Understanding HSMRs

A Toolkit on Hospital Standardised Mortality Ratios

Version 9: July 2014

dr foster®
intelligence

Contents

Introduction

In 1859 Florence Nightingale found that her analysis of in-hospital death rates in London showed wide variation that could not be explained by differences in the health of local populations. She went so far as to state that uniform hospital statistics would "enable us to ascertain the relative mortality of different hospitals as well as of different diseases and injuries at the same and at different ages, the relative frequency of different diseases and injuries among the classes which enter hospitals in different countries, and in different districts of the same country" (Nightingale (1863), p.159). Wide variations in English hospital inpatient death rates have been observed over the many years since, and concerns have been expressed that such variations could reflect important differences in the quality of medical care available in different hospitals.

Dr Foster has been analysing and publishing mortality data for over a decade and believe it to be an effective indicator for the quality of the services that English hospitals provide, and increasingly those provided by international hospitals. Looking forward, our ambition is to be able to include indicators of quality that incorporate patient experience data as well as data that provide alternative views.

There is a long-running debate amongst academics and NHS organisations on HSMRs, which is one that we welcome as it encourages the continued scrutiny of hospital performance and methods of measuring it. This toolkit was written in this context and it lays out the history, methodology and correct way to interpret and use HSMRs to improve quality.

HSMRs continue to be a useful indicator when used effectively. HSMRs should not be used in isolation. They provide an indication of where a problem might exist and should be used as a trigger for investigation. Issues such as coding, variation in palliative care activity and under-reporting of comorbidities can lead to high HSMRs, but where these have been ruled out it is important to note there may be a problem with the quality of care delivered by that organisation.

Diagram of the causes of mortality April 1854 – March 1855 by Florence Nightingale
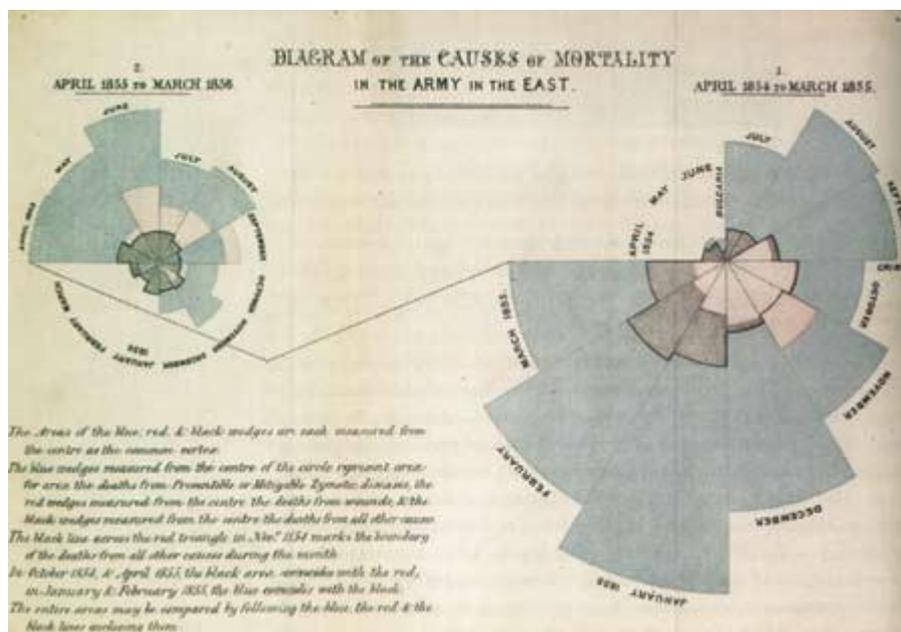
Source: Cohen, (1984) Florence Nightingale, Scientific American 250 p, 128-137

The history of HSMRs

Dr Foster has been analysing mortality data since 2000. In 2001 we published our original *Hospital Guide*, which included the first national publication of standardised hospital death rates in the world. We continue to publish these data each year, helping to fulfil the legacy of the inquiry into children's heart surgery at the Bristol Royal Infirmary[1]. The HSMR was conceived by Professor Sir Brian Jarman, director of the Dr Foster Unit at Imperial College, London, who was also a member of the Bristol Royal Infirmary inquiry.

Since the Bristol Inquiry, Professor Jarman and the Dr Foster Unit at Imperial College London have continually refined and improved the methodology for calculating HSMRs. For example, recent modifications include standardisation for palliative care, refined casemix adjustments and a reconsidered classification system to improve the identification of conditions included in the HSMR population. Today, over 70% of NHS acute trusts use HSMR analysis to monitor clinical outcomes in their hospitals via Dr Foster's Quality Investigator tool.

What is the HSMR?

The HSMR is a calculation used to monitor death rates in a trust. The HSMR is based on a subset of diagnoses which give rise to around 80% of in-hospital deaths. HSMRs are based on the routinely collected administrative data often known as Hospital Episode Statistics (HES), Secondary Uses Service Data (SUS) or Commissioning Datasets (CDS).

Measuring hospital performance is complex. Dr Foster understands that complexity and is clear that HSMRs should not be used in isolation, but rather considered with a basket of other indicators that give a well-rounded view of hospital quality and activity.

HSMRs and the Mid Staffordshire investigation

The Healthcare Commission[2] investigation into Mid Staffordshire Hospital NHS trust brought HSMRs into the news once again. The Commission notes that it was only after the publication of the 2007 Dr Foster Hospital Guide, where the trust was named as having a significantly high HSMR that the trust and the SHA took notice and began to investigate the problem. The trust was criticised for assuming that data anomalies were causing the high rate when in fact the Healthcare Commission found failings in the quality of care. Indeed they conclude:

"Trusts [must] be able to get access to timely and reliable information on comparative mortality and other outcomes, conduct objective and robust reviews of mortality rates and individual cases, rather than assuming errors in data."

