

Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away

Standardised mortality rates are a poor measure of the quality of hospital care and should not be a trigger for public inquiries such as the investigation at the Mid Staffordshire hospital, say **Richard Lilford** and **Peter Pronovost**

Death is the most tractable outcome of care—it is easily measured, of undisputed importance to everyone, and is common in hospital settings. Mortality rates, especially overall hospital mortality rates, have therefore become the natural focus for measurement of clinical quality. In England a high death rate “attracted the attention of the [Healthcare Commission] (HCC) and caused it to launch its investigation” into the Mid Staffordshire NHS Foundation Trust.¹

So what is the problem with measuring clinical performance by comparing hospital mortality rates and what alternatives can we offer?

Hospital mortality as a measure of quality: scientific issues

The problem stems from the ratio of a low signal (preventable deaths) in relation to high noise (deaths from other causes). A common but naive response is to argue that risk adjustment to produce a standardised mortality ratio (SMR) solves this problem. However, the idea that a risk adjustment model separates preventable from inevitable deaths is wrong for two reasons.

Firstly, risk adjustment can only adjust for factors that can be identified and measured accurately.² Randomised trials are preferable to observational studies with statistical controls for this reason. The error of attributing differences in risk adjusted mortality to differences in quality of

care is the “case-mix adjustment fallacy”.³

Secondly, risk adjustment can exaggerate the very bias that it is intended to reduce. This counterintuitive effect is called the “constant risk fallacy” and it arises when the risk associated with the variable on which adjustment is made varies across the units being compared.⁴ For example, if diabetes is a more powerful prognostic factor in Glasgow than in Four Oaks, then adjusting for the average effect of diabetes will deflate expected diabetic deaths in Glasgow and inflate them in Four Oaks. Depending on the distribution of diabetes, this effect could tilt the playing field against Glasgow. The extent to which comorbidities are recorded in hospital statistics (the coding depth) is a potent source of this form of bias.⁵

Moreover, SMRs vary by about 60% across UK hospitals.⁶ The proposal that variance of this magnitude can be attributed to differences in the quality of care is not clinically intuitive and does not align well with the following observations:

(1) Mant and Hicks showed that differences in the quality of care could explain only half the observed variance in heart attack mortality even in a fictitious scenario where adherence to the tenets of good care varied from 0% to 100%.⁷

(2) The famous Harvard malpractice study found that 0.25% of admissions resulted in avoidable death.⁸ Assuming an overall hospital death rate of about 5% this implies that around

one in 20 inpatient deaths are preventable, while 19 of 20 are unavoidable. We have corroborated this figure in a study of the quality of care in 18 English hospitals (submitted for publication). Quality of care accounts for only a small proportion of the observed variance in mortality between hospitals. To put this another way, it is not sensible to look for differences in preventable deaths by comparing all deaths.

(3) Little or no correlation exists between how well a hospital performs on one standard of safe and effective care and how well it performs on another; differences in the quality of care within hospitals are much greater than differences between hospitals.^{9,10} This finding does not support the prevailing notion of large scale systematic differences in quality at the institutional level and suggests that while commercial organisations such as Enron fail corporately, hospitals are more likely to fail on the specifics—pathology in Liverpool¹¹; paediatric cardiac surgery in Bristol¹²; radiation therapy in Missouri.¹³

In view of these findings, it is not surprising that a systematic review of empirical studies showed very little correlation between measured quality of care and SMRs.¹⁴ In short, hospital mortality rates are a poor diagnostic test for quality and SMRs do not identify preventable deaths.^{15,16} Given the fragility of SMRs, they should not be used to calculate excess deaths resulting from



The potential injustice arises from the questionable veracity of a public review carried out with subjective tools in the expectation of finding faults under an atmosphere of escalating emotion

poor care, yet Mid Staffordshire hospital was blamed for 400 excess deaths on this precarious basis.¹⁷

