

What should we report to medical error reporting systems?

S M Dovey, R L Phillips

A satisfactory definition of “medical error” still eludes us

The Netherlands is the latest country to announce the development of a national medical error reporting system.¹ Australia has had one since 1989, Denmark has one, the UK introduced theirs in 2001, Canada announced their plans in 2003, and the USA has a proliferation of error reporting systems, including several that have been going for a number of years and that have a well developed body of knowledge steering their use and development—for example, the Medical Event Reporting System for Transfusion Medicine (MERS-TM)² and the US Pharmacopeia’s MEDMARX Reporting System³. Developed western countries do therefore seem to have “bought into” the message that medical error reporting systems are a very “good thing”—although there is little evidence that Johnson’s⁴ pragmatic cautions have been well considered in setting them up.

The reporting of “medical errors”—whatever they are—is still an embryonic endeavor and, before national and international medical error reporting systems get well under way, some crucial topics—such as defining what we are to report to error reporting systems—should be addressed with clarity. An occasional error reporting system has dealt with ambiguity over what needs to be reported by adopting a list of explicitly defined events “that should never happen”,⁵ but most are far less precise. Hopefully, the national medical error reporting systems of different countries will ultimately (if not initially) use the same definitions. We raise some issues here that expose the complexity of defining “medical error” and demonstrate just how peculiar, unnatural, and un-useful are some of the terms and definitions in current use.

Starting from first principles, it seems clear that a medical error reporting system should hold reports of “medical errors” and any dictionary will tell readers that “medical” means “relating to medicine”. This seems reasonably

straightforward but tends to have been narrowly interpreted by many medical error reporting systems as medical care provided by doctors and nurses to patients in hospitals. In developed countries most medicine is provided and received outside hospitals and a huge number of different occupational groups are involved—from accountants to microbiologists to social workers. The inclusion of medical errors happening and/or observed outside hospitals is crucial, and we therefore support a systems perspective that would draw attention to errors in medicine related policy, regulation, payment, and management as well as medical care delivery.

“Error” is an even more problematic word than “medical”. At its simplest it means “mistake”. It also has other distinct meanings in mathematical and sports contexts. Over the last few years it has started to look as if “error” does not mean simply “mistake” when preceded by “medical” but, as in mathematics, it has a context-specific meaning. In our view, this meaning has not yet been authoritatively defined *simply, clearly, usefully, and inclusively*.

Simple definitions of “medical error” do exist. What could be more simple than “underuse, overuse, or misuse”? But try applying it—categorizing events as one or other of these three—and you soon find that underuse of one thing is often overuse or misuse of another. When faced with coding and classifying reported errors,⁶ we found similar problems with “slips”, “lapses”, “knowledge-based mistakes”, and “rule-based mistakes”.⁷ In “lapsing” or forgetting to do one thing, a “slip” would happen—so which one was it? There were problems with “knowing” about “rules” that confounded coders. Ultimately, the reported events we dealt with just looked like mistakes (in the plain English sense) to us. We found that, despite their theoretical coherence, we could not use these simple definitions.

A similarly simple definition is “harm” or “adverse event”; that is, a mistake doesn’t qualify as a “medical error” unless it causes harm to patients and/or people are worse off than they would have been had the mistake not occurred. The debunked notion that medical malpractice claims are synonymous with medical errors is probably the genesis of this definition. Many favour it because, if something causes no harm (they argue), it is not worth being concerned about and because they maintain the “harm” definition is *clear* as well as *simple*. We disagree. By excluding medical errors that did not cause harm, medical error reporting systems will miss enormous opportunities to improve health care and create safer healthcare environments. Most error reporting system developers seem to agree with us on this. But they also seem to favour using another odd term—the “near miss”—to define these events. “Near miss” is an expression borrowed from the aviation industry and it makes intuitive sense in that context. One can imagine an aeroplane swooping close to a building, for example, and “missing”—but only just. “Near hit” captures the same idea. The translation of “near miss” to the industry of medicine is a bit of a stretch, though. One can never really know how “nearly” a medical error “misses” (presumably) causing harm. In aviation measures of “nearness” are objective—a kilometre is far enough away not to be regarded as “near” but a metre is not. There is no equivalently objective measure in medicine to determine whether an event qualifies as a “near miss” or not.

