



OPEN ACCESS

Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study

Alan J Girling,¹ Timothy P Hofer,² Jianhua Wu,³ Peter J Chilton,¹ Jonathan P Nicholl,⁴ Mohammed A Mohammed,¹ Richard J Lilford¹

► An additional supplementary appendix is published online only. To view this files please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2012-001202>).

¹Department of Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK

²Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan, USA

³Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

⁴Medical Care Research Unit, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

Correspondence to

Prof Richard J Lilford, Department of Public Health, Epidemiology and Biostatistics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; r.j.lilford@bham.ac.uk

Accepted 17 September 2012
Published Online First
16 October 2012

ABSTRACT

Risk-adjustment schemes are used to monitor hospital performance, on the assumption that excess mortality not explained by case mix is largely attributable to suboptimal care. We have developed a model to estimate the proportion of the variation in standardised mortality ratios (SMRs) that can be accounted for by variation in preventable mortality. The model was populated with values from the literature to estimate a predictive value of the SMR in this context—specifically the proportion of those hospitals with SMRs among the highest 2.5% that fall among the worst 2.5% for preventable mortality. The extent to which SMRs reflect preventable mortality rates is highly sensitive to the proportion of deaths that are preventable. If 6% of hospital deaths are preventable (as suggested by the literature), the predictive value of the SMR can be no greater than 9%. This value could rise to 30%, if 15% of deaths are preventable. The model offers a ‘reality check’ for case mix adjustment schemes designed to isolate the preventable component of any outcome rate.

INTRODUCTION

Hospital mortality rates are widely used as a measure of quality in developed countries. The Health Care Financing Administration (HCFA) released risk-adjusted mortality data, in the form of standardised mortality ratios (SMRs), on all Medicare patients admitted to hospitals in the USA in 1986.¹ After years of criticism,^{2–7} and attempts to improve the estimates,^{8–11} HCFA abandoned the effort and for a while hospital mortality data fell into disuse as a way to assess quality of care, outside of a single high-volume surgical procedure—coronary bypass surgery. However, there has been a renewal of interest and a dramatic increase in the number of North American hospitals publishing mortality data.^{12–16} In England, adjusted mortality statistics for NHS hospitals have been published

since 2001 by the Dr Foster organisation,^{16 17} and the Department of Health has recently sponsored a similar approach—the Summary Hospital Mortality Index (SHMI).^{18 19} However, there is little new evidence to support such metrics as indicators of care quality, and further evidence questioning their validity.^{20 21} A recent paper concluded that an observed lack of agreement between different methods for calculating hospital-wide mortality rates may result from ‘fundamental flaws in the hypothesised association between hospital-wide mortality and quality of care’.²² Such a fundamental flaw would arise if the signal (preventable deaths) was small in relation to the noise (inevitable deaths). To further explore this possibility, we model the correlation between overall risk-adjusted mortality and mortality due to poor care.

Since a review of the relationship between mortality and quality of care found no empirical studies that directly report the relationship between SMRs and preventable mortality,²¹ we develop the argument from a number of plausible scenarios using a mathematical model for hospital mortality rates.

DERIVATION OF THE MODEL

The model, though generic to any outcome, is explicated with respect to hospital mortality. For each hospital we assume that the rate of in-hospital mortality (*M*) can be partitioned into two components:

$$M = U + V$$

where *U* denotes the mortality rate arising from deaths that could not have been avoided even under optimal care, and *V* denotes the mortality rate arising from deaths due to suboptimal care. The SMR is defined as the ratio of the actual number of deaths to the number of deaths

expected after case mix is taken into account within a risk-adjustment scheme.

The proportion of the variance in SMRs attributable to preventable mortality—and the correlation between these quantities—depends on the contribution of preventable mortality to overall hospital mortality rates, and on the performance of the risk-adjustment scheme in eliminating variation due to differences in case mix.

