

Broadening the view of evidence-based medicine

Broadening the view of evidence-based medicine

D M Berwick

New methods of learning, new guidelines for publication

Scholars in the last half of the 20th century forged our modern commitment to evidence in evaluating clinical practices. They were courageous people, iconoclasts for their time, insisting that the scientific method was a necessary and plausible tool for judging the value of what we did for and to patients. Scientific evaluation of clinical practice was necessary, they argued, because unguided human observers are frail meters of truth—too prone to see what they expect to see, too likely to confuse effort with results or to attribute outcomes to visible causes rather than hidden ones, too trusting in small numbers and local opinion. Only formal scientific designs and strong statistical methods, they claimed, can protect the human mind from its own biases and adjust for hidden uncontrolled influences, sorting signals from noise. Scientific evaluation of practice is plausible, they argued, because the hypothetico-deductive method and proper statistical theory can be applied, with only modest adjustments, to the world of clinical process, just as it can be in a laboratory. And they taught us how to do that.

Their arguments were not welcomed at first. Today they are heroes, honored in the history of clinical science, but Archie Cochrane,¹ Alvin Feinstein,² Frederick Mosteller,³ Tom Chalmers,⁴ David Sackett,⁵ and others had first to play the role of outsiders, essentially pestering the center of health care to get serious about evaluating its work. They had thick skins, these critics, because they were—and had to be—change agents.

But they had evidence for their assertions, and they systematically accumulated more as they built their case. The risks of unguided impression were documented well in studies of the emergence and persistence of clinical practice of little or no value, once studied. Gastric freezing,⁶ radical mastectomy,⁷ theophylline for asthma,⁸ and dozens of other common practices wilted under the microscope of properly designed clinical trials, proving no better than simpler practices or out-and-out harmful. Beliefs and evidence simply did not always correspond.

A normative framework emerged for judging the value of evidence, a heraldry made clear in works such as the monumental volumes on effective care in pregnancy and childbirth compiled by Iain Chalmers and his colleagues,⁹ the reports of the Canadian Task Force on preventive medicine,¹⁰ Mosteller's magisterial Institute of Medicine report on assessing medical technologies,¹¹ Feinstein's texts,² Sackett's series in the *Canadian Medical Association Journal*,¹² and in the United States Task Force on Clinical Preventive Services.¹³ The Crown Prince of methods was the randomized, double blind, prospective, controlled clinical trial—the "RCT"—which stood second to no other method in protecting the scientist and the reader against bias, confounding, and other generators of false conclusion. Below the RCT stood methods of less nobility, graded in their evidence value from the properly designed cohort and case-control studies of epidemiology to the lowly case series, the suspect expert opinion, and the bestial anecdote. Systematic reviews that allow concatenation of results across multiple studies of varying design and quality, and meta-analysis that uses high powered statistics to combine quantitative findings from across multiple comparison studies, have further ramped up the power of controlled protocol-driven evidence.¹⁴

The leaders of evaluative clinical science fostered their young, and a generation of new scholars emerged in healthcare academia, founding their careers on evaluation of practice and on the progression of methods for evaluation. One of the most impressive success stories of 20th century medicine was how these people and these views—the entire field of "clinical epidemiology" and "evaluative clinical sciences"—not only survived but thrived, and eventually placed its leaders—scholars of the caliber of John Eisenberg, Christine Cassel, Harold Sox, and many others—in positions of the highest influence in departments of medicine, journal editorships, and professional societies, honoring work that a few decades before would not even have been understood to be about health care.

The benefits of evidence-based medicine, thus defined, have been immense. Patients today can count on a growing proportion of the tests, diagnostic processes, surgical procedures, and other costs and risks in care to have been subjected to proper systematic evaluation. The very definition of "quality" in health care has now come to incorporate the use of scientific evidence in practice; that is what the Institute of Medicine meant in its call for improvement of "effectiveness" as a key aim for improving care.¹⁵ Gaps between science and practice remain wide, but we seem increasingly committed to closing them. That is good.