A further reason that we find the “harm” definition of medical error unsatisfactory is that the people who are the intended reporters to medical error reporting systems are poor judges of harm.⁸ Healthcare providers tend to discount consequences such as patients having extended waiting times, having to spend extra money on travel or taking time off work, or being emotionally upset. They tend instead to negatively weight outcomes that patients may actually regard as benefits rather than harms, such as death. The notion of “harm” is far too subjective and lacking in clarity for it to be rolled into a definition of “medical error”.

So, if the simple medical error definitions are problematic, perhaps it is inevitable that more complex definitions will have to be used. A number of these have been proposed such as: “the failure, for reasons that are preventable, of a planned action to be completed as intended (error of execution) or the use

of a wrong plan to achieve an aim (error of planning)”;⁹ “active failures that occur at the sharp end of a continuum of decisions, environmental factors, and actions that affect patient care”;¹⁰ and “anything small or large, administrative or clinical, that you identify as something to be avoided in the future, that happened in your own practice that should not have happened, that was not anticipated, that you don’t want to happen again”. These long definitions are summarized here—they become even longer and more detailed if readers go back to their sources. The longer a definition, the greater the chance something untoward and unhelpful will be included in it. The third of our examples, for instance, is a definition we developed while working with general practitioners and family physicians.⁶ It includes a phrase that became difficult: “... *that was not anticipated*”. Many of the primary care doctors we worked with encountered medical errors so regularly and frequently that they had trouble identifying errors that were not anticipated. In fact, the reporting system we were developing aimed to capture exactly these regular, frequent, and anticipated errors, so we had to revise the error definition we used in later work to remove the offending phrase.

If the beginning of wisdom is knowing what to call things, defining “medical error” is a beginning that has not

yet been completed. An internationally shared definition will be important because, just as the problems of mathematics are not the concern of any single country or constituency, neither are the problems of patient safety. Perhaps the most useful learning opportunities from overarching national reporting systems will come from international comparisons: there may be transferable characteristics of a country’s healthcare system that protect patients from certain kinds of harm and other characteristics that unnecessarily constrain patient safety. No country (let alone any organization or person) holds moral authority to unilaterally propose a “medical error” definition for general use. However, there are enough definitions already in circulation to inform fruitful discussions about what we are to report to national medical error reporting systems. Rather than more unilateral attempts to create the *best* definition, we look forward to consensus activities that will eventually deliver a sound definition we can all work with—patients, doctors, nurses, planners, policymakers, researchers, and others encountering medical errors in hospitals, primary care clinics, research units, government departments, ambulances, and anywhere else they occur.

Qual Saf Health Care 2004;13:322–323.
doi: 10.1136/qshc.2004.011791

.....
Authors’ affiliations

S M Dovey, Professorial Research Fellow, Health Sciences Faculty, University of Otago, P O Box 913, Dunedin, New Zealand
R L Phillips, Assistant Director, The Robert Graham Center, 1350 Connecticut Ave NW, Washington, DC 20036, USA

Correspondence to: Dr S Dovey, Pharmacy School and Dunedin School of Medicine, University of Otago, PO Box 913, Dunedin, New Zealand; sdovey@aafp.org

REFERENCES

- 1 Sheldon T. Netherlands plans system for reporting errors. *BMJ* 2004;329:68.
- 2 Kaplan HS, Callum JL, Rabin Fastman B, et al. The Medical Event Reporting System for Transfusion Medicine: will it help get the right blood to the right patient? *Transfus Med Rev* 2002;16:86–102.
- 3 MEDMARX. <http://www.usp.org/medmarx> (accessed 23 July 2004).
- 4 Johnson CW. How will we get the data and what will we do with it then? Issues in the reporting of adverse healthcare events. *Qual Saf Health Care* 2003;12(Suppl II):ii64–7.
- 5 National Quality Forum. http://www.qualityforum.org/safe_practices_report.html (accessed 23 July 2004).
- 6 Dovey SM, Meyers DS, Phillips RL, et al. A preliminary taxonomy of medical errors in family practice. *Qual Saf Health Care* 2002;11:233–8.
- 7 Reason J. Understanding adverse events: human factors. *Qual Health Care* 1995;4:80–9.
- 8 Woolf SH, Kuzel AJ, Dovey SM, et al. A string of mistakes: the importance of cascade analysis in describing, counting, and preventing medical errors. *Ann Fam Med* 2004;2:317–26.
- 9 Sheikh A, Hurwitz B. A national database of medical error. *J R Soc Med* 1999;92:554–5.
- 10 Medical Event Reporting System – Transfusion Medicine. <http://www.mers-tm.net> (accessed 23 July 2004).