The critical quantities are:

- ξ: the average proportion of deaths that are preventable (the ‘preventability index’);
- c_v: the coefficient of variation (defined as SD÷mean) of the preventable mortality rate across hospitals;
- c_M: the coefficient of variation of the total in-hospital mortality rate;
- R²: the proportion of the variance in total mortality rates explained by the risk-adjustment process;
- Q: the correlation coefficient between the hospital SMR and the preventable mortality rate.

The performance of the SMR as a proxy for preventable mortality is governed by the numerical value of Q, and Q² can be interpreted as the proportion of the variation in SMRs attributable to preventable mortality. In the online appendix an upper bound for Q is derived under two assumptions. The first of these deals with the possibility that a high rate of natural (unavoidable) death (U) in a hospital might go hand in hand with a high rate of preventable death (V). In practice, the presence of such positive correlation is entirely plausible since patients at high intrinsic mortality risk are also those for whom medical error is likely to have the most catastrophic consequences. As stated, the assumption (A1) implies that all such correlation between U and V can be accounted for in terms of case mix factors that reflect that intrinsic risk. The second assumption is concerned with the variation in mortality rates among hospitals with identical case mixes. The simplest version of the assumption (A2) says that such variation, as measured by the statistical variance, is the same whatever the case mix happens to be. As demonstrated in the online supplementary appendix, this leads unequivocally to the bound on Q used throughout this paper.

It follows that Q² will not exceed:

$$\left(\frac{c_v}{c_M}\right)^2 \frac{\xi^2}{1 - R^2} \tag{1}$$

This result enables us to explore the conditions under which SMRs may provide a useful indication of preventable mortality.

Assumption A2 may be a sensible first approximation, but it is open to question as an exact description of reality. Put simply, there is more scope for variability in rates at case mixes when the mean rate is high than at case mixes when it is low. For this reason an alternative assumption (A2') has been entertained, which posits a

proportional relationship between the variance and the square of the case mix-specific mean. (This could arise if case mix differences impact on the relative mortality risk between hospitals.) Under A2' there may be a modest inflation in the bound for Q, leading to an estimated increase of up to 5% (or 10% for Q²) in the base case described below (see online supplementary appendix). Such increases are not large enough to disturb the general conclusions of the paper.

POPULATING THE MODEL

The coefficient of variation of the overall mortality rate (c_M) for 143 Acute Hospital Trusts in England in 2007/8 was 0.19;²³ here a base case value for c_M of 0.2 has been assumed.

In the SHMI scheme proposed for the NHS, the proportion of the variance explained by risk adjustment has been estimated as 81%;¹⁸ a value of R²=0.8 has therefore been assumed.

There appears to be no published study describing the variation in preventable death rates across hospitals. However, the variance of the between-hospital component of preventable adverse events is given as 0.15 by Zegers *et al.*²⁴ We used this figure, together with information in their paper, to compute an approximate between-hospital SD of 0.42 for the logarithm of the rate of such events. This roughly corresponds to a coefficient of variation on the natural scale and informs our base case choice for c_v of 0.4 which, as it happens, is exactly twice the base case for c_M.

Studies of hospital deaths describe the proportion of deaths that may have been caused by clinical error (often at low probability),^{25–27} rather than the proportion of deaths that were preventable. An exception is the direct estimate (ξ=6%; 95% CI 3.4% to 8.6%) given by Hayward and Hofer and used here.²⁸

SMRS AS A PROXY FOR RATES OF PREVENTABLE MORTALITY

The expression (1) imposes a severe constraint on the correlation between hospital SMRs and preventable mortality rates when the parameter values described above are used. For example, if 6% of deaths are preventable (as estimated by Hayward and Hofer), and base case assumptions are made (c_M=0.2, c_v=0.4, R²=0.8), it follows that Q² cannot exceed:

$$\left(\frac{0.4}{0.2}\right)^2 \frac{0.06^2}{1 - 0.8} = 0.072$$

(or about 0.079 if the alternative assumption A2' is preferred). Hence it seems that preventable mortality can account for no more than 8% of the variation in SMRs.