But, we now have a problem: we have overshot the mark. We have transformed the commitment to "evidence-based medicine" of a particular sort into an intellectual hegemony that can cost us dearly if we do not take stock and modify it. And because peer reviewed publication is the sine qua non of scientific discovery, it is arguably true that hegemony is exercised by the filter imposed by the publication process. The failure of the publication filter to accommodate the kind of discovery that drives most improvement in health care—and the failure of those working in healthcare improvement to reconfigure the filter appropriately—is the message of the paper on publication guidelines by Davidoff and Batalden in this issue of *QSHC*.¹⁶ This paper is important, not only because it addresses the narrower issue of publication standards but also because it provides important support for an epistemology of a new and broadened understanding of the evidence needed for the improvement of care.

The argument for that epistemology is not a simple one, but its intuitive force is somewhat easier to uncover with a simple question: "How much of the knowledge that you use in your successful negotiation of daily life did you acquire from formal scientific investigation—yours or someone else's?"

Did you learn Spanish by conducting experiments? Did you master your bicycle or your skis using randomized trials? Are you a better parent because you did a laboratory study of parenting? Of course not. And yet, do you doubt what you have learned?

Broadly framed, much of human learning relies wisely on effective approaches to problem solving, learning, growth, and development that are different from the types of formal science so well explicated and defended by the scions of evidence-based medicine. Although they are far from RCTs in design, some of those approaches offer good defences against misinterpretation, bias, and confounding. In the world of clinical care, especially in the

quest for improvement of clinical processes, is it plausible that those approaches—the ones we use in everyday life—might have value too, used well and consciously, to help us learn?

The answer is “Yes”. And yet, the very success of the movement toward formal scientific methods that has matured into the modern commitment to evidence-based medicine now creates a wall that excludes too much of the knowledge and practice that can be harvested from experience, itself, reflected upon. The iconoclasts of the past now have power, and they can define who will be seen as iconoclasts of the present.

There is a way out. It involves curiosity. The methods of observation and reflection on the basis of which most human learning occurs and, frankly, on the basis of which many modern industries and enterprises are building their futures, are systematic, theoretically grounded, often quantitative, and powerful. They do not include RCTs, but they honor RCTs in their proper place. They perhaps deserve some honor in return, or at least the serious open minded scrutiny that marks true scholarship.

My close friend and mentor Tom Nolan PhD uses a felicitous term to denote these methods of learning: “pragmatic science”.¹⁷ Here are a few of the elements of the methods of pragmatic science:

- tracking effects over time, especially with graphs (rather than summarizing with statistics that do not retain the information involved in sequences);
- using local knowledge—the knowledge of local workers—in measurement (rather than relegating measurement to people least familiar with the subject matter and work);
- integrating detailed process knowledge into the work of interpretation (inviting observers to comment on what they notice rather than

“blinding” them to protect them against what they know);

- using small samples and short experimental cycles to learn quickly (rather than overpowering studies and delaying new theories with samples larger than needed at the time); and
- employing powerful multifactorial designs (rather than univariate ones when the better questions for the time are formative, not summative).

Pragmatic science of this type is alive and well. It thrives in the halls of continual improvement of care now engaging the energies of thousands of healthcare leaders worldwide. It thrives in brilliant texts by theoreticians who have been teachers in sectors of the economy other than health care.^{18 19} But, to our great expense, it remains largely trapped on the far side of a publication wall well guarded by academicians who may, I think, have overlearned the crucial lessons of the courageous clinical methodologists of the past few decades. Today’s evaluation methodologists guard not only the portals of our journals, but also our curricula and the minds of our young professionals. Health care has much to gain if those portals now open again to a new wave of disciplined methods of learning from reflective practice, and disciplined methods of sharing the learning through transparent, accurate, and complete published reports—such as the use of publication guidelines—as explained and defended here by Davidoff and Batalden. Health care is too important and too fragile to deny it the benefits of disseminating the hard won fruits of systemic learning, however this learning takes place.