Pay for performance

.....
Pay for performance: the best worst choice

M L Millenson

.....
A new concept in healthcare reimbursement that links payment and adherence to safety and quality standards

Pay for performance (“P4P”) is the latest catch phrase to cross over from the world of commerce to the work of clinicians. The basic concept is simple: rather than paying for care by the piecemeal method (fee for service) or using administered price arrangements (for example, daily rates, fee schedules and capitation), reimbursement should be linked at least in part to adherence to safety and quality measures.

According to the American Academy of Family Physicians, typical measures

center on utilization and cost management (for example, average number of emergency department visits per patient per year); clinical quality/effectiveness (for example, the percentage of patients with asthma on controller medications); patient satisfaction (for example, the percentage of patients who would recommend the physician to a family member or friend); administrative (for example, the practice’s level of information technology); and patient safety (for example, the percentage of patients questioned about allergic drug reactions).¹

P4P programs offered by health maintenance organizations (HMOs) in the US already affect more than 30 million people (or nearly a third of all HMO members), according to one survey. Physician practices participating in these programs find that 1–40% of their annual income is involved in a P4P bonus or withhold, with an average of 10%. More to the point, the percentage of state governments, employer coalitions, and health plans sponsoring these programs was projected to increase from 40% in 2003 to about 80% by 2006.²

Crucially, one of the new participants is likely to be the federal Medicare program. The Medicare Prescription Drug, Improvement and Modernization Act of 2003, which established a drug benefit for seniors, also directs the Institute of Medicine (IOM) to develop a strategy for aligning payment and clinical performance. Medicare and its sister programs for the poor and for children together account for close to a third of all US health care spending.

In the UK, meanwhile, the National Health Service’s current contract for general practitioners provides financial

incentives for the achievement of a set of 76 clinical quality indicators covering 10 disease groups.

The popularity of P4P can be credited to a combination of considerations. On a practical basis, efforts to force doctors and hospitals to practice in a certain way, rather than to provide incentives, have generally failed. In a consumerist society, meanwhile, the idea of paying for performance resonates with patients as well as providers. More broadly, P4P is being proffered at a time when soaring healthcare costs are regarded by many countries as a national crisis. As a result, all parties are more willing to explore new payment methods, particularly one based on the thesis that, in health care as in other industries, higher quality (or “better performance”) equals lower cost.

Although the high quality/low cost relationship in health care has yet to be conclusively demonstrated,³ it is hardly a novel concept. In the first part of the 20th century Boston surgeon Ernest Amory Codman defined a hospital’s “product” as the extent to which it reliably produced cured patients. Codman wondered why businessmen trustees did not make their institutions more “efficient” producers. In the early 1970s British epidemiologist Archie Cochrane placed efficiency into a wider social context. He considered that, because of limited societal resources, only healthcare services shown to be effective should be provided to patients. Cochrane’s work formed the basis for the evidence-based medicine movement.

A strategy for translating efficiency theory into market place practice, however, seems first to have been articulated by Walter McClure, a Minnesota physician turned health policy activist. His “Buy Right” program in the early 1970s enlisted corporate purchasers to the cause of quality improvement by insisting that incentives rather than doctor and hospital greed was the problem. He considered that providers would cooperate in meeting efficiency measures as soon as large purchasers—the employers who purchase health care on behalf of workers—threatened to either “buy right” or “buy cheap”.⁴

Alas, most US purchasers still preferred to buy cheap—particularly since doctors and hospitals insisted that quality measures were unreliable. “Buy Right” was reborn only after the managed care backlash of the mid 1990s forced managed care organizations to back pedal on tough provider utilization controls and price negotiations. Renamed P4P or “value purchasing” (to differentiate it from “buy cheap”),

it first surfaced with the formation in late 2000 of the Leapfrog Group, a coalition of large corporations. Leapfrog used highly specific purchasing requirements whose benefits were backed up with academic research—for example, the use of computerized physician order entry.

In 2001 a seminal report by the Institute of Medicine gave pay for performance an important professional endorsement. In their report “*Crossing the Quality Chasm*” they noted that “even among health professionals motivated to provide the best care possible, the structure of payment incentives may not facilitate the actions needed to systematically improve the quality of care, and may even prevent such actions”. They stated that private and public purchasers must modify their payment methods to “recognize quality, reward quality and support quality improvement”.⁵ Quality problems—defined as overuse, underuse and misuse—were said to be both common and expensive.

