

## Medical school differences: beneficial diversity or harmful deviations?

I C McManus

Many factors may explain why graduates from different medical schools differ in their professional competence.

British medical students enter medical school at the age of 18 and qualify 5 or 6 years later at the age of about 24. By that time they have spent their whole adult life—and, indeed, a quarter of their entire life—in one educational establishment. So all-encompassing, so involving, so potentially overwhelming is a medical school that sociologists such as Erving Goffman might classify them with other “total institutions” such as prisons, asylums, monasteries, and barracks where daily life is regimented, where the social world primarily revolves around the institution and its members, and whose ultimate goal is a plan structured as much for the ultimate benefit of society as for the needs of the inmates.

If educational environments really matter, then medical schools should be the ideal place for spotting an effect. So what demonstrable effect is there? Five years at 300 working days a year, eight hours a day (and never mind the nights), is more than the notional 10 000 hours which are said to be necessary to become an expert in a skill. Indeed, there are 2000 hours left over to become pretty skilled in a host of other activities such as playing rugby, drama, music, or any other avocation.

Anecdotes abound about the experience of medical school, and a useful anthology is *“My medical school”* edited by the doctor-poet Dannie Abse.<sup>1</sup> In his introduction, Abse comments on how many aspects of medical education seem common to all medical schools, and that there are more similarities than differences. An educationalist, however, reads the accounts very differently. When Lord Platt described how in Sheffield during the First World War there were only 12 students in the year, one wonders how the experience must have differed from that of Sir Derrick Dunlop in 1920s Edinburgh where “the crowd was so great that it was often necessary to stand on a bench to catch a glimpse of the patient under discussion”. Another doctor-poet, Edward Lowbury, contrasts 1930s Oxford where the emphasis was

on “the growing points and gaps in knowledge, the disciplines of research and the critical reading of original papers” with The London Hospital where “the approach was more dogmatic, and the unwary might have imagined that all knowledge was wrapped up in their textbooks and lecture notes”. Is it possible that so many formative years in such different institutions resulted in doctors who are almost indistinguishable?

Despite its importance and obviousness, few studies have assessed the key question of the extent to which different educational environments—be they differences in philosophy, method of delivery, content, approach, attitudes, or social context—produce different sorts of doctor. Folk mythology certainly believes that medical schools make a difference—as the generic version of one joke goes: “You can tell a St Swithin’s man but you cannot tell him much”. However, an extreme contrarian view says that none of these things matter. After all, “cream rises to the top” irrespective of its container, and bright motivated creative medical students will pick out what matters from the multifarious raw material presented by their medical school. When multitalented ability is allied with a professionally driven motivation and a wealth of clinical experience (and, as Abse says, “Every patient ... teaches his physician about the subject of medicine”), then caring, compassion, and clinical competence are surely inevitable, whatever the educational environment. Hard data to refute that strong position are difficult to find, mainly because few studies have compared the educational effects of medical schools. There are many reasons for this, not the least of which is that institutions do not like being compared. When medical schools are weighed in the balance then some may be found wanting, so a sophisticated, self-congratulatory, mutually supporting culture of educational protectionism has arisen. Woe betide this person who asks such questions—they can only make enemies and few will thank them, even

should the institutions themselves have originated the study.

Such an educational context, along with the intrinsic interest of understanding malpractice, makes the paper by Waters *et al*<sup>2</sup> in this issue of *QSHC* of double interest. Information on malpractice claims of individual doctors is publicly available in some American states, making it possible—without the need for schools themselves to cooperate—to ask whether the graduates of some schools are more likely to be sued than others. Of course, whenever institutions can be ranked in order then some are inevitably higher than others—after all, even random numbers differ in size—and the challenge is to demonstrate convincingly that such differences are genuine. Perhaps most crucial, as here, is the demonstration of long term stability; schools producing a higher proportion of graduates with malpractice claims at one time tend to be those that also have a higher proportion at another time. What might cause such systematic differences in the graduates of different institutions?

Many things, is the simple answer. Different sorts of applicant apply to different sorts of schools for different reasons,<sup>3</sup> and different schools probably use different criteria and methods for selecting their entrants from among those applicants.<sup>4,5</sup> Medical schools differ in their social worlds and in their philosophy, outlook and approach to teaching,<sup>6</sup> and students at different schools have different amounts of clinical experience.<sup>7</sup> The net result of these and other differences is that graduates of different medical schools end up in different careers.<sup>8</sup>

“... the medical student is the aggregate of a range of influences ...”

