.bristol-inquiry.org.uk/
3. Knaus WA, Wagner DP, Draper EA *et al*. The APACHE III prognostic system. *Chest* 1991; **100:**1619-36.
4. Copeland GP, Jones D, Walter M. POSSUM: a scoring system for surgical audit. *British Journal of Surgery* 1991; **78**:356-60.
5. Hayes C, Murray GD. Case mix adjustment in comparative audit. *J Eval Clin Prac* 1995; **1:**105-11.
6. Spiegelhalter D, Evans SJW, Aylin P*et al*. Overview of statistical evidence presented to the Bristol Royal Infirmary Inquiry concerning the nature and outcomes of paediatric cardiac surgical services at Bristol relative to other specialist centres from 1984 to 1995. *Bristol Royal Infirmary Inquiry* 2000.

THE PUBLIC, THE PHYSICIAN AND THE PATIENT

I. Watt, Professor of Primary and Community Care in the Department of Health Studies, University of York

INTRODUCTION

In keeping with health policy in many other countries, there is increasing emphasis within the NHS on patients becoming more active participants in their own care. The goal is that relationships between doctors and the public should be based on more equal partnerships than they currently are, with people being given more information about their conditions and taking more active roles in decisions about their care when they want to do so. This paper seeks to consider: how the character of patient/physician interactions has changed as medical care has developed; how some of the developments outside individual physician contacts may have influenced the public's relationship with and view of physicians; and the implications of these issues on the future roles of physicians.

There are many classifications of patient/doctor interactions and for the most part they are broadly similar. An early classification is given below in Figure 1.

FIGURE 1
Patient/doctor interaction.

- activity/passivity model – the physician makes the decisions
- guidance/cooperation model – the physician provides instructions, the patient carries them out
- mutual participation model – the physician assists patients in helping themselves; the patient participates in decision-making

Szasz and Hollander[1]

Classifications of patient/doctor interactions broadly cover a spectrum of styles from what one might term paternal, where it is assumed that the doctor knows best, makes decisions on behalf of patients without consulting them and may feel threatened if patients have access to information from elsewhere. At the other end of the spectrum are models where the power relationship is more in favour of the patient with, at the extreme, consumerist models where the patient has very high control and the physician low control. The trend now, at least in terms of stated policy, is to move from the paternal end of the spectrum to models with a greater emphasis on partnership, but for the most part avoiding the extreme consumerist models.

CHANGES IN PATIENT/DOCTOR INTERACTIONS

In briefly considering how patient/doctor interactions have changed over the past 100 years, this section draws on material covered in more detail in Roy Porter's history of medicine *The Greatest Benefit to Mankind*.[2] Even though a consumerist approach to health care is often viewed as a modern phenomenon, one could argue that a type of consumerist model was commonly seen in the nineteenth century before the results of science gave doctors something useful to do. In the nineteenth century, bedside manner and social graces often counted for a great deal. Patients frequently had strong views about their illnesses and the treatment they needed, including, for example, bloodletting, purging and other ways of expelling bad humours. As Shaw remarked:

> . . . the doctor who has to live by pleasing his patients, soon finds himself prescribing water to teetotalers and brandy or champagne jelly to drunkards; beefsteaks and stout in one house, and 'uric acid free' vegetarian diet over the way; shut windows, big fires and heavy overcoats to old Colonels, and open air and as much nakedness as is compatible with decency to young fadists, never once daring to say either 'I don't know', or 'I don't agree'.

At the turn of that century there was little to combat major infections or serious conditions such as diabetes, arthritis, asthma or ischaemic heart disease. However, with the new century came new diagnostic techniques and so even if doctors could not treat effectively, they started to diagnose more accurately. The new diagnostic methods started to exert an impact on the consultation. Diagnostic techniques were being transformed and new aids such as the stethoscope and ophthalmoscope gave rise to new rituals. Diagnosis became all, as the quote below illustrates:

> I would answer that he is a great physician who, above other men, understands diagnosis. It is not he who promises to cure all maladies, who has a remedy ready for every symptom, or one remedy for all symptoms; who boasts that success never fails him, when his daily history gives the lie to such assertion.
>
> Dr Jacob Bigalow 1852

The doctor proficient at diagnosis came to be seen as 'scientific' and was often held in very high esteem. The new diagnostic science may also have allowed doctors to blind their patients with jargon on occasion. This scientific approach was often very popular with patients, perhaps because being examined made it seem as though they commanded the doctor's attention. In addition, science had a largely positive image and 'things scientific' were often popular in their own right. However, cures for many people still remained an unobtainable goal, although it was often difficult for doctors to admit this:

> . . . in some cases I knew, even in the beginning, that my efforts would be futile in the matter of rendering service to anyone . . . Of course, one left some medicine . . . this was largely the bunk, but someone had to pay for the axle grease, and just plain advice never was productive of revenue, unless fortified by a few pills.
>
> Hertzler 1900

Some doctors began to see, however, that even though medicines might not help, their own manner and psychological support might and the 'patient as a person' doctrine became influential in the decades after 1900. One of the great exponents of this doctrine was the physician William Osler:

> The good physician treats the disease but the great physician treats the patient.
>
> It is much more important to know what sort of patient has the disease than to know what sort of disease the patient has.

Throughout the twentieth century the efficacy of medicine, particularly after World War Two, developed rapidly. However, a Harvard physician named Professor Peabody noted in the 1920s that at the very time medicine was improving, a decline seemed to be taking place in the physician/patient relationship. The improvement in the effectiveness of medicine was accompanied by two major trends. Firstly, there was a shift to more specialist care away from the generalist and secondly, fewer patients were seen in their own homes and were more often seen in the doctor's surgery or the hospital. In addition, in developed countries at least, there was a shift in the disease pattern from acute illness to more chronic conditions.

Even though medicine became more effective, people seemed more discontent. Porter has argued that with effective weapons against organic disease, the psychological significance and benefits of doctor/patient consultations were forgotten or at least demoted in importance.[5] On a societal level, it seems that we have arrived at a situation in the modern day where medicine has more effective interventions than before, yet there is a substantial degree of public dissatisfaction with medical systems and establishments. In part, one reflection of this is the rise in popularity of alternative medicine. For example, two in five GPs now refer patients to complementary therapists, and in the US people make 425 million visits to 'unconventional' healers, compared to 388 million visits to primary care physicians.[3]

CHANGES IN THE PUBLIC'SVIEWS OF DOCTORS

The reasons why public perceptions of medicine and health care are so ambiguous at present are complex, but they are likely involve some of the following factors. Firstly, it is likely that changing outlooks and attitudes in western societies have led to a growing unwillingness among the public to unquestionably accept a paternal model of health care and, at least up until recently, many doctors have failed to recognise this. A second factor relates to the research community. From the 1950s medicine has become a legitimate area to study in its own right, particularly by social scientists. It has been regularly characterised as a means of social control, exercising social power and reproducing social norms. As Porter has stated:

> . . . once medicine proved effective, the scourge of pestilence was forgotten, and the physician no longer had to be thanked and could be discouraged as a figure of authority, a tool of patriarchy or a stooge of the state.[2]

These views, and the research underpinning them, can conceivably be argued to have permeated the wider society and influenced the public's view of physicians.

Perhaps one of the most influential factors affecting the public's view of doctors has been the coverage of health care in the media. For most of the twentieth century until the 1970s the media view of medical staff was benign and very much in keeping with Dr Finlay or Dr Kildare. However, the interpretive frameworks used by journalists to construct their stories in the seventies have been added to by frameworks more critical of doctors. This has seen an increase in stories where 'doctors get it wrong'. Over the last few years there have been a number of very high profile stories where the behaviour of doctors has attracted much negative comment. In addition to stories where doctors get it wrong, the media is covering stories highlighting health care rationing and doctors' involvement in rationing decisions. Whilst doctors have for many years been involved in decisions about health care priorities (both at an individual and societal level) it has been largely hidden from public view. The more public highlighting of doctors' involvement in rationing now witnessed in the media may also be having an impact on the public's perceptions of physicians, with a negative effect on it perceptions of trust.

For example, the story of Jaymee Bowen (also known as Child B) was widely reported in 1995 and was one of the first major stories that discussed health care rationing. Jaymee's story reflected the decision not to offer her a second bone marrow transplant and intensive chemotherapy when she developed acute myeloid leukemia after recovering from a non-Hodgkin's lymphoma and acute lymphoblastic leukaemia. Although not strictly about rationing (the treatment was thought unlikely to succeed and to have such distressing side effects as not to be in her best interests), Jaymee's case did give rise to much discussion about priority setting in health care; such coverage could have a very real and negative impact on doctor/patient relationships.[4]

In addition to criticisms of health professionals and the health service, the media is an important source of information about other aspects of health care, particularly medical innovations and progress. One reflection of this is a public increasingly knowledgeable about health care issues. Although the traditional areas of the media such as broadcast organisations and newspapers have been the major source of this knowledge, the internet is increasing its significance in this area. The internet is now being used by the public to access information on health in three main ways:

- for self help;
- for direct advice from professionals; and
- for general health information.

Whilst there are problems with health information on the internet, including concerns about the quality and independence of much of the advice, an increasingly significant proportion of the population is accessing health information via this medium.

Another area of growth in health care information that has the potential to impact on public perceptions of health care in general, and doctors in particular, is the bringing into the public domain of information about the quality of care provided by hospitals, health professionals and health

care organisations. This has been happening in the US for some time, but is now becoming increasingly common in the UK as well. The impact of such information is difficult to predict but a recent review concluded as follows:[5]

> . . . Consumers and purchasers rarely search out the information and do not understand or trust it . . . Physicians are sceptical . . . Hospitals appear to be most responsive to the data. In a limited number of studies, the publication of performance data has been associated with an improvement in health outcomes.

Although the review was unable to detect any major impact of the public release of performance data, it is too early to be certain about the effects such data might have. A medical director at one of California's largest managed care organisations described the impact that he perceived: 'everybody is doing what they are required to do in responding to the quality measurements that are being used. Every ounce of energy is being diverted to responding to these; not one ounce of energy is going to any other aspect of quality'.[6]

CONCLUSIONS

The factors discussed in this paper would seem to indicate changing demands and expectations of individual patients and also changing perceptions of health care at a public/societal level. The implications of these changes are difficult to predict but it seems likely that if physicians are to maintain public confidence they will, amongst other factors, have to adopt more participatory styles in their individual consultations. They will also have to play an increasing role as an interpreter of information for the public, using well developed critical appraisal skills. These changes for physicians are likely to take place within an environment where there is increasing uncertainty about what it means to be a doctor.

With regards to promoting participatory styles, it is important to realise that while some doctors do this automatically, many do not and training will be required. Towle and Godolphin,[7] from the University of British Columbia in Vancouver, have identified eight competencies that physicians require in order to practice in a way which promotes the participation of individual patients in clinical consultations and facilitates their involvement in decisions about their care (Figure 2). They refer to this as process as 'informed shared decision-making'.

With respect to the role of an interpreter of information, it would also be mistaken to think that this is a simple task. To interpret and appraise research based information and present it to patients accurately and in an appropriate and understandable format is a skilled process and the consequences of getting it wrong are increasingly being penalised in the courts.

The relationship between the public, physician and patient is a complex one and is continuing to develop and evolve. It is difficult to predict how, but increased public access to information about health issues and the demand for patients to have greater participation in their care look set to be important influences on the relationship. Doctors face increasing uncertainty about their roles, due in part at least to some of the issues discussed above. As for the public and the individual patient – what do they want?

FIGURE 2
Competencies for physicians for informed shared decision-making.*

1. Develop a partnership with the patient.
2. Establish or review the patient's preferences for information (such as amount or format).
3. Establish or review the patient's preferences for a role in decision-making (such as risk taking and degree of involvement of self and others) and the existence and nature of any uncertainty about the course of action to take.
4. Ascertain and respond to patient's ideas, concerns and expectations (such as about disease management options).
5. Identify choices (including ideas and information that the patient may have) and evaluate the research evidence in relation to the individual patient.
6. Present (or direct the patient to) evidence, taking into account competencies 2. and 3. above, framing effects (how presentation of the information may influence decision-making) etc. Help the patient to reflect on and assess the impact of alternative decisions with regard to his or her values and lifestyle.
7. Make or negotiate a decision in partnership with a patient and resolve conflict.
8. Agree an action plan and complete arrangements for follow-up.

* Towle A, Godolphin W. Framework for teaching and learning informed shared decision making. *BMJ* 1999; **319:**766-9. Reproduced with permission of the BMJ Publishing Group.

At one level there is an irony that the healthier western society becomes, the more medicine it craves. The danger of an increasingly consumerist approach to health care is as Porter[5] has stated, that doctors and the public conspire to believe that everyone has something wrong with them and that everyone and everything can be cured. Despite such concerns, however, perhaps at one level what the public want of their doctors has remained unchanged over the last 100 years – namely to be treated with respect, kindness and compassion.

REFERENCES

1. Szasz TS, Hollender MH. A contribution to the philosophy of medicine. *Arch Int Med* 1956; **97:**585-92.
2. Porter R. *The Greatest Benefit to Mankind.* London: Harper Collins; 1997.
3. Eisenberg D. Unconventional Medine in the United States: Prevalence costs and Patterns of Use. *N Engl J Med* 1993; **328:**246-52.
4. Entwistle VA, Watt IS, Bradbury R *et al*. The media and the message: How did the media cover the Child B case? In: Marinker M editor. *Sense and Sensibility in Health Care*. London: BMJ Publishing; 1996.
5. Marshall M N, Shekelle PG, Leatherman S *et al*. The Public Release of Performance Data. What do we expect to gain? A review of the evidence. *JAMA* 2000; **283:**1866-74.
6. Casalino LP. The Unintended Consequences of Measuring Quality on the Quality of Medical Care. *N Engl J Med* 1999; **341:**1147-50.
7. Towle A, Godolphin W. Framework for teaching and learning informed shared decision making. *BMJ* 1999; **319:**766-9.

SCIENTIFIC BASIS OF CLINICAL EFFECTIVENESS – PROFESSIONALISM IN PRACTICE

P.W. Brunt, NHS Consultant, Aberdeen Royal Infirmary

We should begin by defining the terms that we are using – namely, clinical effectiveness and professionalism. Clinical effectiveness has already been widely dealt with – it is the cornerstone of practice and of the operation of the health service. Suffice it to say that poor clinical effectiveness has important consequences which include a poor quality of service provided, waste of (often precious) resources and sometimes actual danger to the patient. The roots of our profession are ancient and go back one to two millennia before Christ with the Aesculapian and Hippocratic traditions. Despite the immense changes in clinical practice over the time since then, elements of the Hippocratic tradition persist to this very day. The tradition was taken over by Christian (and perhaps also Jewish and Islamic) practice. Even now, some medical schools – including our own in Aberdeen – retain a 'Hippocratic' oath, albeit modified and short, which graduating students are required to assent to at their graduation ceremony.

The use of the term profession in relation to medicine is likewise ancient and seems first to have been used by an almost unknown physician of the first century after Christ – Scribonius Largus – who, in a treatise entitled *Humanitas*, described medicine as an art, a science and a profession – a simple description which is remarkably up to date and apposite. Scribonius did not define exactly what he meant by a profession and its elements have developed slowly as the practice of medicine has evolved over the centuries. Probably Thomas Sydenham, a famous seventeenth century London physician, was the first to really incorporate professional principles into practice, principles which were then espoused by other clinicians, and notably by the founders of our Royal Colleges such as Sir Robert Sibbald, the founder of this august Edinburgh College. Indeed, it has been the influence of the Royal Colleges more than any other which has both established and maintained those clinical standards which underpin the concept of professionalism. In its attempts to define professionalism, the *Oxford English Dictionary* describes it, amongst other things, as the 'maintaining of a proper standard'.[1]

Of course, professionalism is much more than that. Firstly it implies a relationship between doctor and patient, a relationship established by the provision of a service. The idea that the professional is the servant of his or her client or patient may not always be popular amongst practitioners, but it is fundamental.

Inevitably, professionalism also implies an ethical framework which governs practice. In addition, professionalism implies an element of protection, that is, protection for the patient from quackery, incompetence or fraud. But there is also an element of protection for the professional himself or herself inherent in the nature of this relationship. There is the important element of confidentiality, the fact that whatever is transacted between client and professional remains sacrosanct to that relationship. Professionalism should also embrace autonomy – in this instance I am referring to the professional's autonomy – that is, his or her ability to decide what is right and best for the patient without being bound by external constraints. While this does not allow the professional freedom to disregard the financial and other consequences of his or her actions, it does mean that consideration of those actions should not be dominated by any factors which are not in the best interest of the individual patient. Nevertheless, this 'clinical freedom' has become a crucially important issue in the operation of a health service, particularly as the complexity and cost of medical progress has spiralled in recent decades with the inevitable emergence of 'rationing'. Obviously the professional must strike a balance between what is best for his or her patient and what is possible within the service which operates for the good of all. Professionalism also implies 'representation' – the belonging together within a group with ascriptive values. This carries with it responsibilities. These embrace the entrance qualifications and monitoring function of a professional group which self regulates, sets ethical standards, monitors practice and so on, activities which are now inherently bound up with revalidation, clinical governance and professional accountability.

To sum up, then, professionalism implies the application of a knowledge/skill base to the best interests of client or patient in a special contractual relationship which is based upon mutual trust. This relationship may be dominant on one side or the other. Traditionally, this has been on the side of the physician, but in present times there is an increasing expectation of a more balanced relationship. Sadly, in recent times the trust which underpins this professional relationship has been eroded by more visible failings within the profession, accentuated by sometimes unreasonable expectations on the one hand and by increasing constraining pressures on the profession on the other.

During its life the NHS has seen many changes. There are a number of key figures who have featured in these changes. Among these was Thomas McKeown who in the mid-1970s emphasised so clearly, among other things, that the changes which most improved health were not the actions of individual doctors but the social and public health improvements on a wide scale, as indeed has already been emphasised in this symposium. Thus health becomes separated from the rather focused activity of 'cure'. A second person who has greatly influenced professional practice in recent years is Archie Cochrane, about whom we have also already heard and who has brought the medical profession to understand that clinical effectiveness and clinical efficiency depend upon the application of an evidence-based practice.

Hence, among the many influences that have served to change clinical effectiveness in recent times are changes in public expectations, the development of guidelines and

protocols for treatment and the whole concept of audit, that is, monitoring the effectiveness of activity. Public expectation has been undergirded by a huge explosion in knowledge in the public domain, sadly not all of which is entirely accurate. It has nevertheless lead to patients being in many instances more informed about their condition – rightly so – and more questioning of the advice they are offered. They can sometimes arrive in the doctor's consulting room armed with information from the internet which is more up to date than the professional's. This in turn leads to a welcome increase in openness in practice and greater appreciation on the part of professionals of the importance of accountability. Alongside all this there has been a not unwelcome reduction in the traditional paternalism of medical practice along with rather less of the directive counselling. There is less professional stuffiness and a gradual disappearance of what I have called 'morning suit medicine', the appearance that frightens the patient into accepting 'I know what is best for you'. Yesterday in my clinic a patient came in and announced that he had searched the internet and knew almost all that there was to know about hepatitis C. When I questioned him I found he did! He wanted to discuss his treatment, not to be told. An aspect of professionalism which is a barrier to clinical effectiveness is what Conroy has termed 'being locked in traditional profiles'.

In addition, the profession is often seen as having a distinct middle-class culture image. Sadly, it is seen by many to be mainly concerned with the promotion of self interest although, interestingly, most patients in a recent survey would be prepared for doctors to be paid more! Doctors are perceived, however, like many other professionals, to close ranks when under scrutiny. Finally, there is a perception of the profession having rigid vision and an unwillingness to change. I think this is a misconception. Few professions have seen as much change as medicine has over the last 50 to 100 years. Most doctors have been willing not only to accept the change, but also to drive it.

There is, however, a danger that the pendulum will swing too far. The price to pay for an excess of openness and the inherent danger in patient autonomy (as distinct from professional autonomy) is the development of a medicine of 'consumerism'. The shopping about for a best buy and the endless questioning of every professional decision is likely to do more harm than good to the patient's ultimate care. This emphasises again the importance of trust in the relationship between the doctor and patient. The idea of a 'walk-in medical shop', strategically based in, for example, a high street chemist or a supermarket, is doubtless very convenient but it depends little on any form of professional relationship for the delivery of the patient's 'care'.

When it comes to guidelines and protocols and to the monitoring and audit the Colleges have, I believe, a justifiable reason to be proud. The development of a whole range of SIGN guidelines (over 50 have been published so far), driven by the President of this College (Professor James Petrie) and his colleagues, is the jewel in the Scottish Colleges' crown. Such guidelines help us to set practice on firm evidence-based grounds, assist in the maintenance of standards of care and can influence the quality of management. Despite the apparent constraints inherent in guidelines they do allow some independence, albeit within defined boundaries. Furthermore, monitoring and audit stimulate doctors to question what they do, to make them more open to change and above all to make them more accountable and to provide the structures of accountability which are so necessary. But I think, in addition, monitoring one's activity does above all makes us more humble and perhaps rather more realistic in our practice. The doctor who has never made a mistake or exercised flawed judgement has not yet been born.

It will, of course, be asked if these processes of monitoring and audit and the use of guidelines in management actually change practice for the better and improve effectiveness. The evidence is accumulating that what is intuitively likely is indeed the case. The York Group, which has done so much for evidence-based medicine, recently produced the results of a study on the 'effectiveness of effective health care'. The group studied 44 reviews covering a wide range of interventions and included seven Cochrane reviews. It found dissemination of written educational materials and didactic educational sessions to be largely ineffective. Local consensus processes, audit and feedback sessions and what has come to be called 'patient mediated interventions' had somewhat variable effectiveness, and the sort of things that worked best included reminders, interactive educational workshops, multifaceted interventions and educational outreach (for prescribing). The group concluded that while there was no magic about stimulating professional changes in practice, active tailored interventions informed by the evidence base and owned by the local community could indeed be 'effective'. Simply telling people what to do does not work.[2]

Finally, let us return to 'professionalism' and summarise briefly those characteristics which I believe to be the essential nature of professionalism.

1. First I would put Commitment, which lies at the heart and implies the total response which the profession demands.
2. Then I would put Compassion as the undergirding of the clinical professionalism – what Scribonius and others would call 'humanity', perhaps.
3. Thirdly I would put Competence, something which the profession's structures and bodies should regulate and control.
4. Then I would put Consistence, which is the mark of order and reason.
5. After this comes Common sense, surely an essential element in the exercise of any profession.
6. Then there is Contract, which underlines the fact that in any profession there is a special relationship between individuals which is sealed in a contract which is either explicit or implicit.
7. Then there comes Confidentiality, because the relationship which forms the contract is one built on trust, trust which is impossible unless it is rooted in confidentiality.
8. And then there is Conduct, something which is becoming less prominent in the profession these days, to its detriment, and implies such things as presentation, clothing, appearance and etiquette, but even more importantly the integrity and uprightness which mark the professional person (and which again is subject to the profession's regulation – or should be).

9. Finally I would put Clinical judgement, something which sums up all our dealings which are based not simply (though importantly) upon the information database available but also upon all those qualities which we have listed as making for professionalism, and which together result in that complete relationship.

Recently, Robin Downie and Jane MacNaughton produced an excellent little book entitled *Clinical Judgement – Evidence in Practice*.[3] They look first of all at the question of evidence in practice in relation to the underpinning science and then they turn to the clinical judgement which they say is founded upon the less easily definable aspects of a professional relationship. It is not simply the science and its evidence that should determine our practice but also the art of medicine. They give examples of how the use of the evidence base may be tempered to take account of the patient's own desires and beliefs in the efficacy of any particular intervention, so that prescribing becomes an art as much as a science.

