

Monitoring hospital mortality

A response to the University of Birmingham report on HSMRs

Dr Paul Aylin

Dr Alex Bottle

Professor Sir Brian Jarman

Dr Foster Unit at Imperial, Department of Primary Care and Social Medicine,
1st Floor, Jarvis House, 12 Smithfield Street, London EC1A 9LA

09/01/09

Contents

- 4** Summary
- 6** Overview
- 7** Can coding depth affect HSMR?
- 10** Does place of death (ie in community establishments) affect HSMR?
- 11** The failing hospital hypothesis
- 12** The quality of care hypothesis
- 13** The validity of the Dr Foster methodology and the constant risk fallacy
- 14** References

Summary

In June 2008, a team from the University of Birmingham, led by Dr Mohammed A Mohammed, published a report commissioned by West Midlands Strategic Health Authority (SHA) entitled 'Probing Variations in Hospital Standardised Mortality Ratios in the West Midlands'. The report was highly critical of Hospital Standardised Mortality Ratios (HSMR).

The report explores a number of explanations for variations in HSMR:

- Coding depth
- Community provision
- The failing hospital hypothesis
- The quality of care hypothesis
- The 'constant risk fallacy'

Coding depth

The report claims a significant negative correlation in three of the four hospitals examined with an increase in the average Charlson index associated with a drop in HSMR.

Contradicting their claims, results given within the report show only two out of the four hospitals with a weak but significant relationship between HSMR and the Charlson index ($p < 0.05$). The report's own 'bias corrected' HSMRs (estimates

adjusted for coding bias) do not alter the fact that the hospitals concerned remain outside 99.8 per cent control limits. There is a much stronger relationship between property prices and HSMRs, illustrating the fallacy of assuming a causal relationship from a correlation of temporal trends. Our findings using national data suggest only a weak relationship between coding depth and HSMR.

Community provision

The report finds a negative correlation between HSMR and the proportion of deaths occurring in community establishments.

There was no mention of statistical significance in this chapter. Brian Jarman's original 1999 BMJ HSMR paper looked at the issue of community provision and found that adjusting for this made only very small differences to the HSMR. A more recent analysis of all deaths (including deaths outside of hospital) shows a very strong correlation ($R^2=0.922$) of 30-day HSMRs, with HSMRs calculated using in-hospital mortality.

The failing hospital hypothesis

The report looks at the relationship between HSMRs and some potential indicators chosen by the authors of a

'failing organisation', and concludes there is little evidence supporting a link between these indicators and HSMR.

Although for many variables the report found no relationship, it did suggest a relationship between staff members' views and attitudes towards their workplace. The report highlights a negative relationship between patient survey variables and mortality, particularly 'respect and dignity shown' (ie low respect shown = high mortality). Clearly these are interesting results, and further work is required to explain them.

The quality of care hypothesis

The authors look at the relationship between case-note reviews in six hospitals for stroke and fractured neck of femur (#NOF) and deaths in 'low risk' patients at one trust in the West Midlands. They conclude there is little evidence of a link between process of care measures and HSMR.

None of the process of care measures for stroke and #NOF take into account *C-difficile*, wound infections, bed sores, missed antibiotics, poor fluid control, hospital acquired chest infection rates, suture line leaks, etc. The review of 'low risk' patients defined those with a risk of death predicted by the risk models of less

than 10 per cent. We would not regard a patient with a predicted risk of death of 9 per cent as at a low risk of death, and the assumption that under the Imperial College risk model only 14 cases were expected to die is unreasonable. The authors have a ‘glass half full’ interpretation of their data. The worrying figure is the 33 per cent of deaths where there were areas of concern about patient care which may have contributed to, or did in fact cause, the patient’s death. Forty per cent of these had a hospital acquired infection.

There are other external indications about the process of care at some of the hospitals contributing to the report. The hospital that contributed to the ‘low risk’ case-note review was reported to have one of the highest proportions of deaths involving *C-difficile* infections in England (Health Statistics Quarterly, 2008). One of the other hospitals with a high HSMR, and contributing to the report’s case-note reviews, has been severely criticised by the Healthcare Commission for its emergency care.

The validity of the Dr Foster methodology and the constant risk fallacy

This final chapter suggests that the ‘constant risk fallacy’ can bias results. The chapter focuses on at least two issues

that might contribute to this constant risk fallacy: information bias and the proportionality assumption. It provides HSMR estimates ‘adjusted’ for bias which show reduction in two of the highest HSMR hospitals and it suggests that the HSMR methodology is ‘riddled’ with the constant risk fallacy.