Implicit in any such view of medical education is an “additive model” whereby the medical student is the aggregate of a range of influences that model him or her in the way that clay is moulded by the fingers of a sculptor. That, though, is only part of the story. Medical schools are dynamic social institutions in which the students interact with each other, each year or class developing its own personality which is far from predictable from the sum of its parts. This becomes painfully apparent to medical school examiners who find a far higher proportion of failures in one year than previously, despite the course being the same, the examination being similar, and the average social and educational qualifications of the students seeming to be equivalent. Some years are “good” and others “bad” because students, like peers in general, influence each other in their attitudes and approaches to education. A more subtle

version of this argument suggests that graduates of different institutions differ in relation to the diversity (the variance) in the individuals in their classes. A rich and complex social, ethnic, and class mix among the students is said, with some supporting evidence,<sup>9</sup> to result in more socially able graduates who can interact more effectively in complex, modern social worlds.<sup>10</sup>

Whether any or all of these factors are responsible for the differences in malpractice found by Waters *et al*<sup>2</sup> is far from clear at present. What is clear is that graduates from different medical schools not only differ in their propensity to cheer for Light Blue rather than Dark Blue or some other colour, but also in their professional competence (or, more precisely, their incompetence).

Understanding the reasons for this will tell us both about malpractice and about the enduring effects of different approaches to education.

*Qual Saf Health Care* 2003;**12**:324–325

#### Author's affiliation

**I C McManus**, Professor of Psychology and Medical Education, Department of Psychology, University College London, London WC1E 6BT, UK; i.mcmanus@ucl.ac.uk

#### REFERENCES

- 1 **Abse D**. *My medical school*. London: Robson Books, 1978.
- 2 **Waters TM**, Lefevre FV, Budetti PP. Medical school attended as a predictor of medical malpractice claims. *Qual Saf Health Care* 2003;**12**:330–6.
- 3 **McManus IC**, Winder BC, Sproston KA, *et al*. Why do medical school applicants apply to particular schools? *Med Educ* 1993;**27**:116–23.

- 4 **McManus IC**. Factors affecting likelihood of applicants being offered a place in medical schools in the United Kingdom in 1996 and 1997: retrospective study. *BMJ* 1998;**317**:1111–6.
- 5 **McManus IC**, Richards P, Winder BC, *et al*. Medical school applicants prefer interviewing to non-interviewing schools? *Adv Health Sci Educ* 1999;**4**:155–65.
- 6 **Wakeford R**. The medical school learning milieu: a study of students' perceptions of twenty-five British and Irish medical schools. *Studies Higher Educ* 1984;**9**:139–49.
- 7 **McManus IC**, Richards P, Winder BC. Clinical experience of UK medical students. *Lancet* 1998;**351**:802–3.
- 8 **Lambert T**, Goldacre M, Parkhouse J. Career preferences and their variation by medical school among newly qualified doctors. *Health Trends* 1996;**28**:135–44.
- 9 **Crosby FJ**, Clayton S, Downing RA. Affirmative action: psychological data and the policy debates. *Am Psychologist* 2003;**58**:93–115.
- 10 **Bollinger LC**. The need for diversity in higher education. *Acad Med* 2003;**78**:431–6.

Monitoring outcomes in primary care

## Monitoring clinical outcomes in primary care

**R Baker**

Is monitoring clinical outcomes in primary care possible and, if so, is it worthwhile?

Several arguments can be readily raised against monitoring outcomes in primary care. Many patients attending primary care have minor self-limiting illnesses, and the only outcomes that might be of interest for monitoring would be significant adverse events such as drug reactions. At the practice or practitioner level the numbers of patients with significant disorders are small, and distinguishing the impact of variations in quality of care from case mix and random variation is difficult. Research evidence is often incomplete and the relationship between the process of care and its outcome is not well established. The interval between the delivery of healthcare interventions and eventual outcome can be many years, as in the case of illness prevention strategies. Primary care is provided by teams, and it is frequently impossible to ascribe a particular outcome to the care provided by an individual doctor or nurse. Some family doctors would also argue that their role is more complex than simply the achievement of desirable clinical outcomes following illness episodes; family doctors have long term relationships with their patients, and ultimately all their patients die.