This concept is, of course, not new and was perhaps more percipiently appreciated in days when the knowledge base was less than it is now. Sir William Osler, arguably the finest physician of the last couple of centuries, spoke to his students in Toronto on the occasion of the opening of the new Medical School Science Building in 1903 on this. You may find his aphorism a little paternalistic and old-fashioned by today's standards, but you cannot fail to be impressed by the force of his argument that medicine is not simply a science but an art:

> The practice of Medicine is an art not a trade, a calling not a business, a calling in which your heart will be exercised equally with your head. Often the best part of your work will have nothing to do with potions and powders but with the exercise of an influence of the strong upon the weak, the righteous upon the wicked, of the wise upon the foolish.[4]

But his concept is much older even than the dawn of Christian influence on medicine and I end with this definition of the professional physician – *'Scire potestates herbarum usumque medeni maluit et mutas agitare inglorius artes'* ('It was his part to learn the powers of medicines and the practice of healing, and careless of fame, to exercise that quiet art').[5]

REFERENCES

1 *Oxford Illustrated English Dictionary*. Oxford: Oxford University Press; 1970.
2 Grant A. Personal communication. 2000.
3 Downie R, Macnaughton J. *Clinical Judgement – Evidence in Practice*. Oxford: Oxford University Press; 2000.
4 Osler W. *The Master word in Medicine Aequanimitas*. Philadelphia: Blakiston and Son; 1904; 386.
5 Virgil. *Aeneid*. **XII:** 396.

QUALITY OF HOSPITAL CARE: MEASURING PATIENTS' EXPERIENCES

A. Coulter, Chief Executive, Picker Institute Europe, Oxford

Ensuring that treatment is clinically effective requires attention to patients' subjective experiences of health care as well as to the technical aspects of treatment. Understanding how things look through the patient's eyes should be a central part of any quality improvement programme. Health care providers in Europe and the US have been measuring patient satisfaction for many years, but often these surveys have been conceptually flawed and methodologically weak.[1] They have tended to focus on managers' or clinicians' agendas rather than on the topics which are most important to patients. They usually use satisfaction rating scales which provide little information about specific problems and cannot be used to target priorities for improvement. The complexities of modern health care and the diversity of patients' expectations and experiences cannot be reliably evaluated by asking global rating questions such as 'How satisfied were you with your care in hospital X?' or by focussing solely on food and amenities while ignoring patients' concerns about their illness and clinical care.

More rigorous approaches are required if quality improvement efforts are to become truly patient-centred. The Picker Institute has pioneered the use of carefully designed instruments to obtain detailed reports of patients' experience.[2] Following extensive qualitative research to find out what patients thought about the way they were treated and what the problems were from their point of view, questionnaires were designed to focus on specific dimensions of care, including information and communication, coordination of care, respect for patients' preferences, emotional support, physical comfort, involvement of family and friends, and continuity and transition. Instead of asking patients to provide satisfaction ratings, the Picker surveys ask patients whether or not certain processes and events occurred during the course of a specific episode of care. This approach has been used since 1987 in hospitals in the US, since 1994 in the UK, and since 1997 in Germany, Sweden and Switzerland.

Table 1 shows results from recent surveys we organised in Scotland. Data came from postal surveys carried out in four acute care hospitals. Questionnaires were mailed to patients' home addresses within one month of discharge. Up to two reminders were sent to non-responders. Completed questionnaires were received from 2,249 patients, a response rate of 65%. Responses to detailed questions about specific topics have been transformed into problem scores, indicating the proportion of patients who reported less than satisfactory experience in response to a particular question.

The most commonly reported problems concern communications. For example, many respondents to these surveys said they were given insufficient information when they were waiting in the emergency room, the results of tests were not properly explained to them and staff sometimes gave them conflicting information. A high proportion of patients wanted more information about their treatment and clear explanations about the potential benefits and harms. They wanted to know how they were likely to feel after a surgical operation, what medication side-effects to look out for and the danger signs they should be aware of after they have been discharged from hospital. Some said it was hard to find doctors or nurses who were willing, or had the time, to discuss their anxieties and fears. Many patients felt their families and friends hadn't been given sufficient information about their illness and treatment so they could feel involved and able to help in the recovery period. Poor communication between staff and lack of coordination of care and treatment processes also caused problems for patients.

Focus groups conducted while planning the surveys provided qualitative information on how things looked from the patient's point of view. For example, Table 2 gives examples of patients' reports of failures in staff communication and coordination of their care.

The advantage of asking specific factual questions about detailed aspects of patients' experience is that answers to such questions are easier to interpret than the rating questions commonly included in patient satisfaction surveys. It is interesting to note that few patients responded critically when asked to rate overall quality of care, in striking contrast to their responses to specific questions. Knowing that, say, 9% of your patients rated their care as 'fair' or 'poor' doesn't give a manager or clinician much of a clue about what they need to do to improve the quality of care in their hospital. On the other hand, knowing the proportion of patients who wanted more information about danger signals to watch out for after they had left hospital, for example, and monitoring trends in this and similar indicators over time, is much more useful when it comes to setting priorities for quality improvement. Focusing on the details of patients' experience can help to pinpoint the problems much more precisely.

Seeing how your hospital or your ward performs in comparison with others can be illuminating for staff and board members. Comparisons – between countries, between hospitals, or even between departments within a hospital – can provide useful benchmarks against which to judge progress. People who commission Picker surveys are encouraged to compare their results against the best from our surveys in Europe and the US, giving them a 'stretching target' for improvement. They can then aim to be better than the best.

Our surveys provide further confirmation that the quality of care is often sub-optimal when viewed through the patients' eyes. Why haven't staff done more to address this problem? The answer may lie in the lack of incentives for change. Until recently, attempts to understand the patient's perspective on care depended on the efforts of a few enthusiasts. In well managed hospitals where staff are truly committed to a patient-centred approach there have

TABLE 1
Picker surveys in Scotland: problem scores.

Information and education	**%**
Insufficient information in A&E	52·3
Delay to go to ward not explained	5·7
Doctors' answers to questions not clear	27·7
Nurses' answers to questions not clear	23·6
Test results not clearly explained	34·0

Coordination of care	**%**
Emergency care not well organised	37·0
Admissions process not well organised	20·1
Long wait to go to ward	15·0
No doctor in overall charge of care	15·1
Staff gave conflicting information	22·9
Scheduled tests or procedures not done on time	21·4

Physical comfort	**%**
Didn't get help to go to bathroom/toilet	15·0
Had to wait too long after pressing call button	0·9
Had to wait too long for pain medicine	4·5
Staff did not do enough to control pain	14·6
Given too little pain medicine	6·5

Emotional support	**%**
Doctor didn't discuss anxieties or fears	34·1
Didn't always have confidence in doctors	18·7
Nurse didn't discuss anxieties or fears	32·3
Didn't always have confidence in nurses	19·8
Not easy to find someone to talk to about concerns	30·3

Respect for patient preferences	**%**
Doctors sometimes talked as if I wasn't there	29·4
Nurses sometimes talked as if I wasn't there	14·6
Didn't have enough say about treatment	59·4
Not always treated with respect and dignity	19·5

Involvement of family and friends	**%**
Family didn't get opportunity to talk to doctor	32·3
Family not given enough information about condition	11·6
Family not given information needed to help recovery	38·5

Continuity and transition	**%**
Purpose of medicines not fully explained	23·1
Not told about medication side effects	36·0
Not told about danger signals to watch for at home	60·3
Not told when to resume normal activities	60·9

Surgery specific	**%**
Explanation of risks and benefits not clear	21·9
Answers to questions about surgery not clear	17·2
Not told accurately how could feel after surgery	49·0
Results of surgery not well explained	30·9

Overall impression	**%**
Courtesy of admissions staff not good	6·9
Availability of doctors not good	30·0
Courtesy of nurses not good	5·6
Availability of nurses not good	29·8
Doctor/nurse teamwork not good	8·1
Overall care received not good	8·5
Would not recommend this hospital to friends/family	7·8

TABLE 2
Patients' reports of failures in communication and coordination.

- I was called three different names, the worst being prior to surgery by the anaesthetist who thought I had a completely different medical history – frightening!
- There wasn't anybody who would actually say: 'Right, do something about this'. It just got shipped round. Everybody just passed the buck.
- There were a couple of instances where the medical staff changed my drug therapy without either discussing it with me or telling me they'd changed it. That did disturb me.
- When I needed pain relief after the birth I had to ask three different members of staff.

been good examples of improvements as a direct result of feeding back results from surveys documenting patients' experience,[3] but these successes are by no means universal. Patient surveys have often been used simply as a marketing tool, with providers making claims on the basis of poorly designed and badly conducted surveys that '95% of our patients are satisfied'.

The situation is beginning to change in many European countries. Professional regulators and those providing hospital accreditation are starting to demand evidence that providers have sought the views of their patients and acted on them. In some countries national surveys are being organised to facilitate comparisons between providers and produce national benchmarks. For example, in England the government has introduced a series of national surveys of patients' experience. In addition to our work with individual provider units, Picker Institute Europe is a partner in the research consortium contracted by the Department of Health to carry out the national surveys of NHS patients in England. Two of these large-scale surveys have been

completed (general practice patients: sample size 100,000, response rate 64%; coronary heart disease patients: sample size 113,000, response rate 74%) and a survey of cancer patients is under way.[4] These surveys reveal considerable variations between hospitals in the quality of patient care provided to their patients.

National surveys have the advantage of providing data to facilitate comparison between hospitals, but they are relatively cumbersome to organise and results are not as timely or as relevant as those obtained in locally organised surveys. Local surveys can be carried out relatively quickly and can be tailored to local circumstances, but providers cannot use the results to compare their performance against others unless there is some standardisation of methodologies and questionnaires. The ideal would be to adopt a combination of approaches. National sample surveys could be developed to produce national benchmarks against which the results of local surveys could be compared. National surveys could also be used to develop survey methods and population benchmarks for specific national priorities, e.g. diabetes, mental health or care of older people. Instead of basing these on the very large samples which are necessary to produce results which are statistically valid at the level of individual Trusts, these surveys could be much smaller. Modules on specific topics could be developed nationally for local application if and when required. These could include disease-specific modules, questions about medical errors and measures of quality-of-life and patient-assessed health outcomes. The validated instruments or modules could then be incorporated into local survey programmes and integrated into local clinical governance initiatives.

At a local level trusts could be encouraged, or required, to carry out generic (as opposed to disease-specific) surveys of a sample of their patients each year. A basic template of questions could be developed for inclusion in every local survey to provide comparable data for use in monitoring performance at a national level. Trusts would then be free to add one or more of the nationally developed modules described above, or their own locally designed questions. It would be very important to ensure that local surveys conform to rigorous methodological standards to ensure validity and reliability.

It is now fashionable among politicians and policy-makers to talk about making the NHS more patient-centred. The rhetoric about patient-centred care will not become a reality until it is backed up by robust systems to help providers look critically at their services through the patient's eyes, coupled with clear incentives to use the results in quality improvement programmes. Carefully conducted patient surveys can make an important contribution towards the overall goal of quality improvement and clinically effective care for all.

REFERENCES

1 Cleary PD. The increasing importance of patient surveys. *BMJ* 1999; **319:**720-1.

2 Bruster S, Jarman B, Bosanquet N *et al*. National survey of hospital patients. *BMJ* 1994; **309:**1542-9.

3 Cleary PD, Edgman-Levitan S, Walker JD *et al*. Using patient reports to improve medical care: a preliminary report from 10 hospitals**.** *Quality Management in Health Care* 1993; **2:**31-8.

4 The National Survey of NHS Patients. http://www.doh.gov.uk/public/nhssurvey.htm

EVIDENCE-BASED MEDICINE: A HISTORICAL AND INTERNATIONAL OVERVIEW

S. Woolf, MPH Professor, Department of Family Practice, Virginia Commonwealth University, US

Since Ancient Greece, medicine has been 'evidence-based'. Physicians have always familiarised themselves with science and have tried to keep abreast of the literature. A century ago William Osler was training physicians to cite supporting literature to justify their views on proper tests and treatments.

What distinguishes the modern discipline of evidence-based medicine (EBM) from the earlier approach is an organised effort to explicitly link clinical and public health policy to a systematic examination of the quality of supporting scientific evidence. Its emphasis is on explicitness, in clarifying when practice recommendations are based on opinion and not science and, when supporting evidence is found, in documenting the quality and strength of that evidence.[1] In examining the science base, EBM emphasises comprehensiveness and applies systematic criteria to ensure that all relevant evidence is considered, rather than being cited selectively, and that the quality of studies is evaluated fairly, regardless of preconceived biases.

HISTORICAL CONTEXT

The ascendancy of EBM in most countries began 10–15 years ago in response to a common set of historical challenges confronted by health care systems of the late twentieth century. At that time, the costs of delivering tests, treatments and procedures were escalating dramatically, fuelled by stunning advances in technology and pharmacology and by the chronic health conditions plaguing an ageing population. The inability of health care budgets to keep pace with such growth was increasingly apparent. Limited resources, especially in countries with cash-limited health care systems, made priority-setting a singular priority. Concurrent with these developments, mounting evidence pointed to significant geographic practice variations in the delivery of services[2] and to the ineffectiveness (or even net harm) of widely used tests and treatments.[3] It became clear that some services, many of them costly, were overused; others were underused; and a large proportion were misused (i.e. applied to the wrong patients or in the wrong settings).[4] The evidence suggested that these patterns of care were not optimising patient outcomes.

Evidence-based medicine – an effort to improve the quality of information about the proven benefits of health care services and their relative effectiveness – offered policymakers, physicians and patients a timely resource for addressing these challenges. For policymakers, it offered a knowledge-based resource for setting priorities. For physicians and patients, it provided a framework for making informed choices about available treatment options and their likely outcomes. Evidence-based medicine has matured with passing years, but it has evolved and gained acceptance in different ways and at different paces in countries with diverse health care systems and cultural priorities.

Despite these differences, EBM in most countries has generally diversified into four areas of work: (1) the critical appraisal of individual studies, and (2) the systematic review of groups of studies focus on characterising, systematically and objectively, what the science tells us. Many of these efforts stop short of setting policy; (3) the development of evidence-based practice guidelines goes further to define what should be done, and (4) evidence-based systems of care focus on implementation.

CRITICAL APPRAISAL OF INDIVIDUAL STUDIES

One hallmark of EBM has been the promotion of standardised criteria for evaluating the quality of individual studies and the publication of critical appraisals based on such assessments, both in specialised journals devoted to this purpose (e.g. *Evidence-Based Medicine, ACP Journal Club, Best Evidence, Bandolier*) and in mainstream medical journals. Such reviews not only describe the methods and results of a study but also scrutinise its internal and external validity. To judge the quality of randomised controlled trials, for example, reviewers check for concealment of allocation, blinding, attrition, intention-to-treat analysis and other design features.[5] Guidance on how to conduct such assessments for various types of studies has been widely disseminated, through medical journals (e.g. the 'Users Guide' series in *JAMA*[6]) and web sites, such as that of the Centre for Evidence-Based Medicine (http://cebm.jr2.ox.ac.uk).

SYSTEMATIC REVIEWS

Systematic reviews examine the body of studies that pertain to a particular issue, using explicit methods to ensure that the search of the literature is comprehensive and that the quality of studies is evaluated by standard criteria.[7] Reviews feature evidence tables that detail the design, results and limitations of studies and, when appropriate, meta-analyses or modelling analyses that summarise effect sizes from pooled data. The products of the Cochrane Collaboration[8] and the databases catalogued in the Cochrane Library (www.cochrane.de/) and the Database of Abstracts of Reviews of Effectiveness (http://agatha.york.ac.uk/darehp.htm) are signature examples of systematic reviews. Increasingly, however, superb systematic reviews are produced by other centres and are appearing with growing regularity in medical journals and in white papers by governmental bodies and task forces. Evidence, even without recommendations, is a useful tool for policymakers. Recognising this, the US government has established 12 Evidence-Based Practice Centers to produce systematic reviews (http://www.ahcpr.gov/clinic/epc). Similar programmes exist in other countries.

EVIDENCE-BASED PRACTICE GUIDELINES

Practice guidelines and technology assessments take the next step, translating science into policy for clinical practice

and public health.[9] The EBM movement has given rise to evidence-based practice guidelines which offer clinical guidance but also document the quality of evidence or opinion supporting the recommendations. Interventions supported by good evidence receive higher ratings than those for which the evidence is poor.[10] Evidence-based guidelines are developed according to explicit methods that emphasise clear explication of the topic and question, a systematic review of the relevant evidence, rules of evidence for translating science into policy, documentation that permits users to evaluate the rationale for recommendations and potential conflicts of interest and peer review.[11] Governments, professional organisations and health care systems have invested considerably in funding guideline panels to produce such reports.[10]

Programmes that have been engaged in this effort for some time have made important advances in refining the methods used to develop and critically appraise such guidelines. Examples include the pioneering work of the Scottish Intercollegiate Guideline Network (SIGN) (http://www.sign.ac.uk), the North of England project,[12] the NHS National Institute for Clinical Excellence (NICE) (http://www.nice.org.uk), the Council of Europe, the US Agency for Healthcare Research and Quality, formerly the Agency for Health Care Policy and Research (www.ahcpr.gov/clinic/cpgonline.htm), and the US Preventive Services Task Force.[13] The EU-funded AGREE (Appraisal of Guideline Research and Evidence for Europe) collaborators have developed standard criteria for appraising guidelines. The rapidly growing number of practice guidelines now available around the world has made it necessary to establish internet resources to locate guidelines on topics of interest, such as those established in the US (www.guideline.gov) and Germany (www.azq.de).

Evidence-based technology assessments, which derive from a similar approach, have emanated from Scandinavian countries, Spain, Canada and other programmes around the world. This vibrant international network is exemplified by the global membership of national agencies in the International Society for Technology Assessment in Health Care (www.istahc.org).

IMPLEMENTING EVIDENCE-BASED PRACTICE POLICY

To create the conditions for implementing evidence-based practice policy, systems that deliver clinical and public health services are refocusing their efforts on developing the tools for implementing new practice policies and behaviour change. That such tools are necessary has become more apparent over time, as the early EBM movement learned that reporting the evidence or issuing practice guidelines did little by itself to change practice behaviour.[14] The great successes of EBM – in clearly articulating the compelling evidence for such treatments as beta-blockers following acute myocardial infarction or warfarin for atrial fibrillation, which significantly reduce mortality – have been tempered by disconcerting studies reporting that such therapies are not offered to a large proportion of eligible patients. The most sobering lesson of the EBM movement has been that evidence is not enough, that active implementation strategies are essential to modify behaviour.

The movement has learned more specific lessons: to effect change, there must be knowledge, acceptance, ability, and reinforcement. Practice guidelines, evidence reports, and other educational interventions provide knowledge, but physicians do not change behaviour unless there is acceptance that the evidence is valid and that the recommendations are reasonable standards of care for individual patients. Even if there is acceptance that the recommendations represent good care, they cannot be put into practice without ability: time, personnel, equipment, training and other resources. Finally, even if such ability exists, optimal practice is usually difficult to maintain without reinforcement.

Countries with mature EBM movements now find themselves confronting these challenges and investing heavily to develop tools to 'translate evidence into practice'. They invest in strategies to promote acceptance, such as local adaptation, engaging local opinion leaders, 'academic detailing' and specialty endorsement. Strategies aimed at enhancing ability address time constraints, skills, access to equipment and personnel, information technology, delivery systems and patient compliance. Reinforcement tools include reminder systems (e.g., standing orders and decision support software), audit and feedback reports. Governments, research bodies and health care systems have invested heavily in research and demonstration projects to determine whether these measures are effective in narrowing the gap between what should be, and what is, done in patient care.

Over the years, each of these strategies has been promoted as the 'magic bullet' for improving the quality of care. Many hospitals, trusts and health systems have seized on a single-minded approach focused on a favoured strategy, such as academic detailing, total quality management or peer review. There is no magic bullet, however.[15] Experienced managers have learned, and research has demonstrated,[16] the need to target implementation strategies to the specific obstacles faced in practice.[17] Issuing a practice guideline accomplishes little if the information it contains is already known and the challenge at hand is the inability, due to system constraints, to implement the recommendations. Conversely, investing in system changes will not alter behaviour if clinicians are unaware of new evidence or do not accept its validity.

CONCLUSION

An irony of the rapid expansion in the volume of information generated by EBM is the growing difficulty of applying what has been learnt and locating relevant evidence in real time at the bedside and when other decisions are made. The 'information mastery' movement[18] has concentrated on giving busy physicians the tools to narrow their focus to the evidence that matters most and to determine whether the findings justify a change in practice behaviour. New technologies, often utilising palmtop devices (http://www.medicalinforetriever.com/inforetriever.cfm) and web sites (http:www.//cebm.jr2.ox.ac.uk; www.infopoems.com), help clinicians access relevant evidence and make patient-specific calculations to determine, on an individual basis, which treatment choices the evidence suggests are best for patients. Increasingly, physicians are considering the number-needed-to-treat (NNT), likelihood ratios, and other metrics popularised by the EBM movement in making health care decisions.

The story of EBM in most countries features, at some point in its evolution, a period of resistance within the

medical community and from patients. Often out of misunderstanding, EBM is seen as 'cook book medicine' and is misperceived as advocating care only if supported by randomised controlled trials and meta-analyses. This tends to be an early reaction to EBM, which dissipates as medical cultures gain experience with EBM and see examples in which their misconceptions are overturned.[19] However, one source of anxiety about EBM is that criticisms of the science base for tests and treatments will encourage policymakers to limit services. This is a legitimate concern, especially for physicians and patients who believe those services are beneficial. From a societal perspective it can be argued that such knowledge, by applying resources more wisely, is ultimately to the good of all, but such utilitarian arguments may ring hollow for the individual patient or frustrated physician for whom care is rationed.

Perhaps the greatest contribution of the scrutiny of EBM is to draw attention to gaps in the evidence, thereby helping to shape a research agenda that focuses on answering the questions of greatest importance to medicine. Among the most urgent lines of investigation in which government and health care systems can invest is to fill the current evidence vacuum about the health effects of widely used tests and treatments. In an increasingly global health care community, countries share the same pivotal questions about such interventions, and high-quality studies that provide definitive answers to these questions, regardless of where they were performed, can influence health care around the world. Evidence-based medicine will fulfil its mission if, by stimulating high-quality research, it helps patients and societies to make better choices and thereby optimise patient outcomes and public health.

REFERENCES

1 Woolf SH, George JN. Evidence-based medicine: interpreting studies and setting policy. *Hematol Oncol Clin North Am* 2000; **14:**761-84.

2 Chassin MR, Kosecoff J, Park RE *et al*. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA* 1987; **258:**2533-7.

3 Eddy DM. Clinical policies and the quality of clinical practice. *N Engl J Med* 1982; **307:**343-7.

4 Chassin MR, Galvin RW, the National Roundtable on Health Care Quality. The urgent need to improve health care quality: Institute of Medicine National Roundtable on Health Care Quality. *JAMA* 1998; **280:**1000-5.

5 Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Technol Assess Health Care* 1996; **12:**125-208.

6 Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992; **268:**2420-5.

7 Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997; **126:**376-80.

8 Bero L, Rennie D. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; **274:**1935-8.

9 Field MJ, Lohr KN editors. *Guidelines for Clinical Practice: From Development to Use*. Washington, DC: National Academy Press; 1992.

10 Cook DJ, Guyatt GH, Laupacis A *et al*. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest* 1995; **108(4):**227-30.

11 Shekelle PG, Woolf SH, Eccles M *et al*. Developing guidelines. *BMJ* 1999; **318:**593-6.

12 Eccles M, Freemantle N, Mason J. North of England evidence-based guideline development project: summary version of guidelines for the choice of antidepressants for depression in primary care. North of England Anti-depressant Guideline Development Group. *Fam Pract* 1999; **16:**103-11.

13 U.S. Preventive Services Task Force. Guide to Clinical Preventive Services, 2nd ed. Baltimore: Williams and Wilkins, 1996.

14 Davis DA, Thomson MA, Oxman AD *et al.* Changing physician performance: a systematic review of the effect of continuing medical education strategies. *JAMA* 1995; **274:**700-5.

15 Oxman AD, Thomson MA, Davis DA *et al.* No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995; **153:**1423-31.

16 Bero LA, Grilli R, Grimshaw JM *et al*. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ* 1998; **317:**465-8.

17 Woolf SH. Changing physician practice behavior: the merits of a diagnostic approach. *J Fam Pract* 2000; **49:**126-9.

18 Slawson DC, Shaughnessy AF. Becoming an information master: using POEMs to change practice with confidence. *J Fam Pract* 2000; **49:**63-7.

19 Sackett DL, Rosenberg WM, Gray JA *et al*. Evidence-based medicine: what it is and what it is not. *BMJ* 1996; **312:**71-2.