Consequences of using mortality rate to (mis)judge quality of care

Mortality rates, like knives and nuclear particles, are neutral; it is the use to which they are put that has moral salience and that will determine the balance of benefits and harms. We believe that it is not collection of mortality rates per se that is wrong, but rather the use of mortality rates as a criterion for “performance management” (that is, as the basis for sanction or reward). Instituting penalties (or withholding rewards) on the basis of hospital mortality rates is correctly perceived as unjust. Moreover, hospital mortality rates are silent about where any problem might lie. This combination of unfairness and non-specificity is a toxic mix, inducing what has been called “institutional stigma”—a feeling of helpless frustration.³ Human beings are strongly motivated by stigma and will take the shortest route to avoid it,¹⁸ even if to do so involves “gaming” the system—for example, by upgrading risk assessments.¹⁹ Furthermore, a focus on hospital mortality may lead to overly aggressive care, which is inhumane and drives up costs.²⁰

Defining the principled use of mortality rates

There is an argument for use of hospital mortality rates not as the basis for judgment leading to sanction or reward, but as a signal to identify where further investigation is necessary. This is a seductive argument because at one level it is self evidently correct; there are good arguments for keeping mortality rates under review, so that interesting differences can be investigated. Comparing practices in different places has provided crucial insights, from Semmelweis’s discovery of antiseptics to more recent evidence about the optimal age for surgery to repair congenital mitral valve disease.²¹

However, controlled studies of this type are entirely different from single site investigations called down on organisations by regulatory bodies. In Mid Staffordshire “concerns about mortality” prompted first the HCC investigation,²² then two reviews commissioned by the Department of Health,^{23 24} and finally an independent inquiry set up by the secretary of state for health.¹ Use of mortality rates to prompt such examinations is not politically neutral; the investigation is itself a sanction. The potential injustice arises from

the questionable veracity of a public review carried out with subjective tools in the expectation of finding faults under an atmosphere of escalating emotion.²⁵ Notably, although the majority of people who came forward to give evidence to the Mid Staffordshire inquiry made negative comments, the hospital was certainly not an outlier on the survey of patient opinion that is carried out in all English hospitals every two years.²⁶

Public inquiries may be needed from time to time, but like other interventions they have their own dangers. Our point is not that the Mid Staffordshire hospital was blameless, but that other unidentified hospitals might have been found equally deficient had they been investigated to the same degree in response to the same initial trigger. The signal that initiates an inquiry is therefore of crucial concern. Once a public inquiry is instigated it takes on a life of its own. Far from vindicating the use of SMRs, as some have claimed, finding problems becomes an almost self fulfilling prophecy.²⁷ Thus a twofold risk exists. Firstly, a management team that is no worse than many others will be unjustly singled out by a process that, like the persecution of witches, was well intentioned but informed by the wrong theory.²⁸ Secondly, a myopic focus on a single institution is a distraction from the types of systematic inquiry that may lead to pervasive improvement across a service, such as reducing bloodstream infections in all hospitals.²⁹

Why is use of mortality statistics for performance monitoring so enduring?

Few, if any, of the above arguments are original. Yet the monster survives like the mythical Hydra, which sprouted new heads as quickly as the tireless Hercules could hack them off. We hypothesise that the practice is kept alive by well meaning decision makers who want the idea that mortality reflects quality to be true. Partly this is because stopping people from dying evokes a clear sense of moral purpose, and partly because mortality is such a convenient parameter to collect—for example, it does not require questionnaires for which psychometric properties must be tested, or expensive review of case notes.

If not mortality, then what?

We are not arguing that healthcare providers are exempt from accountability or that patients should not be protected. On the contrary the search for robust measurements should not be impeded by fixing prematurely on a parameter that offers false hope. Two broad types of meas-

urement should be considered: outcomes (other than hospital mortality) and clinical processes.

A few outcome measures appear to be sensitive to quality. Firstly, mortality rates associated with high risk procedures that are heavily dependent on technical skill, such as intrauterine transfusion and heart surgery, seem to have strong signal to noise ratios.³⁰ Use of outlier status to identify poor practice may be credible in these circumstances but the signal will become progressively less informative (random error will make up a larger part of the variance) as poor practice is weeded out.³⁰ Secondly, a limited number of non-mortality outcomes are heavily influenced by the quality of care—rates of hospital acquired bloodstream infection, for example.

However, the responsible use of outcomes in performance management is limited to the few circumstances in which they are valid proxies for quality. We strongly advocate direct measurement of quality by observing clinical process (error rates). That is to say, adherence to the tenets of evidence based safe care should be audited as recommended by the House of Commons Select Committee.³¹ Such a policy has two significant advantages over hospital mortality rates, as the basis for quality improvement.