Intermediate outcomes have offered one approach for responding to these problems. Monitoring of immunisation

rates has long been routine in the healthcare systems of many countries, and in recent years disease control markers such as glycosylated haemoglobin or blood pressure have begun to be included in monitoring schemes. The validity of intermediate outcome measures depends in large measure on the completeness of the data. While information about immunisation in a population of children may be relatively easy to compile, the creation and maintenance of an accurate list of all patients with diabetes or ischaemic heart disease is much more difficult. When financial incentives are attached to the achievement of intermediate outcome targets, the problem of obtaining complete and accurate data can increase. Monitoring of deaths avoids some of the problems inherent to monitoring of intermediate outcomes. In most countries, central systems to collect and record information about all deaths are in place, and the data are much more likely to be complete. However, the disadvantages of monitoring mortality in primary care have been regarded as insurmountable, and include the small annual number of deaths in the small populations of primary care teams and practitioners, the delay between the delivery of care and death, the sharing of care between

different practitioners and between primary and secondary care, and the impact of case mix. Beaumont and Hurwitz, in this issue of *QSHC*, have identified another problem.<sup>1</sup> General practices have considerable difficulty in obtaining details about all the deaths among their patients, and they recommend that a central system should be established to fulfil this task. I have not been able to identify any developed country in which death rates in populations cared for by primary care teams are routinely studied to monitor performance at the level of the team.

" . . . always monitor key outcomes"

The discovery that a general practitioner was able to murder more than 200 patients over a period of 20 years and not be detected until he decided to forge the will of one of the patients he had killed has caused practitioners in the UK to question the belief that monitoring mortality in the small populations of primary care is not worthwhile.<sup>2</sup> Family doctors in other countries would be wise to take note of the UK experience and consider whether they too should rethink the widely held view on the use of mortality data. A monitoring system is to be set up in the UK following a recommendation made as a consequence of a review of the clinical practice of the GP murderer.<sup>3</sup> Problems will be encountered, but new methods of analysis suitable to small numbers will overcome some of these. Local knowledge of the populations of patients served by different primary care teams and the patterns of work of different doctors will help to overcome other problems. If information about other outcomes such as stroke or myocardial infarction were included in the monitoring system, it may prove possible to monitor the impact of clinical policies and identify primary care teams

that could do more to prevent death and other serious non-fatal events. The prospect of managing performance in primary care to improve key outcomes is attractive but speculative; more evidence about the sensitivity of monitoring and its feasibility in routine use is required. Nevertheless, there is an important lesson for all with an interest in safety and quality improvement—always monitor key outcomes. If you don't, you won't

know when the outcomes are poor. Why then should your patients trust you?

*Qual Saf Health Care* 2003;12:325–326

.....  
**Author's affiliation**

**R Baker**, Director, University of Leicester School of Medicine, Clinical Governance Research and Development Unit, Faculty of Medicine, Leicester LE5 4PW, UK; rb14@le.ac.uk

Ethnography in health care  
 .....

## What can ethnography do for quality and safety in health care?

**M Dixon-Woods**  
 .....

Used carefully, ethnography can identify errors in health care and provide explanations for their occurrence.

Originally developed within anthropology, ethnography is one of the most longstanding social science research approaches. Its emphasis is on the description and analysis of “the everyday”—routine behaviours in their natural settings. Many would characterise ethnography as the process of querying understandings and practices that are taken for granted: it renders the everyday world problematic by making the “ordinary” into the “extraordinary”. It is best understood as a holistic approach that does not rely on any single method of data collection. Observation, which may be unstructured “hanging out” or more structured and purposeful scrutiny of situations to look for particular things, is perhaps the defining feature of ethnography. These observations are often supplemented by interviews (sometimes very informal and part of the “hanging out” process) or documentary materials collected from the setting (e.g. posters, internal memos, reports of meetings), photographs, artefacts, and so on.

The interpretation of these data is very much a function of the researcher's own skills and judgement,<sup>1</sup> and will usually involve searching for themes and patterns in the data and generating explanations and theories grounded in them. Like literary criticism, there is no prescribed set of procedures or techniques but there are expectations of good practice. For example, researchers are required to be reflexive—that is, to reflect on and be able to give an account of how they produced their interpretations—and to be able to show that their interpretation is warranted by the data.