THE HISTORY OF CLINICAL EFFECTIVENESS

U. Tröhler, Professor of Medical History, The University of Freiburg, Germany

In 1992, the *Journal of the American Medical Association* published the following manifesto:

> A new paradigm for medical practice is emerging. Evidence-based medicine de-emphasizes intuition, unsystematic clinical experience and patho-physiological rationale as sufficient grounds for clinical decision-making and stresses the examination of evidence from clinical research.[1]

Indeed, many of the issues discussed in current debates about evidence-based medicine and other ways to discover 'the truth' (debates that encompass philosophy, clinical practice, morals, economics and the law) reflect the three classic 'cultures' in the search for evidence in medicine. These cultures may be called:

1. the systemic-pathophysiological;
2. the individual-clinical; and
3. the statistico-analytical tradition.

Although today we associate the systemic-pathophysiological approach with experimental laboratory science, its origins lie in the medical systems of the classical doctor-philosophers of Antiquity, e.g. Hippocrates (*c.* 400 BC) and Galen (*c.* AD 200), while, indeed, some later comprehensive pathophysiological systems such as Samuel Hahnemann's homeopathy (*c.* 1800) and Rudolf Steiner's anthroposophy (*c.* 1920) are still in vogue.[2] If you move within any system it yields certain and objective knowledge. Partisans see this as an advantage.

Clinical observations made on individual patients can also yield certain, necessarily subjective, knowledge. In both of these cultures knowledge is arrived at through the rather vague procedure of 'medical judgement'.

By contrast, the third culture – the statistico-analytical, i.e. the analysis of aggregated data obtained from groups of patients – yields objective, yet probable, results; and it does so by applying a set of defined, non-medical, but statistical rules. Various alliances and enmities among these three positions are possible and have indeed been historically relevant.

In therapeutics, in particular, these debates about types of evidence are also linked to two basic approaches, the dogmatic and the empirical. The dogmatic approach suggests that if you discover the cause of a disease and its pathophysiology, therapy can be derived rationally. It follows that if the cause of illness is correctly deduced, the therapy is bound to work and therefore the outcome does not need to be validated. This unidirectional approach has conferred certainty, particularly when it has been combined with authority, tradition or deduction from higher 'insight'. Its key question is 'Can this therapy work?' Medical history has abounded with dogma since Antiquity – and it still does.[3] Indeed, the mark of distinction of a learned man or woman has long been the certainty of his or her knowledge, and this makes it positively disadvantageous to test this kind of knowledge empirically, as any acknowledgement of uncertainty would undermine their authority.

As long as certainty of knowledge (in accordance with the wisdom of the Ancients and, if possible, the Holy Scriptures) was essential, evidence for therapeutic effectiveness was not an issue. Failures could always be explained away by the incomplete or untimely observance of the system, i.e. they were either the patient's fault or the fault of external circumstances. Personal experience, or the trial and error implicit to the empirical approach, were a matter for craftsman surgeons and mountebanks, i.e. persons considered to be of lower standing.

In contrast to dogma the empirical approach, linked to the notion of experience, has held that clinical symptoms could lead directly to practical therapy regardless of any theoretical considerations. An essential feature of this approach is that effectiveness must be assessed in order to answer the key question, 'Will this therapy work?', and to provide the feedback necessary to gain experience. In order to do so, the Ancients adopted the classical empirical triad of observation, comparison of the data with the literature and conclusion by analogy. They further held that observations must be repeated in order to yield valuable experience and that negative results be considered as well. However, this approach fell into oblivion even in Antiquity.[4] Had not Hippocrates himself, in the first of his aphorisms, called experiment and experience 'treacherous'?[5] The quest was for certain knowledge, logically deduced from unquestioned first principles. Debates were about the truthfulness of those principles rather than the effectiveness of the recommendations deduced from them. The empiricist, which literally means the 'testing doctor', fell into disrepute. A doctor did not 'test', he or she 'knew'. In fact, the empirical approach fell into general discredit in all the learned disciplines during the Middle Ages.

The rise of scientific interest in empirical experience based on new observations and interventions started again in the sixteenth century. However, this could yield only probable results which had a difficult standing alongside the traditional dogmatic and certain knowledge.[6] Around 1600, Francis Bacon (1561–1626) distinguished between 'ordinary experience,' based on chance observations and therefore subjective, and 'ordered experience,' based on the results of methodological investigation and aspiring to a certain form of objectivity.[7] All experience (whether subjective or objective) is empirically gained knowledge and henceforth experience, much maligned before, became the core of a new empiricism. However, within medicine, where it coincided with the study of anatomy, physiology, botany and even patients, the development was particularly slow in therapeutics as physicians (who constituted the academic profession up to the nineteenth century)

continued to regard empiricism as the sphere of quacks and surgeons.

It is true that there was a lot of confusion about the notion of experience. How could we otherwise explain the everlasting praise of panaceas, old and new, or of blood-letting-vampirism and secret medicines, or the recommendation of contradicting therapies in the name of 'experience'? This confusion was finally confronted in the eighteenth century when an elite of authors started to differentiate 'pseudo experience', as they called it, from true experience in therapeutics, be it 'ordinary experience' or 'ordered experience', as according to Bacon.[8] There was an important reason for this insight: by the end of the seventeenth century one result of the voyages of discovery was that pharmacopoeias were revised to incorporate the additional herbal drugs that had been collected during circumnavigations of the world. The most prominent of these was Peruvian bark (from the cinchona tree), which we know now to contain quinine. It became recommended for 'fever', the most prevalent 'disease' at that time. But throughout the eighteenth century, other new therapies were introduced both in medicine and surgery, challenging traditional ones.[9] In that time quantitative assessment became a feature in many fields of life, particularly in Britain. Not surprisingly, in medicine too, strong cases were made for individual clinical observation, as well as for the critical analysis of observations assembled from groups of patients. The latter method was in itself considered an innovation as it implied evaluation by comparison and numerical presentation of the data. However, such statistics were a thorn in the side of many 'dogmatic' doctors because external statistical rules seemed to supersede 'medical judgement' and 'medical authority'.[10]

Unlike the multinational origins of empirical evaluation of surgical therapy, the introduction of an empirical approach to the evaluation of medical treatments was a largely British initiative and the principal actors were medical graduates of Edinburgh University. The most celebrated of these graduates is probably James Lind (1716–94). His prospective clinical experiment, performed in 1747 aboard a ship of the British Navy, is widely known. In order to test the validity of his strong intuition about the best treatment for scurvy (which was killing more sailors than military action), Lind took 12 patients, 'as similar as I could have them', and assigned two each to one of six treatments, all of which were justified either by dogma or 'ordinary experience'. The two sailors who received two oranges and a lemon daily improved dramatically and much more quickly than the other ten. Lind's reason for implementing Bacon's concept of 'ordered experience' was his awareness of (what we call today) selection bias, observer bias and publication bias: he held that results of treatment should be compared with the natural course of a disease in comparable patients. He understood the role of imagination and suggestion in the healing effect, and postulated that all observations from a series of unselected cases rather than 'the habit of publishing individual successful cases only' should yield the evidence for therapeutic recommendations.[11]

This milestone in therapeutic evaluation did not come 'out of the blue'. It took place in both a medical and a social context. Other pioneers in the British Navy, Army and civilian life initiated systematic approaches to the assessment of therapy during the second half of the eighteenth century. Among the new features, which facilitated the emergence of these initiatives, were dispensaries (out-patient facilities for ambulant patients) and the new type of (voluntary) hospitals, both military and civilian, which were unique to British health care at that time. Some of these cared for all kinds of 'curable' ailments; others were specialised, for example in fevers, childbirth, children's ailments and rheumatic diseases.

Record keeping which was propagated in these establishments for administrative purposes was soon recognised as indispensable for medical advances and improved clerical methods were developed. Quantitative methods were thus discussed and implemented for analysing the sometimes huge amounts of data, e.g. for testing the efficacy of smallpox inoculation and vaccination, old and new drugs (such as foxglove or arsenic), operations, bathing, and some of the nastier, yet dogmatically 'well founded', systemic treatments such as bleeding and purging. The research methods used on ships, in the dispensaries and hospitals and also in private practice ranged from (comparative) retrospective analysis of case series including historical and concomitant controls to prospective experiments with untreated or placebo controls. Proportions, ratios and averages were calculated and probabilities estimated, and the limitations of these investigations, including selection biases, insufficient numbers of observations and frank cheating were actively discussed. The approach as a whole was termed 'medical arithmetic'.[10]

However, it would be erroneous to believe that dogmatism had vanished by 1800, even in Britain. In the whole of Europe this was the heyday of new rational systems such as those of the Scot John Brown, of Anton Mesmer in Vienna and Paris, or those of German romantic medicine as well as of homeopathy.[12-16]

Formal opposition to the empirical and at the same time to the probabilistic approach did not manifest itself in Britain; it happened in Paris in the 1830s, where the approach known as the 'méthode numérique' was championed by Pierre Charles Alexandre Louis (1787–1872). In the Académie Royale des Sciences and the Académie Royale de Médecine, two formal debates took place on the issue around 1835.[17, 18] The arguments in favour of statistics in clinical medicine were the same as those advanced by Lind in the eighteenth century. The opponent's arguments were particularly interesting. They were of two kinds. The first, which we may call the 'old school,' wondered whether the old certainty could possibly ever be replaced by probability. Would this not mean that medicine would become a gambling place? The second kind of arguments were 'modern' ones. Some were pragmatic: for example, it was suggested that two sufficiently large groups of comparable patients could never be found, and that such trials were clinically inadequate, because they did not reflect the daily reality of the doctor who always faced an individual rather than a group of patients. Trials were also held to be unethical given the probabilistic nature of the results: one was inclined to follow the results yielded by the majority of patients, but what about those of the minority in which a treatment caused harm, for instance?

By the middle of the nineteenth century the Paris discussions on clinical statistics reached a sort of compromise: statistics with their probabilities were

considered correct for hygiene, epidemiology, or preventative medicine.[17] However, those relatively recent disciplines were seen as linked to clerical office work. This was something quite separate from 'real' clinical medicine from which the statistical empirical approach was only to be banished; it 'killed medical intelligence', as the famous Paris clinican Armand Trousseau (1801–67) put it in 1861.[19] This bipolar position characterised much of the subsequent discussion in other European countries.

While a few retro- and prospective controlled studies continued to be performed in England as well as elsewhere,[20] none of the leading academics championed Baconian 'ordered experience' in the assessment of therapeutics. One intra-medical reason was that 'real' clinical medicine was just about to develop a new dogmatic certainty caused by advances in pathology (e.g. Virchowian pathological anatomy and cellular pathology) and bacteriology which gave the mainstream pathophysiological tradition a new thrust. This was further strengthened by the introduction of the methods of experimental natural sciences, new diagnostic tools and, finally, by the new therapeutic possibilities of modern painless a(nti)septic surgery.[21, 22] In order to show the safety of their interventions, surgeons reported the results in huge unilateral statistics of operated cases only. Nobel Prize-winning surgeon Theodor Kocher of Berne (1841–1917) was an example in point. He published his goitre operations in groups of a thousand (with ever-decreasing intervals) demonstrating the equally decreasing lethality (0·7% in 1909), in order to win the confidence of his non-operating colleagues and of the public.[23]

While surgeons saw them as another expression of 'certainty', such statistics did not say anything on the value of an operation in the healing plan of a doctor. They were a mere illustration of the saying 'The best way of improving on the result of a clinical trial is to leave out the controls.'[24] With the benefit of hindsight we know that this has led to disasters, as well as into fruitless byways, and hindered the timely introduction of valuable therapeutic measures.[25]

After World War II the relative importance of the systemic-pathophysiological tradition further increased for a number of scientific, technical and social reasons, including the successes of laboratory science and the rise of the pharmaceutical industry. However, modern pathophysiology evinced so many pathways, which interacted in a complex way right down to the molecular level, that it has increasingly become impossible to forecast what the overall therapeutic outcome is from influencing one of them. Clinical observation relating to effectiveness is again called for and – in view of historical experience with one-sided applications of pathophysiological systems – there is now a clear quest for comparative experience which is evaluated according to neutral, external rules in order to minimise biases.[26-28] Today's evidence-based medicine is thus a legitimate, albeit late-born, child of the eighteenth century British achievement of clinical 'arithmetic observation and experimentation'.

Among the intra-medical reasons for the recent blossoming of this 'third culture' is the overriding force of historical necessity: neither the pathophysiological nor the clinical traditions alone have been responsible for valuable progress. Indeed, they have often failed to predict the outcome of an intervention. This is also due to the increased complexity of a pathophysiology which is no longer on the organic but on the molecular level. Culturally, evidence-based medicine is consistent with the popular feeling of the need for transparency and for the evaluation of everything from the dishwasher to the cancer drug: medical services are seen as having the economic qualities of products subject to quality standards and cost-effectiveness.

In conclusion, the time has come for the three classical cultures no longer to compete for the top place in a presumed hierarchy of evidence. Rather, history indicates that all three have contributed in specific ways to the progress of health care. Hence historical evidence strongly underpins the quest for their integration, based on respect for the legitimacy of all three of them in our effort to cope with the progress and success of science.[29, 30]

ACKNOWLEDGEMENTS

I thank Iain Milne for pertinent commentaries on earlier drafts of this manuscript and for great help in editing it.

REFERENCES

1 Evidence-Based Medicine Working Group: Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; **268:** 2420-5.

2 Rothschuh K. *Konzepte der Medizin in Vergangenheit und Gegenwart*. Stuttgart: Hippokrates Verlag; 1978; 162, 185-99.

3 Ackerknecht EH. Aspects of the history of therapeutics. *Bull Hist Med* 1962; **36:**389-419.

4 Deichgräber K. *Die griechische Empirikerschule*. 2nd ed. Berlin-Zurich: Weidmann; 1965.

5 Hippocrates. Aphorisms, First section I. In: *Hippocrates, vol. IV* (transl. by W.H.S. Jones). London: Heinemann, (Loeb Classical Library); 1979; 99.

6 Temkin O. *Galenism. Rise and decline of a medical philosophy*. Ithaca: Cornell University Press; 1973.

7 Krüger L. Erfahrung. In: Seifert H, Radnitzky G, editors. *Handlexikon zur Wissenschaftstheorie*. München: Ehrenwirth; 1989; 48-51.

8 Ackerknecht EH. Die therapeutische Erfahrung und ihre allmähliche Objektivierung. *Gesnerus* 1969; **26:**26-35.

9 Maehle AH. *Drugs on trial: Experimental pharmacology and therapeutic experiments in the eighteenth century*. Amsterdam-Atlanta: Rodopi; 1999.

10 Tröhler U. *"To improve the evidence of medicine": The 18th century British origins of a critical approach*. Edinburgh: Royal College of Physicians; 2000.

11 Tröhler U. Towards clinical investigation on a numerical basis: James Lind at Haslar Hospital 1758-1783. In: *Proc. XXVII. Int. Congr. Hist. Med. Barcelona*. Barcelona: Academia de Ciènces Mediques de Catalunya i Balears; 1981; vol. 1; 414-19.

12 Risse G. The quest for certainty in medicine: John Brown's system of medicine in France. *Bull Hist Med* 1971; **45:**1-12.

13 Bynum WF and R Porter, editors. *Brunonianism in Britain and Europe*. London: Wellcome Institute;1988.

14 Schott H, editor. *Franz Anton Mesmer und die Geschichte des Mesmerismus*. Stuttgart: Steiner; 1985.

15 Schwanitz HJ. *Homöopathie und Brownianismus*. Stuttgart: Fischer;1983.

16 Wiesing U. *Kunst oder Wissenschaft: Konzeptionen der Medizin in der deutschen Romantik*. Stuttgart-Bad Cannstadt: Frommann-Holzboog; 1995.

17 Murphy. Medical knowledge and statistical methods in early nineteenth-century France. *Medical History* 1981; **25:**301-19.

18 Matthews JR. *Quantification and the Quest for Medical Certainty*. Princeton, N.J.: Princeton University Press; 1995.

[19] Trousseau A. *Clinique médicale de l'Hôtel Dieu de Paris.* Vol. 1. Paris: Baillière; 1861.

[20] Coleman W. Experimental Physiology and Statistical Interference: The therapeutic trial in 19th century Germany. In: Krüger L, Gigerenzer G, Morgan MS, editors. *The probabilistic revolution. Vol. 2: Ideas in the Sciences*. Cambridge (Mass.), London: MIT Press; 1988; 201-26. *See also*: Hobi BA. 'Die methodisch-klinische Forschung des Internisten Theodor Juergensen (1840-1907).' M.D. thesis. Univ. Basle:1988.

[21] Matthews JR. Major Greenwood versus Almroth Wright: Contrasting visions of "scientific" medicine in Edwardian Britain. *Bull Hist Med* 1995; **69:**30-43.

[22] Tröhler U. Surgery (modern). In: Bynum WF and Porter R, editors. *Companion Encyclopedia of the History of Medicine.* London-New York: Routledge; 1993; 984-1028.

[23] Tröhler U. *Der Nobelpreisträger Theodor Kocher: Auf dem Weg zur physiologischen Chirurgie*. Basle-Boston: Birkhäuser; 1984.

[24] Tröhler U. 'To operate or not to operate?' Scientific and extraneous factors in therapeutical controversies within the Swiss Society of Surgery 1913-1988. *Clio medica* 1991; **22:**89-113.

[25] Silverman WA. *Retrolental fibroplasia: A modern parable*. New York-London: Grune & Stratton; 1980.

[26] Lilienfeld AM. Ceteris paribus: The evolution of the clinical trial. *Bull Hist Med* 1982; **56:**1-18.

[27] Marks HM. *The progress of experiment. Science and therapeutic reform in the United States, 1900-1990*. Cambridge: Cambridge University Press; 1997.

[28] Kaptchuk TJ. Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998;**72:**389-433.

[29] Sackett DL and JE Wennberg. Choosing the best research design for each question. *BMJ* 1997; **315:**1636.

[30] Swales I. The troublesome search for evidence: three cultures in need of integration. *J R Soc Med* 2000; **93:**402-7.

TACKLING BIAS IN ASSESSING THE EFFECTS OF HEALTH CARE INTERVENTIONS: EARLY CONTRIBUTIONS FROM JAMES LIND, ALEXANDER LESASSIER HAMILTON AND T. GRAHAM BALFOUR

I. Milne, Librarian, Royal College of Physicians of Edinburgh and I. Chalmers, Director, UK Cochrane Centre, NHS Research and Development Programme, Oxford

Whenever possible, the development of clinical guidelines and standards should be informed by reference to the results of reliable research assessing the effects of clinical interventions. There has been gradual acceptance of the need to control biases of various kinds in this evaluative research and to reduce the likelihood of being misled by the play of chance. It is widely assumed that the conceptualisation of the ways of achieving these two methodological objectives arose during the twentieth century. In this paper we draw attention to evidence that there was methodological awareness among three Edinburgh graduates much earlier than this.

Reliable clinical guidelines depend on reliable evidence about the effects of treatments. Reliable evidence implies that fair comparisons have been made between different treatments. Fairness requires that like be compared with like. If selection bias has led patients receiving one treatment to be systematically different from patients receiving an alternative treatment, any differences (or similarities) in outcomes may simply reflect differences in the prognoses of the two groups of patients before treatment and not the relative merits of the different treatments. However, this selection bias can be abolished by allocating patients to different treatments at random, or by using strict alternation or rotation in a consecutive series.[1]

There is a widespread belief that these methods of making fair treatment comparisons were first used in the middle of the twentieth century and that they reflected developments in the design of experiments introduced in the 1930s by the statistician Ronald Fisher.[2] In particular, it is believed by many that the celebrated Medical Research Council trial of streptomycin for pulmonary tuberculosis[3] was the first trial to use random allocation, and that this reflected Fisher's influence. Neither of these beliefs is supported by the evidence[1, 4] and this is made clear by the early twentieth century examples included on the Controlled Trials from History website (http://www.rcpe.ac.uk/controlled_trials/index.html).[5] Steps taken to ensure that like would be compared with like in therapeutic comparisons antedated Fisher by decades. One of the earliest well known examples is the trial of anti-diphtheria serum reported by Fibiger in 1898,[6] but there is evidence that clinical researchers were aware of the need to compare like with like in therapeutic comparisions much earlier than this. In this paper we draw attention to relevant passages from the writings of three Scottish graduates – one very well known, the other two hardly known at all.

JAMES LIND (1716–1794)

Our well known example, James Lind, holds an important place in any history of controlled trials. It was the suspicion that Lind's celebrated work was unlikely to be the only early example of British contributions to the development of a critical and quantified approach to therapeutic claims that stimulated Ulrich Tröhler to begin his research.[7] Lind began his career as an apothecary's apprentice, becoming a naval surgeon in 1739. After leaving the Navy, he graduated as a doctor of medicine in 1748 in Edinburgh, where he built up a successful medical practice and became a Fellow and Treasurer of the Royal College of Physicians of Edinburgh. In 1758 he was appointed Physician-in-Charge of the Royal Naval Hospital at Haslar.

Lind's 1753 *Treatise on the Scurvy* (Figure 1)[8] was written in Edinburgh and describes how (as a naval surgeon) he had selected 12 sailors with scurvy and given one of six different treatments to one of six groups of two patients. The two sailors who had been prescribed oranges and limes recovered much more quickly than the five pairs given the other treatments. Lind makes clear that he was aware of the need to compare like with like by noting that, before he administered the six treatments to the twelve patients: 'Their cases were as similar as I could have them . . . They lay together in one place . . . and had one diet common to all.'

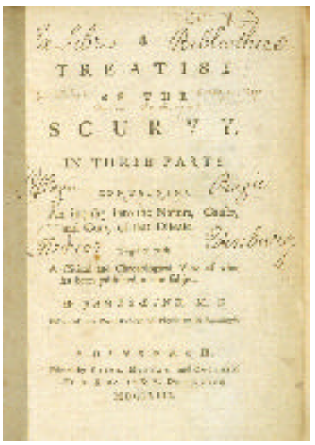
TREATISE
SCURVY.

FIGURE 1

Title page of Lind J. *A treatise of the scurvy*. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson; 1753.

ALEXANDER LESASSIER HAMILTON (1787–1839)

In 1816, a military surgeon with an Edinburgh background – Alexander Lesassier Hamilton – used his Edinburgh University MD thesis on fever (Figures 2 and 3) to describe an experiment that used rotation to assess the effects of bloodletting. Lessasier Hamilton had Edinburgh dynastic

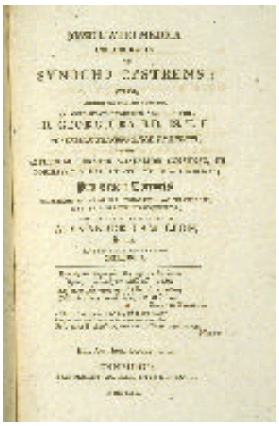

FIGURE 2
Title page of Hamilton's inaugural dissertation. Hamilton AL. *Disseratio Medica Inauguralis De Synocho Castrensi*. Edinburgh: J Ballantyne; 1816.

FIGURE 3
Page from the manuscript translation of *De Synocho Castrensi*. Royal College of Physicians of Edinburgh, Alexander Hamilton Collection 11/70.

connections. Alexander was the son of Pierre Lessasier (a French doctor) and Christina Hamilton (eldest daughter of the Edinburgh University Midwifery Professor, Alexander Hamilton). Lessasier Hamilton's uncle was Alexander Hamilton's son (and successor as Edinburgh University Midwifery Professor), James Hamilton, who was President of the Royal College of Physicians of Edinburgh between 1792 and 1794.

The experiment that Alexander Lesassier Hamilton reports in his MD thesis [9,10] took place in 1809, in Portugal, during the Peninsular War, at the hospital at Elvas, which involved 366 sick soldiers. In the experiment Lessasier Hamilton and two other army surgeons used rotation to generate comparable groups. The crucial paragraph (which can be found at the end of the thesis) reads:

> It had been so arranged, that this number was admitted, alternately, in such a manner that each of us had one third of the whole. The sick were indiscriminately received, and were attended as nearly as possible with the same care and accommodated with the same comforts. One third of the whole were soldiers of the 61st Regiment, the remainder of my own (the 42nd) Regiment. Neither Mr Anderson nor I ever once employed the lancet. He lost two, I four cases; whilst out of the other third [treated with bloodletting by the third surgeon] thirty five patients died.