This leaves very little scope for risk-adjusted mortality to function as an effective proxy for quality of care.

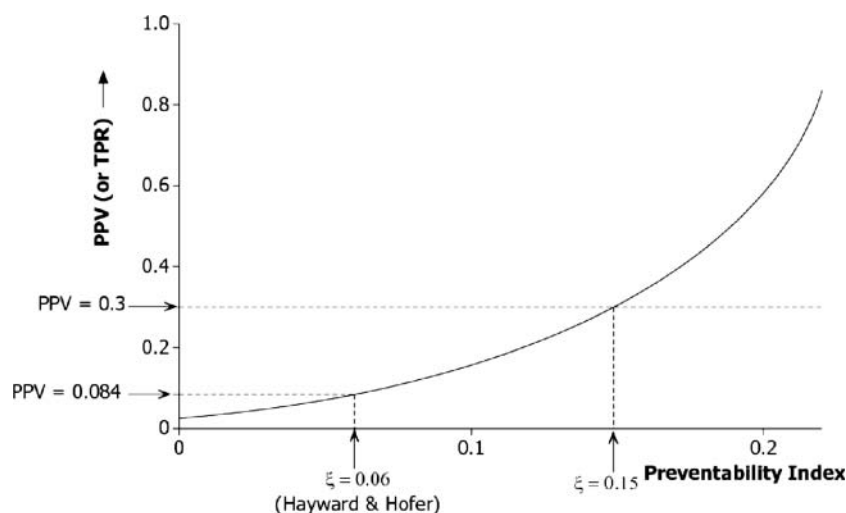
The point is reinforced if the SMR is treated as a formal diagnostic test for high rates of preventable mortality. Suppose that a warning is triggered if the SMR for a hospital places it among the worst 2.5% of all hospitals. This criterion corresponds to a ‘2-sigma’ action limit for the SMR. The diagnostic performance of this test depends on the value of Q^2 . The positive predictive value (PPV) of such a warning for identifying a hospital with high preventable mortality will be very low indeed if $Q^2 < 0.08$, as suggested above. For instance, the PPV for identifying a hospital with a *preventable* mortality rate among the worst 2.5% of hospitals would be no more than 0.09 (ie, 9%). The same applies to the true positive rate (TPR) of the action limit for correctly detecting high preventable mortality. In fact the TPR equals the PPV here because the same fraction (2.5%) has been used to define the action limit and to specify the notion of high preventable mortality. Based on these numbers, at least 10 warnings out of 11 would be false alarms, and at least 10 out of 11 poorly performing hospitals would escape attention.

At higher levels of the preventability index (ξ) the bound on Q^2 in expression (1) becomes less stringent, and effective monitoring using the SMR will be correspondingly more feasible. The effect is illustrated in figure 1, which shows how an increase in ξ translates into potential for an improved PPV (and TPR). Nevertheless, it appears that worthwhile PPV (or TPR) values can be attained only at values of ξ well in excess of what is supported by empirical studies. For example, a PPV of 0.3 would require that more than 15% of deaths are preventable ($\xi > 0.15$).

DEPARTING FROM THE BASE CASE: THE REALISTIC SCOPE FOR SMRS

The argument so far has relied on a base case value for c_V (=0.4) grounded in a study of preventable adverse events.

Figure 1 Diagnostic performance of the standardised mortality ratio (2-sigma upper limit) to detect a hospital among the worst 2.5% for preventable deaths. The curve shows the dependency of the upper bound for positive predictive value (PPV) (or true positive rate (TPR)) on the preventability index under a risk-adjustment scheme accounting for 80% of the variation between hospitals. The base case relationship $c_V = 2c_M$ is assumed.