Ethnography is not for the faint hearted. It is a time consuming and demanding research process and can be a profoundly uncomfortable experience for the researcher. Ethnographers have often identified a natural affinity for their methods with relatively disadvantaged groups, seeing themselves as offering these groups a voice and a means of making explicit systems of oppression and coercion. When the ethnographer's role changes to one where s/he is attempting to explain professional practices, important problems can arise. Access to the field may be difficult to negotiate. When that has been gained, further challenges lie ahead. It can sometimes be difficult to access the people who may be most important to understanding a particular phenomenon: they may be impossible to get hold of, or unwilling to speak or be seen. Covert observation may be ethically unacceptable, but clearly obtaining consent to observations is not always easy and introduces risks of the ethnographer influencing the behaviours under observation. Participants in the process can begin to feel they are being inspected and judged, and to feel disempowered. They may “act up” in the presence of the researcher. The participants may also become hostile or uncooperative, and much depends on the skill of the researcher to overcome these obstacles. Sometimes this is done by creating an “insider” status, identifying the researcher with the group being studied—for example, a nurse studying nurses. However, such groups, having accepted the researcher as “one of us”, may feel betrayed by the research account that is

## REFERENCES

- 1 **Beaumont B**, Hurwitz B. Is it possible and worth keeping track of deaths within general practice? Results of a 15 year observational study. *Qual Saf Health Care* 2003;12:337–42.
- 2 **Baker R**, Jones D, Goldblatt P. Monitoring mortality in primary care. *BMJ* 2003;326:274–6.
- 3 **Baker R**. *Harold Shipman's clinical practice, 1974–1998. A review commissioned by the Chief Medical Officer*. London: HMSO, 2001.

subsequently produced. Some researchers begin to feel voyeuristic and exploitative in some settings, or to experience conflicting loyalties. Ethical dilemmas about when and how to intervene are not uncommon. In addition, ethnographic research will also be subject to criticism of its apparent subjectivity, and researchers can find themselves accused of producing partisan, partial, or misleading accounts.

Many of these difficulties should and can be overcome, particularly in making a contribution to quality and safety in health care. Much can be learned relevant to quality and safety from already published ethnographies. Strong's account of the “etiquette rules” governing face-to-face interactions between parents and doctors<sup>2</sup> demonstrates how difficult it is for patients to raise concerns about possible errors: to do so disrupts the “ceremonial order” of consultations and puts their status as “good patients” at risk. Patients may therefore be silenced when it comes to pointing out things they are concerned about. Pope's work on waiting lists demonstrated the importance of gatekeepers, both clinical and administrative, in controlling access to health care.<sup>3</sup> An ethnographic study of interactions between consumers and pharmacy staff and patient interviews<sup>4</sup> challenges traditional professional assumptions about how the public understands the role of pharmacists, with important implications for medication safety. Findings from ethnography can also be integrated with findings from other study types including quantitative or other qualitative research.<sup>5</sup>

Generally, however, the obvious potential for ethnographic approaches to make a contribution to the study of safety and quality in health care has been under-exploited. Ethnographic research is well suited to identifying conditions of risk, particularly where these involve human performance, organisational and cultural dynamics, and interactions between people and technology. Ethnography is especially good at probing into areas where measurement is not easy, where the issues are sensitive and multifaceted, and where it is important to get at the tacit, not the already evident. It can capture the winks, sighs, head shaking, and

gossip that may be exceptionally powerful in explaining why mistakes happen, but which more formal methods will miss.

The paper by Taxis and Barber<sup>6</sup> in this issue of *QSHC* makes a welcome addition to the canon. It shows how, used carefully, ethnography can identify types of errors and provide explanations for why these occur, through interrogating the everyday understandings that staff have of their own practices and identifying the cultural contexts of practice. It is difficult to see how the latent conditions for error described in this paper, which are likely to be hugely

important, could have been characterised appropriately using any approach other than ethnography. Much more research of this kind is needed, particularly in complex areas where long chains of causation exist in terms of health and safety outcomes.