How is it that this evidence has only recently come to light? Sometime in the 1830s, in a hiatus between two court actions, an Edinburgh solicitor, John Gibson, lodged a trunk containing Lesassier Hamilton's difficult-to-decipher diary and papers in the College.[11] There the trunk remained, undisturbed, but stored safely and listed in the Manuscript Catalogue until 1987, when the then College Archivist, Joy Pitman, catalogued it in detail and wrote about it in the College journal.[12] Ms Pitman uncovered a fascinating story that was so intriguing that an American historian, Lisa Rosner, has written a book based on the papers.[13] Among the manuscripts are papers about Lesassier Hamilton's Edinburgh medical studies, Army Medical Department service in the Peninsular War, graduation from Edinburgh, midwifery practice and his writing as a novelist! But that is not all there is in the archive. The court action was a divorce and it is easy to see why the solicitor representing Lesassier Hamilton's wife wanted to use the material: the papers contain an extraordinary record of his numerous extramarital adventures. As his numerous female partners, creditors and professional colleagues discovered, Dr Lesassier Hamilton was not always entirely reliable.

Lesassier Hamilton's comprehensive diaries do not include any mention of a bloodletting trial in Elvas. Did the trial take place? The evidence of Lesassier Hamilton's MD thesis must be weighed against the evidence of Lesassier Hamilton's diaries. Corroborative evidence also can be considered – there is evidence that Lesassier Hamilton's colleague, Mr Anderson, was at Elvas in 1809,[14] but against that is the fact that pamphlets recommending bloodletting were still being written by serving officers in 1813. Then again, the diaries show that Hamilton liked record keeping and that his work was appreciated by the director of medical services during the Peninsular War – the administrative reformer James McGrigor. Even if the account of the bloodletting trial was fabricated, however, it is still remarkable that Hamilton chose to describe the experiment in the terms that he did, particularly if he judged that his description of alternation and standard conditions would impress his examiners and other readers.

THOMAS GRAHAM BALFOUR (1813–1891)

Thomas Graham Balfour must have been aware of medicine from an early age as his Edinburgh family – the Balfours of Pilrig – had several medical connections.[15] His great-grandfather was the influential Scottish Enlightenment neurologist Robert Whytt, and his cousin, George William Balfour, became an eminent cardiologist and President of Royal College of Physicians of Edinburgh.[16] Balfour studied medicine at Edinburgh University where he was taught by William Pulteny Alison and Robert Christison. After graduating MD in 1834, he joined the army and went on to become Surgeon General, a Fellow of Royal College of Physicians of London, a President of the Royal Statistical Society and a Honorary Physician to the Queen.

Thomas Graham Balfour was appointed surgeon to the Duke of York's Asylum, Chelsea, in 1848. While he was there, there was an epidemic of scarlet fever and Dr Balfour used this opportunity to test the claim that homeopathic belladonna offered protection from the

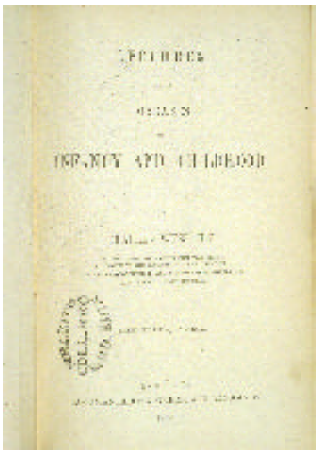

FIGURE 4
Title page of West C. *Lectures on the Diseases of Infancy and Childhood.* London: Longman, Brown, Green and Longmans; 1854.

disease. The report of his experiment appears in the chapter on scarlet fever in the 1854 edition of *Lectures on the Diseases of Infancy and Childhood* (Figure 4), a popular book authored by Charles West, founder of the Hospital for Sick Children in Great Ormond Street.[17] Balfour's account must rate as one of the most succinct and careful accounts of a clinical experiment every written:

> There were 151 boys of whom I had tolerably satisfactory evidence that they had not had scarlatina; I divided them in two sections, taking them alternately from the list, to prevent the imputation of selection. To the first section (76) 1 gave belladonna; to the second (75) 1 gave none; the result was that two in each section were attacked by the disease. The numbers are too small to justify deductions as to the prophylactic power of belladonna, but the observation is good, because it shows how apt we are to be misled by imperfect observation. Had I given the remedy to all the boys, I should probably have attributed to it the cessation of the epidemic.

In these four sentences, Balfour deals with the application of eligibility criteria, the control of selection bias, the problem of Type 2 statistical errors (i.e. false negatives), and the dangers of reliance on uncontrolled case series as a basis for causal inferences about the effects of treatment. It is perhaps particularly noteworthy that he drew attention to the fact that the low numbers of outcome events meant that there was insufficient data to justify firm conclusions – an issue that remains insufficiently appreciated even now.

CONCLUSION

The Hamilton and Balfour reports have been uncovered very recently. This raises the interesting possibility that other reports of studies in which steps were taken to ensure fair treatment comparisons await discovery. If such reports do exist, they are particularly likely to be found among the writings of Scottish graduates during the second half of the eighteenth century and the beginning of the nineteenth century for, as Ulrich Tröhler has shown, these men played a key role in introducing a quantified and evaluative approach in medicine.

In seeking and evaluating these and other documents relevant to the origins of critical thinking about therapeutic evaluation, it will be important to bear in mind the observations of John Vincent, Professor of History at the University of Bristol.[18]

> History is about evidence. It is also about other things: hunches, imagination, interpretation, guesswork. First and foremost, though, comes evidence: no evidence, no history. It is also about intrinsically fallible evidence. In this it resembles medicine and the detection of crime. And it is about fallible evidence as interpreted by fallible people; hence no question of finality can ever arise . . . There is a bias in the creation of evidence, and a bias in the survival of evidence. There may be a bias in access to what survives, too . . .

REFERENCES

1 Chalmers I. Why transition from alternation to randomisation in clinical trials was made. *BMJ* 1999; **319:**1372.
2 Fisher RA. The design of experiments. Edinburgh: Oliver and Boyd; 1935.
3 Medical Research Council. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 1948; **2:**769-82.
4 Chalmers I. Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. *International Journal of Epidemiology* (in press).
5 Chalmers I, Milne I, Douglas S, editors. Controlled Trials from History. www.rcpe.ac.uk/controlled_trials
6 Fibiger J. Om serumbehandling af difteri [On treatment of diptheria with serum]. *Hospitalstidende* 1898; **6:**309-25.
7 Tröhler U. *"To Improve the evidence of medicine": the 18th century British origins of a critical approach*. Edinburgh: Royal College of Physicians of Edinburgh; 2000.
8 Lind J. *A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease.* Together with a critical and chronological view of what has been published on the subject. Edinburgh: Kincaid and Donaldson; 1753.
9 Hamilton AL. *Disseratio Medica Inauguralis De Synocho Castrensi*. Edinburgh: J Ballantyne; 1816.
10 Hamilton AL. *Disseratio Medica Inauguralis De Synocho Castrensi* manuscript translation. Royal College of Physicians of Edinburgh. Alexander Hamilton collection [11/70].
11 Hamilton A. Manuscript journals. Royal College of Physicians of Edinburgh. Alexander Hamilton Collection [11/72].
12 Pitman J. *Alexander (Lessasier) Hamilton: an eventful life. Proc R Coll Physcians Edinb* 1987; **17(4):**286-90.
13 Rosner L. *The most beautiful man in existence: the scandalous life of Alexander Lesassier*. Philadelphia: University of Pennsylvania Press; 1999.
14 Anderson A. *My dear Cath…the story of an army doctors wife*. Selkirk: W. Thomson; 1976.
15 Surgeon General T. Graham Balfour. Obituary. *Lancet* 1891; **1:**228-9.
16 Milne I. George William Balfour in New Dictionary of National Biography. Oxford: Oxford University Press (in press).
17 Balfour TG. Quoted in: West C. *Lectures on the Diseases of Infancy and Childhood*. London: Longman, Brown, Green and Longmans; 1854; 600.
18 John Vincent. *An intelligent person's guide to history*. London: Duckworth; 1995; 52.

THE SCOTTISH INTERCOLLEGIATE GUIDELINES NETWORK (SIGN): DEVELOPING NATIONAL GUIDELINES TO AN INTERNATIONAL METHODOLOGY

J. Miller, Director of SIGN, Edinburgh

Being invited to talk about the work of the Scottish Intercollegiate Guidelines Network (SIGN) at the Clinical Effectiveness symposium in November 2000 was a somewhat daunting experience, the audience comprising a veritable 'who's who' in evidence-based medicine. However, the advantage of having such a knowledgeable audience was that, rather than explaining *what* SIGN does, I had the opportunity instead to highlight some of the more recent developments that we have made in our guideline development methodology. Or rather, I should say our *application* of guideline development methodology, because, in a sense, the central theme of this presentation was that guideline methodology is something that we all share and all contribute to.

INTRODUCTION

The Scottish Intercollegiate Guidelines Network is an essentially collaborative initiative, involving what can be seen as an unprecedented level of cooperation in Scotland between the medical specialties, between doctors and other health care disciplines, and between the professions which 'own' SIGN and the Scottish Executive Health Department which funds it. The membership of SIGN now extends to over 45 individuals representing some 35 Colleges and other professional associations. Of course, there are potential difficulties in managing the competing interests which these diverse groups bring with them, and in this context I must pay tribute to role that the Chairman of SIGN, Professor Jim Petrie, has played in steering the development of the organisation. But equally important is the positive approach and strong sense of common purpose towards achieving improved health care for patients in Scotland which all those involved bring to SIGN. In particular, I would like to acknowledge the enormous contribution to the methodological developments described here by three longstanding members of the SIGN team: Mr Robin Harbour, Dr Moray Nairn and Dr Safia Qureshi.

BACKGROUND

Where did we start and why?

The Scottish Intercollegiate Guidelines Network was established in 1993, one of a number of national guideline programmes created at this time. But why was the introduction of national guidelines thought to be so important and so necessary? The main reasons, at least in Scotland, were to address variation in practice and outcome in the National Health Service (NHS), and to maximise the validity of the guidelines used within the Scottish health service. With hindsight, I think we would add a third important aspect: to avoid duplication of effort. I'll return to this point in a little while, but I want to focus here on the national and local dimensions of guideline development and implementation in Scotland. In particular, it is important to emphasise that, although SIGN is responsible for the development of national guidelines, we are not and cannot be responsible for their implementation into practice. This is a responsibility of each individual NHS Trust, and is now reinforced by the twin 'levers' of clinical governance, that is, the accountability of Trust Chief Executives (for clinical quality) and the Clinical Standards Board for Scotland (which is responsible for setting and monitoring standards of care and accreditation of clinical services in the NHS in Scotland).

But SIGN's remit being limited to guideline development doesn't mean that we aren't concerned about implementation. We recognise that guideline development is only one element in a much larger process and we are keen to take a more active role in facilitating local implementation activities where possible, if only by extending the 'N' of SIGN to provide a networking forum to promote sharing of good practice in implementation.

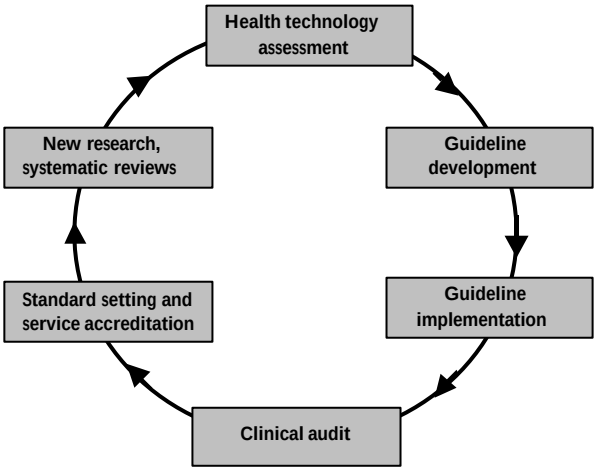

FIGURE 1
The clinical effectiveness cycle.

Similarly, guidelines themselves are of course only part of a series of initiatives to promote improved patient care in the NHS in Scotland. This could also be shown as a cycle; and strengthening the links between the various activities shown in Figure 1 is one of the main challenges for us all in the future.

WHAT HAS SIGN ACHIEVED?

Fifty guidelines and an international reputation . . .

But before we start thinking about future development, I'd like to discuss what we have achieved in SIGN up to now. In particular, I want to emphasise the essentially *practical* nature of this. Fifty SIGN guidelines have now been published, and these 50, and the others in production or under review, have all been accepted into the SIGN programme following a rigorous selection process designed to establish the existence of both evidence of variation in

practice and outcome, and the availability of evidence of effective practice on which to base the guideline recommendations. And, most importantly, the need for these guidelines has been identified from *within* the health service – individual practitioners or specialist societies have come to us with evidence of a problem and it is in response to this identified need that SIGN guidelines are developed. (For more information about the SIGN programme, visit our website: www.sign.ac.uk.)

The key elements of guideline development methodology are very simple, and very powerful. Whatever the context, the resources or subject matter, the basic principles of evidence-based guideline development can be applied, that is:

- involve representatives of all the relevant stakeholders in developing recommendations for change in practice;
- use the evidence available, but identify it systematically, and evaluate it critically; and
- make the basis of the recommendations – whether or not they are based on evidence – explicit and transparent.

Unfortunately, however, the simplicity and obvious appeal of these principles doesn't make them any easier to apply in practice. Over the past five years, the SIGN team have therefore put a lot of effort into structuring the guideline development process in order to guide our multidisciplinary development groups through the many steps involved in defining the questions to be addressed by the guideline, agreeing the scope of the literature search and appropriate selection criteria, assessing the methodological quality of the evidence and synthesising and summarising the findings. And that's just to consider the more objective elements of the process. The outcome of this systematic review will be an evidence table or some other method of collating information on all the validated studies which address a particular question. But the evidence doesn't speak for itself. Nor, if it could, would it speak with one voice. Thus, after (in the case of a SIGN guideline) up to a year of wrestling with the evidence, the hard part is yet to come: the multidisciplinary guideline development group have still to apply their considered professional judgement to reach agreement on what recommendations the evidence base – incomplete and inconsistent as it may well be – can support.

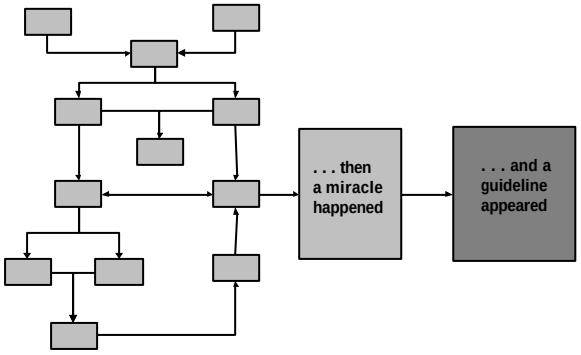

FIGURE 2
Early guideline development.

The problem which SIGN, and many guideline developers, were faced with until a few years ago is illustrated in Figure 2. Despite a sincere commitment to develop valid evidence-based guidelines, even if the guideline was developed according to methodologically robust methods, we would never have been able to prove this, so opaque was the decision-making process. We therefore realised – and thankfully, our funders also accepted – that we would need to manage the guideline development process far more proactively, bringing more of the support activities and project facilitation in-house, in order not only to ensure that the guideline was developed according to a robust methodology, but that the process was fully documented, and quality assured – with, for example, a detailed independently reviewed literature search, standard checklists for methodological appraisal of the evidence and evidence tables compiled including details of all the validated studies, as illustrated in Figure 3.

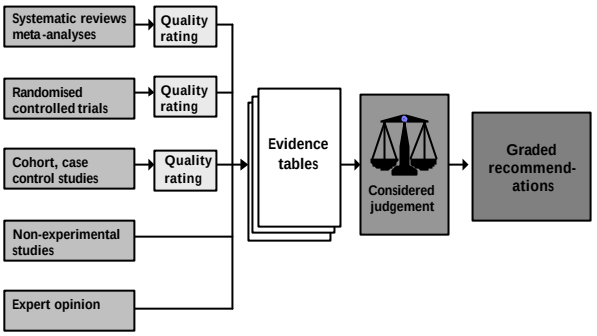

FIGURE 3
Structured guideline development.

We still had a problem, though, with that final, more subjective, stage of the process, for which we have coined the term 'considered judgement'. What we're talking about here is the judgement that the multidisciplinary group have to make about the quality, quantity, applicability and generalisability of the evidence. Added to that, there may be considerations about clinical impact, cost-effectiveness and practicality of the recommendations.

Now, we haven't solved this one yet. We've come up with a method of ensuring that SIGN guideline development groups record how they have considered each of these factors in forming the recommendations, in order to make the process as transparent as possible. But there is clearly more work to be done on how to incorporate these factors, appropriately weighted, into the process of forming guideline recommendations. Like other guideline developers, SIGN has a 'hit list' of aspects of our methodology and procedures which we need, with varying degrees of urgency, to address. But we are well aware (and relieved to know) that we don't have to do this in isolation – many of the issues we are struggling with may already have been tackled elsewhere, and ready-made solutions may be available from other guideline developers, even if a little tweaking to fit our systems may be needed. But if, as I mentioned before, our methodologies for guideline development are so similar, shouldn't we be thinking about how we can share the output of the process, as well as the methodology, in order to avoid unnecessary duplication of effort?

A VISION AND A CHALLENGE FOR THE FUTURE

It's time to start sharing . . .

A key development supporting the potential for greater cooperation in guideline development is the AGREE collaboration, which is in the final stages of producing a common guideline appraisal instrument. The project was initially conceived for Europe but the collaboration has already grown to include involvement from Canada, the US and New Zealand. The AGREE instrument contains 23 validated criteria against which users can assess the quality of clinical practice guidelines (for further details, see the AGREE website: www.agreecollaboration.org). Turn an appraisal instrument on its head – by which I mean, use it *proactively* to guide the development process, rather than retrospectively to assess the finished product – and you have a standard for guideline development. So, assuming we accept and apply the AGREE criteria, in future, not only will SIGN and other guideline programmes in Europe and beyond be developing guidelines following the same basic methodology, but these will also meet an internationally agreed standard. Then all that is needed to add to this is to ensure that the process is fully documented and to make that documentation accessible, and that we be able to use (or adapt) elements of each others' work, just in the way that we use the systematic reviews undertaken by, for example, the Cochrane Collaboration, as quality-assured building blocks for our own literature reviews.

Figure 4 shows the audit trail that SIGN is in the process of building up for the guidelines in our programme. The format in which this information is recorded may differ, but if this type of documentation were to be made widely available, not only would guideline users be able to trace the derivation of the recommendations, but guideline developers would be able to take elements of each others' work and 'unpick' the process in order to adopt or adapt appropriate elements of it into their own guidelines (with potentially significant resource savings, if we consider the year-long systematic review process mentioned earlier).

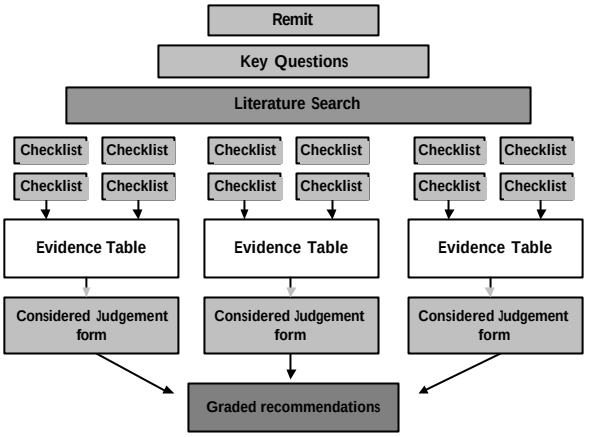

FIGURE 4
The guideline audit trail.

In conclusion, I should stress that I am certainly not attacking the principle of individual guideline programmes developing their own national or regional guidelines. Quite the opposite: the more local involvement in guideline development, the more relevant the recommendations will be to the working practices of potential users, and the more ownership they will feel over the guidelines – all of which have been shown to increase the likelihood of successful implementation into practice. But if we can share appropriate elements of the development process between our various national and regional guideline programmes, then, as well as creating our own guidelines, we will be contributing to an immensely valuable data resource for guideline developers around the world, and so maximising the potential benefit to all patients from our investment in guidelines.

THE ESTABLISHMENT OF THE NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE (NICE)

P. Littlejohns, Clinical Director, NICE, London

INTRODUCTION

The National Institute for Clinical Excellence (NICE) was established as a Special Health Authority in April 1999. Its role is to provide guidance to the National Health Service (NHS) in England and Wales on the clinical and cost effectiveness of clinical interventions. It achieves this through appraising new and existing technologies, developing clinical guidelines and supporting clinical audit. It is part of the government's approach towards improving the quality of the NHS outlined in the White Paper *The new NHS: modern and dependable*[1] with further details provided in *A first class service: quality in the new NHS*.[2] Standards are set at a national level for service configurations through the creation of National Service Frameworks and for clinical issues through the guidance issued by NICE.[3] The aim of the Institute's work programme is to reduce inappropriate variation in clinical practice in the NHS.

While its responsibilities are limited to the creation and dissemination of guidance it is also keen to support the implementation of its products. The evidence on how to encourage the use of guidance has identified that a range of reinforcement strategies are required to support implementation. This approach is reflected in the changes put in place on a nationwide basis seeking to encourage local ownership and responsibility. Renewed emphasis has been placed on establishing effective professional self-regulation and making continuing professional development an individual and collective professional responsibility. Managerial commitment to quality improvement is sought through the development of clinical governance. Research has also identified the role of monitoring and feedback and this is reflected in the establishment of a range of new mechanisms, which included creating the Commission for Health Improvement (CHI), and working within a National Performance Framework (see Figure 1). While the approach was initially presented as a 'Quality Improvement' model with education and support being the key driving forces, over the last year mixed messages have been expressed by politicians. Breakdown of public and professional confidence in some of the current systems in place to assure professional standards has resulted in tougher messages from the government.

However, the institutions involved – CHI, NICE and the National Clinical Governance Support Team – have gone out of their way to emphasise that their role is to support the NHS in addressing the daily challenges of delivering health care. This is particularly important when expectations and demand frequently exceed what is likely to be feasible within current service configurations. In this context the Institute views its main role as the provision of guidance on controversial health issues when there was evidence that the lack of clarity had resulted in inappropriate variation in the care provided by the NHS. This guidance should then be incorporated into local clinical governance mechanisms via its use in local guidelines and protocols.

The National Institute of Clinical Excellence as a special Health Authority has a board consisting of executive and non-executive members, its chairman being appointed by the Secretary of State for Health. It meets in public (peripatetically around the country) every two months.

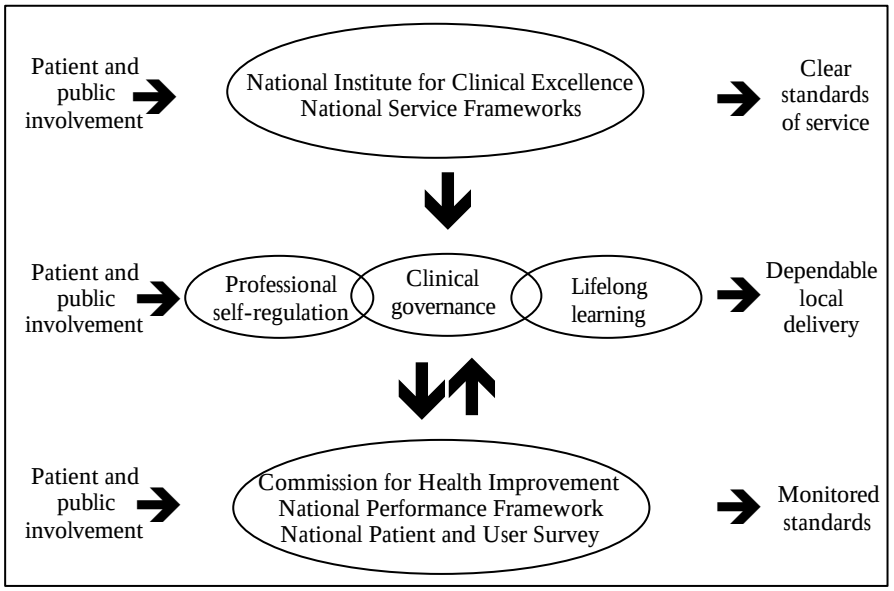

FIGURE 1
Setting, delivering and monitoring standards.