Under this condition, preventable mortality rates vary four-fold between the 5th and 95th centiles (figure 2). A lower value of c_V will tighten the constraint (1) on the correlation Q and thus reduce the scope for the SMR to diagnose poor care. This would be the case if, for example, the coefficient of variation was the same for preventable as for overall mortality. For the SMR to function as an effective proxy for preventable mortality rates, c_V would need to be higher—indeed, much higher—than the base case value. For example, when $\xi = 0.06$, as in Hayward and Hofer, a PPV of 0.3 could be achieved only if c_V were increased from 0.4 to 1.0. This would mean that in a group of 20 randomly chosen hospitals the preventable death rate is, on average, more than 15 times higher in the worst hospital compared with the best one (figure 2). It is doubtful whether such wide discrepancies among similar institutions in the same healthcare system are plausible.

Satisfactory diagnostic performance could perhaps be achieved if both ξ and c_V were to exceed their base case values, but by smaller amounts. For example, case (b) in figure 2 envisages a value of ξ (=8.6%) at the upper confidence limit in Hayward and Hofer.²⁸ Then c_V need be no more than 0.7 (case (b)) to achieve the same effect on the upper bound for Q^2 as in case (c). Here preventable mortality rates still would need to differ by an arguably implausible factor of 8.0 between the 95th and 5th centiles.

DISCUSSION

The results indicate that worthwhile correlations between case-mix adjusted SMRs and rates of preventable mortality are not attainable unless rates of preventable mortality are either (a) much higher than current estimates suggest, or (b) implausibly variable between different hospitals. It can be argued that specificity is not crucial to the performance of SMRs, since they are used as a screen for hospitals requiring further

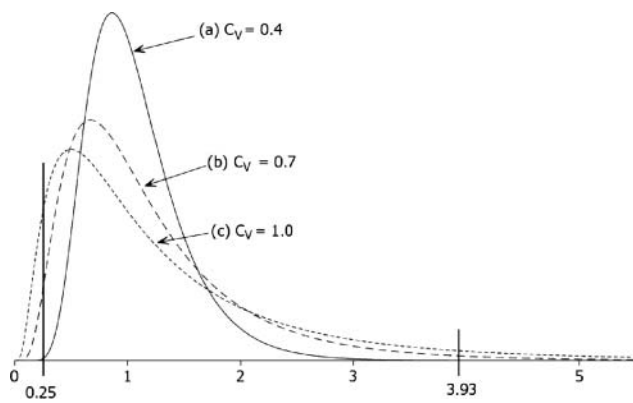


Figure 2 Three candidate distributions to describe variation among hospitals in rates of preventable mortality. The distributions are scaled to unit median and a log-normal model is assumed. Under the base case ($c_v=0.4$) the hospital in the 95th centile would have about four times the preventable mortality rate of the hospital at the 5th centile. Under the most dispersed distribution ($c_v=1.0$) the ratio between the 5th and 95th centiles (ie, 0.25 and 3.93) is more than 15, an implausibly large range across a random sample of 20 hospitals.

investigation, not as a diagnostic trigger for sanction. However, there is always a trade-off between sensitivity and specificity even for screening tests. In the case of SMRs, high false positives waste resources, stigmatise hospitals, and lead to gaming;^{29 30} while false negatives provide false reassurance, thereby deflecting attention away from quality issues across the whole healthcare system.³¹ Until risk adjustment can explain much larger proportions of the variance in mortality it is unsafe to use high SMRs to identify poor quality of care in any particular hospital, as has been done recently for some English hospitals.^{29 30}

Aside from quality of care, possible sources of variation in hospital SMRs include: differences in discharge policies leading to variations in underlying mortality rates; differences in recording practices for primary diagnoses or comorbidities;²⁰ sampling fluctuations in observed mortality rates; and failures of the risk adjustment scheme to adequately account for patient differences. However, a full discussion is beyond the scope of this paper.