Following the 'Mid Staffs' investigation the government set up an independent enquiry chaired by Robert Francis QC and reported in February 2010[3]. There is a specific section on mortality statistics which is very supportive of the sharing of these data, both in the form of publishing and working with clinicians and managers to understand outcomes. It fully endorses the principles that underpin the work of Dr Foster:

"The development and publication of comprehensive, reliable and clearly understood, statistically based information about the performance of hospitals is clearly vital not only to the NHS to assist in the management and provision of high quality health service, but also to enable the public to judge for themselves the standard of performance achieved, to inform their own healthcare choices"

---

[1]Available at: http://www.bristol-inquiry.org.uk/
[2]Now Care Quality Commission
[3]Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_113018

intelligence dr foster ®

"It is therefore particularly important that such information should be available from unimpeachably independent and reliable sources, and that it should be accompanied by clear explanations of what any figures mean, and, just as importantly, what they do not mean."... "The contribution made in this field by Professor Jarman's Unit and Dr Foster Intelligence is considerable."

HSMR: an international indicator

The HSMR is gaining in currency as a useful indicator of patient safety. In the USA, the Institute for Healthcare Improvement (IHI) has adopted HSMR analyses in their campaigns to improve the safety of patients. These include the Move Your Dot™ initiative, which gives advice and guidance to US hospitals on how to lower mortality rates. The IHI views this as "one of many current approaches being used to improve healthcare safety"[4]. In England the Patient Safety First Campaign being led by the National Patient Safety Agency (NPSA) is using HSMRs as a high level tracking measure. HSMRs are also monitored routinely in other countries such as Canada and the Netherlands.

---

[4] Reducing Hospital Mortality Rates (Part 2)'. J Whittington, T Simmonds, D Jacobsen. IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement; 2005. Available: www.IHI.org

Understanding HSMRs

How the HSMR is calculated (for full methodology see p 27)

The HSMR is a method of comparing mortality levels in different years, or for different sub-populations in the same year, while taking account of differences in casemix. The ratio is of observed to expected deaths (multiplied conventionally by 100). Thus if mortality levels are higher in the population being studied than would be expected, the HSMR will be greater than 100.

For all of the 56 diagnosis groups, the observed deaths are the numbers that have occurred following admission in each NHS Trust during the specified time period.

The expected number of deaths in each analysis is the sum of the estimated risks of death for every patient.

Benchmarks are updated annually to include values for the most recent full financial year.

Adjustment for case mix

Risks take into account those patient characteristics that are most strongly correlated with death and which reflect the patient's risk profile rather than the way in which the hospital has treated them. These factors are:

- Sex

- Age on admission (in five year bands up to 90+)

- Interactions between age on admission (in five year bands up to 90+) and Charlson co-morbidity score

- Admission method (non-elective or elective)

- Socio-economic deprivation quintile of the area of residence of the patient (based on the Carstairs Index)

- Diagnosis/procedure subgroup

- Co-morbidities (based on Charlson score)

- Number of previous emergency admissions in the preceding 12 months

- Year of discharge (financial year)

- Palliative care received

- Month of admission

- Source of admission

We currently adjust for the presence of palliative care episodes by including it in the risk adjustment model. If any episode in any of the spells in the superspell has treatment function code 315 or contains Z515 in any of the diagnosis fields, then it is defined as "Palliative", all others are termed "Non-palliative".

Ethnicity was removed from the model in 2011 due to variable coding between trusts. In some cases, up to 50% of ethnicity fields were recorded as unknown and this unequal recording may have been very slightly biasing some of our indicators.

Bandings and statistical processes

Usually we display each HSMR on a funnel plot. Funnel plots (a type of statistical process control charts) are a graphical method used to assess variation in the data and are used to compare different trusts over a single time period. Funnel plots are so named because they use control limits which form a 'funnel' around the benchmark and reflect the expected variation in the data.

Each funnel plot has three lines:

- A centre line, drawn at the mean (the National average, RR=100)

- An upper control-limit (drawn three sigma above the centre line, upper 99.8 per cent control limit)

- A lower control limit (drawn three sigma below the centre line - lower 99.8 per cent control limit)

Data points falling within the control limits are consistent with random or chance variation and are said to display 'common-cause variation'; for data points falling outside the control limits, chance is an unlikely explanation and hence they are said to display 'special-cause variation', that is, where performance diverges significantly from the national rate and the trust is classified as an outlier.

In classifying HSMRs as "high", "low" or "within the expected range", we use statistical banding to account for random chance and minimize false positives. We use 99.8 per cent control limits to determine whether an HSMR is high or low. This means that if an HSMR is outside the control limit there is only a small possibility (0.2 per cent) that this is due to chance. Only hospitals that 'pass' this control limit test are grouped as high or low and all others are classed as within the expected range.

In order to ascertain statistical significance:

- To be high, a hospital must have an HSMR above 100 and have this value above the upper control limit. A hospital with an HSMR above 100 but with the data point within the control limits is classed as 'within the expected range.'

- To be low, a hospital must have an HSMR below 100 and have this value below the lower control limit. A hospital with an HSMR below 100, but with the data point within the control limits, is classed as 'within the expected range.'

Confidence Intervals vs Control Limits

Dr Foster Intelligence only publishes data using the 99.8 per cent control limit statistical test. In our Quality Investigator tool we use the 'more liberal' 95 per cent confidence interval banding. This is to give clinicians and managers an early warning of potential problems.

To achieve statistical significance using confidence intervals:

- To be high, a hospital must have an HSMR above 100 and the lower confidence interval must also be above 100. A hospital with an HSMR above 100 but with the lower confidence interval below 100 is classed as 'within the expected range.'

- To be low, a hospital must have an HSMR below 100 and the upper confidence interval must also be below 100. A hospital with an HSMR below 100 but with the upper confidence interval above 100 is classed as 'within the expected range.'