Details of all the Institute's activities are available on its website (www.nice.org.uk). Its specific guidance is incorporated into the broader organisational standards set by the National Service Frameworks. It is a small organisation (36 employees) and undertakes its work by commissioning and liaising with a range of professional, specialist and patient organisations. It is supported by its Partner's Council (which includes representatives from all its stakeholders, including the health care industries), a series of advisory committees and has formal links with a number of universities and the National Research and Development Programme. It works closely with local trusts and clinical governance professionals to ensure support for those responsible for implementing its guidance. This includes providing audit advice to accompany its guidance.

THE NICE APPROACH

The details of NICE's main work programmes for 1999/2000 and 2000/2001 have been published.

Appraising health technologies

In the past the Department of Health (DoH) and the National Assembly for Wales selected technologies for appraisal by NICE based on a number of criteria:

- Is the technology likely to result in a significant health benefit, taken across the NHS as a whole, if given to all patients for whom it is indicated?
- Is the technology likely to result in a significant impact on other health-related government policies (e.g. reduction in health inequalities)?
- Is the technology likely to have a significant impact on NHS resources (financial or other) if given to all patients for whom it is indicated?
- Is NICE likely to be able to add value by issuing national guidance? For instance, in the absence of such guidance is there likely to be significant controversy over the interpretation or significance of the available evidence on clinical and cost effectiveness?

However, a new approach is being developed in which the Institute has a more active role in defining its work programme. This consists of 'mapping' the clinical priority areas with each of the national disease specific clinical directors.

The National Institute of Clinical Excellence follows a transparent and structured process for its appraisals (outlined in Figure 2), giving appropriate interested parties the opportunity to submit evidence, to comment on draft conclusions, and to appeal to a panel independent of those involved in the original judgement in cases where NICE is alleged to have: failed to act fairly; to have exceeded its powers; or to have acted perversely in the light of the evidence submitted. This is a dynamic process and the approach has recently been reviewed in the light of the experiences of the first year of appraisals. In the future the first deliberations of the appraisal committee will also be placed on the website.

The National Institute of Clinical Excellence's function in relation to appraisals, as set out in the Secretary of State's directions, is 'to appraise the clinical benefits and the costs of such health care interventions and to make recommendations'. It assesses the evidence of all the clinical and other health-related benefits of an intervention, taking this in its widest sense to include impact on quality of life, relief of pain or disability etc. as well as any impact on likely length of life – to estimate the associated costs, and to reach a judgement on whether on balance this intervention can be recommended as a cost-effective use of NHS resources (in general or for specific indications, subgroups etc.). Where there is already a cost-effective intervention for the condition, the appraisal should appraise the net impact on both benefits and costs of the new intervention relative to this benchmark. The Institute's

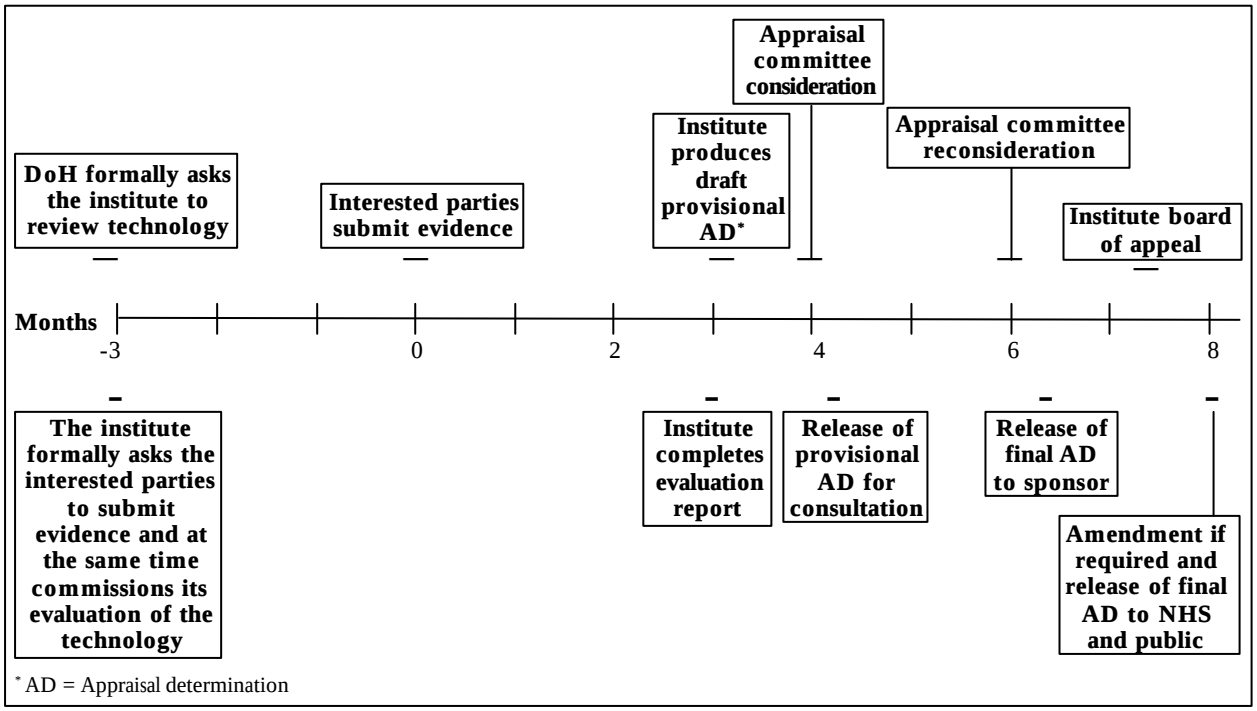

FIGURE 2
Diagrammatic representation of the appraisal process.

position in relation to cost-effective methodology has recently been published after a series of workshops bringing together experts in the field.

The National Institute of Clinical Excellence is also required to ensure that, in carrying out its statutory functions, it is sympathetic to the longer-term interest of the NHS in encouraging innovation.

Evaluation documentation is commissioned from expert groups (working closely with the Health Technology Assessment arm of the National Research and Development Programme). This report is combined with submissions from patient groups, professional organisations and industry to form an Assessment Report that is given to the Appraisals Committee (Figure 3). This committee carries out the appraisals.

FIGURE 3
The Institute's technology appointments advisory committee.

Chairman
Professor David Barnett

Membership

- Vice chair x 1
- Community nurse x 1
- Health economists x 3
- Pharmacist x 1
- Biostatistician x 1
- Patient advocates x 2
- Health managers x 3
- Hospital physicans x 2
- Pharmaceutical physician x 1
- Surgeon x 1
- Diagnostic pathologist x 1
- General practitioners x 2
- Public health physician x 1
- Hospital nurse x 1

Additional ad hoc *members may be appointed 'for the day' by either the chairman, the vice-chairman or the chief executive*

In view of the increased workload generated by the addition of all new cancer drugs to the appraisal programme, the Institute has created an additional committee working under the same chairman. The National Institute of Clinical Excellence produces guidance to commissioners and clinicians on the appropriate use of the intervention alongside current best practice. This guidance covers:

- an assessment of whether or not the intervention can be recommended as clinically effective and as a cost-effective use of NHS resources for NHS use, either in general or in particular circumstances (first or second line treatment, for particular subgroups, for routine use or only in the context of targeted research where appropriate, any priorities for treatment, recommendations on any questions requiring further research to inform clinical practice etc.);
- an assessment of any wider implications for the NHS; and
- a concise summary of the reasoning behind NICE's recommendations and the evidence considered.

The National Institute of Clinical Excellence also prepares guidance for users and carers, consulting with appropriate patient groups on the best format and means of dissemination. This guidance explains the nature of the clinical recommendations, the implications for the standards which patients can expect and the broad nature of the evidence on which the recommendations are based.

Clinical guidelines

The Institute is also charged with developing and disseminating 'robust and authoritative' clinical guidelines. In constructing its clinical guidelines, the Institute is expected to take into account both clinical and cost effectiveness. Where relevant, the Institute seeks to produce parallel clinical guidelines for patients and their carers.

KEY PRINCIPLES

Ten key principles underpin the way in which the Institute handles clinical guideline developments on behalf of the NHS. While there will be many differences between the various kinds of clinical guidelines produced by the Institute, the key principles should be relevant to our approach to all clinical guideline developments.

TEN KEY PRINCIPLES FOR NHS CLINICAL GUIDELINES.

- The objective of clinical guidelines is to improve the quality of clinical care by making available to health professionals and patients well-founded advice on best practice.
- Quality care is based on clinical effectiveness – the extent to which the health status of patients can be expected to be enhanced by clinical interventions.
- Quality of care in the NHS necessarily includes giving due attention to the cost effectiveness of health care interventions.
- The National Health Service clinical guidelines are relevant to the care provided by the NHS throughout England and Wales.
- The National Health Service clinical guidelines are *advisory*.
- The National Health Service clinical guidelines are based on the best possible research evidence, expert opinion and professional consensus.
- The National Health Service clinical guidelines are developed using methods that command the respect of patients, the NHS and NHS stakeholders.
- While clinical guidelines are focused around the clinical care provided by clinicians, patients are to be treated as full and equal partners along with the relevant professional groups involved in a clinical guideline development.
- All those who might be affected by a clinical guideline deserve consideration within the clinical guideline development (usually including clinicians, patients and their carers, service managers, the wider public, government and the health care industries).
- NHS clinical guidelines should be both ambitious and realistic in nature. They should set out the clinical care that might reasonably be expected throughout the NHS.

The methododology that the Institute will use to commission its guidelines has been finalised after extensive consultation and is available on the website. It has involved the establishment of six national collaborating centres based on Royal Colleges to produce the guidelines. In addition, two further units have been commissioned. Firstly, a technical support unit to work alongside the collaborating centres to facilitate the development of guidelines. This unit will also initiate a research and development programme into guideline methodology, concentrating on how to produce guidelines that address cost effectiveness as well as clinical effectiveness. The second unit will support patient and carer involvement in all aspects of guideline development.

Referral advice

One of the key reasons for the establishment of NICE was to reduce variation in the quality of care provided by the NHS. Variation of the quality of care can be manifested in many ways and one important area is access from primary care into specialist services. There is now considerable research describing the variability in general practitioner outpatient referral rates, but less understanding on what the underlying reasons are. Key to an effective and efficient health service is the appropriate and timely referral of those patients who will benefit from specialist intervention. In October 1999 the DoH and the National Assembly for Wales invited NICE, in addition to its other programmes, to produce an initial set of out-patient referral advice. These would offer advice to general practitioners on when to refer patients to specialists. The guidance was to be available by April 2000 and a range of topics was proposed. These included atopic eczema in children, acne, psoriasis, acute low back pain, osteoarthritis of the hip, osteoarthritis of the knee, glue ear in children, recurrent episodes of acute sore throat in children, dyspepsia, varicose veins, urinary tract (outflow) symptoms and menorrhagia.

This was a challenging timetable in a field with little research evidence on how this should be approached. Drawing on experience of developing local protocols in both the south Thames region and Newcastle, a methodology was designed that adhered to the Institute's generic principles of rigor, transparency and inclusion of all stakeholders.

A steering group was convened to oversee the project. Advisory groups were created to modify and adapt advice created by the NICE project team. Each group consisted of general practitioners, specialists and patient advocates. These documents are designed to provide advice on when patients should be referred. They are not clinical guidelines on how to manage patients. In the future, when the Institute's guidelines programme is established, all of its clinical guidelines will include 'referral advice'. This guidance is currently being piloted nationally and will be rolled out to all Primary Care Groups during 2001.

Service Guidance

In 1998 the Chief Medical Officers for England and Wales published a report on the appropriate configuration of cancer services. The Clinical Outcomes group within the DoH commissioned Professor Bob Howard to create a number of cancer-specific evidence-based service configuration models in order to implement the Calman Hine report. The first guidance for commissioners was for breast cancer which is currently being updated. The National Institute of Clinical Excellence has inherited the responsibility for carrying this work forward in conjunction with Leeds University. In future this service guidance will be published as part of the NICE portfolio.

Other NICE initiatives

In addition to the above main strands of work the Institute has also taken responsibility for supporting clinical audit in the NHS and a number of major programmes. These include the National Confidential Enquiries and Prodigy computer support system. There are currently four enquiries, and the Institute commissioned an independent external review by Professor Sir John Grimley Evans who reported in June 2000. His brief was to review their past, present and future contributions to health; to assess whether the current organisational arrangements represent best value for money; whether the anonymity provided to participating health professionals is a necessary component; whether the methodological approaches require change in order to enhance their contribution to public health; and whether they should be extended to other areas of clinical practice.

An implementation committee was established under the chairmanship of Professor Peter Littlejohns and changes will be introduced in 2001/2. The aim is to build on the strengths of the individual enquiries whilst allowing the sharing of good practice and a more efficient use of common resources. Particularly important is the need to support dissemination of the conclusions and ensure that they influence clinical practice.

Prodigy is a computerised prescribing system in primary care that is being rolled out nationally to all general practitioners. The Institute is working closely with the creators, the Sowerby Informatics Centre in the University of Newcastle (www.schin.ncl.ac.uk) to create robust guidance on the prescribing of drugs.

EARLY LESSONS LEARNT

The first 18 months of the Institute have been controversial. It was inevitable that in seeking to change longstanding clinical practice that some decisions would 'ruffle the feathers' of individual stakeholders. The initial decisions on the use of zanamivir for the treatment of influenza upset the pharmaceutical manufacturing industry, the subsequent update upset some sections of general practice.[4] Patient and professional groups responded vociferously against the Institute's guidance on beta-interferons.[5] The cautious approach the Institute took to laparoscopic surgical interventions in treatment of hernias and colorectal surgery was not welcomed by the enthusiasts within the profession. The general role of the Institute in contributing to reducing variation in health care has also being questioned.[6] The debate about the appropriate balance between central and local responsibilities is the most important and is likely to continue, not only in terms of the 'prioritisation' issue, but also interestingly on the role of evidence, its interpretation and its contribution to health care.

However, there has also been considerable support. Organisations throughout the world seeking to address similar issues to NICE have sought further information on how the Institute functions. Following representations from France, Germany, the US, Canada, New Zealand, Greece, Italy, Sweden and Japan there is an ongoing

dialogue with a view to establishing collaborative working. The Institute takes an active role in the AGREE collaboration in order to encourage the dissemination of good guideline development practice.[7]

The government itself has demonstrated its support for the Institute through a series of ministerial speeches and responses in the house during Prime Minister's Question Time. It has increased the Institute's annual budget by 20% to allow expansion of its work programme. Ironically, the strength of this support has occasionally had a detrimental effect, as both the public and professionals have questioned the independence of the Institute. On a number of occasions the Institute has deemed it necessary to re-affirm its independence in keeping with its status as a NHS special health authority and that it issues 'guidance' not 'directives'. At the beginning of all guidance issued is the following statement:

The guidance represents the views of the Institute's Appraisal Committee (or Guideline Development groups) which was arrived at after careful consideration of the available evidence. Health professionals are expected to take it fully into account when exercising their clinical judgment. The guidance does not, however, override the individual responsibility of health professionals to make appropriate decisions in the circumstances of the individual patient, in consultation with the patient.[8]

The Institute is now established. In the future it will assess its performance in terms of whether health care professionals and their patients view its output as being useful to them on a day-to-day basis.

REFERENCES

1 Department of Health. *The new NHS: modern and dependable*. London: HSMO, 1997.
2 Department of Health. *A first class service: quality in the new NHS*. London: HSMO, 1998.
3 Rawlins M. *In pursuit of quality: the National Institute for Clinical Excellence*. *Lancet* 1999; **353:**1079-82.
4 Smith R. The failings of NICE. *BMJ* 2000; **321:**1363-4 and reply from Chairman *BMJ* 2000; **321:**489-91.
5 Ellis SJ. Doctors treating patients with multiple sclerosis will lose confidence in NICE. *BMJ* 2000; **321:**489-91.
6 Sculpher M, Drummond M, O'Brien B. Effectiveness, efficiency, and NICE. *BMJ* 2000; **322:**943-4.
7 www.agreecollaboration.org.
8 *Compilation summary of guidance issued to the NHS in England and Wales*. Vol. 2. London: NICE, 2001.

COUNCIL OF EUROPE – GUIDELINE DEVELOPMENT IN EUROPE

M. Mäkelä, Research Professor, FinOHTA, Finland

The Council of Europe nominated in 1999 an expert group to prepare guidelines for producing clinical practice guidelines. The work was completed in December 2000, and the recommendations of the group were submitted for political acceptance at the meeting of the European Health Committee (CDSP). The work has raised considerable interest in the 41 member states of the Council of Europe, and there is a clear need for guidance in this area.

The Expert Group work resulted in a three-part document. The first part is a recommendation to be adopted and presented by the committee of ministers as their statement on the matter. This part contains a number of fairly flexible recommendations to the governments of the member states, for example that they should 'adopt policies and create structures that support the production and use of nationally relevant, evidence-based guidelines targeting important issues in health care'.

In the Appendix to these recommendations, more specific statements on how guidelines should be optimally developed and implemented are listed. These suggest more specific tasks; for example, that 'Guidelines should be produced by multi-professional groups in a systematic, independent and transparent fashion, using appropriate quality criteria'. The main statements are listed in Table 1 (see over).

The third and most extensive part of the work is the explanatory memorandum, which also will be adopted by the Committee of Ministers. This is based on literature as well as the experience drawn from several of the most active national guideline programmes in Europe. The memorandum is structured around the various stages of guideline production: topic selection, development, dissemination, implementation, evaluation and updating.

The expert group relied on previous high quality documents, such as the groundbreaking work by the Institute of Medicine on essential features of good guidelines.[1] Results from recent European collaborative projects were used, including the work done on guideline implementation by the CPP (Changing Professional Practice) project[2] and the internationally created and validated instrument for appraising guidelines by the AGREE collaboration.[3, 4, 5]

A section describing the current state of guideline development in Europe is based on a ProGuide survey carried out in 1999. In most European countries guidelines are produced and used in a multitude of manners.[6] Several governmental, non-governmental and scientific organisations are involved in the development, dissemination and implementation of guidelines. A few countries have programmes for producing and updating guidelines, but in most cases these activities are widespread and rather unsystematic. A study looking in more detail at the work done by guideline development programmes and guideline clearing houses will provide up-to-date information of activities this year.[4]

Among the main points in the discussions were the need for ready availability of guidelines in daily practice, easy readability, and practical applicability. Professional ownership of the guideline was seen as a crucial basis to effective implementation. In many guideline programmes, patient versions are based on and co-produced with the professional version. Educated patients do bring guidelines and other materials to their physicians for discussion. The expert group saw an important role for professionals in interpreting guideline recommendations to individuals and adapting them skilfully into specific patient situations.

Merely developing a guideline, no matter how well it is done, was considered a waste of resources unless the other stages were also planned and secured. To facilitate expert work, sufficient support structures for literature searching, critical appraisal methodologies, statistical expertise, editing and distribution need to be ascertained and provided, both for the primary development and the regular updates of each guideline. But guideline dissemination, implementation and evaluation also require many types of resources: printed information, facilitators, training programmes etc. In very few countries do the guideline programmes include funding for the whole process from production to evaluation.

The expert group easily agreed on technical questions about guidelines. The most complex topics for discussion were legal issues and the relationship between guidelines and clinical standards. This is understandable as the legal and political environments and the implementation of guidelines in the different European countries vary notably. Patient participation, cost-effectiveness and evaluation were also discussed extensively in the group.

Recommendations on guideline development will hopefully be used by the member states of the Council of Europe to support health policy formulation. One of the political recommendations will very likely be to 'support international networking between guideline organisations, clearing houses, and other agencies producing evidence-based medical information'.[7] Thus there will be increasing need for collaboration between national guideline organisations. During the process of developing the recommendations, as well as in some scientific projects, the guideline houses have established fruitful personal contacts and are benchmarking each other's methods.

TABLE 1
Appendix to the recommendation.

I Guidelines in support of health care

- The main aim of clinical practice guidelines is to support and promote good clinical practice.
- Guidelines are produced and used in a complex environment of ethical, economic, legal and other aspects; these aspects need to be taken into consideration in each country.

II Topic selection

- Guideline topics should be selected for development to support and assist decision-making on important issues in health care.
- Prioritisation of guideline topics may be based on incidence of health problems, health inequalities, variations in quality of care, emergence of new technologies, or other factors that create a need for high-quality, updated information.
- The existence of previous evidence-based guidelines should be considered in the prioritisation of topics for development.

III Guidelines development

- Guidelines should be produced by multi-professional groups in a systematic, independent and transparent fashion, using appropriate quality criteria.
- End user involvement through a wide review and/or testing of the pilot version is necessary before adopting a guideline for implementation.
- If guidelines are adapted from other countries or areas, they must be re-edited and reviewed or tested for applicability in the new environment.

IV Dissemination of guidelines

- The funding for guideline dissemination, implementation, evaluation and updating must be carefully considered at the same time as the decision is made to develop the guideline.
- Guidelines should target multiple audiences (professionals, patients and policy-makers) and be available in suitable formats for these different groups.
- Guideline dissemination should be planned, active, sustainable, and ensure high accessibility.
- Guideline clearing houses or guideline production programmes facilitate the accessibility of multiple guidelines and may increase guideline quality.

V Guideline implementation

- For the most effective implementation of guidelines, a systematic approach to managing the quality of health care is essential.
- Various guideline dissemination and implementation strategies should be used in combinations to ensure maximum effect.
- Professional, organisational, financial and regulatory incentives and disincentives need to be considered together with other barriers and facilitators of guideline use on both national and local levels (tailored implementation).
- Guidelines must become an essential element in the clinical training of health care professionals as well as in the continuous training of health care teams.

VI Evaluation of guidelines and of their impact

- Tools for evaluating the quality of existing guidelines should be used to decide which guidelines should be implemented.
- Well-planned monitoring of guideline effects is essential, and the impact of guidelines on health outcomes especially needs further development and evaluation.
- Guidelines can include a list of essential indicators that can be used for evaluating the results of guideline implementation.

VII Updating

- The guideline production process must include clear policies on guideline updating.

REFERENCES

1 Field MJ, Lohr KN, editors. *Institute of Medicine: Guidelines for clinical practice*. From development to use. Washington DC: National Academy Press; 1992.

2 Thorsen T, Mäkelä M, editors. *Changing Professional Practice. Theory and practice of clinical guidelines implementation*. Copenhagen: DSI; 1999.

3 The Appraisal of Guidelines, Research and Evaluation in Europe (AGREE) Collaborative group: Guideline development in Europe: An international comparison. *Int J Tech Ass Health Care* 2000; **16:**1039-49.

4 See www.agreecollaboration.org for details.

5 Cluzeau F, Littlejohns P, Grimshaw J *et al*. Development and application of a generic methodology to assess the quality of clinical guidelines. *International Journal for Quality Health Care*, 1999; **11(1):**23-8.

6 Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; **28(20):**1900-5.

7 Developing a methodology for drawing up guidelines on best medical practices. Recommendation and explanatory memorandum. Council of Europe Draft, June 2001.

THE GERMAN GUIDELINES CLEARING HOUSE (GGC) – RATIONALE, AIMS AND RESULTS

G. Ollenschläger, C. Thomeczek, H. Kirchner, U. Oesingmann, F. W. Kolkmann, Agency for Quality in Medicine, Cologne, Germany; R. Kunz, Expert Group Hypertension of the German Guideline Clearing House, Cologne, Germany

INTRODUCTION

The role of clinical practice guidelines (CPGs) as a tool for quality management in health care is now widely accepted in Germany – not only by health professionals, but also in politics. As a consequence, the federal parliament passed a new social code in 1999, stating that 'every year, quality indicators for ten priority health problems – developed from evidence-based guidelines – have to be implemented on a national level'.[1]

However, the number and quality of the available CPGs has become an issue of concern since the German scientific medical societies started their CPG programme in 1995. This is because the majority of the more than 900 CPGs accessible through the world wide web (www.awmf-leitlinien.de) relied on expert opinion more than on scientific evidence. Until 2000, they often lacked a thorough and transparent development process,[2,3] a situation which is comparable to that in other countries.[4] Against this background, the concept for a guideline clearing house was developed by experts from the German self-governing bodies in health care in 1997[5] in order to assure and to improve the quality of CPGs. In 1999, the German Guidelines Clearing house (GGC) was established at the Agency for Quality in Medicine – a joint institution of the German Medical Association (GMA) and the National Association of Statutory Health Insurance Physicians (NASHIP) – in partnership with the German Hospital Federation and the Federal Association of the Statutory Sickness Funds. In spring 2000, the GGC finished its first project – the *National Hypertension Guidelines Clearing Report.* This is the first time that the goals, procedures, methods and results of the GGC are presented in an English language publication.