It is widely acknowledged that all statistical models are flawed (“all models are wrong but some are useful”). Some are less flawed than others, but the authors’ selection of the four trusts at the extremes of the distribution across the region will tend to exaggerate the flaws in any model. However, despite adjusting for the potential bias highlighted in the report, the four hospitals examined still remain in their bands (outside 99.8 per cent control limits).

The HSMR is a summary figure, designed to give an overview of mortality within a trust, and we accept it will hide a considerable number of differences in the risk profiles across different factors in the model, but we do not see why this should decrease the value of the HSMR as a summary figure used in conjunction with other measures. We also looked at direct standardisation as an approach, which does not rely on the proportionality assumption, and found that directly standardised HSMRs are very closely correlated with indirectly standardised HSMRs ($R^2=0.89$).

Overview

In June 2008, a team from the University of Birmingham, led by Dr Mohammed A Mohammed, published a report commissioned by West Midlands SHA entitled 'Probing Variations in Hospital Standardised Mortality Ratios in the West Midlands'. The report was highly critical of Hospital Standardised Mortality Ratios (HSMR). The methodology, devised initially by Professor Sir Brian Jarman (Jarman et al., BMJ, 1999) and further developed by the team at Imperial College London, has been used in several countries, including the US, to monitor adjusted hospital death rates. The Dr Foster Unit (DFU) at Imperial College welcomes criticism and comment, and is looking forward to seeing some of the results of the report published in a peer-reviewed journal. However, we are keen to respond to some of the points set out in the report in more detail than an academic paper might allow, and hence have prepared this document.

To set the report in context, its authors have in the past made their position clear on the fact that they support process measures over outcome measures.

Between them, Mohammed and Richard Lilford have published several papers, arguing the merits of process measures over outcome indicators, and have stated that "although outcome data are useful for research and monitoring trends within an organisation, those who wish to improve care for patients and not penalise doctors and managers should concentrate on direct measurement of adherence to clinical and managerial standards" (Lilford et al., Lancet, 2004).

The report was commissioned by West Midlands SHA, several of whose acute trusts in the area had high standardised mortality ratios.

The report explores a number of explanations for variations in HSMR:

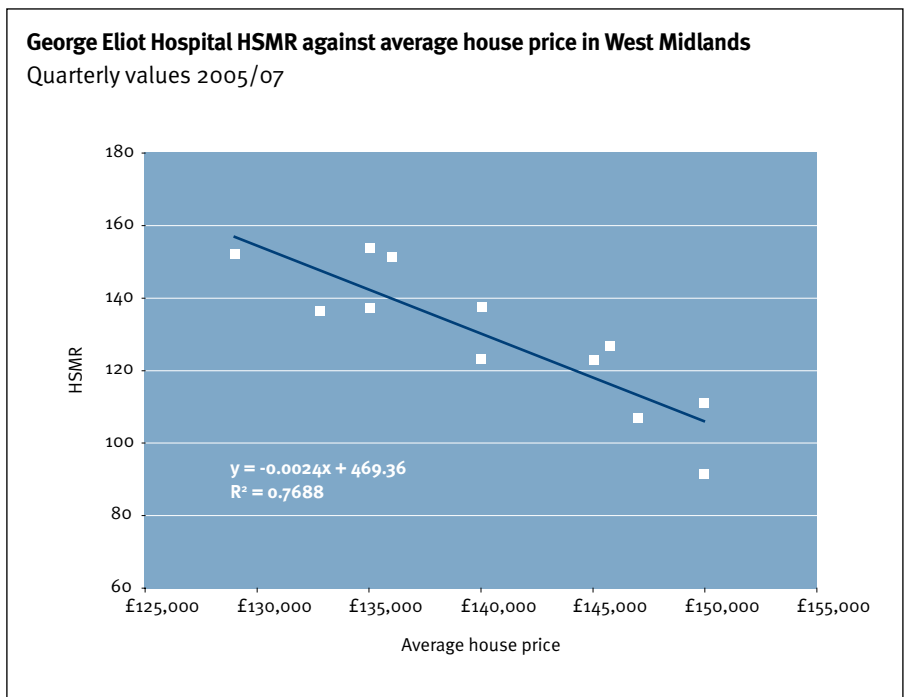
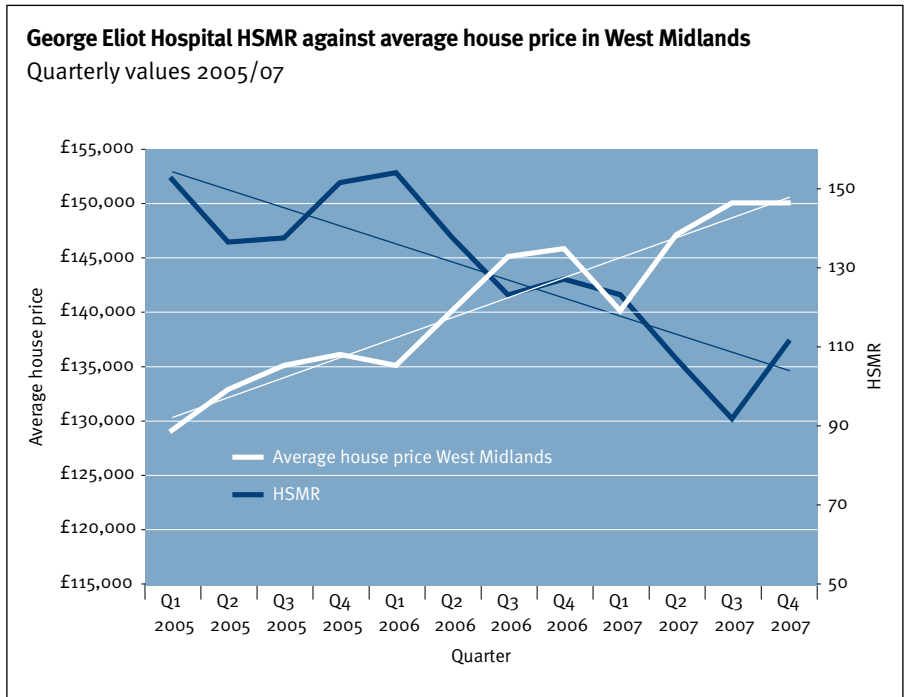
- Coding depth
- Community provision
- The failing hospital hypothesis
- The quality of care hypothesis
- The 'constant risk fallacy'

Can coding depth affect HSMR?

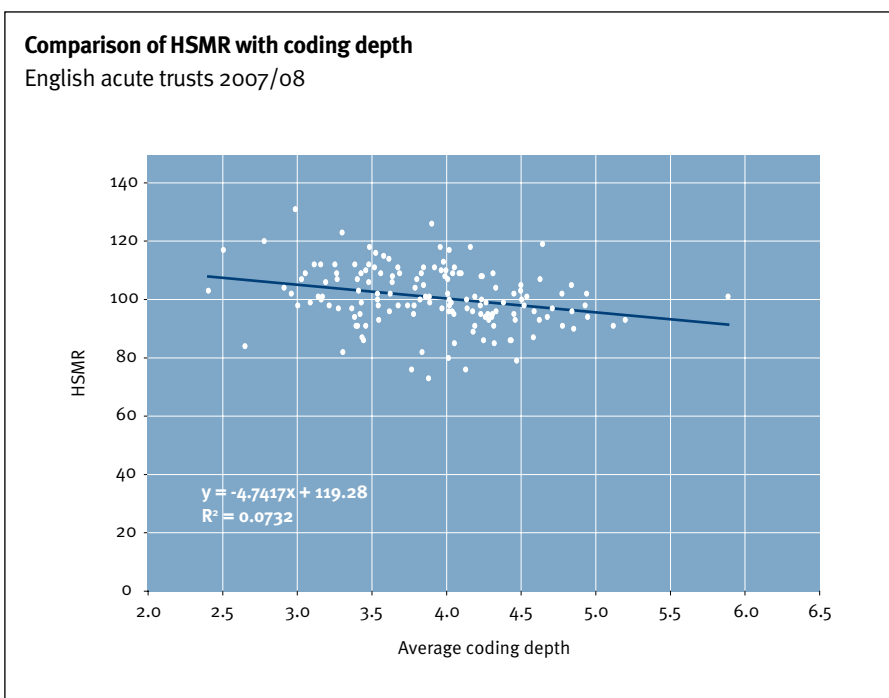
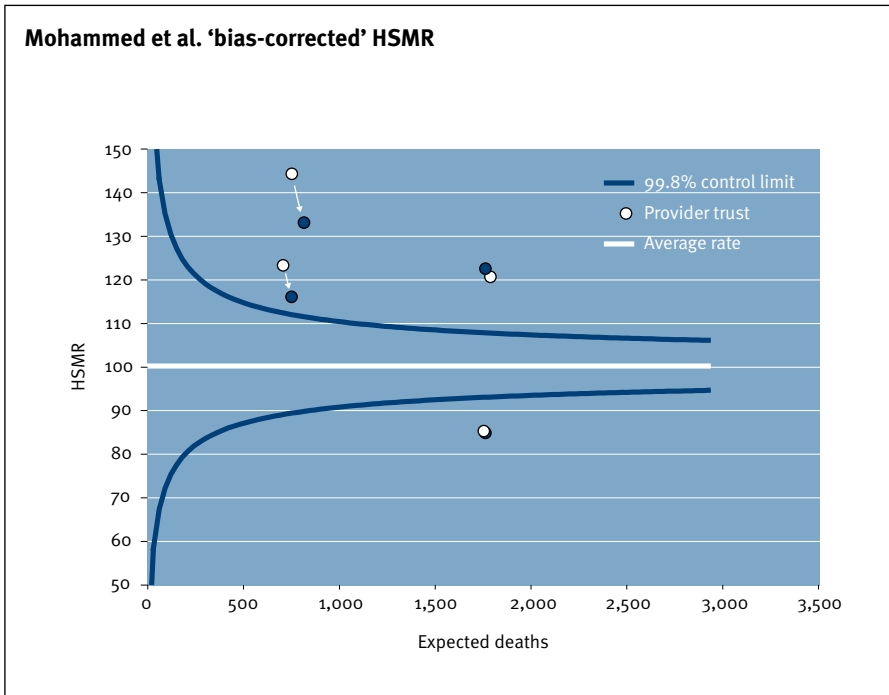
In this chapter, the report looks at four hospitals and examines the relationship between the Charlson index, coding depth and the HSMR. It claims a significant negative correlation in three of the four hospitals examined with an increase in the average Charlson index associated with a drop in HSMR (although their stated p-values show only two out of the four hospitals with p less than 0.05).