Adding to the P4P momentum has been the growth of so-called consumer driven healthcare plans which combine large deductibles with better cost and quality information in order to persuade patients to make more value conscious purchasing decisions.

While P4P may seem like a fairy tale solution that leaves everyone living happily ever after, significant sticking points remain. Although some early research has shown that it is possible to quantify the health gain to a practice population of achieving quality targets,⁶ much more research remains to be done on program implementation, the design features that promote success, and those which serve as barriers.⁷ In addition, there is little standardization across plans in how quality improvement is measured, and incentive payments typically are modest in comparison with providers’ total revenue.⁸ Expectations also must be realistic; even a sympathetic researcher cautions that the task of transforming the current market place into one that promotes quality improvement is a difficult task that may not be completed until “well into the 21st century”.⁹

Along the way, deciding what to measure and how to measure it will be critical. The wrong measures or the wrong kind of measurement can easily erode trust and prompt widespread gaming of the system. Some physicians are optimistic, regarding P4P as an opportunity for broad scale collaboration to improve care.¹⁰ Others are more sceptical. For example, the chairman of the American Medical Association board

denounced P4P as a “scam” designed by “multimillionaire CEOs of health insurance companies” to cut reimbursement by taking advantage of gullible physicians.¹¹

Yet whatever the potential hazards of hastening down the P4P path, one cannot help but be reminded of Churchill’s often quoted remark that “democracy is the worst form of Government except all those others that have been tried”. Every reimbursement system creates some sort of potential conflict of interest. As Rodwin has written, what is needed to reduce the clash between “medicine, money and morals” are policies that hold doctors accountable to patients for fulfilling the profession’s ideals.¹²

Pay for performance is no panacea and implementation challenges abound. Nonetheless, it offers the greatest potential yet for balancing the autonomy that is critical to the practice of medicine with the accountability that is equally critical to patients receiving safe and high quality care.

Qual Saf Health Care 2004;13:323–324.
doi: 10.1136/qshc.2004.011668

Correspondence to: M L Millenson, The Kellogg School of Management, Northwestern University, Evanston, IL, USA; m-millenson@northwestern.edu

REFERENCES

- 1 **Endsley S**, Kirkegaard M, Baker G, *et al*. Getting rewards for your results: pay-for-performance programs. *Fam Pract Manage* 2004;11:45–50.
- 2 **Med-Vantage**. *National study of provider pay-for-performance programs: lessons learned*. San Francisco: Med-Vantage Inc, 2003.
- 3 **Leatherman S**, Berwick D, Iles D, *et al*. The business case for quality: case studies and an analysis. *Health Aff* 2003;22:17–30.
- 4 **Millenson ML**. *Demanding medical excellence: doctors and accountability in the information age*. Chicago: University of Chicago Press, 1997:206–8.
- 5 **Institute of Medicine**. *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press, 2001:194.
- 6 **McElduff P**, Lyratzopoulos G, Edwards R, *et al*. Will changes in primary care improve health outcomes? Modelling the impact of financial incentives introduced to improve quality of care in the UK. *Qual Saf Health Care* 2004;13:191–7.
- 7 **Maio V**, Goldfarb NI, Carter C, *et al*. *Value-based purchasing: a review of the literature*, Report No. 636. The Commonwealth Fund, 2003.
- 8 **Strunk BC**, Hurlley RE. *Paying for quality: health plans try carrots instead of sticks*. Issue Brief No. 82. Center for Studying Health System Change, 2004.
- 9 **Casalino LP**. Markets and medicine: barriers to creating a “business case for quality”. *Perspect Biol Med* 2003;46:38–51.
- 10 **Lee TH**, Meyer GS, Brennan TA. A middle ground on public accountability. *N Engl J Med* 2004;350:2409–12.
- 11 **Plested, WG III**. Fair pay a sounder approach than “pay for quality”. *Am Med News*, 1 March 2004.
- 12 **Rodwin MA**. *Medicine, money, and morals: physicians’ conflicts of interest*. New York: Oxford University Press, 1993:xvii.

The volume-quality relationship: insufficient evidence for use as a quality indicator

T A Sheldon

The relationship between volume and quality is still unclear

Health service planners are increasingly trying to find ways to improve the quality and safety of health care. A wide range of approaches is being used from high level regulatory frameworks, use of clinical guidance and guidelines, to more micro level activity such as audit of care. None of these is easy; all require significant investment of resources, training, time and monitoring. The results are often uneven and result in variations in quality as initiatives diffuse unevenly through the system. It is understandable then that policy makers seek easier ways to deliver these improvements.