BACKGROUND: THE GERMAN GUIDELINES CLEARING HOUSE'S RATIONALE

In Germany CPGs were introduced into the political discussion as early as 1924.[6] Since then, guidelines have been published by a variety of interested groups in the health care system under different names, i.e. 'recommendation' (Empfehlung), 'guideline' (Leitlinie) and 'directive' (Richtlinie). Mandatory national guidelines (i.e. directives) for preventive services in primary care have been implemented within the regulations of the social sickness funds (covering nearly 90% of the German population) since the late sixties.[7] Guidelines are also part of disease management contracts between statutory sickness funds and state associations of statutory health insurance physicians. Contracted physicians are only paid if they are willing to offer and to record the health care specified in guidelines (e.g. for diabetes mellitus[8]). By the use of evaluated implementation tools,[9,10] diabetologists practising in ambulatory care across nearly all of Germany have succeeded in obtaining special contracts for performance of ambulatory diabetology based on CPGs.

However, the Association of the Scientific Medical Societies (AWMF), representing 130 individual medical societies, ignored these activities and started a new, comprehensive CPG programme in 1995 – leading to about 1,000 national CPGs after five years. And more CPGs are published by additional players, such as the regional chamber of physicians, professional associations, hospital groups and networks of private surgeries (for an overview – look at GGC's website: www.leitlinien.de), in the German health care system. All these well-meant and competitive activities between 1995 and 1999 have resulted in an uncontrolled growth of German CPGs with often low methodological quality and sometimes even conflicting recommendations (see Table 1).

TABLE 1
Major problems of German clinical practice guidelines.*

- Negligible information on the evidence of the recommendations.
- Lack of disclosure of development strategy/process/sponsorship/accountability/implementation tool.
- Conflicting guidelines on relevant topics.
- Focus on specialist care, given the academic background of the authors.
- Little to no relevance for primary care.

* until 1999

CHARACTERISTICS OF THE GERMAN GUIDELINES CLEARING HOUSE

Clearing houses may function either simply as directories of available guidelines, or they may provide evaluative services, assessing the methodological quality of guidelines as well as the appropriateness of their recommendations before inclusion in their databases.

Given the specific background (see Table 1), the GGC was built up as a national centre of excellence which:

- appraises the methodological quality and the appropriateness of CPGs;
- networks between guideline programmes, producers, evaluators and users (see Figure 1);
- collaborates with other health care quality programmes (e.g. Health Technology Assessment (HTA) agencies, the Cochrane Collaboration, CME programmes);
- guarantees easy access and transparency; and
- provides practical support for everybody interested in CPGs (see Table 2).

INSTRUMENTS AND METHODS OF GUIDELINE CLEARING

Taking into account published experiences and recommendations in the field of CPG methodology, the instruments, methods and procedures of the GGC were developed by a national committee of experts established

FIGURE 1
The German network for development and implementation of evidence-based guidelines.

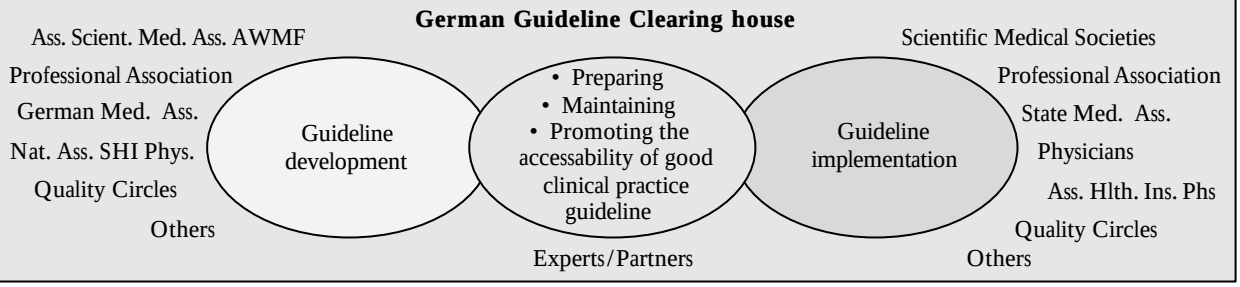

TABLE 2
Characteristics and aims of the German Guidelines Clearing house.

GGC – a national centre of excellence with the following tasks:

- development of CPG methodology;
- critical appraisal of CPGs;
- prioritisation of CPG topics;
- information on CPGs' existence/quality;
- implementation of CPGs on national/regional/local level; and
- coordination of CPG evaluation.

TABLE 3
Instruments of the German Guideline Clearing house.

Aims	Instruments
Development of CPG methodology	• Committee of experts for developing a CPG clearing method
Critical appraisal of CPGs	• German *Guideline for Guidelines*[8] • German checklist for critical appraisal of CPGs[19, 28] • Procedure for appropriateness of appraisal of CPGs' recommendations[18, 19] • Standard guideline clearing procedure[20]
Prioritisation of CPG topics	• Procedure for prioritisation of CPG topics[21]
Information on CPGs existence/quality	• German Guidelines Information Service (GERGIS) (www.leitlinien.de/engl)
Implementation of CPGs on national/regional/local level	• CME tools for CPG implementation[22, 23]

by the GMA and the NASHIP in 1997 (Table 3). This committee represented a broad spectrum of medical specialties, health care services and related fields (anesthesiology; clinical epidemiology; Cochrane Collaboration; emergency medicine; evidence-based medicine; health economics; internal medicine; laboratory medicine; legal counselling; medical education; medical informatics; medical librarians; pathology; pediatrics; primary care; quality management; sickness funds; and surgery). Key documents which served as blueprints or background papers during the instrumental development are *Guideline for Guidelines*;[11, 12] checklist for critical appraisal;[13, 14] prioritisation;[15] appraisal of appropriateness;[16] CME tools for implementation.[17]

PROCEDURES AND RESULTS OF GGC: THE GUIDELINE CLEARING PROJECT – HYPERTENSION 2000

Background and objectives

In 1999, the Guideline Clearing Project – Hypertension 2000 was initiated by the GGC's sponsors in order:

- to identify and review published guidelines on hypertension in German and English language;
- to establish criteria for future guideline development and implementation;
- to familiarise stakeholders in Germany with state-of-the-art management of hypertension;
- to identify key topics for a future national evidence-based guideline; and on the whole
- to develop the quality of hypertension management in Germany.

Methods

Search procedures.

By means of a comprehensive literature search – using Medline, Healthstar, Embase, National Guideline Clearing house (www.guideline.gov) and the guideline databases available through the German Guideline Information Service (www.leitlinien.de) – 132 guidelines on hypertension were identified for the period from 1990 until 1999.

Formal appraisal.

Of those, 34 fulfilled the pre-established inclusion criteria (i.e. German and English language, national guideline, cited references published after 1994) to serve as blueprints for a German evidence-based guideline on hypertension. These CPGs were assessed with the help of the German checklist for critical appraisal of CPGs, and 11 of 34 passed the checklist's minimal quality criteria as a basis for the clearing process.

TABLE 4

Results of the guideline clearing project hypertension 2000 – appraisal of the recommendations on appropriateness and of methodological quality.

	Appraised CPGs (*2)										
Key topics of a German hypertension CPG (*1)	CDN 1	CDN 2	USA 1	USA 2	D 1	UK	SA	USA 3	WHO	NZ	D 2
General management of hypertension • = Benchmark Texts (*3)											
Definition of hypertension						•					
BP-measuring	•					•					•
Medical history, evaluation			•			•					
Case-finding , screening											
Specialist referral						•					
Risk estimation						•				•	
Laboratory tests	•										•
Choice of drug treatment						•					
Non drug treatment		•									
Follow-up/motivation				•		•			•		•
Comorbidity	•					•					
Prevention		•									
Quality assurance, audit				•		•					
Tools for dissemination, implementation	•	•	•	•	•	•		•	•	•	•
Future research		•							•		
Pharmacotherapy											
Indication/risk estimation	•										
Choice of drugs	•								•		
Preferred/avoided therapy	•			•					•		
Drug combination					•	•			•		
Mono vs combination therapy									•		
Specific drugs					•			•			
Obsolete drugs											•
Compliance with drug therapy			•						•		•
Side effects/interactions					•			•			
Drug monitoring								•			
Therapy step-down											•
Economic considerations			•								
Other drug treatment									•		
Emergencies			•		•						
Drugs used in pregnancy			•								

TABLE 4 *continued*

Indicators of methodological quality (*4)	**Results of formal appraisal (*5)**										
	CDN 1	CDN 2	USA 1	USA 2	D 1	UK	SA	USA 3	WHO	NZ	D 2
1. CPG development	15	14	14	11	11	8	7	8	6	6	5
2. CPG content	16	17	15	16	15	14	16	14	13	14	7
3. CPG practicability	4	3	5	6	4	0	0	6	2	0	4

Explanations: *1: Key topics of a national hypertension guideline – as identified by the focus group; ***2**: Best available evidence-based guidelines – as identified by the focus group; ***3**: Links of key topics (***1**) and benchmark texts (***2**); ***4**: Indicators of methodological quality according to the appraisal checklist;[19] ***5**: Sums of 'yes-answers' concerning the quality indicators (highest ranking: left column; lowest ranking: right columns).

Links of key topics of a German hypertension guideline, as recommended by the clearing project, to examples from the 11 appraised evidence-based guidelines that covered each of these topics best[24]

Appraisal of clinical appropriateness of the recommendations.
The appraisal of the appropriateness of the recommendations was performed by a multi-disciplinary focus group of experts from different backgrounds (primary and secondary care, clinical pharmacology and clinical epidemiology), all of whom were familiar with evidence-based medicine. All experts confirmed no conflict of interest; of them, one had been involved in a CPG production programme before or during the clearing process.

Disclosure of appraisal results.
The results of both the methodological appraisal as well as the appraisal of clinical appropriateness were documented on a structured abstract form. The focus group consented on a guidelines clearing report containing structured abstracts, comments on CPG methods and recommendations and methodological features which should be employed for the development of a national guideline on hypertension for Germany (like systematic retrieval and a selection of evidence or explicit links between recommendations and the supporting evidence). This was followed by a detailed description of 16 key topics (including screening, risk stratification, clinical assessment, pharmacological and non-pharmacological therapy, patient education, as well as dissemination and implementation strategies for health care providers and health care users and the requirement of a re-evaluation of the guideline) on content issues. All suggestions and recommendations were accompanied by explicit examples from the 11 existing evidence-based guidelines that covered each of these topics best (see Table 4). The whole report is available for the public on the internet (www.leitlinien.de).

Results (see Table 4)
Methodological appraisal.
Eleven out of 132 guidelines were in accordance with the formal minimal standard with a wide range within the following domains: 'description of the development process'; 'authors' declaration on conflict of interest'; 'explicit link between recommendations and the supporting evidence'; 'management options'; 'tools for implementation'.

For future national hypertension guidelines the focus group suggested the following procedures:

1. the recommendations should be developed using professional association standardised and transparent consensus methods based on evidence retrieved and selected in a systematic way;
2. links should be established between recommendations and supporting evidence;
3. specific guideline versions should be formulated for defined groups of health care professionals as well as for consumers/patients;
4. guideline-based educational tools should be developed; and
5. periodical updates of the hypertension guideline need to be ensured.

Appraisal of guidelines' content.
None of the guidelines comprised information on the key elements of a national hypertension guideline for Germany (as identified by the focus group), such as:

1. definition of hypertension: epidemiology, health care problems, intended guideline users/goals;
2. blood pressure measurement;
3. medical history and physical examination;
4. case-finding/screening;
5. indications for referral;
6. risk-stratification;
7. diagnostic procedures;
8. therapeutic goals/indications for therapy;
9. non-pharmacological interventions;
10. pharmacotherapy;
11. follow-up/patient education/motivation/compliance;
12. comorbidity, hypertension in childhood/elderly/pregnancy;
13. primary prevention;
14. quality assurance/quality management;
15. dissemination/implementation; and
16. open questions/challenges for the future.[25]

HAVE THE AIMS AND GOALS OF THE CLEARING PROJECT BEEN ATTAINED?

Appraisal of guidelines' methodological quality and appropriateness
The above mentioned procedures and results demonstrate that it was possible to establish a systematic approach for critical appraisal of clinical practice guidelines in Germany within an acceptable period of time (three years).

Using internationally accepted quality criteria we were able to identify and to rank the best available guidelines on hypertension in German and English language according to their validity, applicability and clinical appropriateness. Therefore, it was an important step to draw up a methodology for appraising the appropriateness which enhanced the quality development of our national guideline programme.

Following the clearing report, the editors of the two German hypertension guidelines were asked to coordinate their activities in the future. The focus group responsible for the clearing project will supervise the expected collaboration and its results.

Thus the clearing programme will – hopefully – end in an evidence-based national guideline programme for Germany, without producing the guidelines by itself, rather leaving this task to the medical community. The ongoing clearing projects on other topics (tumor pain, back pain, Type 2 diabetes mellitus, asthma) will show whether these activities will have long-term effects.

Network between guideline programmes, producers, evaluators and users

These activities would not have become reality without networking between all parties interested in clinical practice guidelines. The institutionalisation of the GGC in 1999, and the application of its instruments during the project Hypertension 2000 can be regarded as the starting period/point of a national programme for more quality in practice guidelines. Before 1999, the AWMF – the umbrella association of the scientific medical societies – opposed intensively the activities of the self governmental bodies in the CPG field. Since then, a close collaboration between AWMF and the GGC has been taken up with regular joint meetings, joint prioritisation and joint instruments aiming at good clinical practice guidelines.[22]

Collaboration with other health care quality programmes

The development of a new concept for evidence-based postgraduate education in Germany is an additional consequence of the clearing project. The National Academy for Continuing Medical Education has been asked to build up a CME programme on management of hypertension on the grounds of the results, comments, discussions and examples of the hypertension report. Numerous audit groups and networks of physicians – predominantly in ambulatory care – have started to use the clearing report as a tool for continuing medical education on the management of hypertension.

Furthermore, health policy makers are planning to integrate GGC's projects and their results into the framework of activities for continuous quality improvement according to the new Social Law for Germany (see Table 5).

As outlined in Figure 1, the priorities and projects of the GGC programme are planned and performed in close collaboration with several German health care quality programmes, e.g. the German Cochrane Centre (www.cochrane.de), the German Network for Evidence-Based Medicine and its members (www.ebm-netzwerk.de), the German Association for Quality Assurance in Medicine and its members (www.aqs.de) and the German HTA Agency (www.dimdi.de). International partners of the GGC are the Scottish Intercollegiate Guidelines Network (www.sign.ac.uk), the US National Guideline Clearing House (www.guideline.gov) and the AGREE collaboration (agreecollaboration.org).[26]

Easy access of guidelines and practical support

The first and the most relevant of all GGC's projects was to guarantee easy access to guidelines as well as practical support for everybody interested in that field – long before GGC's official establishment in 1999. The homepage of the GGC (www.leitlinien.de) (i.e. the German Guideline Information Service 'GERGIS') has been the first German internet site on guideline programmes. Focusing on the quality of CPGs and linking to all known guideline websites in German, English, French and Dutch language, we were able to implement the ideas of 'good and evidence-based clinical practice guidelines' into German health care, and health care policy, within a relatively short period of time: GERGIS was launched in spring 1997. Only two years later, the state ministers of health and the federal parliament referred to the GGC's activities when advising all parties in health care to use and to implement good and evidence-based CPGs.[27]

THE FUTURE OF GGC

What are the perspectives of all these activities? Is the existence of a CPG clearing house a short running event? Or is such an institution necessary – given the specific situation of the German health care system?

- In winter 2000, 44 months after the first ideas on the GGC,[5] there are more and more German stakeholders interested in collaboration or partnership with GGC.

TABLE 5
Implementation of guidelines in the German health care system.

Federal/state parliaments define	⇒	Legal framework for quality in health care
		⇓
Self-governing bodies define	⇒	Priority health care topics
		⇓
GGC identifies	⇒	Evidence-based and practicable guidelines
		⇓
Self-governing bodies define	⇒	GL based quality assured regulations and budget
		⇓
Regulating self-governing bodies implement	⇒	GL based education and TQM programmes
		⇓
Health professional guarantee	⇒	High quality health care

- The GGC has started to establish a clearing house for patient information (www.patienten-information.de).
- The GGC is preparing electronic technologies to enable country wide implementation of appraised CPGs.

There seem to be major and challenging tasks ahead for the GGC.

ACKNOWLEDGEMENT

The GGC is thankful for the support granted by Professor James Petrie, Ms Juliet Miller and the whole team of SIGN during the development of the projects, described herein. The authors thank Professor Martin Eccles, Newcastle, who kindly commented on the manuscript.

SPONSORSHIP

The staff and the projects of the GGC are sponsored exclusively by the German Medical Association (www.bundesaerztekammer.de), the National Association of Statutory Health Insurance Physicians (www.kbv.de), the German Hospital Association (www.dkgev.de) and the Federal Association of Statutory Health Insurers (www.g-k-v.com).

REFERENCES

[1] Deutscher Bundestag. *Gesetz zur Reform der gesetzlichen Krankenversicherung ab dem Jahr 2000* (GKV-Gesundheitsreform 2000) - § 137 e. Berlin: BMG; 1999. (www.bmgesundheit.de)

[2] Ollenschläger G, Thomeczek C. Ärztliche Leitlinien - Definitionen, Ziele, Implementierung. *Z Ärztl Fortbildg* 1996; **90:**347-53.

[3] Helou A, Perleth M, Bitzer EM *et al*. Methodological quality of CPGs in Germany. *Z Ärztl Fortbild Qualitatssich* 1998; **92:**421-8.

[4] Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed literature. *JAMA* 1999; **281:**1900-5.

[5] Lauterbach KW, Lubecki P, Oesingmann U *et al*. Concepts of a clearing procedure for guidelines in Germany. *Z Ärztl Fortbild Qualitatssich* 1997; **91:**283-8.

[6] Kraus F. Wie ließe sich die ärztliche Behandlung der Kranken angesichts der jetzigen wirtschaftlichen Notlage der Bevölkerung sparsam und doch sachgemäß gestalten? *Dtsch Med Wochenschr* 1924; **50(13):**391-3.

[7] Hutzler D. Ambulatory physician performances in health counseling – possibilities, organization and payment. *Z Ärztl Fortbild* 1995; **89:**57-63.

[8] Berger M, Jörgens V, Flatten G. Health care for persons with non-insulin-dependent diabetes mellitus. The German experience. *Ann Intern Med* 1996; **124:**153-5.

[9] Gerlach FM, Beyer M, Römer A. Quality circles in ambulatory care: state of development and future perspective in Germany. *Int J Qual Health Care* 1998; **10(1):**35-42.

[10] Gruesser M, Hartmann P, Schlottmann N *et al*. Structured treatment and teaching programme for Type 2 diabetic patients on conventional insulin treatment: evaluation of reimbursement policy. *Patient Educ Couns* 1996; **29:**123-30.

[11] Field MJ, Lohr KN editors. *Clinical practice guidelines. Directions for a new program*. Washington D.C. : National Academy Press; 1990.

[12] Field MJ, Lohr KN editors. *Guidelines for clinical practice: from development to use*. Washington D.C.: National Academic Press; 1992.

[13] Scottish Intercollegiate Guidelines Network. *Clinical Guidelines – Criteria for Appraisal for National Use*. Edinburgh: SIGN; 1997.

[14] Cluzeau F, Littlejohns P, Grimshaw J *et al*. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; **11:**21-8.

[15] Field MJ editor. *Setting priorities for clinical practice guidelines*. Washington D.C.: National Academy Press; 1995.

[16] Leape LL, Park RE, Kahan JP *et al*. Group judgement of appropriateness: the effect of panel composition. *Qual Assur Health Care* 1992; **4:**151-9.

[17] Scottish Intercollegiate Guidelines Network. *SIGN Guidelines – An introduction to SIGN methodology for the development of evidence-based clinical guidelines*. Edinburgh: SIGN; 1999.

[18] Bundesärztekammer, Kassenärztliche Bundesvereinigung (BAEK/KBV). Beurteilungskriterien für Leitlinien in der medizinischen Versorgung. *Deutsches Ärzteblatt* 1997; **94:**A2154-5, B1622-3, C1754-5, (www.aerzteblatt.de/archiv/artikel.asp?id=7397).

[19] Agency for Quality in Medicine (AZQ). Checklist Methodological Quality of Medical Guidelines, 2nd version. *Deutsches Ärzteblatt* 2000; **97:**A1170 (www.aerzteblatt.de/archiv (in German); www-leitlinien.de/engl (in English)).

[20] Bundesärztekammer, Kassenärztliche Bundesvereinigung (BAEK/KBV). Das Leitlinien-Clearingverfahren von Bundesärztekammer und Kassenärztlicher Bundesvereinigung in Zusammenarbeit mit der Deutschen Krankenhausgesellschaft und den Spitzenverbänden der Gesetzlichen Krankenversicherungen Ziele und Arbeitsplan. *Deutsches Ärzteblatt* 1999; **96:**A2105-6 (www.aerzteblatt.de/archiv/).

[21] Ollenschlager G. Priority of helth or treatment problems as themes of international guideline programs. Proposition of the Guideline Clearing House on the evaluation of guidelines. *Z Ärztl Fortbild Qualitatssich* 2000; **94:**425-9.

[22] AWMF/AZQ. The German Manual for Clinical Practice Guidelines. *Z Ärztl Fortbild Qualitatssich* 2001; **95(1):** 1-84.

[23] Kirchner H, Ollenschläger G. Implementation of clinical practice guidelines in ambulatory care networks. In: Kassenärztliche Bundesvereinigung (editor) *Handbuch für Netzärzte*. Köln: KBV; 2000.

[24] Agency for Quality in Medicine (AZQ). *Leitlinien Clearingbericht Hypertonie. AZQ-Schriftenreihe, Bd. 5*. Munich: Zuckschwerdt Verlag; 2000. (www.leitlinien.de)

[25] Donner-Banzhoff N, Echterhoff HH, Hense HW *et al*. Critical appraisal of clinical practice guidelines on hypertension – Executive summary and recommendations for a future German hypertension guideline. *Z Ärztl Fortbild Qualitatssich* 2000; **94:** 341-9.

[26] The AGREE Collaboration. Guideline development in Europe: an international comparison. *Int J Technol Assess Health Care* 2000; **16(4):**1036-46 (www.sghms.ac.uk/depts/phs/hceu/biomed.htm).

[27] Conference of the Ministers. Goals for a uniform quality strategy in the health system. Trier: GMK; 1999. (www.aqs.de/gmk-goals.htm)

[28] Ollenschläger G, Helou A, Kostovic-Cilic L *et al*. A check list for the methodological quality of guidelines a contribution to the quality promotion of medical guidelines. *Z Ärztl Fortbild Qualitatssich* 1998; **92:**191-4.

ADVANCES IN SYSTEMATIC REVIEW METHODS

A. Grant, Director, Health Services Research Unit, University of Aberdeen, Aberdeen

Systematic reviews[1] are now widely accepted as one of the cornerstones for clinical effectiveness, clinical guidelines and clinical standards. Traditional narrative reviews are prone to serious biases, and these are avoided in systematic reviews by using protocol-driven systematic methods to: identify the evidence; appraise its quality; synthesise the data; and report the findings. Growing acceptance of systematic reviews has been paralleled by greatly increased accessibility.