Mortality alerts – not the same as HSMRs

The Dr Foster Unit at Imperial College London is an independent academic unit funded in part by Dr Foster. This unit writes to trusts when an alert occurs on cumulative sum charts, another kind of statistical process control chart for a variety of individual diagnosis and procedure groups. These charts are run each month, and alerts are considered with a probability of a false alarm less than 0.1% (the default in the Quality Investigator tool is 1% to give an early warning, but this can be adjusted by the user to match the 0.1% false alarm rate as desired) and other restrictions are also applied to exclude some diagnoses including cancer and vague symptoms and signs. They also exclude diagnostic procedures such as endoscopies and alerts with fewer than five deaths.

The two senior academics at the unit, Professor Sir Brian Jarman and Dr Paul Aylin, examine each alert and decide whether the trust should be notified or not. They look more carefully at alerts from specialist trusts, to examine possible casemix reasons for an alert. The Care Quality Commission is notified of the alert when it is sent to the trust chief executive. These notifications are carried out in confidence and Dr Foster is not party to which notifications are sent out.

Investigating a high HSMR – best practice

HSMR must not be considered a punitive measure but rather as an indicator for organisations as a whole to monitor their mortality. HSMRs can be used to identify potential issues early, thereby giving the organisation an opportunity to make sustainable changes to their service delivery. To facilitate this we recommend that should an organisation be shown as an outlier for HSMR that they use the following investigation pathway:

1. Check coding

Has the trust submitted incorrect data or applied different data codes to other trusts across the UK? Poor depth of coding can also affect the HSMR, i.e. when there are no or few secondary codes.

A trust can improve its coding by encouraging coders and clinicians to work more closely together (some organisations have coders attached to specific specialities) so they can better understand each others' roles and limitations; they could encourage clinicians to use a Körner Medical Records (KMR) to determine the most appropriate primary diagnosis and procedure code ; they also need to ensure that staff inputting data entry such as DOB, Sex, Discharge dates etc. are properly recorded on the PAS system understand the importance of the work they are doing and it impacts on the organisation.

2. Casemix

Has something extraordinary happened within the time frame i.e. an abnormal run of severely ill patients in a short period of time?

Is co-morbidity coding correct? Check the co-morbidity coding to identify the true casemix of the patient. No or poor co-morbidity coding can affect the HSMR.

3. Structure

Does the organisation and its surrounding healthcare partners work in a different way to other trusts across the country? Do they have different care pathways i.e. end of life care in the hospital or NHS funded hospices? Other structural differences such as no weekend discharges or nurse-led discharge teams should be considered too.

4. Process

At this point start considering that there is a potential issue with quality of care. Where service delivery needs to be reviewed, issues can be identified after monitoring and investigating alerts. Information systems such as Quality Investigator can help with this.

5. Individual or team

Very occasionally the investigation will lead you to an individual or team. Where there is a commonality of personnel involved or a particular team, nurse or department, see what extra support they need in order for them to deliver the best possible care.

HSMR checklist

Do's

- Monitor regularly – monthly, bi-monthly

- Report to the board quarterly

- Form or use existing regular reporting groups such as mortality and morbidity meetings and patient safety committees which normally include external delegates including CCGs

- Flag and audit every in-hospital death

- Investigate each alert in an open and transparent way

- Involve your clinicians in any investigation i.e. at mortality and morbidity meetings

- Be open with commissioners and the SHA/Monitor through the formation or use of your patient safety committee

- Use Quality Investigator to drill down if you have a high HSMR and examine procedural and diagnostic SMRs

- Contact Dr Foster for help with the data

Don'ts

- Assume it is just coding

- Ignore the problem

- Think HSMR is the only indicator that matters

- Use HSMRs in isolation

- Try and assign blame

- Be complacent if your HSMR is low, also check SMRs

- Wait for external organisations to raise concerns. Instead, use Quality Investigator and Dr Foster to monitor your HSMR and advise on best practice

- Ignore the need for training on both Quality Investigator and HSMRs

intelligence dr foster®

Monitoring HSMR using Quality Investigator

Successful use of HSMR monitoring can be more effective by the implementation of the Quality Investigator tool within an organisation. A lot of the best practice described below will fall out of the processes introduced as part of this implementation.

It is recommended that you have:

- An internal reporting process set up for monitoring and reporting on Quality Investigator alerts in general that could be used to monitor HSMR; it is important to investigate all Quality Investigator alerts not just the HSMR.

- Monthly monitoring of CUSUM alerts based on the individual diagnoses or procedures are useful in detecting any short term changes, whereas the HSMR itself should be monitored at least quarterly.

- The Clinical Governance, Risk, or similar teams set up to monitor patient outcomes, should take a lead on monitoring the HSMR and Quality Investigator CUSUM alerts and investigation of alerts using the investigation pathway suggested. Outcomes should then be shared with the Clinical Directors, Medical Directors, and Clinical Governance or clinical teams within each directorate affected for audit and comment.

- The inclusion of an mortality agenda item on mortality and morbidity team agendas or at the patient safety committee meeting:

  - Ensure clinical involvement, and limit a 'blame' culture developing

  - Ensure a feeling of transparency and inclusion

  - Patient Safety Committees meetings ensure transparency as they normally include external delegates including representatives from the CCG.

- A steering group that meets regularly as part of the Quality Investigator /HSMR implementation process to:

  - Ensure that reports are being monitored

  - Ensure users have had the training necessary to understand and use Quality Investigator and HSMRs

  - Ensure the process is embedded within the organisation's reporting process

  - HSMR monitoring should be discussed quarterly in preparation for completing a report for the board about changes to HSMR, investigation outcomes and action plans to improve patient outcomes and care pathways.

If you are showing as an outlier or an alert has gone off within a time frame then these should be investigated immediately. If you are an outlier a trend analysis should be carried out to identify if there was a particular point in time when you became an outlier, so you can understand what changed at this point, i.e. coding, new staff, and change in practice. Put in place plans that ensure changes are implemented and monitor any improvement in patient outcomes to ensure these are sustainable.

It is important whilst using HSMR as a performance indicator not to lose sight of your organisation's performance in other areas i.e. have alerts gone off in other diagnoses or procedures and do these have an impact on the HSMR?