The authors appear to argue for a causal relationship between coding depth and HSMR, although their analysis is likely to suffer from what is known as the ecological fallacy. To illustrate this, in a similar time series analysis, we have found much stronger negative correlations between local property prices and HSMRs. We would clearly not want to suggest a causal link in this relationship. With regard to the report's findings, we know that HSMR is decreasing somewhat anyway over time, and we also know that coding is getting better, probably spurred on by payment by results.

We accept that coding can affect mortality ratios. However, the extent to which it does so depends on what fields are affected and by how much. Mohammed et al. have tried to estimate the potential bias (presumably based on their analysis based on time trends, though it is not clear what they have done) resulting from incomplete coding of secondary diagnoses by calculating an 'unbiased' estimate. Although their revised estimate does change some of the point estimates of the HSMR, it does not change the banding of any of the trusts included in their analysis: the high-mortality trusts still have clearly high mortality.

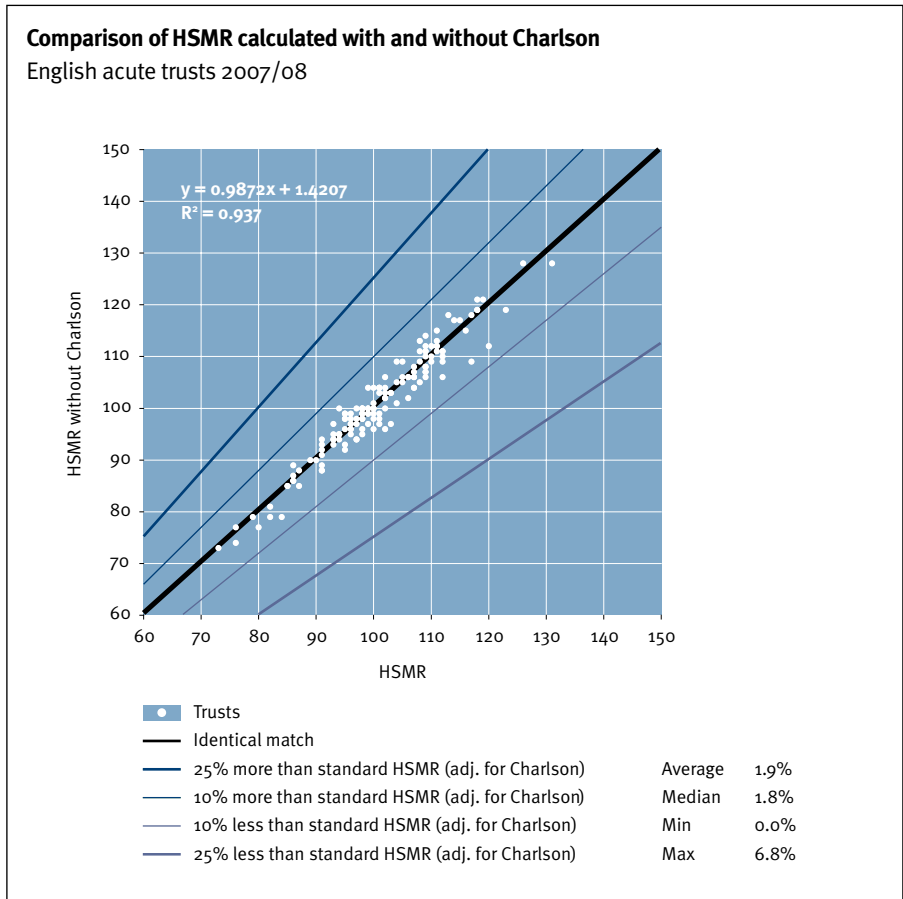


So what is the true extent of bias associated with coding depth and accuracy? Looking at national data (not restricted to one SHA) we find a very small, negative correlation ($R^2 = 0.073$) between average coding depth in the diagnosis groups used within the HSMR and the HSMR itself. The figures suggest an average coding depth of around four diagnoses. If one assumes a causal relationship, this suggests a decrease of less than five points in the HSMR if a trust were to increase its average coding depth by an additional diagnosis. However, this does assume a causal relationship, and there could be other related factors or confounders that might explain this relationship. For instance, hospitals with low mortality due to better quality of care may have better systems all round, including better diagnostics, communication, note taking and IT, and may as a by-product have better clinical coding.



We have also looked at the relationship between HSMRs adjusted for co-morbidities (using the Charlson index), and HSMRs calculated unadjusted for co-morbidity. Although there are some differences, the two measures are highly correlated ($R^2=0.937$).

A positive aspect of a focus on coding depth, as the report by the University of Birmingham suggests, is that there is evidence that trusts, such as Burton Hospitals NHS Foundation Trust and Mid Staffordshire NHS Foundation Trust, have improved their coding since – and perhaps even as a consequence of – the publication of the Hospital Guide (an annual report published by Dr Foster Intelligence).



Does place of death (ie in community establishments) affect HSMR?

The report looks at where deaths occur in primary care trusts supplying West Midlands SHA's acute hospitals. They find a negative correlation between HSMR and the proportion of deaths occurring in community establishments. They suggest perhaps using 30-day mortality instead of in-hospital mortality.

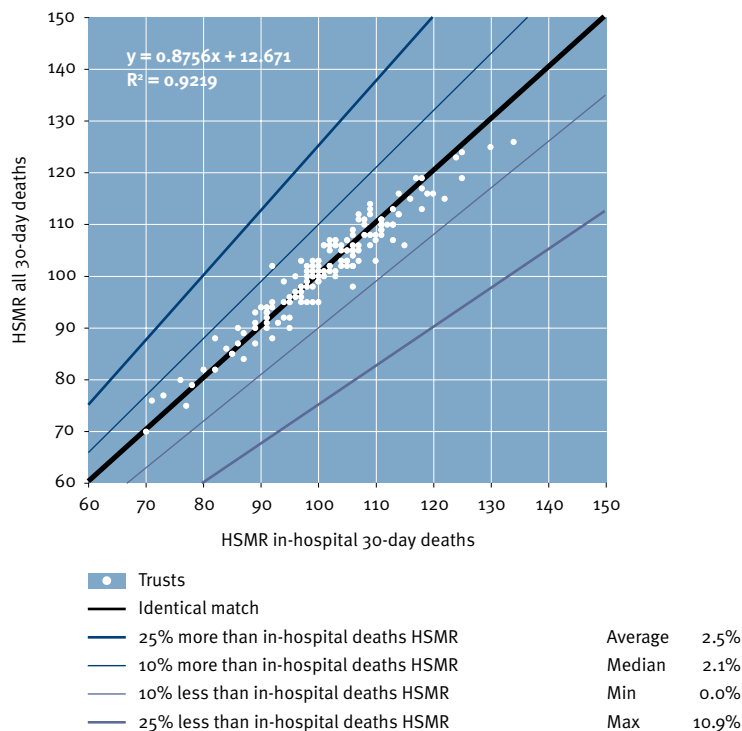
We agree that community provisions may affect HSMR, but to what extent? Within the report, there is no reference to statistical significance in their chapter on place of death, suggesting their results are not statistically significant.