Figure 1 shows the rise over the last five years in reviews and protocols available within the Cochrane Database of Systematic Reviews;[2] all of these address issues of clinical effectiveness. The Cochrane Collaboration[3] is now well established, involving thousands of people worldwide. There are currently 15 Cochrane centres, including most recently ones in Latin America and China. The contributors to the Collaboration look to Cochrane Centres for local support and training. However, it is the 50 Collaborative Review Groups that are the engine room and responsible for the production of reviews. These groups are problem-orientated; the four based in Scotland, for example, are Stroke, Incontinence, Peripheral Vascular Disease and Effective Practice and Organisation of Care. All potential reviews of the effects of health care should now have a place within one of the 50 groups.

The main output of the Cochrane Collaboration is the quarterly updated Cochrane Library,[2] 'the best single source of reliable information about the effects of health care interventions'. As well as being widely available in libraries, this is also distributed on CD-ROM and is accessible through the internet. In the fourth issue of the Library during 2000 there were 923 completed reviews, of which 67 were new and 50 updated since the previous issue three months earlier. There were also 827 protocols, of which 106 were new. These protocols would be expected to be developed into full reviews within the next year. Significant new developments within the Cochrane Library include: an electronic comments and criticism facility allowing users to interact with reviewers; a brief non-technical synopsis within each review aiming to make the information more accessible to a wider readership; indexing in Medline; and many web-based links.

The Cochrane Library also illustrates that the usefulness of systematic reviews depends on the quantity, quality and completeness of reporting of eligible studies. The Cochrane Control Trials Register (CENTRAL) is another of the databases in the Library.[2] This shows that as many as 290,000 controlled trial reports have been identified – 22,000 new since the last issue. Assimilating these within reviews represents a huge task. Nevertheless, many of these trials have little relevance to current health care. Furthermore, many of those that are relevant are of poor methodological quality and typically have small sample sizes. The trial reports that are available represent the agenda for primary research (over which a reviewer has no control), and there is a disproportionately large number of trials testing pharmaceutical interventions, and relatively few trials in some clinical areas. Reflecting this, despite the huge number

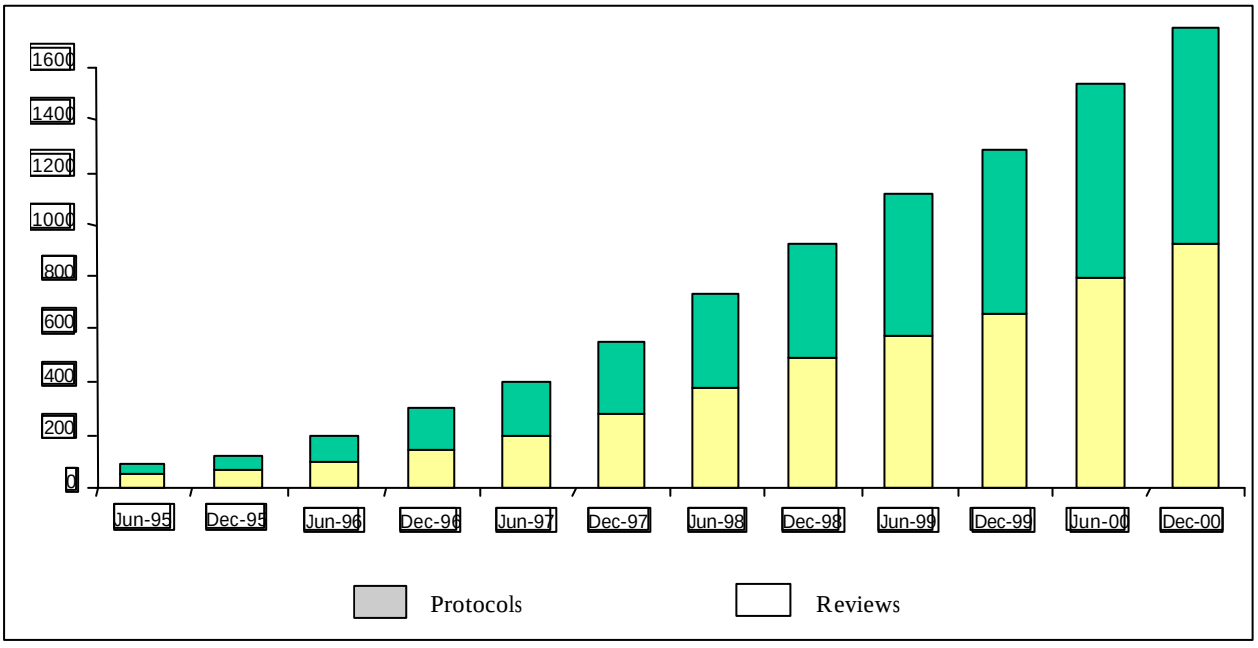

FIGURE 1
Numbers of reviews and protocols in the Cochrane Database of Systematic Reviews. Figure supplied by Dr Phil Alderson of the UK Cochrane Centre.

of reports that have been identified, reviews with few or no eligible trials are not uncommon within the Cochrane Library. For example, Issue 4, 2000 included reviews of vasoactive drugs for acute stroke (65 trials; 11,510 people) and inhaled beclomethasone versus placebo for chronic asthma (52 trials; 3,500 people). But it also included reviews of surgery versus non-surgical treatment for bronchiectasis (no trials) and subjective barriers to prevent wandering of cognitively impaired people (no trials). There is little a reviewer can do if there has been no rigorous primary research on a topic, other than to point this out and call for the research to be done.

The rigour of systematic reviews may also be compromised by incompleteness of reporting.[4] This may take the form of publication bias, that is, failure to publish the results of completed research, or missing data, that is, failure to include relevant data in a report, either because they were not collected or because they were not deemed important enough to present. Methods have been developed to explore these issues formally, such as through 'funnel plots',[4] to see if the estimated effects in relatively small trials are disproportionately 'positive'. There are other common limitations. The outcomes as presented may not meet the review definition; the summary statistics presented may not be suitable for combination, for example because of failure to report a measure of dispersion (such as the standard deviation); and lack of clarity about the numbers of participants at risk of a particular outcome. These problems are more common in reviews where a number of outcomes are of importance, in contrast to situations where there is a single easily defined primary outcome, such as death or disease recurrence.

These kinds of difficulty are one reason for the increasing use of individual patient data (IPD) meta-analysis in systematic reviews. In these, raw data from each trial are re-analysed, concentrating on the pre-specified outcomes, before being incorporated into a meta-analysis.[5] There have been striking examples of the value of IPD reviews in cancer, cardiovascular disease and HIV/Aids. Individual patient data analysis can provide quality assurance in respect of the security of randomisation, the basis of analysis on intention-to-treat, and through data checking. The meta-analyses are more complete because data collected but not reported in a suitable format can be included after re-analysis. Individual patient data analyses are particularly appropriate for 'time to event' outcomes, when varying lengths of follow-up can be accounted for. IPD analyses also allow differential effects in pre-defined sub-groups to be explored in a way which is rarely possible when reviewing published aggregated data. Individual patient data meta-analyses do, however, have disadvantages; there are practical difficulties, the reviews are time consuming, and they are resource intensive. Data are sought from, and subsequently provided by, the individual trialists; the data management and analysis may be complex, and the re-analyses are checked with the trialist before incorporation in a meta-analysis. While IPD analyses definitely have advantages, the pay back from the resources required to undertake them needs further exploration to identify the circumstances in which IPD reviews should be performed.

Despite recent methodological advances in systematic review methods[4] controversy still remains. It is still unclear how best to assess study quality, and the schemes used range in their complexity from a relatively simple assessment of the randomisation procedure through to a comprehensive assessment of a range of methodological issues. What is clear is that the more complex methods do not appear to correlate well with each other, and simpler methods assessing individual components of a study now seem more appropriate than deriving an overarching score. Another controversy relates to the inclusiveness of reviews: 'lumping' versus 'splitting'.[6] It is a matter of judgement to find the right balance between including as much information as possible in a review while avoiding combining data from studies which are too dissimilar for this to be appropriate. Interpreting inconsistency in the results of included studies, 'heterogeneity', is also controversial. Drawing on a wide range of studies performed in different places at different times, perhaps with slightly different versions of the same intervention, should enhance generalisability. Where the review suggests heterogeneity, however, this invites examination of why this might have occurred, such as because of variation between the types of people involved in studies, in the treatments given or in the methods of measuring outcome.

An area where the principles of systematic review have recently been applied is economic evaluation. In this, estimates of clinical effectiveness are related to estimates of resource use (costs). One approach to a summary economic evaluation is through systematic review of all economic evaluations, regardless of methodological quality. The problem with this is that the estimates of effectiveness and/or costs used in these may be so unreliable as to make the underlying evaluation flawed. A second approach aimed at avoiding this is to limit consideration to economic evaluations that are above minimum methodological standards for both effectiveness and cost estimation. The best estimates of effectiveness are, however, likely to come from formal systematic reviews of effectiveness, like the Cochrane reviews discussed above. A third approach is therefore to collect new resource data and relate these to rigorous effectiveness reviews. The problem here is that the collection of new data may be expensive and time consuming. The fourth approach is to combine systematic review of effectiveness data with systematic review of resource use and costs data. This approach assumes that appropriate data will already be available, however, and this is not always the case.

Looking to the future, there is likely to be even wider use of systematic reviews. In respect of clinical effectiveness, there is active discussion about extension to consideration of studies using non-randomised controlled methodologies. The underlying concern about this is the potential for serious bias and this is a major limitation of this approach. Systematic reviews of methods of diagnosis, disease prognosis and aetiology are less common than reviews of clinical effectiveness but are on the increase. Methods to synthesise data from these types of studies are less straightforward and still need development. Like effectiveness reviews, they suffer from the difficulties of publication and missing data bias, such that there is likely to be a move towards IPD analyses in these contexts,v too. A recent development has been systematic reviews of aspects of research methodology, particularly funded through the NHS Reasearch and Development Health Technology Assessment Programme.[7] While these have demonstrated that using systematic methods for this

purpose is far from straightforward, the outputs have proved to be very useful compilations and critiques of currently available information.

Systematic reviews also have a major part to play in determining the research agenda. Increasingly, research funders are demanding that applicants have undertaken a systematic review and incorporated the findings in any new proposal before it is submitted. Systematic reviews may show where there is already sufficient evidence, or where there is promising but inconclusive evidence, or where there is no good evidence at all. Earlier in this paper, I mentioned the fact that a proportion of Cochrane reviews have few or no randomised controlled trials. One way in which these reviews are valuable is in the context of determining future research priorities.

Over the last 20 years systematic reviews have become the pre-eminent way of summarising research evidence, and there has been greatly increased coverage and accessibility, particularly for effectiveness reviews. There is now a much better understanding of the strengths and weaknesses of systematic reviews and greatly enhanced presentation of their findings. The methods for reviewing cost-effectiveness still lag behind, but there is growing recognition of their importance with significant recent methodological advance. As the methods of systematic reviews are used more broadly they generate new methodological challenges. Perhaps the most intractable difficulty remains, however, that the value of a systematic review is largely determined by the primary research agenda, and perverse incentives, over which a reviewer has no control, may drive this.

ACKNOWLEDGEMENTS

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Health Department. However, the opinions expressed in the paper are those of the author alone.

REFERENCES

1 Chalmers I, Altman D editors. *Systematic reviews*. London: BMJ Publishing Group; 1995.

2 The Cochrane Library. Issue 4. Oxford: Update Software; 2000.

3 The Cochrane Collaboration. Website address: http://www.cochrane.org/

4 Sutton AJ, Abrahams KR, Jones DR *et al*. Systematic reviews of trials and other studies. *Health Technology Assessment Monograph Series* 1998; **2:**No. 19.

5 Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine* 1995; **14:**2057-79.

6 Gøtzsche PG. Why we need a broad perspective on meta-analysis. *BMJ* 2000; **321:**585-6.

7 NHS Research and Development Health Technology Assessment Programme. Website address: http://www.hta.nhsweb.nhs.uk/

ECONOMIC ANALYSIS WITHIN NATIONAL GUIDELINES

J. Mason, Senior Research Fellow, Medicines Evaluation Group, Centre for Health Economics, University of York;
M. Eccles, Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

BACKGROUND

In England and Wales, national guidelines are being commissioned for dissemination throughout the NHS. These guidelines are to deliver evidence-based recommendations for treatment in common disease areas, and will have to meet clinical and also political needs. The stated political aims are to promote the quality and consistency of health care delivery and ensure the cost effectiveness of recommended services. Value for money in health care is a contentious issue, and economic analyses of health care interventions have not been previously perceived to have much impact upon the delivery of health care.

METHODS

This paper explores how the concept of cost effective care may be incorporated into guideline development whilst taking account of the current realities of clinical practice. A profiling approach, describing the known costs and consequences of treatment alternatives, is presented as a basis for guideline development group working and contrasted with other methods.

FINDINGS

National guidelines should be developed by groups representing health professionals and patients, should include full consultation with interested parties and should use the best technical methods. They should promote an awareness of the current evidence for treatment options, thus improving health care decision-making. An explicit profile allows health service professionals and patients to weigh up the costs and the various good and bad consequences of treatment alternatives. This approach takes the focus away from the aggregated health gain concept of the Quality-Adjusted Life Year (QALY), in favour of simple quantification of the benefits and risks of health care options described in physical terms.

INTERPRETATION

National guidelines will become available to offer guidance on many areas of clinical care over time. The intention is that these guidelines define socially worthwhile health care, and consequently deliver transparent and defensible treatment recommendations. Such recommendations are intended to provide useful guidance but do not replace the obligation to practice safe and appropriate medicine.

INTRODUCTION

The world is changing for those who conduct economic analyses of health care. It used to be enough to publish decision analyses: complex models drawing together the various facets of treatment and care, pulling out apparently robust answers of cost effectiveness from circumspect data. The fact that the medical profession as a whole has taken little notice of these models has not been a hindrance to what has been largely an academic or commercial pursuit.[1] The creation of the National Institute for Clinical Excellence (NICE) means that economic method must come of age. Recommendations about the provision of care need to be made that satisfy clinical, public and political needs.

NICE has two committees addressing different slices of the task of producing recommendations.[2] The Appraisal Committee is responsible for the timely assessment of new technologies and for providing advice about their reimbursement by the NHS. The Guidelines Advisory Committee advises on the commissioning, and oversees the delivery, of evidenced-based guidelines for use by the NHS: these guidelines will be focussed at specific diseases and will thus cover the range of treatment and care options provided for common conditions. They will be an important tool in clinical governance and form the basis for generating audit criteria.[3] Decisions concerning the reimbursement of new treatments are newsworthy and the appraisal process is likely to continue to receive much attention. However, it is guidelines that will have most scope to impact upon the quality of care received by most patients.

The purpose of modern clinical guidelines is uncontroversial: 'systematically developed statements to assist both practitioner and patient decisions in specific circumstances' valid if 'when followed they lead to the improvements in health status and costs predicted by them'.[4] Developments in guideline methodology have sought to promote a consistently high awareness in professionals of the current evidence for treatment options, thus improving individual health care decisions.[5, 6] Recent guidelines have successfully incorporated the cost implications of treatments alongside the risks and benefits of treatments in the guideline development process.[1, 7, 8] In this paper, the rationale and implications behind the approach to economic analysis in national guideline development are explained.

WELFARE ECONOMICS

Welfare economics seeks to 'formulate propositions by which we may rank, on the scale of better or worse, alternative situations open to society'.[9] Therefore, with any form of social decision-making we need to determine how we value 'better' and whose values are to count when assessing alternatives. Available resources are limited, and when used to improve welfare through one decision we forego other opportunities for their use. A prudent use of resources tries to minimise the opportunity cost of decisions by adopting the most valuable policies.

The discipline of health economics has sought to make this task manageable in health care by amalgamating the very different types of benefits of treatments in different disease areas into a simple measure of health gain, the QALY. With only one metric in the welfare function, the task is

one of maximising QALYs in the health care system within available resources, hence the cost/QALY league tables that first emerged about a decade ago.[10] When making these estimates, the value of being at different levels of distress and disability are valued by sample groups drawn from society as a whole: the values of individual patients have no clear role to play. The strengths and weaknesses of this approach have been discussed.[11-13]

The availability of cost/QALY estimates is limited and their quality variable. Consequently, we do not know (for example) the marginal value of resources spent in cardiology or paediatrics, and there is no technical basis for changing historical allocations between the two. Additionally, economic analysis of individual treatments currently provided in a publicly-funded system may have little value when resources are allocated by speciality and, legally, national recommendations for treatment are only advisory. The potential path to a better use of resources is to engage the health service professionals themselves in the valuation process rather than attempting to impose external solutions.

The desire to aggregate the (frequently) complex consequences of treatment options to provide consistent decision rules may have disenfranchised the real audience. Health service professionals, together with patients, need to discuss and weigh up all the aspects of treatments. These professionals, who use resources and who have a duty of care, cannot be expected to make appropriate health care decisions based on a cost/QALY estimate: physical descriptions of the outcomes of treatment are needed. Furthermore, much health care activity is not outcome- but process-orientated: the quality of the process is itself highly valued but has largely been ignored by economic analysis.

IMPROVING WELFARE IN THE NHS

Clinical guideline methodology has developed dramatically over the last decade and presents an obvious forum for the valuing of health care options. The method allows that a socially representative and accountable group identify the aspects of available treatments that are important and uses evidence-based medicine techniques to quantify each aspect.[1,4,5] Early in each guideline development the group members identify the aspects of management that are important to decision-making in a specified disease area. These may be measures of effect, tolerability, safety, quality of life (broadly defined), resource implications and delivery issues. These are considered from different relevant perspectives (e.g. doctors, patients, carers and relatives for process and outcomes, and the NHS and society for resources). This process sets out the important aspects of treatments for which the group would like to consider evidence, and provides a valuable opportunity to express prior beliefs. The subsequent evidence gathering process will succeed to varying extents in each aspect according to available research. After due consideration of available evidence and of its limitations, the group will have a profile of the known costs and physical consequences of treatment alternatives, as well as uncertainties and unknowns.

Sometimes the profile will provide clearly dominant recommendations for first line care and guidance for what happens next if this fails. The profile may also exclude some activities as inappropriate. In many instances there will be trade-offs between alternatives, and the group may ask for simple models of cost effectiveness to help explore these. Sometimes the quality of evidence will be inadequate to achieve much beyond identifying the key resource issues, and recommendations will be appropriately empirical with a clear research agenda being identified.

Decision-making under this system has two stages. During guideline development the guideline group, with clinician and patient disease-specific knowledge, defines the aspects of treatments that are important and assesses the evidence for these. Their values decide those interventions that society should generally be willing to provide (and in which order) for a given disease and the group makes recommendations accordingly. This information is then available to a clinician and patient to discuss, and their values guide the second stage. The clinician has to treat the patient in a manner consistent with their duty of care and their understanding of the patient's needs. The clinician is aware of current evidence-based recommendations and uses these to augment his or her clinical decision-making.

This process has some important advantages over traditional economic evaluation. First, the presentation of evidence and the language of recommendations are relevant to the users of the guideline – the health care professionals delivering, and the patients receiving, care. Second, the process is transparent and any guideline user can work back from the recommendations to see what interpretation is placed upon available evidence.

A criticism of guideline recommendations might be their crudity, that they don't provide refined cost/QALY rankings or league tables that tell us the relative value for money of alternatives across many diseases. However, such rankings may be largely a deceit: the data are not good enough.[14] The process of identifying the important aspects of treatment alternatives, and how these are valued and by whom, has been defined. The recommendations that arise from the guideline process should value health care alternatives credibly and transparently. This is likely to be both valuable and important given the competing pressures upon any national guideline development process.

REVIEWING EXISTING ECONOMIC STUDIES

In many fields of health care there is an economic literature accompanying the clinical studies. As with a review of available clinical trials, it is feasible to have a summary of published economic analyses.

Unlike protocol-driven prospective clinical trials, economic analyses are usually retrospective and the analyst has the choice of how to construct the model and use the data. Qualitatively there is far greater scope for bias, either explicit or implicit, in the process of model construction, reporting of findings and exploration of uncertainty.[15,16] A guideline development process should lead to the best available presentation of the known costs and various physical consequences of treatment alternatives. These data are unlikely to match the baseline assumptions in any published model.

Each clinical trial presents unique or independent data. The trials together can be summarised to obtain an overview. Different economic analyses take different cuts at the same clinical data. There is no quantitative way to summarise the findings of all the analyses as the data are not independent; thus, there is no 'weight of evidence'.

Published decision analyses are often not transparent and it can be difficult and time-consuming to validate the findings presented. In some areas of medicine, a thorough review of published economic analyses would be a mammoth task with little obvious return.

Sometimes one or two economic analyses may have had considerable influence upon a clinical field. In such circumstances, it may be useful to summarise these as part of a guideline development process and comment on their findings compared with those of the guideline.

PERFORMING ECONOMIC MODELLING IN GUIDELINES

The usefulness of any analysis of costs and benefits hinges on the existence of unbiased, valid clinical trial data comparing the relevant alternative treatments. Available clinical trials tend to feature clinical endpoints rather than patient-orientated health outcomes, inadequate follow-up and limited generalisability. Thus, to understand the long-term overall health impact of treatment alternatives upon a disease process, analysts draw on data from a range of other sources and a number of assumptions. Such analyses are often referred to as 'models', since they simulate reality, rather than the outcomes of real patients followed in trials. There are pros and cons to modelling as a form of decision aid (Figure 1).

FIGURE 1
Some pros and cons of modelling.

Arguments made in favour of modelling are:

- Models can simplify down complex decisions with many influences to a simple presentation of cost-effectiveness
- Modelling allows the long-term health impact of health care interventions to be explored using limited clinical data
- Presentation of findings as a cost/QALY estimate allows comparison with other treatments in other disease areas
- Modelling is better than more informal decision making

Objections made are:

- Models are value-laden, and consequently different analysts can get different answers
- Findings may appear over-precise because uncertainties are inadequately explored
- Some models are by necessity complex in their methods and difficult to validate, thus the findings have to be taken on trust
- Models may simplify the aspects and requirements of clinical decision-making inappropriately

The profile of costs and consequences emerging from a guideline development process could be used as the basis of a new modelling exercise, e.g. generating cost/QALY estimates for different treatment alternatives. This might be thought simply to augment the guideline process with some formal economic analysis, and allow policy analysts to assess whether guideline recommendations are consistent with the perceived value for money of treatments provided elsewhere in the NHS. However, there are some important challenges to this approach. When the impact of assumptions and data uncertainties are adequately explored, it is commonly apparent that a wide range of answers is possible. To settle for the analyst's best guess is inappropriate, as the average value has no clinical or statistical interpretation (unlike the findings of a prospective trial). Thus, each best guess from each model involves different methods and types of assumption; there is no valid basis for comparing the best guesses from different models.

The purpose of most economic analyses has been to simplify the basis of clinical decisions by considering one primary outcome (cost effectiveness analysis) or by mapping diverse effects upon quality and length of survival on to a QALY score (cost-utility analysis). Cost/QALY estimates are unlikely to be transparent to users of the guideline and may be seen as untrustworthy. A separate and valid consideration is that a health system might wish to consider making cost/QALY calculations external to the guideline for policy purposes. The exception to this discussion is where long-term estimates of quality-adjusted survival are measured directly in trials; such data should be considered on their merits like any other outcome measure. In (most) less researched areas of medicine, cost/QALY estimates will not be informative.

Modelling *per se* should not be ruled out of a guideline development process. If, having described the known costs and consequences of treatments in physical terms, questions arise in the group about how this information may be interpreted or summarised, it may be useful to offer and present simple models which further understanding in the group. Modelling then should be group initiated rather than methodologically imposed. A responsibility rests upon a participating health economist to help a group explore how modelling may be useful in this context.

ASSESSING THE COST IMPACT OF A GUIDELINE

Policy makers may routinely wish to know the cost impact of implementing a guideline. For example, in the instance where a new and expensive treatment is recommended, it might be possible to assess the net cost to the NHS of different levels of uptake of recommendations, alongside the expected benefits. However, the longer-term costs and consequences of treatments are often circumspect, making overall cost impact assessments inherently uncertain. These may be of less value to policy-makers than they suppose. Cost impact assessments, like cost/QALY calculations, could be performed for policy reasons, but it would seem unhelpful and possibly damaging to include these in the guideline development process.

Whilst such presentations might be of interest to a guideline development group, their use within a guideline development group needs to be handled with care. They are potentially problematic since they may be perceived to take the focus away from improving individual treatment decisions and towards health care policy traditionally concerned with budgets. Doctors and patients may perceive that the guideline recommendations are underpinned by 'affordability' rather than 'value', potentially discrediting the guideline medium.