(1) The action that should follow is inherent in the criterion itself. For example, if audit of clinical processes shows that anticoagulants are not being given before hip replacement surgery, then it is clear where the hospital should direct its attention. The stigmatising effect is minimal, because the organisation knows what to do to improve.^{9 10}

(2) Since the contingent action is usually clear cut, process audit can provoke improvements wherever there is headroom for better performance, including where the hospital is an average or above average performer. For example, repeated audit and feedback has resulted in improved care for patients with acute heart attack, such that some sort of treatment to unblock the coronary artery is now used in nearly 100% of cases in England.³² Process audit can therefore shift distributions of performance across many criteria and across many organisations. Shifting the mean in this way yields more health gain than can be realised simply by truncating the tail of a distribution.³

Where management and statistical theories collide

Process monitoring is not a panacea. Process is generally more expensive to monitor than

outcome, since numerator and denominator data have to be harvested, often from case notes (this process should become easier as electronic records become more widespread). The continuous evolution of evidence requires measures to be constantly updated, and process audit is subject to many types of measurement error, discussed in detail elsewhere.³³ Comparative measurement of clinical process will be biased if observers are party to information that may prejudice them, if different observers review different organisations, or if organisations are reviewed in series where learning and fatigue effects may manifest. In short, scientific principles must be followed to avoid bias.

Bias due to differences in measurement in different institutions is of little or no consequence when data are used in bottom-up, staff driven, quality improvement exercises, as in the Veterans' Administration QUERI system.³⁴ However this sort of bias is crucially important when the results of audit are used as the basis for reward (such as a financial payment) or sanction (such as an external inquiry).

We incline towards a bottom-up agenda for quality improvement and would advocate performance management at one remove, by ensuring that clinical teams have systems in place to monitor quality rather than collecting large amounts of poorly calibrated information centrally. However, the issue of how best to improve quality for a given unit of resource and how best to be accountable to the public remains open. We do not pretend to have all the answers; the science needs to mature, not only to improve the measurement of quality, but also to learn how to use the (inevitably imperfect) measurements so that they do more good than harm. Deeper understanding will depend not on statistical or organisational studies carried out in isolation, but on synthesis of both subjects, in the tradition of Shewhart and Deming over half a century ago.¹⁶ While the optimal solution remains elusive, we have argued that certain suboptimal solutions can be identified; performance management of medical care by hospital mortality belongs in this latter category.

Richard Lilford professor of clinical epidemiology, Public Health, Epidemiology and Biostatistics, University of Birmingham, Edgbaston, Birmingham B15 2TT

Peter Pronovost anaesthesiologist and critical care physician, Department of Anesthesiology and Critical Care Medicine, Quality and Safety Research Group, Johns Hopkins University School of Medicine, 1909 Thames St, Baltimore, MD 21231, USA

Correspondence to: R Lilford r.j.lilford@bham.ac.uk

Accepted: 2 April 2010

We thank Alan Girling (University of Birmingham), Cor Kalkman (UMC Utrecht), and John Wright (Bradford Teaching Hospitals NHS Foundation Trust) for their comments, and Peter Chilton (University of Birmingham) for help in preparation of the manuscript.

Contributors and sources: RL produced the first draft of the paper, with subsequent iterations between both RL and PP. RL is guarantor. This article was written as a result of the authors' longstanding interest in safety and quality of care and the provocation arising from the publication of the Mid Staffordshire report in the week commencing 24 February 2010, and the subsequent media coverage—for example, the BBC Panorama programme of 8 March 2010.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that: (1) RL had financial support for methodological work from the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) for Birmingham and Black Country, and the Multidisciplinary Assessment of Technology Centre for Healthcare (MATCH) project, although views expressed are his own. PP had no financial support for the submitted work from anyone other than his employer; (2) No financial relationships with commercial entities that might have an interest in the submitted work; (3) No spouses, partners, or children with relationships with commercial entities that might have an interest in the submitted work; (4) PP receives speaking honorariums from not-for-profit hospitals for speaking on safety and quality measurement. RL has no non-financial interests that may be relevant to the submitted work. No other conflicts of interest declared by either author.

Provenance and peer review: Not commissioned, externally peer reviewed.

