

The findings of the Mid-Staffordshire Inquiry do not uphold the use of hospital standardized mortality ratios as a screening test for ‘bad’ hospitals

M.A. MOHAMMED¹, R. LILFORD¹, G. RUDGE¹, R. HOLDER¹ and A. STEVENS²

From the ¹Primary Care Clinical Sciences, University of Birmingham, Edgbaston and ²Public Health, Epidemiology and Biostatistics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Address correspondence to Dr M.A. Mohammed, Primary Care Clinical Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. email: m.a.mohammed@bham.ac.uk

Received 21 February 2013 and in revised form 28 March 2013

Summary

The Mid-Staffordshire Public Inquiry has published its findings. The initial investigations were triggered by an elevated hospital standardized mortality ratio (HSMR). This shows that the HSMR is being used as a screening test for substandard care; whereby hospitals that fail the test are scrutinized, whilst those that pass the test are not. But screening tests are often misunderstood and misused and so it is prudent to critically examine the HSMR before casting it in the role of a screening test for ‘bad’ hospitals. A screening test should be valid, have adequate performance characteristics and a clear post-test action plan. The HSMR is not a valid screening test (because the empirical relationship between clinically avoidable mortality and the HSMR is unknown). The HSMR has a poor performance profile (10 of 11 elevated HSMRs would be

false alarms and 10 of 11 poorly performing hospitals would escape attention). Crucially, the aim of a post-test investigation into an elevated HSMR is unclear. The use of the HSMR as a screening test for clinically avoidable mortality and thereby substandard care, although well intentioned, is seriously flawed. The findings of the Mid-Staffordshire Public Inquiry have no bearing on this conclusion because a ‘bad’ hospital cannot uphold a bad screening test. Nonetheless, HSMRs continue to pose a grave public challenge to hospitals, whilst the unsatisfactory nature of the HSMR remains a largely unacknowledged and unchallenged private affair. This asymmetric relationship is inappropriate, unhelpful, costly and potentially harmful. The use of process measures remains a valid way to measure quality of care.

Introduction

The Mid-Staffordshire Public Inquiry has recently published its findings,¹ which together with an earlier independent Inquiry,² report on the quality of care at Mid-Staffordshire NHS Foundation Trust (MSFT) and the role of those responsible for performance monitoring of the Trust. According to the first Inquiry² (p. 23), the investigations were triggered by an elevated hospital standardized mortality ratio

(HSMR), implying that a less elevated HSMR might have escaped attention—a sobering thought. This shows that the despite methodological concerns about the construction of the HSMR,^{3–8} it is being used as a population screening test for substandard care,^{4,9} where hospitals that fail the test are scrutinized whilst those that pass the test are paid less attention and/or are quietly exonerated. Nonetheless, as Grimes and Schulz warn us, ‘Screening tests are ubiquitous, double edged

swords that are used unwittingly by the well intentioned...screening remains widely misunderstood and misused'.¹⁰ So, despite the findings of the Inquiry, it is necessary to critically examine the HSMR before casting it in the role of a screening test for 'bad' hospitals.

A useful screening test provides, amongst other things, information that addresses a core question (i.e. is a valid test); has adequate performance characteristics and has a clear post-test action plan.¹⁰ After providing a historical perspective to the HSMR, we examine (i) the extent to which the HSMR is a valid test (i.e. whether it addresses the core question of clinically avoidable deaths); (ii) the performance characteristics of the HSMR as a screening test for 'bad' hospitals and (iii) the clarity of subsequent actions following an elevated HSMR.

Historical perspective

The present interest in comparing outcomes across healthcare providers can be traced back to much earlier times with notable names, such as Ignaz Semmelweis (ca. 1847), Florence Nightingale (ca. 1857), James Simpson (ca. 1869) and Ernest Codman (ca. 1914), who amongst others used mortality comparisons (crude and adjusted) as a lever to deliver improvements in the quality of care.^{11–13} The modern era of monitoring outcomes in healthcare can, arguably, be traced back to the first Health Care Finance Administration report in 1987¹⁴ with hospital-specific standardized mortality ratios comparing observed and expected deaths for Medicare patients in the USA. Hospitals with higher than expected mortality were flagged up as hospitals with potential quality of care problems (i.e. a screening test). In England, the Dr Foster HSMRs were first published in 2001, based on methodology developed by Jarman *et al.*¹⁵ and have been adopted by other countries such as Canada,¹⁶ the Netherlands,¹⁷ Sweden¹⁸ and Japan.¹⁹ The National Health Service (NHS) of England has recently published its own HSMR known as the Summary Hospital-level Mortality Indicator.²⁰ Our use of the term HSMR is not restricted to one particular supplier, but rather to the range of methods employed to derive an HSMR-type metric, especially in the NHS.