Brian Jarman's original paper (Jarman et al., BMJ, 1999) looked at community provision and found that the number of NHS facilities per head of population in the district surrounding the hospital was a predictor of in-hospital mortality – the more facilities, the lower the hospital standardised mortality ratio – so this is not a new finding. However, the effect was small, with the standard deviation of the change of HSMR related to the variable being +/-1.8.

As suggested in Mohammed's report, we have looked at HSMRs based on 30-day mortality (including in- and out-of-hospital deaths) in England using ONS linked data, and have found a very strong correlation ($R^2=0.922$) with HSMRs calculated using in-hospital mortality.

Although we agree that, ideally, one would like to calculate HSMRs using all deaths (both in- and out-of-hospital deaths), unfortunately the delay involved in linking death certificate data and hospital data means that results would be out of date before they could be published.

Comparison of HSMR calculated using 30-day in-hospital deaths with HSMR using all 30-day deaths English acute trusts 2004/05



The failing hospital hypothesis

The University of Birmingham report looks at the relationship between HSMRs and some potential indicators chosen by the authors of a 'failing organisation'. The report examines the relationship between HSMR in 150 non-specialist acute hospital trusts, the NHS staff survey and the NHS hospital inpatient survey. Although for many variables the report found no relationship, it did suggest a relationship between staff members' views and attitudes towards their workplace. The report highlights a negative relationship between patient survey variables and mortality, particularly 'respect and dignity shown' (ie low respect shown = high mortality).

These are interesting findings and ones that are supported by work independently carried out by Professor Sir Brian Jarman, who also found significant ($p < 0.001$) associations between HSMR and the following questions in the National Survey

of NHS Patients (with the poorer, more dissatisfied responses corresponding to higher mortality):

- "If you had any anxieties or fears about your condition or treatment, did a doctor discuss them with you?"
- "If your family or someone else close to you wanted to talk to a doctor, did they have enough opportunity to do so?"
- "Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?"
- "Did a member of staff tell you about medication side-effects to watch for when you went home?"
- "Would you recommend this hospital to your family and friends?"

Clearly these are interesting results, and further work is required to explain them.

The quality of care hypothesis

The authors look at the relationship between case-note reviews in six hospitals for stroke and fractured neck of femur (#NOF) and deaths in 'low risk' patients at one trust in West Midlands SHA. They caution that these are preliminary findings. For stroke they found no relationship between process of care indicators and SMRs for stroke in the six hospitals. For #NOF they found a significant relationship between 'do not resuscitate' (DNR) orders and high mortality. They found that the hospital that had the lowest proportion of patients operated on within 24 hours (38 per cent) had the highest crude mortality for #NOF. However, there were no clear relationships between process of care and mortality.

None of the process of care measures for stroke and #NOF take into account *C-difficile*, wound infections, bed sores, missed antibiotics, poor fluid control, hospital acquired chest infection rates, suture line leaks, etc. The process measures examined are of interest, but other specific and systematic failures that could affect mortality were not considered. One could equally argue from the report that their process measures were not suitable indicators of quality of care and that the authors' conclusions might be revised from "there is no systematic relationship between quality of care and SMR" to "there is no systematic relationship between a limited number of process indicators for a narrow range of diagnoses and SMR". Without taking into account some of the other factors above, and looking at more diagnoses, one would not necessarily expect to see a relationship in any case.

The review of 'low risk' patients defined those with a risk of death predicted by the risk models of less than 10 per cent. We would not regard a patient with a predicted risk of death of 9 per cent as at a low risk of death. Comparisons were

made with a set of arbitrary risk categories defined by the case-note assessor. Although the grading of each patient by the reviewer is subjective and quantitative details of what is considered a high risk or low risk patient are not given, it is not surprising to find that a high proportion of the case notes were assessed as moderate-high risk.