HSMR as a data group should not be used to look at overall performance in outcomes such as length of stay, readmissions or day cases as HSMR looks at a very specific, restricted set of patients and could lead to some misleading results.

Recalibrating the benchmark and risk models

Each year, usually in September, Dr Foster and the Dr Foster Unit at Imperial College London recalculate the expected values and the risk estimates which are used to produce HSMRs. This is to take into account the changing patterns of in-hospital deaths and volume of admissions which alter year on year. The reasons for this include:

- Adding another year of data into the model

- Improving the risk adjustment

- Refreshing historic data

Due to the natural reduction in mortality all trusts will see their most recent HSMR increase following this update. This means a few trusts may change 'banding' and some may find their HSMR becomes significantly higher than expected when provisional results have indicated that the HSMR will be 'within the expected range'. We make contact with those trusts likely to be affected and please make contact us if you require further information.

The debate

HSMRs have been subject to much peer-review and are now widely used. Nevertheless, they remain the subject of a long-running debate in relation to their use and interpretation.

The HSMR is a measure of overall mortality, but it should be used in conjunction with other indicators in the assessment of the quality of care. Analysis of mortality in individual diagnoses and procedures, as well as the examination of other outcome and process indicators is invaluable in explaining and exploring variations between trusts.

Dr Foster is committed to continuing our work in making these data public. We aim to do this in a developmental way, helping trusts to understand their figures and ultimately improve patient care where necessary. The *Hospital Guide* annually publishes the names of trusts that have been determined as 'outliers', which means their results are significantly different to what is expected. An outlier is a data point that falls outside the control limits. These limits are set at 99.8 per cent, so it is unlikely the outliers are caused by chance. Therefore they are said to display 'special cause variation', where performance diverges significantly from the national rate.

HSMRs have been criticised by some. Below we set out the arguments made by two of the most prominent of these – Professor Nick Black and Dr Mohammed A Mohammed – and our detailed responses to them.

1. Criticism of HSMRs by Professor Nick Black

On March 4th 2014, the BBC broadcast a programme *File on Four* which presented a range of arguments for why mortality measurement and Hospital Standardised Mortality Ratios (HSMRs) should not be used for monitoring quality in health services. Professor Nick Black (London School of Hygiene and Tropical Medicine) argued that the public should 'ignore' mortality rates. Subsequently there have been a number of statements arguing against the use of administrative data to measure performance and against the use of standardised mortality rates or mortality rates as a metric. Essentially, six arguments against HSMRs have been made:

Claim 1: there is no association between mortality rates and other measures of quality of care

One argument made against HSMRs was that other ways of measuring quality of care produced inconsistent results and that, therefore, the mortality measurement must be inaccurate. Comparisons were made between the HSMR and:

• standardised mortality rates, in which the number of patient deaths is compared to the number we would expect from the patients being treated;

• avoidable deaths measures, in which a case note review is used to try to identify the number of patients where errors or omissions led to their death; and

• process measures, which count the proportion of patients who received care in line with particular recommended practices.

During the File on Four programme broadcast by the BBC on March 4th 2014, Professor Black stated: "The study we've already completed found no association between the HSMR a hospital had and the proportion of avoidable deaths and in this regard it was consistent with four or five other studies from North America and from the Netherlands which also failed to find any connection between the two."

The studies he quotes are about mortality rates at both hospital level (HSMR) and at the level of individual clinical areas. They are:

1. Best WR, Cowper DC. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. *Med Care* 1994; 32(4): 390-400.
2. Gibbs J, Clark K, Khuri S, Henderson W, Hur K, Daley J. Validating risk-adjusted surgical outcomes: chart review of process of care. *International Journal for Quality in Health Care* 2001; 13(3): 187-196.
3. Dubois RW, Rogers WH, Moxley JH, et al. Hospital inpatient mortality. Is it a predictor of quality? *N Engl J Med* 1987; 317(26): 1674-80.
4. Guru V, Tu JV, Etchells E, et al. Relationship between preventability of death after coronary artery bypass graft surgery and all-cause risk-adjusted mortality rates. *Circulation* 2008; 117(23): 2969-76.

The table below comments on each of these:

Study	Number of participants	Comment
Best 1994	222	This small study, looking at only 222 deaths, showed "a slight shift in distribution toward better care in low-ratio hospitals was not statistically significant". The study was not big enough to show differences.
Gibbs 2001	739	This study concludes: "The absence of a relationship between most of our measures of process of care and risk adjusted outcomes may be due to an insensitivity of chart reviews to hospital-level differences in quality of care." It points out that: "On some of the secondary measures, patient care was rated higher for hospitals with lower than expected operative mortality". Avoidable mortality was one of these measures but the relationship was not significant (P = 0.13). When combined with other measures of quality it was significant.
Dubois 1987		"In summary we found differences in quality between the high and the low [mortality] outliers when we used clinicians' subjective assessments, but found no significant difference for any condition when we used structured process criteria." Adjusted mortality rates did agree with clinicians' views of quality of care. But attempts to capture that in structured data on particular processes of care did not produce an association. In other words, out of three measures of quality, two agreed (mortality rates and clinical views of quality) and one disagreed (the structured process measures). Any assumption that the structured process measures are correct and the other two incorrect is unwarranted. Indeed the paper highlights a possible explanation for this result – the possibility that their structured approach may have lacked adequate sensitivity to detect a difference. For example, the paper cites the example: "One patient died soon after a feeding tube was inadvertently introduced into the right lung instead of the stomach, but the structured review had only 125 criteria and did not have one that dealt with such a case". Even given this weakness, they did still find that there were more preventable deaths in high mortality hospitals concluding: "The adjusted death rate model identified outlier hospitals that as a group had both sicker patients and more possibly preventable deaths".
Guru 2008	347	This is a small case note review study authored by a cardiac surgeon looking at 40 deaths in each of nine cardiac units and comparing with mortality rates, a total of 347 deaths. He found a very high rate of preventable deaths (111 out of 347 deaths: 32%). There are a number of issues with this paper: - The deaths for case note review were taken from 1998 to 2004. These were compared with mortality rates for a different time period (April 2000 and March 2002).