The HSMR is the ratio of observed to expected deaths multiplied by 100 for each hospital. The expected deaths are estimated from a statistical model. If the SMR (within the limits of statistical inference) for a given hospital is <100, then that hospital has a favourable mortality experience; a ratio of >100 is unfavourable; and a ratio of 100 is neutral.

The difference between observed and expected deaths is known as excess mortality.

The premise underlying the use of the HSMR is that the mortality seen at a given hospital, after adjustment for patient risk factors, reflects the quality of care at that hospital through higher/lower than statistically predicted deaths. This premise needs substantial caveats, e.g. assuming similar admission and discharge policies; consistently good quality data; similar provision of palliative care in the community; reliable case-mix adjustment; and the absence of unknown confounding factors, which have been discussed elsewhere (under the term 'case-mix adjustment fallacy').²¹ Nonetheless, whilst these points are sufficient to caution against any naive interpretation of the HSMR, the use of the HSMR as a screening test for 'bad' hospitals is discussed below. A 'bad' hospital is one that experiences an unusually high number of clinically avoidable deaths.

Validity of the HSMR as a screening test

The fundamental reason for focusing on hospital mortality is to identify and ultimately reduce the number of clinically avoidable deaths. The definition of a clinically avoidable death is one where experts conclude that substandard care contributed to death and, that absent those failures, the patient might have survived. Indeed expert review is our imperfect master diagnostic test (although this Inquiry along with the Shipman Inquiry,²² has demonstrated a possible role for testimonial evidence from patients/relatives—something which is not without its own methodological challenges)²³ to determine if a death was clinically avoidable.

For the HSMR to be a valid screening test, the empirical relationship between excess mortality (observed–expected) and clinically avoidable mortality needs to be determined. In other words, the correlation between the HSMR and clinically avoidable deaths is needed. Since such studies have not yet been undertaken the claim that the HSMR is a screening test for clinically avoidable mortality and poor quality of care is hollow. Nonetheless, it is all too easy to misinterpret excess mortality (observed–expected deaths) as being synonymous with clinically avoidable mortality. Excess mortality does not equate to clinically avoidable mortality. Why? Consider a hospital with an elevated HSMR of 125, based on 1000 observed deaths and 800 expected deaths. To determine which of the observed deaths were clinically avoidable we would have to subject all deaths (i.e. 1000 cases) to expert review and only then could we determine

the number of clinically avoidable deaths. The excess mortality ($1000 - 800 = 200$) associated with the elevated HSMR does not tell us how many deaths were clinically avoidable. Therefore, regardless of whether two hospitals have similar or quite different HSMRs we are unable to infer the number of clinically avoidable deaths without undertaking a review of all the deaths in each hospital. Unfortunately, despite this incisive distinction between clinically avoidable deaths and excess mortality, the confusion between the HSMR (or excess mortality) and clinically avoidable deaths is pervasive.

To illustrate the muddle, contrast the correct position from the Inquiry with misleading media reports. The first Inquiry² ‘...concluded that it would be unsafe to infer from the figures that there was any particular number or range of numbers of avoidable or unnecessary deaths at the Trust’ (p. 23). Some media outlets, however, used the upper confidence limit of the HSMR to construct their headlines. For example, the Guardian claimed that ‘It was the biggest scandal of NHS care in years. Several hundred, possibly as many as 1200, patients died at Stafford Hospital between 2005 and 2009 after suffering neglect, indignity and shoddy care’.²⁴ Recently, the Telegraph reported that the Secretary of State for Health, Jeremy Hunt, warned that ‘Medical failures that saw up to 1200 patients die needlessly at Stafford hospital are occurring across the NHS’.²⁵ Even an editorial in the British Medical Journal perpetuated the misconception when inviting readers to spot the odd-one-out: ‘Mid Staffordshire is particularly curious if you compare the death tolls associated with the scandals: Alder Hey (0), Bristol (30 to 35), Shipman (probably 250), Mid Staffordshire (400 to 1200)’.²⁶ The comparison to Mid-Staffordshire hospital is invalid because the other death tolls were determined by expert review of cases, whilst Mid-Staffordshire’s avoidable mortality remains unknown.