Qual Saf Health Care 2005;14:315–316.
doi: 10.1136/qshc.2005.015669

Correspondence to: Dr D M Berwick, Institute for Healthcare Improvement, 20 University Road, 7th Floor, Cambridge, MA 02138, USA; dberwick@ihi.org

REFERENCES

- 1 **Cochrane AL.** *Effectiveness and efficiency: random reflections on health services.* London: Royal Society of Medicine Press, 1999.
- 2 **Feinstein, AR.** *Clinimetrics.* New Haven: Yale University Press, 1987.
- 3 **Bunker JP,** Barnes BA, Mosteller F. *Costs, risks, and benefits of surgery.* Oxford: Oxford University Press, 1977.
- 4 **Chambers TC.** Meta-analysis in clinical medicine. *Trans Am Clin Climatol Assoc* 1987;**99**:144–50.
- 5 **Sackett DL,** Rosenberg WM. On the need for evidence-based medicine. *Health Econ* 1995;**4**:249–54.
- 6 **Edmonson JM.** Gastric freezing: the view a quarter century later. *J Lab Clin Med* 1989;**114**:613–4.
- 7 **Fisher B,** Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med* 2002;**347**:1270–1.
- 8 **National Asthma Education and Prevention Program.** Key clinical activities for quality asthma care: recommendation of the National Asthma Education and Prevention Program. *Morbidity and Mortality Weekly Report: Recommendations and Reports* 2003;**52**(RR-6).
- 9 **Chalmers I,** Enkin M, Keirse M. *Effective care in pregnancy and childbirth.* New York: Oxford University Press, 1989.
- 10 **Woolf SH,** Battista RN, Anderson GM, et al. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. *J Clin Epidemiol* 1990;**43**:891–905.
- 11 **Institute of Medicine.** *Assessing medical technologies.* Washington, DC: National Academy Press, 1985.
- 12 **Sackett DL.** The hypertensive patient: 1–6. *Can Med Assoc J* 1979;**120**:1319–20, 1477–8; **121**:7–11, 145–8, 259–61, 397–8.
- 13 **US Preventative Services Task Force.** *The guide to clinical preventative services: report of the United States Preventative Services Task Force,* 1st edn. New York: Lippincott Williams & Wilkins, 1989.
- 14 **Egger M,** Smith GD, Altman DG. *Systematic reviews in health care: meta-analysis in context.* London: BMJ Books, 2001.
- 15 **Institute of Medicine.** *Crossing the quality chasm.* Washington, DC: National Academy Press, 2001.
- 16 **Davidoff F,** Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 2005;**14**:319–25.
- 17 **Brock WA,** Nolan KM, Nolan TW. Pragmatic science: accelerating the improvement of critical care. *New Horizons* 1998;**6**:61–8.
- 18 **Juran JM,** Godfrey AB. *Juran’s quality handbook,* 5th edn. New York: McGraw-Hill, 1999.
- 19 **Langley GL,** Nolan KM, Nolan TW, et al. *The improvement guide.* San Francisco: Jossey-Bass Publishers, 1996.

Consensus publication guidelines: the next step in the science of quality improvement?

R G Thomson

The time is ripe for a formal structured review of guidance on quality improvement reports

Samuel Beckett wrote “*Ever tried. Ever failed. No matter. Try again. Fail again. Fail better.*” Fiona Moss and I tried some time ago to produce a structure for publication of quality improvement reports on behalf of this journal that would facilitate and encourage their publication;¹ the *BMJ* subsequently adopted the structure for their authors.² Now, several years on from these first attempts, Davidoff and Batalden suggest new publication guidelines that build on this earlier version.³ These should be welcomed—in the spirit of improvement and intellectual evolution, it would be very surprising if the first attempt were to remain unchanged and unchallenged.

Before commenting on their proposals, it is worth reflecting on the original purpose of the development of quality improvement reports. This was based on the demand, within an emerging science and practice of quality improvement, for shared learning and dissemination of good practice. Those in the field knew that there were many excellent examples of successful projects where real changes had been made with demonstrable impact upon patient care, but that these examples were rarely disseminated such that others could draw upon this experience.