		- The study only looked at deaths from isolated first time CABGs so the numbers are very low. There was no allowance for small numbers in looking at the variation in mortality rates. - The deaths/volume numbers underlying the mortality rates are not provided. Given this, the lack of any significant association is hardly surprising. The finding is directly contradicted by another similar study using similar methodology. 'The Peer Review Organization found that 18 of 40 deaths (45%) reviewed from high-outlier hospitals had a quality-of-care problem, whereas 1 of 23 deaths (4.4%) reviewed from low-outlier hospitals had a quality problem.' Hannan EL, Kilburn H Jr, O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York State: an analysis of risk factors and hospital mortality rates. *JAMA*, 1990; 264: 2768–2774.

Since the broadcast Professor Black has published his study, referred to in his comments, which he offered as evidence of a lack of correlation between 'avoidable deaths' measures and standardised mortality rates.[5] As with earlier studies, the paper acknowledges that it looks at too few patients to be able to establish a connection between avoidable mortality measures and standardised mortality measures. The paper does not give the detailed numbers, but appears to be looking at on average 100 patients in ten hospitals which would yield a total of three-eight "avoidable deaths" in each – far too few to reliably inform any conclusions.

The focus of these studies is the comparison of risk-adjusted outcomes with other measures of quality, including process measures as well as measures of avoidable deaths from case note reviews. In the *File on Four* programme, it was suggested that there were no studies that found an association between risk-adjusted mortality and other measures of quality. *The reverse is true* as the below demonstrates. We have identified nearly 20 recent papers all of which did find an association between mortality measurement and other quality metrics. These are listed in Box B available on page 8 of this report.

One widely quoted review tries to make the case that process measures and outcome measures do not agree.[6] This paper was produced by the same research team from the University of Birmingham that carried out the analyses for West Midlands Strategic Health Authority in 2008, which argued that HSMRs were unreliable at the same time that concerns were being raised about Mid-Staffordshire NHS Foundation Trust. In this later study, in the team's continued pursuit of arguments against measurements of mortality, they made a number of methodological errors.

The paper summarises the author's conclusions as follows:

> *'A positive correlation between better quality of care and risk-adjusted mortality was found in under half the relationships (26/51 51%) but the remainder showed no correlation (16/51 31%) or a paradoxical correlation (9/51 18%).*
>
> Conclusion: *The general notion that hospitals with higher risk-adjusted mortality have poorer quality of care is neither consistent nor reliable.'*

Aside from the obvious point that 51% is not "under half", the first thing to say is that the review takes no account of the quality of the studies examined, particularly regarding sample size. Many of

---

A toolkit on Hospital Standardised Mortality Ratios

the studies in which no relationship was found were too small to detect an effect with any statistical significance.

More importantly, many of the papers cited as finding a paradoxical relationship between quality of care and mortality, actually suggested the contrary. We give quotes from each of the papers in Box A on page 17.

There is also a fundamental problem with the approach to the interpretation of these studies. The key question is: if the outcome metric is accurate what level of correlation would one expect to find between an individual process metric and an outcome metric? There are two reasons why such studies will not find a correlation even if the mortality rate is an accurate measure of quality.

1. Some of the process measures are either unrelated to mortality, or their relationship is too small to detect.

2. Even where the process measure relates directly to mortality, the relationship at a hospital level may be weak. Thus there is a strong correlation at the level of the individual patient between a process metric, such as time to surgery for a fractured hip, and the likelihood of the patient surviving. Those patients who wait longer for surgery are more likely to die. However, that relationship is often not evident at a hospital level. The correlation between the average time to surgery for a hospital and the mortality rate at the hospital can be weak or non-existent. This is not surprising since the average survival of patients at a hospital will be affected by many more issues than just time to surgery. It would be wrong to conclude that either the process measure or the outcome measure was somehow wrong simply because there was no correlation at a hospital level.

Often, a better understanding can be achieved by using mortality in combination with other metrics – for example looking at mortality alongside data about reoperations. A US study of 84,730 patients who had undergone inpatient general and vascular surgery found that hospitals with either very high mortality or very low mortality had similar rates of overall complications (24.6% and 26.9%, respectively) and of major complications (18.2% and 16.2%, respectively). However, mortality in patients with major complications was almost twice as high in hospitals with very high overall mortality as in those with very low overall mortality (21.4% vs. 12.5%, P<0.001).[7] The same has been found in the UK.[8]

Hence it is wrong to go beyond what the evidence suggests and to use these studies to argue that mortality measurement is unreliable.

---

[7] Ghaferi A, Birkmeyer J, Dimick J. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* 2009; 361:1368-1375.

[8] Almoudaris A, Mamidanna R, Bottle A, Aylin P, Vincent C, Faiz O, Hanna GB. Failure to rescue patients after reintervention in gastroesophageal cancer surgery in England. *JAMA Surg* 2013;148(3):272-276.

Box A. Some quotes from Pitches' nine so-called 'paradoxical' papers

As discussed on page 16, many of the papers cited as finding a paradoxical relationship between quality of care and mortality actually suggested the contrary, as the quotes below demonstrate:

'These results suggest that the HCFA adjusted hospital mortality rate and the PRO-confirmed problem rate are related methods to compare hospitals on the basis of quality of care'
Hartz AJ, Krakauer H, Kuhn EM, Young M, Jacobsen SJ, Gay G, Muenz L, Katzoff M, Bailey RC, Rimm AA. Hospital characteristics and mortality rates. *New England Journal of Medicine* 1989, 321(25):1720-5.