One message from our study is that the diagnostic value of institution-level outcome data is critically dependent on the preventability index (ξ) as demonstrated in figure 1. In the base case ($\xi=6\%$), PPVs are very low, although they start to climb steeply once ξ exceeds about 15%. Most of the relevant literature provides an estimate of the proportion of deaths for which a potentially preventable factor was present, rather than an estimate of the proportion of deaths that were in fact preventable. For example, Zegers *et al*³² found an error that may have been causal in 4% of deaths, while Hogan

*et al*³³ found such a factor that was causative on the balance of probabilities in 5.2% of deaths. Only Hayward and Hofer have attempted to measure the critical quality—the proportion of deaths that were preventable.²⁸ More research into this issue is urgently needed. The model demonstrates that the predictive value of SMRs is also highly sensitive to the extent to which preventability varies across hospitals. An indirect estimate suggests a coefficient of variation in the rate of preventable deaths in the region of 40% ($c_v=0.4$), consistent with a fourfold variation in rates across a representative sample of hospitals. However, there appear to be no direct estimates of this quantity in the literature. It may even be that our estimate is too high for a mortality measure in which eclectic performance across individuals and departments is aggregated at the hospital level. In such circumstances variation in the performance of individual clinical units may be diluted when these are combined together.

Other things being equal, standardised outcome rates will discriminate well for specific conditions in which preventability rates are high (eg, pressure ulcers, maternal deaths, deaths following elective surgery).^{23 34 35} Moreover, the variance-dilution effect will be absent when performance is measured at the level of the individual clinical unit. For these reasons standardised mortality rates are likely to be more reliable as a reflection of preventable mortality rates when the unit of comparison is relatively discrete (eg, clinical specialities) than when it is a consolidated measure across separate entities (eg, hospital-wide outcome rate).

When the preventability index is low, as it is for hospital mortality and many other outcomes in healthcare, it may be necessary to fall back on direct measurement of process and outcome by examining individual cases in detail—for example, by case-note review. Currently, this method is expensive, though it may become easier as sophisticated electronic records become widespread. It is also subject to classification errors (eg, judgment of the preventability of deaths varies by case reviewers). Yet it may remain the only viable option for measuring preventable mortality rates unless there are further improvements in risk-adjustment technology.

What this study adds

- ▶ Hospital standardised mortality ratios (SMRs) are markers of poor care only to the extent that they correlate with preventable mortality rates.
- ▶ A mathematical model populated by empirical estimates for critical parameters suggests that this correlation is low.
- ▶ If preventable deaths make up less than 15% of all deaths, then SMRs are poor diagnostic tests for suboptimal care.

Acknowledgements We thank Michael Langman, FRCP, FFPM, FMedSci, Yen-fu Chen, PhD and Semira Manaseki-Holland, PhD (University of Birmingham) for helpful comments.

Contributors RJL conceived the idea for the paper and drafted the initial core manuscript; AJG derived the mathematical argument (with input from JW) and drafted the Methods and Results sections; AJG, TPH, JW, PJC, JPN and MAM contributed text and critically reviewed and commented on the document. RJL is the guarantor.

Funding AJG, JW, PJC, RJL acknowledge financial support for the submitted work from the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) for Birmingham and Black Country; the West Midlands Quality Institute; and the EPSRC Multidisciplinary Assessment of Technology Centre for Healthcare (MATCH) programme (EPSRC grant GR/S29874/01).

Competing interests All authors have completed the unified competing interest form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years, no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review The corresponding author (RJL) has been involved in statistical aspects of SMRs and the conduct of individual case note review. This prompted the idea that a mathematical model could be constructed to link preventable deaths revealed through detailed scrutiny of individual cases and overall death rates used to compare hospitals statistically. He discussed this idea with his more algebraically accomplished coauthors, and together the argument was developed.