Members of guideline groups might perceive a cost impact assessment to be important at a local level. For example, it may be helpful to know how much a new screening programme will cost, organised at various levels, such as the GP or PCG where delivery issues

and implementation are seen as integral to the recommendations.

CONCLUSIONS

National guidelines, as currently being developed by NICE, are for use by health service professionals and patients and aim to improve the quality of care. Experience with this type of process will help representative groups to develop a perspective of the value for money of treatment choices using a language appropriate to clinical decision-making and to set appropriate recommendations for colleagues. The advantage of starting with a profiling approach is that to make defensible recommendations it does not require all the evidence to be arbitrarily weighted together in a summary cost effectiveness estimate.

It is not possible to say that the widespread use of national guidelines will only ever improve the benefits received by patients. It is possible that some patients will be denied expensive treatments of marginal health benefit that they personally value highly and would have received had a guideline not been produced. It is, however, likely that decision-making will become more consistently directed towards socially defined 'worthwhile' health care and thus make best use of scarce resources.

ACKNOWLEDGEMENT

The authors are grateful to Diane Dawson, Centre for Health Economics, University of York who provided helpful and challenging comments: views expressed remain those of the authors.

COMPETING INTERESTS

J. Mason is a member, and M. Eccles Chairman, of the NICE Guidelines Advisory Committee. J. Mason is seconded as a health economics advisor to NICE. However, the views expressed in this article are solely those of the authors.

J. Mason has received fees and expenses for research and consultancy work in a range of methodological and disease-related areas from pharmaceutical companies, from the UK Department of Health and from medical charities.

REFERENCES

1 Mason J, Eccles M, Freemantle N *et al*. A framework for incorporating cost-effectiveness in evidence-based clinical practice guidelines. *Health Policy* 1999; **47:**37-52.

2 NHS Executive. *A first class service: quality in the new NHS*. London: Stationery Office; 1998.

3 Rosen R. Improving quality in the changing world of primary care. *BMJ* 2000; **321:**551-4.

4 Field MJ, Lohr KN, editors. *Guidelines for clinical practice: from development to use.* Washington DC: National Academy Press; 1992.

5 Eccles M, Freemantle N, Mason JM. Methods of developing guidelines for efficient drug use in primary care: North of England Evidence Based Guidelines Development Project. *BMJ* 1998; **316:**1232-5.

6 Eccles M, Grimshaw J. Clinical guidelines: from conception to use. Oxon: Radcliffe Medical Press; 2000.

7 Eccles M, Mason J, Freemantle N. Developing valid cost effectiveness guidelines: a methodological report from the North of England evidence based guideline development project. *Quality in Health Care* 2000; **9:**127-32.

8 Eccles M, Mason J. How to develop cost-conscious guidelines. *Health Technol Assess* 2001; **5:**16.

9 Mishan EJ. *Welfare economics: ten introductory essays*. New York: Random House; 1969.

10 Mason JM, Drummond MF, Torrance GW. Some guidelines on the use of cost-effectiveness league tables. *BMJ* 1993; **306:** 570-2.

11 Loomes G, McKenzie L. The use of QALYs in health care decision making. *Soc Sci Med* 1989; **28:**299-308.

12 Rawles J. Castigating QALYs. *J Med Ethics* 1989; **15:**143-7.

13 Smith A. Qualms about QALYs. *Lancet* 1987; **16:**1134-6.

14 Mason JM, Drummond MF. The DH register of cost-effectiveness studies: content and quality. *Health Trends* 1995; **27(2):**50-6.

15 Freemantle N, Mason JM. Publication bias in clinical trials and economic analyses. *Pharmacoeconomics* 1997; **12:**10-16.

16 Mason JM. The generalisability of pharmacoeconomic studies. *Pharmacoeconomics* 1997; **11:**503-14.

EVIDENCE-BASED IMPLEMENTATION OF EVIDENCE-BASED GUIDELINES

J. Grimshaw, Programme Director and Professor of Health Services Research and A. Walker, Senior Health Psychologist, Health Services Research Unit, University of Aberdeen

High quality clinical guidelines support medical practice by allowing doctors to identify the most effective treatments and health care interventions rapidly. The development of valid guidelines (based on a systematic review and synthesis of research evidence) is most efficiently done at national or international level. Over the last 20 years, there has been considerable progress in both the methods of developing clinical guidelines and in their acceptance by policy-makers and doctors. As a consequence of this, a growing range of high quality guidelines is becoming available to support practice. This, however, is only the beginning of the process. Once guidelines have been developed, how should we go about getting them into practice?

Traditionally a simple dissemination model has been used most commonly to get guidelines into practice. The guideline is published and made widely available through a variety of means, often including targeted mailing to clinicians. Unfortunately, this is rarely sufficient to change clinical practice. A recent systematic review of the effectiveness of dissemination of written educational materials (including guidelines) by publication in professional journals or mail to targeted clinicians concluded that 'the effects of printed educational materials . . . are at best small . . . and of uncertain clinical significance'.[1]

In the real world, clinicians encounter a range of potential barriers that may prevent them from putting guideline recommendations into practice. These may occur at a variety of levels including: structural (e.g. financial disincentives); organisational (e.g. inappropriate skill mix, lack of facilities or equipment); peer group (e.g. local standards of care not in line with desired practice); individual (e.g. knowledge, attitudes, skills); or professional-patient interaction (e.g. problems with information processing). Passive dissemination of guidelines may serve an important function in increasing awareness of the most effective treatments and interventions, but it does little to address barriers that may exist at other levels. Further implementation strategies are usually needed. A number of different strategies are available, e.g. audit and feedback, educational outreach visits, reminders; and there is increasing interest at policy level in the use of more active implementation strategies. However, our knowledge about their effectiveness in different settings remains limited. More often the choice of a strategy is based on familiarity rather than any scientific evidence that it will make a difference. Research efforts in evidence-based medicine should therefore be complemented by research into how best to implement this evidence into normal practice. As Richard Grol put it: 'evidence-based medicine should be complemented by evidence-based implementation'.[2]

There are three key prerequisites for effective guideline implementation. The first is the use of effective and efficient dissemination and implementation strategies to ensure the adoption of effective interventions in practice. The second is adequate resourcing to provide the most effective interventions. The third is a supportive culture and organisational structure which enables clinicians to make changes in their routine practice. These prerequisites mean that, in contrast to guideline development, guideline implementation is essentially a local activity. It is not guideline developers who need information about the effectiveness of implementation strategies, but those responsible for the quality of care within provider organisations. In the UK, this is most likely to be clinical governance staff in acute and primary health care trusts.

A wide range of research methods has been used to investigate the effectiveness of different strategies for getting research evidence into practice. These include case studies, before-and-after studies, time-series studies and randomised controlled trials. Of these, rigorous evaluations (mainly randomised controlled trials) are likely to provide the best evidence of effectiveness of different interventions. The best source of information to inform decisions about the choice of implementation strategies comes from systematic reviews. Systematic reviews of rigorous evaluations of implementation interventions increase the precision of findings by combining results from different studies, exploring the consistency of findings between different studies and, most importantly, investigating the generalisability of findings across different settings, professional groups and behaviours. The Cochrane Effective Practice and Organisation of Care review group (EPOC) aims to undertake systematic reviews of interventions to improve the quality of care.[3] These include: reviews of professional interventions (e.g. continuing medical education, audit and feedback); financial interventions (e.g. professional incentives); organisational interventions (e.g. the expanded role of pharmacists); and regulatory interventions. To date, EPOC has developed a specialised register of over 2,000 primary studies. Over 100 collaborators from 12 countries have been involved in the production of 18 reviews, with a further 20 protocols published in the Cochrane Library. In addition, EPOC staff have conducted an overview of previously published systematic reviews of provider behaviour change, summarising the current state of knowledge.[4]

The objective of the overview was to use systematic methods to identify, appraise and summarise the findings of systematic reviews of professional behaviour change. Fifty-one reviews published up to the end of 1998 were identified and their findings included. The implementation strategies that had been considered in these reviews could be divided into three broad categories: strategies that are ineffective in most studies; strategies that are effective in some studies but not in others and strategies that are effective in most studies.

Generally, ineffective strategies are passive dissemination (for example, mailing educational materials to targeted clinicians) and a didactic educational session in which

information is presented to a passive audience. These strategies may raise awareness of an issue, but they have little or no impact on practice. Strategies that vary in their effectiveness are audit and feedback, local consensus conferences, opinion leaders and patient mediated interventions. These may be effective in some circumstances, but it is not yet clear which circumstances these are. The only strategies that are largely effective are reminders, educational outreach visits (for prescribing), interactive educational workshops and multifaceted interventions.[4]

This overview of systematic reviews suggests that a variety of implementation strategies have the potential to be effective under certain conditions. However, the available evidence is sparse for many interventions, and the reviews themselves were of variable quality with common methodological flaws. Nonetheless, it is clear that there are 'no magic bullets'[5] when it comes to choosing implementation strategies. Most interventions are effective under some circumstances, but none are effective under all circumstances. Interventions based on a prior assessment of potential barriers to change may be more effective, and multifaceted interventions targeting different barriers to change are more likely to be effective than single interventions, but they are also likely to be more costly.

In conclusion, guideline development is most efficiently undertaken at national or international level. In contrast, guideline implementation is largely a local activity. The recognition that passive dissemination of guidelines is not enough to improve the quality of health care has led to a greater emphasis on active implementation strategies. These implementation activities need to be undertaken at local level and the choice of strategy to be used should be based on evidence about their effectiveness. There is a substantial body of rigorous evidence that can be used to inform decisions about implementation strategies. This needs to be integrated with other types of knowledge to maximise the likelihood of successful implementation, and the extent to which evidence base is used when planning local implementation activities is uncertain.

ACKNOWLEDGEMENTS

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Health Department. The opinions expressed in the paper are those of the authors and may not be shared by the funding body.

REFERENCES

1 Freemantle N, Harvey EL, Wolf F *et al*. *Printed educational materials: effects on professional practice and health care outcomes*. Cochrance Review. The Cochrane Library. Issue 3. Oxford: Update Software; 1996.

2 Grol R. Beliefs and evidence in changing clinical practice. *BMJ* 1997; **315:**418-21.

3 Bero L, Grilli R, Grimshaw JM *et al*, editors. *Cochrane effective professional and organisation of care group*. The Cochrane Library. Issue 2. Oxford: Update Software; 2002. (Further information about EPOC available at www.abdn.ac.uk/hsru/epoc/)

4 NHS Centre for Reviews and Dissemination. Getting evidence into practice. *Effective Health Care Bulletin* 1999; **5(1):**1-16. (Also available from: www.york.ac.uk/inst/crd/ehc51.pdf)

5 Oxman AD, Thomson MA, Davis DA *et al*. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ* 1995;**153:**1423-31.

FROM EVIDENCE-BASED MEDICINE TO CLINICAL STANDARDS

D.R. Steel, Chief Executive, Clinical Standards Board for Scotland[*]

THE SCOTTISH APPROACH TO CLINICAL EFFECTIVENESS

The Scottish approach to clinical effectiveness has been built upon a long tradition of excellence and innovation in medicine and in health care generally. Nonetheless the achievements of the last 15 years are particularly impressive. They have been centred around the Clinical Resource and Audit Group (CRAG) which is one of the lasting legacies of the previous government's White Paper *Working for Patients*.[1] Under CRAG's auspices, a substantial programme of national audits has been undertaken, and support and encouragement has been given to the service to develop extensive programmes of local audit.

The Clinical Resource and Audit Group was also instrumental in responding to the initiative of the Scottish Colleges in coming together to establish the Scottish Intercollegiate Guidelines Network (SIGN) in 1993. The Scottish Intercollegiate Guidelines Network is a particularly good example of the Scottish approach: collaboration between the professions and government; multi-disciplinary; and rigorous in its methodology but at the same time largely informal in the way in which it has operated.

The result has been 52 guidelines to date, virtually all of which at the time of publication have been regarded as models of excellence, and an evolving methodology that has developed apace over the last seven years.

Alongside these developments, Scotland has pioneered the publication of clinical outcome indicators. That this has been possible reflects the quality of much of the data, the benefits of record linkage, and the maturity of relationships within the NHS family in Scotland. Under the auspices of the Clinical Outcomes Working Group, five reports have been published containing 44 indicators, broken down on a named trust basis.

Space does not permit a full evaluation of the Scottish approach, but alongside these achievements it is necessary to place a number of continuing challenges. First, both nationally and locally there is a multiplicity of bodies responsible for taking forward the agenda, each with its own remit and its own acronym. This has the benefit of targeting effort but poses real challenges in coordination. To a significant extent also the agenda has been driven by enthusiasts pursuing their own particular interests and causes. In this mode the NHS is frequently at its best; but it does mean that the emphasis has not always reflected the service's priorities.

The need to pull all this together and to give it a stronger strategic focus prompted the development in the 1990s of the Priorities and Planning Guidance including what became known as the common core work programme.[2] Initially, this was allied to the internal market; more recently it has underpinned the development of Health Improvement Programmes and Trust Implementation Plans. To complement these developments, a Clinical Effectiveness Strategy Group has recently been established at the centre, providing a forum in which the key players can agree priorities and share information about their work programmes and progress.

Second, there is the challenge of involving patients and the public in this agenda. This is not a criticism of the past; the agenda had first to be owned professionally. But the nature of the issues requires patient and public involvement and increasingly this is what they expect not just of the health service but of all public services. That is why the bold decision was taken in 1994 to publish clinical outcome indicators; that is why SIGN established a patient information and participation subgroup a few years ago. But SIGN would be the first to admit that this represents only a start.

Third, and most important of all, is the challenge of implementation. What has all this effort produced in terms of improvements in the quality of the care and treatment provided to patients? What changes in clinical practice have resulted from audit and the publication of guidelines? Without doubt, this is the area of greatest weakness, which has prompted two recent initiatives:

- the introduction of clinical governance; and
- the establishment of the Clinical Standards Board for Scotland (CSBS).

CLINICAL GOVERNANCE

The Health Act of 1999 imposed on the board of each NHS body in Scotland a statutory responsibility for monitoring and improving the quality of the health care it provides. This duty of clinical governance is designed to make clinical effectiveness:

- mandatory, in the sense that it is no longer a voluntary or optional activity but is required of every part of the NHS and everyone who works in it;
- mainstream, in the sense that it is no longer something for evenings and weekends but is an integral part of working life; and
- managed, in the sense that it is targeted upon the priorities of the NHS and that it is properly planned.

Local implementation of clinical governance serves as an essential complement to the national initiative spearheaded by the CSBS. The two processes feed off each other and the chances of either succeeding are greatly enhanced by the fact that they are being put into effect together.

Why clinical standards?

If clinical governance potentially is such a powerful lever

[*] *Dr David R. Steel gave the Al Hammadi Lecture*

for change, why is there a need also to identify and monitor clinical standards? Clinical standards enable the identification of the essentials that need to be right in the treatment of particular conditions if outcomes for patients are to be optimised. This is important in itself to target efforts to improve the quality of care, but it also serves other important purposes: those of promoting public confidence and public accountability.

It was for this reason that the Acute Services Review recommended in 1998 that a national system of quality assurance and accreditation should be established, in effect combining external quality assurance with the setting of standards.[3] Its aim was to promote public confidence that the services of the NHS meet nationally agreed standards, and to demonstrate that, within available resources, the NHS is delivering the highest possible standards of care. Members of the public have the right to know that, wherever they cross the threshold of the NHS, they will receive not identical care and treatment – that would be impossible and undesirable – but care and treatment that accords with what reasonably they can expect in terms of evidence of good practice and available resources.

Following acceptance of this recommendation by ministers, the CSBS was established in April 1999 under the chairmanship of Lord Patel to develop and operate a quality assurance and accreditation process. The system that the Board is developing is set out in the Quality Assurance and Accreditation Manual which was issued in August 2000.[4] The system itself comprises three processes:

1. setting standards;
2. undertaking external peer review of performance against these standards; and
3. reporting findings to both the service and the public.

However, there is a fourth element that is crucial to the effectiveness of the system: self-assessment of performance locally in relation to the standards. This will give clinicians and managers involved in the service under review opportunities to assess and develop their own practice and systems, and it will ensure that the Board's external reviews are part of an ongoing process of quality improvement. The more effective self-assessment is, the lighter touch the Board's external check can be.

CSBS STANDARDS

The normal tests applied to standards – that they are unambiguous, measurable and so on – are well-known. The standards that the Board is seeking to develop for the NHS in Scotland, however, have some distinctive features.

First, the challenge is to produce clinical standards that focus explicitly on the care and treatment that is provided by health care professionals. Most previous quality assurance and accreditation schemes have addressed the organisation of services and the environment in which they are delivered. The Board needs to take account of these factors too because in many cases they have an impact on the quality of care, but its challenge is to look explicitly at what health care professionals do in treating particular conditions.

Second, in doing so, standards need to be evidence-based. Unashamedly, the Board is part of the movement towards evidence-based health care. But, whilst there must be evidence underpinning all its standards, levels and types of evidence will vary and the quality of the evidence does not necessarily correlate with the impact of a standard upon outcomes.

Third, the Board's standards are results orientated in terms of improved patient outcomes. That said, many of them relate to process and some to structure, in part because of the considerable difficulties in using outcomes as a measure of clinical performance, particularly in the short term. The essential test, however, is that any structure or process standard needs to be linked through evidence to outcomes.

Fourth, the standards are patient focused, following as far as possible the journey of care or experience of patients as they go through the system. This represents one of the most exciting challenges facing the Board. It would be much easier to look at the services provided by primary care or hospitals on their own.

But this would not represent reality for patients as they move from one part of the service to another. Moreover, that is just the time that things frequently go wrong. But there is another dimension to the patient focus of the Board's work which is important. Patient focused standards do not, as has been the case so often in the past, relate only to 'softer' issues such as the provision of information, privacy, dignity and so on. A patient perspective needs to be integrated with the evidence on effective care and treatment, applying what evidence there is on patients' needs and preferences and involving patients in the groups that are developing the standards.

Fifth, alongside all these challenges, there is the comfort that the Board is building upon extensive foundations. Other organisations have produced – and are producing – standards upon which the Board can draw and its work is significantly accelerated, for example, by the existence of clinical guidelines, particularly those that have been produced through a methodology that is as rigorous as that developed by SIGN.

Sixth, some of the Board's standards are generic in that they apply to most, if not all, clinical services, whereas others are related to particular conditions or diagnoses.

And, finally, some are categorised as essential in that it is expected that they will be met wherever a service is provided; others as desirable in that they will promote continuous quality improvement as they are being met in some parts of the service and demonstrate levels of quality to which other providers of a similar service should strive to achieve.

CSBS EXTERNAL PEER REVIEW

The setting of standards – and the way in which they are set, involving both health care professionals and lay people – is crucially important. But it is only the first part of the quality assurance and accreditation process. Self-assessment too is important, particularly if the Board is going to sustain the active support of those working in the service. But, however well self-assessment works, there is a need for external quality assurance. This is based on peer review. Indeed, the Board believes so strongly in peer review that it has decided to break convention by retaining this term even though lay people are also members of its review teams. In part this can be justified semantically on the ground that users are *de facto* peers in that they have experience of the service under review. But, more importantly, whilst they add a vital element to the review process, the intention – confirmed by experience to date

– is that they do not alter the fundamental character of the process that is based upon the principles and values of peer assessment.

The Board's reviews combine objective assessment of written evidence, including the results of self-assessment and visits by multi-disciplinary teams, including the public, to the locations where a service is being provided. Both elements are crucial. As far as possible, hard data about performance is required, and documentation that establishes beyond doubt, for example, that protocols are being put into practice. But this cannot – and should not – be merely a paper exercise. For the foreseeable future, there are many areas where neither data nor documentation exists; and even were this not the case, there are many standards where an assessment of performance depends upon the kind of judgement that can only be provided by someone with experience of the service in question on the basis of observation and dialogue.

Moreover, the Board is determined that its reviews will be constructive and supportive. It will not shy away from identifying and reporting on unacceptable or inadequate performance but, wherever possible, it will do so in a way that is designed to encourage improvement. The intention is that the ethos of visits should be focused not on catching people out, but on a positive and helpful discussion, identifying constraints that impact upon performance and highlighting and sharing good practice.

It is for this reason primarily that the Board has decided not to follow the normal pattern of appointing a small number of people to review services across the whole country. Rather, it is developing an inclusive process that will involve as many people as possible so that the benefits of being a reviewer as well as being reviewed can be shared widely. This also has the advantage of reducing the demands that are made upon any one health care professional, but it does of course present particular challenges in terms of maintaining consistency of assessment which the Board is having to address in other ways.

CHALLENGES AHEAD

What other challenges lie ahead for the Board?

A great deal seems to have been achieved over the last year or so and the response has generally been very positive. The standards issued in relation to coronary heart disease, four cancers and mental health have also been well received in that they are felt to address the right issues and to have been pitched at the right level; they have been the subject of wide-ranging debate including some people who have not normally been involved in this type of exercise; and the 50 or so review visits to date have generally succeeded in meeting the criteria outlined above.

Moreover, probably the most difficult part of the process – and the one on which the Board will be most readily judged – lies ahead; that of producing reports which set out an honest assessment of performance in a manner that does not undermine the work of staff and which promotes public confidence rather than public alarm.

One of the lessons learned already – and experience to date has demonstrated just how serious the problem is – is the gaps in the data required to monitor clinical performance. Failure in the past to develop cohesive and clinically focused information systems, and to do so on a national basis with agreed definitions so that comparisons can be made, means that all too often in the early days it will not be possible to monitor whether a standard is being met or not.

The challenge is to define what data are needed and then to put systems in place that will enable them to be captured effectively. Information technology has a crucial role to play but there is also the human dimension of data capture and analysis as part of routine clinical practice or, if this is not possible, by suitably trained people under the supervision of clinicians. Investment in clinical infrastructure may not capture the headlines but it is essential to targeting effort on gaps and shortcomings, and demonstrating to the public that the NHS is delivering the best services that it can within the resources available.

The White Paper *Designed to Care*, published in December 1997, said all the right things about quality as the key driver of change.[5] But, whilst progress has certainly been made, many would question whether the step change has been achieved to place clinical quality at the centre of the NHS agenda. This is an issue about organisational culture, and it will percolate through to the work of individuals. In particular, the agenda of quality assurance and accreditation needs to be seen not as an additional burden but as essential and integral.

Patient and public involvement, too, must not be an 'add-on'. It needs to be built in from the beginning. The Board has made an encouraging start. Fifty per cent of its members come from outwith the NHS; there have been members of the public on each of its project groups developing standards, and on each of the teams undertaking the review visits. But this is only a start. For everyone involved, there needs to be greater clarity as to what lay people are being asked to do so that appropriate support can be provided to enable them to fulfil this role.

There is also a wider dimension to all this. The Board will soon start putting into the public domain reports on the performance of the service against clinical standards. It is essential to develop better public understanding of what conclusions can be drawn from this data and, equally important, what conclusions cannot be drawn. The experience of publishing clinical outcome indicators is encouraging. On the whole, media coverage of these reports has been remarkably responsible. Without doubt, a bumpy road lies ahead, but this is a journey upon which the NHS needs to embark. The gulf between public understanding of what can be delivered with the resources the health service is given and the reality which clinicians and managers face in their working lives is one of the most serious problems facing the NHS today.

The CSBS will be asking a lot of the service over the coming months and years. It needs to subject itself to the same sort of scrutiny that it is asking of others. Are its accreditation processes effective? Are they sufficiently rigorous? And it also applies to their impact. Is the CSBS adding value? Is its work contributing to quality improvement. Is it 'making it happen'?

REFERENCES

1 *Working for Patients* (Cm 555). London: HMSO; 1989.

2 *Priorities and Planning Guidance for the NHS in Scotland 1997/98*. Edinburgh: NHS MEL; 1996; 59.

3 *Acute Services Review Report*. Edinburgh: The Stationery Office; 1998.

4 *Quality Assurance and Accreditation Manual*. Edinburgh: Clinical Standards Board for Scotland; 2000.

5 *Designed to Care: Renewing the National Health Service in Scotland* (Cm 3841). Edinburgh: The Stationery Office; 1997.

CONTENTS .....