Why was this? As Davidoff and Batalden³ point out, one reason was the nature of the people responsible for quality improvement work—they are often highly committed people whose primary incentives are delivering and improving good clinical care; for most of them the incentive and the perceived rewards of publication are low, and the often thankless task of writing, submission, revision, rejection, and so on was a distraction from the next patient or the next round of quality improvement.

Alongside this were other interwoven issues. Firstly, the nature of original publication in scientific journals (in contrast to articles such as reviews, opinion pieces, editorials) is largely focused on original research articles, where the aim is

to report generalisable results. Secondly, the structure and guidance for writing up original research in the internationally accepted IMRaD (introduction, methods, results and discussion) format was designed to meet the needs of reports of original research and not of quality improvement. Thirdly, editors and reviewers were largely socialised into a mindset that gave a predominance to original research, compounded by the structure and guidance available for peer reviewers and authors. Thus, although not explicit in the original arguments, creating a new structure for quality improvement reports also acknowledged that such reports were different and gave them a focus and identity to enable them to escape the shackles of the traditional journal article.

Times have moved on since then but, even with the availability of the new structure and increased journal capacity to publish such articles, it is still a struggle to get such reports written and submitted; they remain very much a minority of all articles published in this journal. Will the revised proposal help with this?

Before it can do so, I think there are several points to consider. Firstly, exactly what sort of activity should be reported in quality improvement reports? I am not sure that the article by Davidoff and Batalden³ is quite clear about that. It seems to consider not only reports of effective quality improvement projects (as do the current guidelines), but also studies of the efficacy of quality improvement methods. This needs clarification. Studies of the efficacy of methods require, in their purest form, robust intervention studies such as randomised controlled trials (probably cluster randomised) in order to produce generalisable results. And guidelines in this area already exist (e.g. CONSORT⁴).

Furthermore, such studies are likely to be best applied to methods that can be generalised across a range of settings and topics—for example, the use of statistical process control charts. But

the original concept of quality improvement reports, at least to my mind, was to enable practitioners to share and learn from practical examples of projects—for example, a clinician who wants to undertake a project to improve the quality of acute treatment of myocardial infarction in his or her hospital would seek to find examples of others who have done the same in order to be able to apply and/or adapt their methods and experience to his or her own circumstances. This is very much about sharing experience and learning rather than sharing results. Indeed, we argued originally that the methods of quality improvement reports might be more generalisable than the results.¹

Secondly, the suggested reversion to the IMRaD structure is worthy of challenge. Does this really fit the purpose of quality improvement reports? The answer may well be “yes” if considering studies of methodological efficacy, but I am not convinced this is the case for reporting practical examples of quality improvement. The authors need to justify this further, not least by explaining why it might be preferable to the present accepted structure? I don’t believe the use of IMRaD is justified on the basis of incorporating “several important additional topics”—those listed in the article (such as prior information available on the problem area and assessment of the project’s limitation) are topics that are clearly covered within already published quality improvement reports using the present structure. These could be made more explicit by a revision of the original guidance without necessitating reversion to IMRaD.

Finally, as the authors point out, most quality improvement work is never made publicly available. This is undoubtedly true, but one only has to think about how many quality improvement projects may be in process in a single acute hospital, and then multiply that up by all acute hospitals internationally, to recognise that that will always be the case. Thus, publication in academic journals is only likely to be a limited, albeit valuable, method for dissemination of such practice; it needs to be part of a suite of methods of publication and dissemination. Included in the former, one might argue that (web based) databases of projects with very limited information but providing contact details for others to communicate with the project leads could be a major development—web sites such as the IHI site⁵ and the recently released saferhealthcare site⁶ can contribute here. In the latter, methods such as clinical networks and quality improvement cooperatives can fulfil a similar purpose.