Research since the late 1970s seemed to point in the direction of a relatively constant relationship in health care—increased use of a hospital procedure reduces the mortality associated with it. The message emerging from a large number of studies, mainly from the USA, was that patients treated in hospitals which (or by clinicians who) managed high volumes of patients with the same condition had better outcomes than those with lower volumes. This was summed up by Luft and others in an influential report in 1990.¹ Hundreds of studies have been published, many of which are based on analysis of large US administrative databases, and most report an inverse relationship between the volume of activity and mortality or other poor outcomes. These studies have been reviewed and summarised in several publications in the last few years.²⁻⁵ Although there was not full agreement, in general the reviews—especially those published in the USA—support the hypothesis of a volume-outcome relationship and the existence of volume-quality thresholds.

The policy response was predictable. Several national and state regulators and professional associations set volume thresholds for hospital based procedures and for hospital accreditation and pushed for the regionalisation of services. Policy makers like simple messages and interventions more under their control which do not depend

heavily for their success on the compliance of clinicians. The idea that volume directly influences outcome has become so accepted that at times volume is used as a proxy for—and even confused with—outcome. The influential Leapfrog Group for Patient Safety recently produced a fact sheet on evidence based hospital referral,⁶ giving advice on the criteria to be used to select high quality hospitals in which they state that:

“Another measure of surgical outcomes is **volume** – how many procedures of a given type a hospital performs each year. ... Choosing the right hospital is not just important in surgery. For example, babies with very low birth weight or major congenital abnormalities are much more likely to survive if they are treated at hospitals with large neonatal care units.”

The consensus around the relationship between volume and outcomes and its potential use as a measure of quality is, however, based on shakier foundations than many leading researchers in the field care to acknowledge. The quality of the studies underpinning the commonsense view has often been poor and their interpretation at times disingenuous; concerns expressed a decade ago were largely ignored.²⁻⁷⁻⁸ However, in the last couple of years the pendulum has begun to swing back and a recent methodological critique⁹ and new analyses¹⁰⁻¹² have raised concerns about both the evidence for this relationship and its use for policy making.

One concern with the research in this field is that it consists mainly of cross sectional studies, often using information from administrative databases with relatively little clinical detail. If units with higher volume were admitting patients who, on average, were less seriously ill (indeed, this might be a consequence of having larger capacity), then a straight comparison would

automatically show high volume units to have lower mortality than low volume units. The association would, however, be spurious—the result of the confounding effects of case mix. Only if these studies take into account variations in the distribution of prognostic factors in patients across the units—risk adjustment—can the associations be trusted. This was shown, for example, in a meta-analysis of studies of coronary bypass surgery (one of the most studied procedures and strongly regulated by volume thresholds) which showed that studies which poorly adjusted for case mix reported much stronger associations between volume and outcome than those which more adequately took case mix into account.⁸

The quality of research in this area has improved over the last few years with some interesting results.⁴ The largest and most recent study based on a high quality clinical database using very good risk adjustment did not show a strong association between volume and quality.⁹ The authors concluded that the use of hospital procedural volume of coronary artery bypass surgery is of limited value in discriminating between better or worse risk adjusted mortality outcomes.⁹ Another high quality study of the mortality of very low birth weight infants in the US showed that the volume of the neonatal intensive care units “cannot prospectively identify high quality providers”.¹⁰ These two studies strongly suggest that the sort of guidance given and thresholds applied by the Leapfrog Group and others are unjustified and do not warrant the description “evidence based”.

This does not mean that there is no association between volume and quality; for some procedures (such as AIDS treatment, surgery for pancreatic and oesophageal cancer, abdominal aortic aneurysm, and paediatric cardiac procedures) there does appear to be reasonably strong evidence of a relationship between volume and improved patient outcomes.²⁻⁴ In other areas this seems only to be relevant in high risk patients or very low volume providers.

However, even if there is a valid statistical association between volume and outcome, this does not tell us whether it is causal. In other words, and most importantly, it does not provide evidence as to whether a policy of hospitals increasing their volumes will result in an improvement in their clinical outcomes. To establish the impact of changing volume on outcomes needs a prospective and preferably experimental design.¹¹ In their review, Sowden *et al*⁷ identified only two longitudinal studies which examined changes in outcomes as volumes varied

over time. Even though these studies found cross sectional associations between volume and in-hospital mortality, there was no relationship between changes in volume and outcome over time. The absence of reliable prospective data raises serious concerns about our ignorance of the likely impact of policies to concentrate services.