*Qual Saf Health Care* 2003;**12**:326–327

.....  
**Author's affiliation**

**M Dixon-Woods**, Senior Lecturer in Social Science and Health, Department of Health Sciences, University of Leicester, Leicester LE1 6TP, UK; md11@leicester.ac.uk

Clinical databases  
.....

## Basing decisions on better quality routine clinical data

**T A Sheldon**  
.....

Clinical databases, properly used, have the potential to contribute to quality improvement.

Over the last decade there has been increasing pressure to make health care more accountable and to identify variations suggesting poor quality care which can be improved. While special studies can be conducted with these aims in mind, quality improvement activity will only become part of the fabric of health services if it can draw on data that are routinely collected. In North America extensive hospital data have been available; however, because they are generated as byproducts of administrative systems collected for other (usually financial) purposes, they are severely limited. Without the ability to adjust for patient case mix, for example, incorrect inferences about the cause of variations in patient outcomes are likely.<sup>1</sup> The sort of detailed clinical data needed to make sensible analyses of patient outcomes are rarely generated from routine systems, and it may not be feasible or cost effective for this to be a core administrative function. Instead, bespoke clinical databases have been developed in parallel, often as the result of the cooperative efforts of clinicians seeking reliable data on their practice and ways in which quality might be monitored and improved. However, as one finds with other sources of data collection, the validity and reliability of clinical databases are variable. As the number of such databases grows, it is important to ensure

that we have enough information to be able to interpret the results of studies based on them and to decide if they are fit for the purpose. As with the conduct and reporting of trials<sup>2</sup> and systematic reviews,<sup>3</sup> we need common high standards for establishing, running and reporting clinical databases, an appraisal tool to assess the degree to which these standards are being applied, and an accessible source of information on clinical databases and their quality.

The Directory of Clinical Databases (DoCDAT) described by Black and Payne<sup>4</sup> in this issue of *QSHC* is the first important step in providing such a resource in the UK. By providing key information on and critical appraisals of clinical databases which provide individual data, it will help people to find databases suitable to their needs and, more importantly, it will act as a pressure to improve standards. Inevitably there are ways in which this resource will need to mature. The methods of development and validation of the assessment instrument are probably not as robust or as explicit as that developed for the reporting of trials or systematic reviews.<sup>2,3</sup> Ideally there should be double checking of the appraisals and independent verification of the claims of the database custodians. There is also room for further conceptual development, possibly including a clearer separation of information and criteria about the internal

## REFERENCES

- 1 **Hammersley M**, Atkinson P. *Ethnography: principles in practice*. 2nd ed. London: Routledge, 1995.
- 2 **Strong PM**. *The ceremonial order of the clinic: parents, doctors and medical bureaucracies*. London: Routledge and Kegan Paul, 1979.
- 3 **Pope C**. Trouble in store: some thoughts on the management of waiting lists. *Social Health Illness* 1991;**13**:193–212.
- 4 **Hassell K**, Noyce P, Rogers A, et al. Advice provided in British community pharmacies: what people want and what they get. *J Health Serv Res Policy* 1998;**3**:219–25.
- 5 **Tashakkori A**, Teddlie C. *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage, 1998.
- 6 **Taxis K**, Barber N. Causes of intravenous medication errors: an ethnographic study. *Qual Saf Health Care* 2003;**12**:343–7.

validity of the data (accuracy and susceptibility to bias), usefulness for comparative work, and national representativeness. With sufficient investment this directory could develop into an important national and possibly international resource.

While the development and proliferation of clinical databases will inevitably contribute to audit and research, there is a danger that they will be misused. In England, for example, the Department of Health has made a commitment to publish cardiac surgeon specific mortality data based on the UK Cardiac Surgical register developed by the Society of Cardiac Surgeons. Sampling of this database revealed it to be both incomplete and unreliable in its ability to yield accurate, risk adjusted outcomes data.<sup>5</sup> Thus, publication is likely to result in misleading information about individual performance entering the public domain, with potentially damaging results. Even if the data were accurate and risk adjustment sufficient, it is still not obvious that the individual surgeon is the right unit of analysis or that publishing the data is the optimal way to promote quality improvement. Clinical databases, if properly used, have the potential to contribute to quality improvement. However, if they are used by those who lack sufficient scientific intelligence or are exploited for political ends, then clinicians will become cynical, data quality will fall, and the public interest will not be well served.