'There is some evidence that targeting hospitals with consistently high death rates over periods longer than one year may better identify potential quality problems'
Park RE, Brook RH, Kosecoff J, Keesey J, Rubenstein L, Keeler E, Kahn KL, Rogers WH, Chassin MR. Explaining variations in hospital death rates: randomness, severity of illness, quality of care. *Journal of the American Medical Association* 1990, 264:484-490.

'Low-mortality hospitals start aspirin within six hours of presentation more often than intermediate or high-mortality hospitals'
Romano PS, Remy LL, Luft HS. Second Report of the California Hospital Outcomes Project (1996): Acute Myocardial Infarction Volume Two: Technical Appendix – chapter 014. Validation study of acute myocardial infarction study: results. Center for Health Services Research in Primary Care, University of California, Davis 1996 [http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1017&context=chsrpc].

'Admission to a hospital ranked high on the list of "America's Best Hospitals" was associated with lower 30-day mortality among elderly patients with acute myocardial infarction'
Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Do "America's Best Hospitals" perform better for acute myocardial infarction? *New England Journal of Medicine* 1999, 340:286-292.

'Risk-standardized 30-day mortality rates were lower for patients treated at higher-rated than lower-rated hospitals (21.9% 1-star vs 15.9% 5-star, P=.001)'
Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented internet health care report card – the risk of quality ratings based on mortality data. *Journal of the American Medical Association* 2002, 287:1277-1287.

'There was evidence for a relationship between some process variables and outcomes at hospital discharge, but the relationships were generally weak'
McNaughton H, McPherson K, Taylor W, Weatherall M. Relationship between process and outcome in stroke care. *Stroke* 2003, 34(3):713-7.

Teaching hospitals had a higher ICU death rate – 'More interventions do not necessarily improve outcomes in critically ill patients'
Metnitz PG, Reiter A, Jordan B, Lang T. More interventions do not necessarily improve outcome in critically ill patients. *Intensive Care Medicine* 2004, 30(8):1586-93.

'The results of this project suggest that there are substantial opportunities to improve the quality of care for CABG in Oklahoma'
Venkatappa S, Murray CK, Bratzler DW. Coronary artery bypass grafting surgery in Oklahoma: processes and outcomes of care. *Journal – Oklahoma State Medical Association* 2003, 96(2):63-69.

'Ratings of overall quality of care did not differ significantly between patients from hospitals with higher and lower than expected mortality and morbidity'
Gibbs J, Clark K, Khuri S, Henderson W, Hur K, Daley J. Validating risk-adjusted surgical outcomes: chart review of process of care. *International Journal for Quality in Health Care* 2001, 13(3):187-196.

Box B: List of recent papers (published since 2006) showing an association between mortality and other quality metrics (as discussed on page 16)

Brooke BS, Perler BA, Dominici F, Makary MA, Pronovost PJ. Reduction of in-hospital mortality among California hospitals meeting Leapfrog evidence-based standards for abdominal aortic aneurysm repair. *Journal of Vascular Surgery* 2008; 47(6): 1155–1163.

Nguyen HB, Corbett SW, Steele R, Banta J, Clark RT, Hayes SR, et al. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. *Critical Care Medicine* 2007; 35(4): 1105–1112.

Ferrer R, Artigas A, Levy MM, Blanco J, González-Díaz G, Garnacho-Montero J, et al. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *JAMA* 2008; 299(19): 2294–2303.

Levy MM, Dellinger RP, Townsend SR, Linde-Zwirble WT, Marshall JC, Bion J, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Critical Care Medicine* 2010; 38(2): 367–374.

Schramm GE, Kashyap R, Mullon JJ, Gajic O, Afessa B. Septic shock: a multidisciplinary response team and weekly feedback to clinicians improve the process of care and mortality. *Critical Care Medicine* 2011; 39(2): 252–258.

Weiser TG, Haynes AB, Dziekan G, Berry WR, Lipsitz SR, Gawande AA. Effect of a 19-item surgical safety checklist during urgent operations in a global patient population. *Annals of Surgery* 2010; 251(5): 976–980.

Shahian DM, Meyer GS, Mort E, Atamian S, Liu X, Karson AS, et al. Association of National Hospital Quality Measure adherence with long-term mortality and readmissions. *BMJ Quality & Safety* 2012; 21(4): 325–336.

Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health affairs* 2007; 26(4): 1104–1110.

Heidenreich PA, Lewis WR, LaBresh KA, Schwamm LH, Fonarow GC. Hospital performance recognition with the Get With The Guidelines Program and mortality for acute myocardial infarction and heart failure. *American Heart Journal* 2009; 158(4): 546–553.

Wang TY, Nallamothu BK, Krumholz HM, Li S, Roe MT, Jollis JG, Jacobs AK, Holmes DR, Peterson ED, Ting HH. Association of door-in to door-out time with reperfusion delays and outcomes percutaneous coronary intervention. *JAMA* 2011; 305 (24): 2540-2547.

McNamara RL, Herrin J, Wang Y, Curtis JP, Bradley EH, Magid DJ, et al. Impact of delay in door-to-needle time on mortality in patients with ST-segment elevation myocardial infarction. *The American Journal of Cardiology* 2007; 100(8): 1227–1232.

Hoeks SE, Scholte Op Reimer WJM, Lingsma HF, van Gestel Y, van Urk H, Bax JJ, et al. Process of care partly explains the variation in mortality between hospitals after peripheral vascular surgery. *European Journal of Vascular and Endovascular Surgery* 2010; 40(2): 147–154.

Peterson ED, Roe MT, Delong ER, Lytle BL, Brindis RG, Smith SC, et al. Association between hospital process performance and outcomes among patients with acute coronary syndromes. *JAMA* 2006; 295(16): 1912-1920.

Lingsma HF, Dippel DWJ, Hoeks SE, Steyerberg EW, Franke CL, van Oostenbrugge RJ, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. *Journal of Neurology, Neurosurgery, and Psychiatry* 2008; 79(8): 888–894.

Kim MM, Barnato AE, Angus DC, Fleisher LA, Fleisher LF, Kahn JM. The effect of multidisciplinary care teams on intensive care unit mortality. *Archives of Internal Medicine* 2010; 170(4): 369–376.