Just as a high HSMR is misinterpreted (in lieu of clinically avoidable mortality), a low HSMR (i.e. lower than expected number of deaths) can also be misinterpreted to mean the number of lives saved. Here is an example from one hospital Trust that reduced its HSMR from 84 to 71. ‘The mortality improvement equated to about 255 patients. In other words, there are 255 people still walking around attending family weddings, grandchildren’s christenings and so on who would otherwise be dead if this action had not been taken’.²⁷ How could they know? By what method? What if a strategy for lowering the HSMR was to ensure that palliative care patients were less frequently admitted to

hospital and more frequently discharged to die elsewhere? This does not constitute saving lives.⁴

The situation is further exasperated by the public release of HSMR figures (usually in isolation from other performance metrics) intimating a link between quality of care and the HSMR with headlines such as ‘best performing’ and ‘worst performing’ hospitals or ‘hospitals with worryingly high death rates’ or ‘persistently high death rates’. This practice is premature (absent the results from any subsequent diagnostic test), misleading (because the correlation with poor quality care is unknown) and reckless (because the consequences are grave for all concerned). Unfortunately, such shortcomings from using the HSMR as a screening test are not mitigated by incorporating them in a basket of performance measures, because ultimately most deaths do not reflect quality problems and most quality problems do not cause death.⁴ This is underscored by the findings of a systematic review of the relationship between quality of care and case-mix adjusted mortality rates, which concluded that ‘The general notion that hospitals with higher risk-adjusted mortality have poorer quality of care is neither consistent nor reliable’.²⁸

Performance of the HSMR as a screening test

Girling *et al.*²⁹ used a mathematical model to determine the screening properties of the HSMR to estimate the proportion of the variation in HSMRs that can be accounted for by variation in clinically avoidable mortality. They found that worthwhile correlations between HSMRs and avoidable mortality are not attainable unless avoidable mortality is either (i) much higher than current published estimates using ‘gold-standard’ methodology suggest (around 6%)³⁰ or (ii) implausibly variable between different hospitals. The positive predictive value for identifying a hospital with an avoidable mortality rate (among the worst 2.5% of hospitals) would be no more than 9%. In other words, at least 10 of 11 elevated HSMRs signals would be false alarms, and more worryingly perhaps, at least 10 of 11 poorly performing hospitals would escape attention.

These poor performance characteristics underscore why it is misleading to claim any general conclusions that a high HSMR implies poor quality of care, or that a low HSMR implies good quality of care. MSFT provides a stark illustration.³¹ The Dr Foster HSMR for MSFT for the period 2005/06 was 127, when the Trust was rated by the Healthcare Commission as having ‘fair’ quality of services and ‘fair’ use of resources. The Healthcare Commission

became very concerned about patient safety in May 2008, but by then the Dr Foster HSMR for MSFT was 105 and falling.³²

Course of action after an elevated HSMR

An elevated HSMR, as opposed to a neutral or depressed HSMR, usually attracts the greatest attention and the action that usually follows is a *post hoc* investigation. Whilst the notion of a *post hoc* investigation is sound, the underlying aim of the investigation remains crucially unclear.

This issue is readily illustrated using evidence from the Shipman Inquiry,³³ which identified 12 general practitioners (GPs) as having unacceptably high death rates (after case-mix adjustment). One of the GPs was the murderer Dr Harold Shipman. The Solicitor to the Shipman Inquiry requested further investigations of the remaining 11 GPs who were subsequently investigated and exonerated because they looked after a preponderance of patients in nursing homes.^{34,35} In other words, 11 of the 12 elevated mortality signals were false alarms (similar to estimates suggested by Girling *et al.*).²⁹ The modus operandi of an elevated HSMR tends to be 'guilty until proven innocent' wherein unsubstantiated accusations made via mortality statistics are seriously investigated, yet the poor performance of the mortality statistics is, to a large extent, ignored.