*Qual Saf Health Care* 2003;**12**:327–328

.....  
**Author's affiliation**

**T A Sheldon**, Department of Health Sciences, University of York, York YO10 5DD, UK; tas5@york.ac.uk

## REFERENCES

- 1 **Sowden A**, Sheldon TA. Does volume really affect outcome? Lessons from the evidence. *J Health Serv Res Policy* 1998;**3**:187–90.

- 2 **Moher D**, Schulz K, Altman D. The Consort statement: revised recommendations for improving the quality of reports of parallel group randomised trials. *JAMA* 2001;**285**:1987–91.
- 3 **Moher D**, Cook DJ, Eastwood S, *et al*. Improving the quality of reports of

- meta-analyses of randomised controlled trials: the QUORUM statement. Quality of reporting meta-analysis. *Lancet* 1999;**354**:1896–900.
- 4 **Black N**, Payne M. Directory of clinical databases: improving and promoting their use. *Qual Saf Health Care* 2003;**12**:348–52.

- 5 **Fine L**, Keogh BE, Cretin S, *et al*. How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK cardiac surgery experience. *BMJ* 2003;**326**:25–8.

Clinical practice guidelines

## Deriving recommendations in clinical practice guidelines

M Eccles

The content of guideline recommendations is influenced by a range of factors including the published evidence, cost, and the composition of the guideline development group. Increasing the sophistication and validity of guideline recommendations will lead to an increase in their length which will need a corresponding increase in the sophistication of the thinking and methods of those responsible for implementing them.

Most people who pick up a clinical practice guideline will only ever read the recommendations. Even though guidelines are an efficient exercise in research synthesis and summary, the full documentation can run to hundreds of pages—a volume that is beyond all but the most committed reader. In an attempt to address this, guideline programmes produce summary versions. These are usually formed solely of the guideline recommendations and presenting them in this way removes the possibility of the reader checking the validity of any or all of the recommendations. This means that guideline developers, and the programmes that they work with, must ensure that their processes produce recommendations that accurately reflect, not only the content of the scientific evidence, but also the appropriate range of clinical factors. In this issue of *QSHC* Manna *et al*<sup>1</sup> present an analysis of the degree to which guideline recommendations reflect clinically important facets of ethnicity and conclude that they do so to a variable extent.

The content of guideline recommendations is influenced by a range of factors. An assessment of the nature of the published evidence (accepting that some may come from the unpublished “grey” literature) addresses the methodological quality of the evidence in terms of threats to its validity. Assessing the applicability of the evidence to the population of interest deals with the issues of generalisability—the extent to which the subjects enrolled into studies reflect the

population with which the guideline will be used. The criteria for making judgements on the nature of the published evidence are best developed for studies examining a cause and effect relationship (effectiveness studies) and are reflected in many of the evidence and recommendation grading systems currently in use by guideline development agencies. Epidemiological studies are an important component of any guideline as they are the evidence that will characterise the range and complexity of care that has to be addressed within it. They will identify the important characteristics of individuals that should be reflected both in the inclusion and exclusion criteria of the effectiveness studies and the content of the recommendations. However, unlike effectiveness studies, they are one of the areas of a guideline that are more likely to draw on a national, as opposed to international, literature. Furthermore, epidemiological studies represent a challenge to guideline methodologists to develop the evidence grading systems which appropriately reflect the important dimensions of validity and avoid the current situation of their automatic downgrading by systems designed to grade effectiveness studies.

Aside from the published evidence, three other factors will influence the content of a recommendation. The first of these is costs. Assuming that issues of harm, a potential cost, will be dealt with from the published evidence, then costs usually encompass two considerations: (1) the relative costs of competing

technologies set against their effects, and (2) the likely cost impact of the guideline viewed in the context of the overall budget available for health care. This dimension of the guideline depends crucially on how cost and effect are balanced, the quality of the primary data contributing to this, and the nature and scale of the assumptions that have to be made.<sup>2</sup>