- Francis R. The Mid Staffordshire NHS Foundation Trust Inquiry. Independent Inquiry into care provided by Mid Staffordshire NHS Foundation Trust. January 2005-March 2009. Volume 1. HC375-1. Stationery Office, 2010.
- Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;335:648-50.
- Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-54.
- Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 2007;61:1010-3.
- Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens AJ, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009;338:b780.
- Dr Foster. The Dr Foster hospital guide 2009: how safe is your hospital? Dr Fosters, 2009
- Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;311:793-6.
- Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 1991;324:370-6.
- Jha AK, Li Z, Orav EJ, Epstein AM. Care in U.S. hospitals—the Hospital Quality Alliance program. *N Engl J Med* 2005;353:265-74.
- Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Braunholtz D, et al. The Leeds University Maternity Audit Project. *Int J Qual Health Care* 2002;14:175-81.
- Redfern M, Keeling JW, Powell E. The Royal Liverpool Children's Inquiry Report. Stationery Office, 2001.
- Aylin P, Alves B, Best N, Cook A, Elliott P, Evans SJ, et al. Comparison of UK paediatric cardiac surgical performance by analysis of routinely collected data 1984-96: was Bristol an outlier? *Lancet* 2001;358:181-7.
- Bodganich W, Ruiz RR. Missouri Hospital Reports Errors in Radiation Doses. *New York Times* 2010 Feb 25; Sect A:17.
- Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 2007;20:91.
- Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care* 1996;34:737-53.
- Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;357:463-7.
- Kmietowicz Z. Relatives demand answers on care at Mid Staffordshire. *BMJ* 2010;340:c656.
- Goffman E. Stigma: notes on the management of a spoiled identity. Prentice-Hall, 1986.
- Simborg DW. DRG creep: a new hospital-acquired disease. *N Engl J Med* 1981;304:1602-4.
- Pronovost PJ, Colantoune E. Measuring preventable harm: helping science keep pace with policy. *JAMA* 2009;201:1273-5.
- Uva MS, Galletti L, Gayet FL, Plot D, Serraf A, Bruniaux J, et al. Surgery for congenital mitral valve disease in the first year of life. *J Thorac Cardiovasc Surg* 1995;109:164-74.
- Commission for Healthcare Audit and Inspection. Investigation into Mid Staffordshire NHS Foundation Trust. Healthcare Commission, 2009.
- Alberti G. Mid Staffordshire NHS Foundation Trust: a review of the procedures for emergency admissions and treatment, and progress against the recommendation of the March Healthcare Commission report. Department of Health, 2009.
- Colin-Thome D. Mid Staffordshire NHS Foundation Trust: a review of lessons learnt for commissioners and performance managers following the Healthcare Commission investigation. Department of Health, 2009.
- Fischhoff B. Hindsight not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. *J Exp Psychol Human* 1975;1:288-99.
- Care Quality Commission. Inpatient survey 2008. Mid Staffordshire Foundation Trust. Picker Institute Europe, 2008.
- Hilgartner S, Bosk CL. The rise and fall of social problems: a public arenas model. *Am J Sociol* 1988;94:53-78.
- Ben-Yehuda N. The European witch craze of the 14th to 17th centuries: a sociologist's perspective. *Am J Sociol* 1980;86:1-31.
- Pronovost PJ, Goeschel CA, Colantuoni E, Watson S, Lubomski LH, Berenholtz SM, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. *BMJ* 2010;340:c309.
- Ettinger WH, Hylka SM, Phillips RA, Harrison Jr LH, Cyr JA, Sussman AJ. When things go wrong: the impact of being a statistical outlier in publicly reported coronary artery bypass graft surgery mortality data. *Am J Med Qual* 2008;23:90-5.
- House of Commons Health Committee. Patient Safety. Sixth report of session 2008-09. Volume I. HC-151-1. Stationery Office, 2009.
- MINAP Steering Group. How the NHS manages heart attacks. Eighth public report 2009. National Institute for Clinical Outcomes Research, 2009.
- Brown C, Hofer T, Jhal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Qual Saf Health Care* 2008;17:170-7.
- McQueen L, Mittman BS, Demakis JG. Overview of the Veterans Health Administration (VHA) Quality Enhancement Research Initiative (QUERI). *J Am Med Inform Assoc* 2004;11:339-43.

Cite this as: *BMJ* 2010;340:c2016

See EDITORIAL, p 933 and FEATURE, p 950