The assumption that under the Imperial College risk model only 14 cases were expected to die is unreasonable. The individual risk of death is based on a logistic regression analysis of national data, and is intended to be used as a casemix adjustment tool, not for risk prediction. However, given that it is based on a national 'average', it is not surprising to see that higher numbers of deaths are found than would be expected in a hospital with one of the highest HSMRs in England. The researchers only selected patients who had died post hoc. Under the Imperial College risk models, you would need to make a selection of all patients (both alive and dead) in order for the risk model to accurately predict numbers of deaths. The researchers have carried out the equivalent of rounding up 250 lottery winners, each with six correct numbers (post hoc) and concluding that the predicted probability of winning the lottery (1 in 14 million) is wrong, as in the sample of 250 the winning rate is 100 per cent.

The authors have taken rather a 'glass half full' interpretation of their data. They cite the figure of 67 per cent of cases where quality of care was either adequate or non-contributory to the eventual outcome. The far more worrying figure is the remaining 33 per cent of deaths where there were areas of concern about patient care which may have contributed to or did in fact cause the patient's death. Forty per cent of these had a hospital acquired infection. This is troubling, and the fact that these factors would not have been picked up in

the case-note reviews and the examination of process of care casts more doubt on their analyses of stroke and #NOF.

It is interesting to note that there are other indications about the process of care at some of the hospitals contributing to the report. The hospital that contributed to the 'low risk' case-note review was reported to have one of the highest proportions of deaths involving *C-difficile* infections in England (Health Statistics Quarterly, 2008). It also had one of the highest HSMRs in England. From our analysis of hospital episode data from that trust, of the thousand or so deaths occurring in 2005/06, 8 per cent had a mention of *C-difficile* as a diagnosis. Recent work (Jen et al., 2008), comparing *C-difficile* rates within HES and HPA figures, suggests that HES under-records *C-difficile* by around 50 per cent, meaning the actual figure for this trust could be much higher.

One of the other hospitals with a high HSMR, and contributing to the report's case-note reviews, has been severely criticised for its emergency care. In May 2008, Healthcare Commission representatives met with the trust and outlined serious concerns about the A&E department. These were about low staffing levels in relation to medical and nursing staff, poor leadership, the structure and operation of the department, and the governance arrangements to ensure the quality of care and to protect the safety of patients. The Commission wrote to the trust detailing its concerns and asking for immediate action to address the issues. The trust has since responded to the concerns and developed an action plan. This included seeking expert advice from neighbouring hospitals and reviewing its model of care in A&E (Healthcare Commission press release, 25 September 2008).

The validity of the Dr Foster methodology and the constant risk fallacy

This final chapter of the report examines the three hospitals with the highest HSMR and the one with the lowest HSMR within the SHA. It suggests that the ‘constant risk fallacy’ (Jon Nicholl’s term) can bias results (Nicholl, *Epidemiol Community Health*, 2007). It provides HSMR estimates ‘adjusted’ for bias which shows reduction in two of the highest HSMR hospitals, and it suggests that the HSMR methodology is ‘riddled’ with the constant risk fallacy. The discussion criticises league tables and the language of Dr Foster Intelligence as describing hospitals with high HSMR as “poorly performing”.

It is widely acknowledged that all statistical models are flawed (“all models are wrong but some are useful”). Some are less flawed than others, but the authors’ selection of the four trusts at the extremes of the distribution across the region will tend to exaggerate the flaws in any model.

In the present setting, the ‘constant risk fallacy’ occurs when the relation between a casemix factor and the outcome (death in this case) differs across hospitals, and there are various potential causes, for example information bias (such as poor coding) or the use of proxies or subjective measures – see Nicholl (2007) for a review. The key assumption in multiplicative models such as indirect standardisation and logistic regression is of constant relative risk (better known as *homogeneity* or *proportionality*). We assume that between any given hospital H and the reference population (which here is all English hospitals combined) the risk in each stratum (combination of age and sex, etc) at H, multiplied by a constant (the HSMR or other estimate of relative risk), equals the risk in the same stratum in the reference population. For example, if H’s HSMR is 120, then the assumption is made that hospital H has 1.2 times the risk of death compared with the English average for all ages, both sexes and every level of other casemix factors. If this is not met (and it can be tested statistically) then bias occurs.

One could therefore report separate HSMRs for each set of risk factor levels for which the assumption is met, although Greenland and Rothman’s view is that one should only do this in the face of clear evidence that it is not met (Greenland and Rothman, 1998), in the interests of ease of analysis and reporting.