The two other factors that influence the content of recommendations are interlinked, less tangible, and not usually defined. They both refer to the members of the guideline development group and are their knowledge of the health-care system and their general and healthcare specific beliefs and values. The guideline development group is responsible for interpreting the published evidence in the light of their knowledge and experience. This provides an important reality check on the content of recommendations, making it more likely that feasible recommendations are written. It also allows the group to check that the guideline is appropriately addressing healthcare issues for all of the relevant population. These functions mean that the composition of the guideline development group is of fundamental importance. As Shekelle *et al*<sup>3</sup> observed: “There is good evidence that, when presented with the same evidence, a single specialty group will reach different conclusions than a multidisciplinary group, with the former being systematically biased in favor of performing procedures in which the specialty has a vested interest.”<sup>4,5</sup> For example, the conclusions of a group of vascular surgeons favored the use of carotid endarterectomy more than did a mixed group of surgeons and medical specialists.<sup>6</sup> An international collaborative group, the GRADE group, has reviewed all of the current major guideline grading systems and is currently piloting a new system aimed at making both the treatment of evidence and value judgements more explicit and the decision making in deriving recommendations more transparent.

Value judgements about content are also made at another point in the process—when the original scope of the guideline is defined. Although a guideline development group can argue to vary the content of a guideline, they will usually have the parameters of the guideline set for them by some sort of formal process. For instance, the National Institute

for Clinical Excellence (NICE) in England and Wales has a formal process for defining the content of a guideline before starting its development. Here the reality check includes a process of external consultation with those who have an interest in the guideline.

Increasing the sophistication with which guidelines address the validity of their recommendations will not, however, be without cost. The more detailed and specific recommendations are, the longer the guideline will be. Within the context of a full version of a guideline this will not necessarily make a discernable difference as the recommendations form only a small part of the whole. Where the difference will be noticeable is in the summary versions. In the recent "short form" version of the guideline on schizophrenia produced by NICE,<sup>7</sup> 20 of the 64 pages are given over solely to the recommendations. Further specification within the recommendations could easily double this length. Such summaries are already way beyond a "single side of A4", the beloved utopia of guideline development. If the product of greater validity and further sophistication is more detailed (and thus longer) recommendations, there will need to be a corresponding increase in the

sophistication of the thinking and methods of those responsible for implementing them. Given that the short form guideline on schizophrenia produced by NICE also contains a corresponding 19 page representation of the guideline for "people with schizophrenia, their advocates and carers, and the public", we can be sure that reduction of the guideline to a single sheet for clinicians is not an appropriate response.

Having been handed a scope, guideline development groups will always struggle with interpreting the published evidence. Given the potential for the guideline content to be influenced by the range of factors described, it is too late to leave consideration of this to the point when recommendations are being written. If the aim is to produce a guideline that appropriately addresses the needs of all of a population (including those determined by ethnicity<sup>1</sup>), the implications of this for writing the recommendations should be one of the first—not the last—things to consider. However, such improvements will present challenges to all involved in developing, implementing, and using the guidelines that result.

*Qual Saf Health Care* 2003;**12**:328–329

#### Authors' affiliations

**M Eccles**, Professor of Clinical Effectiveness & The William Leech Professor of Primary Care Research, Centre for Health Services Research, School of Health Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4AA, UK; martin.eccles@ncl.ac.uk

#### REFERENCES

- 1 **Manna DR**, Bruijnzeels MA, Mokking HGA, *et al*. Ethnic specific recommendations in clinical practice guidelines: a first exploratory comparison between guidelines from the USA, Canada, the UK, and the Netherlands. *Qual Saf Health Care* 2003;**12**:353–8.
- 2 **Eccles M**, Mason J. How to develop cost-conscious guidelines. *Health Technol Assess* 2001;**5**(16).
- 3 **Shekelle PG**, Woolf SH, Eccles M, *et al*. Developing guidelines. *BMJ* 1999;**318**:593–6.
- 4 **Kahan JP**, Park RE, Leape LL, *et al*. Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Med Care* 1996;**34**:512–23.
- 5 **Coulter I**, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels—a comparison of a multi- and single disciplinary panel. *Health Serv Res* 1995;**30**:577–91.
- 6 **Leape LL**, Park RE, Kahan JP, *et al*. Group judgments of appropriateness: the effect of panel composition. *Qual Assur Health Care* 1992;**4**:151–9.
- 7 **National Institute for Clinical Excellence (NICE)**. *Schizophrenia. Core interventions in the treatment and management of schizophrenia in primary and secondary care*. London: NICE, 2002.