Crucially, the subsequent investigations proceeded on the basis of two distinct questions. (1) Why were the adjusted death rates elevated? (2) Is the care provided by these GPs safe? The two questions were addressed using two different methods. Question (1) involved a desktop statistical methodology based on the Pyramid Model of Investigation (which places quality of care in the search for credible explanations, after data and case-mix related explanations have been ruled out); and question (2) involved expert case-note review (which prioritized quality of care and malpractice). Whilst it is easy to see that the two questions are distinct, requiring different methods and different skills, it is difficult to see which question an elevated HSMR actually asks. In the light of the distinction between clinically avoidable deaths and statistically expected deaths, there is little apparent justification for prioritising question (2) over question (1). Yet even if an elevated HSMR is credibly explained by a benign factor (e.g. clinical coding problems)³¹ a hospital with an elevated HSMR will likely need to expend considerable efforts to remove any potential stigma regarding its quality of care—a grave challenge to any hospital. Moreover, even if quality of care is

found to be substandard we still cannot use this to infer the merits of the HSMR because, as is often the case, no investigation was undertaken of hospitals with lower HSMRs. In the few instances where such controlled comparisons have been undertaken, the link between the HSMR and quality of care has been found to be unreliable.^{36,37}

Alternatives to the HSMR

The HSMR is an outcome-based approach to measuring quality of care.³⁸ However given the problems with the HSMR, it is necessary to identify alternatives. In respect of determining avoidable mortality our master, albeit imperfect and costly, method is independent retrospective case-note review by experts.³⁹ The primary reason why case-note review is superior is because it actually measures the quality of care, whilst the HSMR does not.³⁸ Nonetheless urgent work is required to determine how to implement case-note review and process based measures into routine assessments of quality of care.

Conclusion

The widespread use of HSMRs as a screening test for clinically avoidable mortality and thereby substandard care, although well intentioned, is seriously flawed. The findings of the MSFT inquiry have no bearing on this conclusion because a 'bad' hospital cannot uphold a bad screening test. Nonetheless HSMRs continue to pose a very sombre public challenge to hospitals regarding the quality of their care, whilst the unsatisfactory nature of the HSMR remains a largely unacknowledged and unchallenged private affair. This asymmetric relationship is inappropriate, unhelpful, costly and potentially harmful. Using the HSMR to identify 'good/bad' hospitals is analogous to the practice of dowsing—the search for water without scientific apparatus—it is time to abandon this screening test and search for a better one. Meanwhile, the use of process measures remains a valid way to measure quality of care.

Acknowledgements

M.A.M., A.S. and R.L. conceived the idea. R.H. and G.R. provided methodological guidance and support. All authors contributed to the final manuscript. M.A.M. is the study guarantor.

Conflict of interest: None declared.