The chapter focuses on at least two issues that might contribute to this constant risk fallacy: information bias and the *proportionality* assumption. Certainly, information bias, including poor coding, will have an impact on HSMRs. It is the extent to which this can affect HSMR which is important. While the report appears to estimate its effect based on the flawed analysis on coding depth in the early chapter, we have shown only a slight (although statistically significant) effect. Interestingly, despite purporting to adjust for the potential bias highlighted in the paper (although to date, Mohammed has been unable to provide us with details of his methodology of how he did this), the four hospitals examined still remain in their bands (outside 99.8 per cent control limits).

The HSMR is a summary figure, designed to give an overview of mortality within a trust, and will hide a considerable number of differences in the risk profiles across different factors in the model. This will inevitably affect the HSMR to a certain extent. It would be perfectly possible for a trust to have a low HSMR, with some disease groups (or age groups) actually having a higher than expected mortality within that figure. Conversely, it would also be possible to have a high HSMR, with some subgroups underpinning this figure with quite low mortality. This is not in dispute, and makes comparisons of HSMR between trusts difficult, but we do not see why it should decrease the value of the HSMR as a summary figure used in conjunction with other measures. We have also looked at direct standardisation as an approach, which does not rely on the *proportionality* assumption,

and therefore would not be subject to the constant risk fallacy. Directly standardised HSMRs are very closely correlated with indirectly standardised HSMRs ($R^2=0.89$), and therefore the extent of this potential bias does not seem to have a large impact on our ‘indirectly’ standardised HSMRs (supported by the report’s own results).

We would agree that the HSMR could potentially be affected by a number of factors, including data quality, admission thresholds, discharge strategies and underlying levels of morbidity within the population, but maintain that quality of care must also be considered as a contributing factor. Where a hospital has a high HSMR then further investigation is merited in order to exclude or identify quality of care issues. Hospitals that have taken this approach in the US, UK and other countries have gained a useful insight into mortality at their institution and this has been associated with documented falls in mortality (Wright et al., *J R Soc Med*, 2006; Jarman et al., *BMJ*, 2005). Such a reduction in mortality rates can only be good for patients.

Dr Foster Intelligence does caution against the use of HSMRs in isolation, and suggests that they be used in conjunction with other evidence:

“Our aim in publishing these data is, as ever, to encourage dialogue between clinicians and managers around improving the quality of care, and to help them track changes over time and assess the impact of clinical governance. Good information combined with good leadership is effective in improving quality of care sufficiently to reduce hospital mortality. Experience tells us that the effort must be community-wide and must include good local evidence, as well as accurate, reliable data from across each trust.” (Hospital Guide, 2007)

“No measure is perfect and there are always risks that poor coding of data may affect the figure.” (The Daily Telegraph, 24 April 2007)

References

- Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, Hurwitz B, Iezzoni LI. Explaining differences in English hospital death rates using routinely collected data. *BMJ*, 1999; 318:1515–1520.
- 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.
- Health Statistics Quarterly, 2008. URL available at: http://www.statistics.gov.uk/downloads/theme_health/HSQ38_MRSA_CDif.pdf
- Jen MH, Holmes AH, Bottle A, Aylin P. Descriptive study of selected healthcare associated infections using national Hospital Episode Statistics data 1996–2006 and comparison with mandatory reporting systems. *J Hospital Infection*, 2008; 70:321–327.
- Wright J, Dugdale B, Hammond I, Jarman B, Neary M et al. Learning from death: a hospital mortality reduction programme. *JR Soc Med*, 2006; 99:303–308.
- Jarman B, Bottle A, Aylin P, Browne M. Monitoring changes in hospital standardised mortality ratios. *BMJ*, 2005; 330:329.
- Healthcare Commission press release. Healthcare watchdog triggers action to address safety concerns at Mid Staffordshire's A&E department. Published: 25 September 2008.
- Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health*, 2007; 61:1010–1013.
- Greenland S, Rothman KJ. Measures of effect and measures of association. In KJ Rothman and S Greenland, editors, *Modern Epidemiology*. Pages 47–64. Lippincott–Raven. Philadelphia, 2nd ed, 1998.
- West Midland house prices based on Land Registry data, by district, from 1996 (quarterly) 1-5. URL available at: <http://www.communities.gov.uk/housing/housingresearch/housingstatistics/housingstatisticsby/housingmarket/livetables/>