In summary, I welcome the proposal for enhancement of current guidelines

for publication of quality improvement reports and the authors' suggestion for a more structured and formal approach to refining guidance is very sensible. The original guidance developed by *QSHC* involved a very similar informal process to that described by Davidoff and Batalden.³ The time is now ripe for a more formalised approach, and experience from other groups such as CONSORT⁴ or, perhaps more relevant, from the International Patient Decision Aids Standards (IPDAS)⁷ collaboration could be very helpful in developing the next stages proposed by the authors.

Both content and structure should be addressed.

Qual Saf Health Care 2005;14:317–318.
doi: 10.1136/qshc.2005.015727

Correspondence to: Professor R G Thomson, Director of Epidemiology and Research, National Patient Safety Agency and Professor of Epidemiology and Public Health, Newcastle upon Tyne Medical School, UK; Richard.thomson@newcastle.ac.uk

REFERENCES

- 1 Moss F, Thomson R. A new structure for quality improvement reports. *Qual Health Care* 1999;8:76.
- 2 Smith R. Quality improvement reports: a new kind of article. *BMJ* 2000;321:1428.
- 3 Davidoff F, Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 2005;14:319–25.
- 4 Moher D, Schulz K, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Ann Intern Med* 2001;143:657–62.
- 5 Institute for Healthcare Improvement. <http://www.ihc.org/IHI> (accessed 13 August 2005).
- 6 Saferhealthcare. <http://www.saferhealthcare.org.uk/ihc> (accessed 13 August 2005).
- 7 International Patient Decision Aids Standards collaboration (IPDAS). <http://ipdas.ohri.ca/> (accessed 13 August 2005).

Clinical Evidence—Call for contributors

Clinical Evidence is a regularly updated evidence-based journal available worldwide both as a paper version and on the internet. *Clinical Evidence* needs to recruit a number of new contributors. Contributors are healthcare professionals or epidemiologists with experience in evidence-based medicine and the ability to write in a concise and structured way.

Areas for which we are currently seeking authors:

- Child health: nocturnal enuresis
- Eye disorders: bacterial conjunctivitis
- Male health: prostate cancer (metastatic)
- Women's health: pre-menstrual syndrome; pyelonephritis in non-pregnant women

However, we are always looking for others, so do not let this list discourage you.

Being a contributor involves:

- Selecting from a validated, screened search (performed by in-house Information Specialists) epidemiologically sound studies for inclusion.
- Documenting your decisions about which studies to include on an inclusion and exclusion form, which we keep on file.
- Writing the text to a highly structured template (about 1500–3000 words), using evidence from the final studies chosen, within 8–10 weeks of receiving the literature search.
- Working with *Clinical Evidence* editors to ensure that the final text meets epidemiological and style standards.
- Updating the text every six months using any new, sound evidence that becomes available. The *Clinical Evidence* in-house team will conduct the searches for contributors; your task is simply to filter out high quality studies and incorporate them in the existing text.
- To expand the topic to include a new question about once every 12–18 months.

If you would like to become a contributor for *Clinical Evidence* or require more information about what this involves please send your contact details and a copy of your CV, clearly stating the clinical area you are interested in, to Klara Brunnhuber (kbrunnhuber@bmjgroup.com).

Call for peer reviewers

Clinical Evidence also needs to recruit a number of new peer reviewers specifically with an interest in the clinical areas stated above, and also others related to general practice. Peer reviewers are healthcare professionals or epidemiologists with experience in evidence-based medicine. As a peer reviewer you would be asked for your views on the clinical relevance, validity, and accessibility of specific topics within the journal, and their usefulness to the intended audience (international generalists and healthcare professionals, possibly with limited statistical knowledge). Topics are usually 1500–3000 words in length and we would ask you to review between 2–5 topics per year. The peer review process takes place throughout the year, and our turnaround time for each review is ideally 10–14 days.

If you are interested in becoming a peer reviewer for *Clinical Evidence*, please complete the peer review questionnaire at www.clinicalevidence.com or contact Klara Brunnhuber (kbrunnhuber@bmjgroup.com).