Arabi YM, Haddad S, Tamim HM, Al-Dawood A, Al-Qahtani S, Ferayan A, et al. Mortality reduction after implementing a clinical practice guidelines-based management protocol for severe traumatic brain injury. *Journal of Critical Care* 2010; 25(2): 190–195.

Cuschieri J, Johnson J, Sperry J, West M et al. Benchmarking outcomes in the critically injured trauma patient and the effect of implementing standard operating procedures. *Ann Surg* 2012; 255(5): 993–999.

Weiss CH, Moazed F, McEvoy CA, Singer BD, Szleifer I, Amaral LAN, et al. Prompting physicians to address a daily checklist and process of care and clinical outcomes: a single-site study. *American Journal of Respiratory and Critical Care Medicine* 2011; 184(6): 680–686.

Bradley EH, Elbel B, Mcnamara RL, Magid DJ, Wang Y, Normand ST, et al. Hospital quality for acute myocardial infarction-correlation among process measures and relationship with short-term mortality. *JAMA* 2006; 296(1): 72-78.

Claim 2: avoidable deaths is a more reliable measure than standardised mortality rates

One of the assumptions underlying the criticism of the use of standardised mortality rates is that an 'avoidable deaths' measure is inherently more reliable (a gold standard). An avoidable deaths measure is based on a review of patient case notes to try to identify those patients where a failure in care led to their death. The evidence from attempts to do this does not support the view that it would be a more reliable form of measurement. Indeed there is strong evidence outlined below that an 'avoidable deaths' measure will fail to identify important aspects of quality of care.

Weak levels of agreement between note reviewers as to the degree of avoidability

Most studies attempting to identify whether or not a death was avoidable find only moderate levels of agreement between reviewers[9]. Typically measures of consistency between reviewers (ICC – Intraclass Correlation Coefficient) are in the 0.4 to 0.6 range. It is therefore not surprising that studies about the levels of avoidable deaths show wide variations in the levels of deaths regarded as avoidable.

The higher rates of reliability are achieved by setting strict criteria for the reviewers as to what should be considered avoidable.   However, this approach has the weakness that any such criteria will inevitably exclude important cases. For example, in one study (Dubois 1987, described above), 125 categories of avoidable death were listed, but there were still events that clearly qualified as avoidable deaths which did not fall within these defined categories.

Evidence of bias in assessment of avoidability of death

There is strong evidence that reviewers are prejudiced against seeing avoidability among more vulnerable patients. In most studies, reviewers are *less* likely to regard a death as avoidable if the patient is more vulnerable. So, for example, older patients or patients in a critical condition are less likely to be regarded as having suffered an avoidable death. This is counter-intuitive because these patients are the most susceptible to errors or omissions in treatment. As Hogan et al.[10] say:

> *The observation that patients were more likely to experience a problem in care if they were less functionally impaired, were elective admissions and had a longer life expectancy on admission was inconsistent with studies in other countries and might reflect a bias among reviewers towards discounting problems in the most frail, sick patients.*

At the extreme, a case-note audit carried out on deaths at Mid-Staffordshire NHS Foundation Trust prior to the independent Inquiry concluded that only one death was avoidable. While this was no doubt the sincere view of the reviewer, it is wholly inconsistent both with the evidence of independent studies of levels of avoidable death and it is at odds with the subsequent evidence from the Inquiry in which the details of individual patient stories revealed a much higher level of avoidable deaths.

---

[9] Lilford R, Edwards A, Girling A, et al. Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Policy* 2007;12:173e80.
   Hayward RA, Hoffer TP. Estimating hospital deaths due to medical errors: Preventability is in the eye of the reviewer. *JAMA* 2001 – vol 286, No 4.
   Hutchinson A, Coster JE et al. Assessing quality of care from hospital case notes: comparison of reliability of two methods. *Qual Saf Health Care* 2009. doi:10.1136/qshc.2007.023911.

[6] Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Relationship between preventable hospital deaths and other measures of safety. *Int J Qual Healthcare* (in press).

dr foster®
intelligence

Limitations of 'avoidability' as a concept

The idea of avoidability of death is too limited when considering the impact of quality of care on mortality.

As an example, Brooke et al.[11] examined 140 hospitals performing elective open abdominal aortic aneurysm repair in California, and showed that hospitals that implemented a policy for perioperative β-blockers were found to have an estimated 51% reduction of in-hospital mortality compared with control hospitals. Although still a matter for debate, even if the effect is genuine for this procedure, it is unlikely that a case note review would define any death as avoidable, simply because they didn't receive peri-operative β-blockers.

Delays in treatment for a heart attack increase the likelihood of death. McNemara et al.[12] found that in-hospital mortality was lower with shorter door to needle times. Patients waiting more than 45 minutes for fibrinolytic therapy had a 37% higher odds of death compared to those waiting 30 minutes or less. There was no cut off point at which delay becomes fatal, but on average a patient who waits longer for treatment is less likely to survive than one who is treated in less than 30 minutes. It is rarely possible to say for any given patient that extra time waited contributed to their death. However, we know that in aggregate, across a group of patients who wait longer, more of them will die and that some of them would have survived if treatment had been provided earlier. Such studies can then be replicated in practice by other units. For example, University Hospitals Coventry and Warwickshire NHS Trust redesigned their revascularisation process and reduced their door-to-balloon time and saw their in-hospital mortality rate fall. An audit of patients prior to the change would not have identified a higher rate of 'avoidable' deaths from a case note review, but the change in the process of care resulted in more patients surviving[13].

For these reasons it is wrong to assume that any lack of correlation between risk-adjusted mortality rates and case-note reviews of avoidability is due to weaknesses in the accuracy of risk-adjusted measures. It is equally likely that it reflects weaknesses in the review methodology that is failing to accurately capture avoidability of death.