REFERENCES

- Department of Health and Human Services. *Medicare Hospital Mortality Information 1986*. Washington, DC: Government Printing Office, 1987.
- Blumberg MS. Biased estimates of expected acute myocardial infarction mortality using MedisGroups admission severity groups. *JAMA* 1991;265:2965–70.
- Greenfield S, Aronow HU, Elashoff RM, *et al*. Flaws in mortality data. The hazards of ignoring comorbid disease. *JAMA* 1988;260:2253–5.
- Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care* 1996;34:737–53.
- Park RE, Brook RH, Koseoff J, *et al*. Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. *JAMA* 1990;264:484–90.
- Rosen HM, Green BA. The HCFA excess mortality lists: a methodological critique. *Hosp Health Serv Adm* 1987;32:119–27.
- Sanazaro PJ, Mills DH. A critique of the use of generic screening in quality assessment. *JAMA* 1991;265:1977–81.
- Hannan EL, Arani DT, Johnson LW, *et al*. Percutaneous transluminal coronary angioplasty in New York State. Risk factors and outcomes. *JAMA* 1992;268:3092–7.
- Hannan EL, Kilburn H Jr, Lindsey ML, *et al*. Clinical versus administrative data bases for CABG surgery. Does it matter? *Med Care* 1992;30:892–907.
- Hartz AJ, Kuhn EM, Kayser KL, *et al*. Assessing providers of coronary revascularization: a method for peer review organizations. *Am J Public Health* 1992;82:1631–40.
- lezzoni LI, Ash AS, Coffman GA, *et al*. Predicting in-hospital mortality. A comparison of severity measurement approaches. *Med Care* 1992;30:347–59.
- 111th United States Congress. Patient Protection and Affordable Care Act. H.R. 3590. 2009.
- Cassel JB, Jones AB, Meier DE, *et al*. Hospital mortality rates: how is palliative care taken into account? *J Pain Symptom Manag* 2010;40:914–25.
- Department of Health. *Health and Social Care Bill 2010–11 (55/1)*. Norwich, UK: The Stationery Office, 2011.
- Department of Health. *Equity and excellence: liberating the NHS. (CM7881)*. Norwich, UK: The Stationery Office Limited, 2010.
- Dr Foster Limited. *Dr Foster Hospital Guide 2010. What Makes a Good Hospital?* <http://www.drfoosterhealth.co.uk/docs/hospital-guide-2010.pdf> (accessed 28 May 2012).
- Jarman B, Aylin P, Bottle A. Hospital mortality ratios. A plea for reason. *BMJ* 2010;340:c2744.
- Campbell M, Jacques RM, Fotheringham J, *et al*. Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. *BMJ* 2012;344:e1001.
- National Quality Board. Report from the Steering Group for the National Review of the Hospital Standardised Mortality Ratio (Gateway Reference: 15066). http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_121328.pdf (accessed 29 May 2012).
- Mohammed MA, Deeks JJ, Girling A, *et al*. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009;338:b780.
- 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;7:91.
- Shahian DM, Wolf RE, lezzoni LI, *et al*. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med* 2010;363:2530–9.
- The Health and Social Care Information Centre. HESonline: freely available data. <http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=889> (accessed 28 May 2012).
- Zegers M, de Bruijne MC, Spreeuwenberg P, *et al*. Variation in the rates of adverse events between hospitals and hospital departments. *Int J Qual Health Care* 2011;23:126–33.
- Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med* 1988;109:582–9.
- Hayward RA, Heisler M, Adams J, *et al*. Overestimating outcome rates: statistical estimation when reliability is suboptimal. *Health Serv Res* 2007;42:1718–38.
- Rubenstein LV, Kahn KL, Reinisch EJ, *et al*. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. *JAMA* 1990;264:1974–9.
- Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA* 2001;286:415–20.
- 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. London, UK: The Stationery Office, 2010.
- Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010;340:c2016.
- Lilford R, Mohammed MA, Spiegelhalter D, *et al*. Use and misuses of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147–54.
- Zegers M, de Bruijne MC, Wagner C, *et al*. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care* 2009;18:297–302.
- Hogan H, Healey F, Neale G, *et al*. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf* Published online first 7 July 2012. doi:10.1136/bmjqs-2012-001159
- Pronovost P, Needham D, Berenholtz S, *et al*. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;355:2725–32.
- Bergstrom N, Braden B, Boynton P, *et al*. Using a research-based assessment scale in clinical practice. *Nurs Clin North Am* 1995;30:539–51.