Even if the research showed a causal link, we would need to know more about the mechanism of action so that it could be implemented effectively. For example, if there is an effect, what are the relative contributions from volume at the hospital, unit, team or clinician level? Rarely do studies consider the relative effects of the unit of analysis and their possible interactions. Importantly, Urbach and Baxter, using data from Ontario, Canada, recently reported that the inverse association they found between procedure volume and postoperative mortality risk was not specific to the volume of the procedure being studied!¹² As with several previous studies, they found that higher volume was associated with lower 30 day mortality from repair of abdominal aortic aneurysm, oesophagectomy, pancreaticoduodenectomy, major lung resection for lung cancer, and not from colorectal cancer surgery. However, the reduction in mortality from all these procedures except colorectal cancer was also associated with higher volumes in the other procedures, sometimes more than with its own volume. Although the study can be criticised for using volume thresholds rather than treating volume as a continuous measure and for inadequate case mix adjustment

compared with some of the best, the findings raise serious questions about what we are really measuring when such associations are calculated. Simple empirical associations are not a sufficient basis for policy making in this field. We need to understand more about what is going on in units with different levels of quality. As the authors suggest, we need to “revisit the conceptual framework underlying volume based regionalisations”.

This is not just a technical debate. The drive to increase volumes can have adverse effects which would need to be outweighed by greater benefits to be justifiable. For example, hospitals worried about accreditation or losing business might inflate volumes by artificially lowering the threshold for treatment, admitting patients with less need and so reducing efficiency. More importantly, the policy of concentrating services can result in reduced local access for services, more travel time, and possibly suboptimal local treatment. We have been urged to “move ahead” on the question of volume and outcome,¹³ however, perhaps with the exception of a few high risk procedures where the evidence is clear, we still do not know enough about the meaning of the relationships found to be able confidently to use volume as an effective policy instrument.

Qual Saf Health Care 2004;**13**:325–326.
doi: 10.1136/qshc.2004.012161

Correspondence to: Professor T A Sheldon, Department of Health Sciences, University of York, Heslington, York YO10 5DD, UK; tas5@york.ac.uk

This editorial was inspired by an article originally published in *BMJ* 2004;**328**:737–40.

REFERENCES

- 1 **Luft HS**, Garnick DW, Mark DH, et al. *Hospital volume, physician volume and patient outcomes. Assessing the evidence*. Michigan: Health Administration Press, 1990.
- 2 **Sowden A**, Aletras V, Place M, et al. Volume of clinical activity in hospitals and healthcare outcomes, costs and patient access. *Qual Health Care* 1997;**6**:109–14.
- 3 **Halm EA**, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002;**137**:511–20.
- 4 **Grandjour A**, Bannenberg A, Lauterbach KW. Threshold volumes associated with higher survival in health care: a systematic review. *Med Care* 2003;**41**:1129–41.
- 5 **Finlayson SR**, Birkmeyer JD. The association between provider volume and characteristics and patient outcomes in colorectal surgery. *Seminars in Colon and Rectal Surgery* 2003;**14**:53–61.
- 6 **The Leapfrog Group for Patient Safety**. Evidence-based hospital referral, Fact Sheet. www.leapfroggroup.org, 2004.
- 7 **Sowden AJ**, Sheldon TA. Does volume really affect outcome? Lessons from the evidence. *J Health Serv Res Policy* 1998;**3**:187–90.
- 8 **Sowden AJ**, Deeks JJ, Sheldon TA. Volume and outcome in coronary artery bypass graft surgery: true association or artefact? *BMJ* 1995;**311**:151–5.
- 9 **Sheikh K**. Reliability of provider volume and outcome associations for healthcare policy. *Med Care* 2003;**41**:1111–7.
- 10 **Peterson ED**, Combs LP, DeLong ER. Procedural volume as a marker of quality for CABG surgery. *JAMA* 2004;**291**:195–201.
- 11 **Rogowski JA**, Horbar JD, Staiger DO. Indirect vs direct hospital quality indicators for very low-birth-weight infants. *JAMA* 2004;**291**:202–9.
- 12 **Urbach DR**, Baxter NN. Does it matter what a hospital is “high volume” for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *BMJ* 2004;**328**:737–40.
- 13 **Epstein A**. Volume and outcome – is it time to move ahead? *N Engl J Med* 2002;**346**:1161–4.