References

1. The Mid Staffordshire NHS Foundation Trust Public Inquiry. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. 2013. <http://www.midstaffspublicinquiry.com/> (24 February 2013, date last accessed).
2. The Mid Staffordshire NHS Foundation Trust Inquiry. *Independent Inquiry into Care Provided by Mid Staffordshire NHS Foundation Trust January 2005–March 2009*, vol I. London: The Stationary Office, 2010. http://www.midstaffsinquiry.com/assets/docs/Inquiry_Report-Vol1.pdf (24 February 2013, date last accessed).
3. Normand SLT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc* 1997; **92**:803–14.
4. Shojania KG, Forster AJ. Hospital mortality: when failure is not a good measure of success. *CMAJ* 2008; **179**:153–7.
5. Mohammed MA, Deeks J, 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.
6. Shahian DM, Normand SL. Comparison of “risk-adjusted” hospital outcomes. *Circulation* 2008; **117**:1955–63.
7. van den Bosch WF, Kelder JC, Wagner C. Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission. *BMC Health Serv Res* 2011; **11**:57.
8. Mohammed MA, Manktelow B, Hofer T. Comparison of four methods for deriving hospital standardised mortality ratios from a single hierarchical logistic regression model. *Stat Methods Med Res*, doi:10.1177/0962280212465165 [Epub 6 November 2012].
9. Jarman B. Stafford scandal: hindsight is a luxury we can ill afford. *The Telegraph*, 27 January 2013. <http://www.telegraph.co.uk/health/heal-our-hospitals/9828966/Stafford-scandal-hindsight-is-a-luxury-we-can-ill-afford.html> (24 February 2013, date last accessed).
10. Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet* 2002; **359**:881–4.
11. Spiegelhalter DJ. Surgical audit: statistical lessons from Nightingale and Codman. *J Roy Stat Soc A Sta* 1999; **162**:45–58.
12. Best M, Neuhauser D. Ignaz Semmelweis and the birth of infection control. *Qual Saf Health Care* 2004; **13**:233–4.
13. Neuhauser D. Surgical experience, hospital size and severity adjusted mortality: James Y Simpson, 1869. *Qual Saf Health Care* 2005; **14**:67–8.
14. Normand SLT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc* 1997; **92**:803–14.
15. Jarman B, Gault S, Alves B, Hilder A, Dolan S, Cook A, *et al*. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; **318**:1515–20.
16. Canadian Institute for Health Information. *HSMR: A New Approach for Measuring Hospital Mortality Trends in Canada*. Ottawa, CIHI, 2007.
17. Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B, Westert G. Measuring and explaining mortality in Dutch hospitals; The Hospital Standardized Mortality Rate between 2003 and 2005. *BMC Health Serv Res* 2008; **8**:73.
18. Köster M, Jürgensen U, Spetz C-L, Rutberg H. Standardiserad sjukhusdödlighet« som kvalitetsmått i hälsooch sjukvård. *Läkartidningen* 2008; **105**:1391–6.
19. Miyata H, Hashimoto H, Horiguchi H, Matsuda S, Motomura N, Takamoto S. Performance of in-hospital mortality prediction models for acute hospitalization: Hospital Standardized Mortality Ratio in Japan. *BMC Health Serv Res* 2008; **8**:229.
20. Campbell MJ, Jacques RM, Fotheringham J, Maheswaran R, Nicholl J. Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. *BMJ* 2012; **344**:e1001.
21. 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.
22. Shipman Inquiry. Third report—death certification and the investigation of deaths by coroners. July 2003. <http://www.official-documents.gov.uk/document/cm58/5854/5854.pdf> (24 February 2013, date last accessed).
23. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. *NEJM* 2013; **368**:201–3.
24. Campbell D. Neglect and indignity: Stafford hospital inquiry damns NHS failings. *The Guardian*. 1 December 2011. <http://www.guardian.co.uk/society/2011/dec/01/stafford-hospital-inquiry-nhs-failings> (24 February 2013, date last accessed).
25. Ross T. Jeremy Hunt warns of ‘Stafford’-type neglect across the NHS. *The Telegraph*. 21 January 2013. <http://www.telegraph.co.uk/news/politics/9814134/Jeremy-Hunt-warns-of-Stafford-type-neglect-across-the-NHS.html> (24 February 2013, date last accessed).
26. Delamothe T. Repeat after me: “Mid Staffordshire”. *BMJ* 2010; **340**:c188.
27. Jameson A. Lives saved in hospitals which acted on death rates figures. *The Telegraph*. 28 March 2009. <http://www.telegraph.co.uk/health/heal-our-hospitals/5066809/Lives-saved-in-hospitals-which-have-acted-on-death-rate-figures.html>. (24 February 2013, date last accessed).
28. Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with high-risk adjusted mortality rates provide poor quality care? A systematic review of the literature. *BMC Health Serv Res* 2007; **7**:91.
29. Girling A, Hofer TP, Wu J, Chilton PJ, Nicholl JP, Mohammed MA, *et al*. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. *BMJ Qual Saf* 2012; **21**:1052–6.
30. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA* 2001; **286**:415–20.
31. Mohammed MA, Deeks J, Girling A, Rudge G, Carmalt M, Stevens A, *et al*. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. Authors reply. 3 April 2009. <http://www.bmj.com/rapid-response/2011/11/02/authors-reply-11> (24 February 2013, date last accessed).
32. Mohammed MA, Deeks J, Girling A, Rudge G, Carmalt M, Stevens A. Standardised mortality ratios. Author’s reply. *BMJ* 2009; **338**:b1750.
33. The Shipman Inquiry: independent inquiry into the issues arising from the case of Harold Fredrick Shipman. <http://>

- www.archive2.official-documents.co.uk/document/cm63/6394/ (24 February 2013, date last accessed).
34. Mohammed M, Rathbone A, Myers P, Patel D, Onions H, Stevens A. An investigation into general practitioners associated with high patient mortality flagged up through the Shipman inquiry—retrospective analysis. *Br Med J* 2004; **328**:1474–7.
 35. Billett J, Kendall N, Old P. An investigation into GPs with high patient mortality rates: a retrospective study. *J Public Health* 2005; **27**:270–5.
 36. Mohammed MA, Mant J, Bentham L, Rafter J. Comparing processes of stroke care in high- and low-mortality hospitals in the West Midlands, UK. *Int J Qual Health Care* 2005; **17**:31–6.
 37. Mohammed MA, Lilford RJ (eds). Probing variations in hospital standardised mortality ratios in the West Midlands. University of Birmingham, Birmingham, UK, 2008.
 38. Lilford R, Brown C, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007; **335**:648–50.
 39. Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf* 2012; **21**:737–45.