Claim 3: the signal is too weak

The extent to which any indicator is useful depends on the degree of 'signal' vs 'noise'. With SMRs, noise is created by random variation in the number of patients for whom death was unrelated to quality of care issues or which was due to other aspects not identifiable in the data. This problem is exacerbated by imprecision or inaccuracies in the data.

At the extreme, it has been argued that there are simply too few avoidable deaths to account for the degree of variation in HSMRs. As Professor Black put it: 'only' 5% of deaths are avoidable according to some estimates.

A number of studies have attempted to estimate the rate between signal to noise in various measures of mortality. One study[14] estimated that the Positive Predictive Value (PPV) may be as low as 0.3 – in other words, only one-third of the hospitals in the high mortality group are genuinely worse than average. They conclude that the information should therefore be ignored. However these figures underestimate the true rate because they are based on the estimates of avoidable mortality that, as discussed above, will underestimate the level of mortality attributable to variations

---

[11] Brooke BS, Perler BA, Dominici F, Makary MA, Pronovost PJ. Reduction of in-hospital mortality among California hospitals meeting Leapfrog evidence-based standards for abdominal aortic aneurysm repair. *Journal of Vascular Surgery* 2008; 47(6): 1155–1163.
[12] McNamara RL, Herrin J, Wang Y, Curtis JP, Bradley EH, Magid DJ, et al. Impact of delay in door-to-needle time on mortality in patients with ST-segment elevation myocardial infarction. *The American Journal of Cardiology* 2007; 100(8): 1227–1232
[8] Presentation from Peter Glennon, Consultant Cardiologist University Hospitals Coventry and Warwickshire NHS Trust. Global Comparators Conference, Boston 2012.
[14] 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.

in the quality of care.   In effect, the 'low PPV' argument is simply restating the argument that 'avoidable mortality' is a better measure of care.

The actual PPV for HSMRs is unknown. The best evidence we have of the actual level is the extent to which hospitals with high mortality rates have been found wanting. That evidence is thin but suggests that they are more reliable as an indicator of quality of care than the estimated models predicted. Of the 14 trusts inspected for quality of care on the grounds of high mortality rates by Sir Bruce Keogh in 2013 (Keogh Review into the Quality of Care and Transparent Treatment by 14 Hospital Trusts in England), 11 were regarded as being sufficiently poor to warrant being placed under special measures. This would equate to a PPV of 0.79. The question about PPVs is an issue that has been raised about HSMRs rather than about mortality measurement in general. The use of mortality measures advocated by Dr Foster is that, along with other measures, including process measures, mortality rates should be considered in a range of formats including:

- HSMRs, SHMIs and other organisation level metrics;
- Mortality rates for individual procedures and diagnoses;
- Alerts generated by control chart analyses designed to identify a succession of deaths that look unlikely to be due to chance; and
- Targeted mortality rates, such as Failure to Rescue or Deaths in Low Risk Diagnoses, which look at deaths among patients where death was not expected.

All of these, in turn, should be looked at in the context of other relevant data – such as whether or not staff at a hospital would recommend the care to members of their family.

The question of whether one particular metric has a low PPV is interesting but must be considered in the light of the cumulative PPV achieved by assessing all these data together.

Independent experts including Robert Francis, Mike Richards and Don Berwick have, for these reasons, recognised the potential value of mortality rates. As Don Berwick put it:

"Unless and until a better metric is developed, the NHS should use mortality rate indicators like the Hospital Standardised Mortality Rate or suitable alternatives as one of its ways to detect potentially severe performance defects worth investigating further."

Page 28 of the Berwick Report "A promise to learn – a commitment to act; Improving the Safety of Patients in England", 2013.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/226703/Berwick_Report.pdf

Claim 4: you can manipulate the indicator by manipulating the data

One of the arguments put forward by Professor Black on *File on Four* was that it was possible to manipulate data and thereby alter the indicator. The programme cited the case of University Hospitals Birmingham NHS Foundation Trust, which reduced its reported mortality rate by increasing the level of palliative care coding.

This argument is equivalent to arguing that financial accounting rules should be abandoned because some people commit fraud. Instead of arguing that monitoring of quality is not possible, we should insist on greater audit and scrutiny of the accuracy of data. Enforcement of high standards in the recording of data is essential for the effective running of the NHS as well as ensuring the accuracy of outcome measurement.

Dr Foster is one of the few organisations that has consistently called for this to happen. Dr Foster was the first organisation to highlight the way in which palliative care coding was impacting on mortality measures. It was work by Dr Foster in 2013 that highlighted the inaccurate coding at Bolton NHS Foundation Trust – errors that had a significant impact on their reported mortality rates.

The NHS has to date failed to act adequately in response to this problem.

File on Four covered the events at Bolton NHS Foundation Trust, but concluded that the management could not be said to have done 'anything wrong' on the basis of an inquiry which found that the errors in the data were not motivated by an intention to deceive. This attitude contributes to the unreliability of data in the NHS. In any other sphere of life, the incorrect recording of official information would be a serious issue regardless of whether it was done intentionally.

The problem caused by palliative care coding is another example of the failure of the NHS to respond with sufficient speed and seriousness to problems of data recording.

In monitoring mortality rates, it makes sense to treat patients who are admitted specifically for palliative care differently from those patients who are admitted with an intention to treat and transfer home. However the coding of such patients is unreliable. That is why Dr Foster has recommended looking at rates adjusted and unadjusted for this factor in order to understand the impact. In addition, Dr Foster monitors and publishes rates of palliative care coding at different trusts.

The coding of palliative care remains unreliable because the rules around coding do not distinguish between patients admitted for the purposes of palliative care as opposed to patients admitted for treatment who subsequently deteriorated and received palliative care before their death.

Dr Foster continues to call upon NHS England and the Health and Social Care Information Centre to introduce new coding rules that will enable more accurate understanding of the care received by patients.

Claim 5: clinical audit data are better

Another argument put forward against the use of HSMRs and SMRs in general has been about the particular data source used to generate these indicators. This is not a criticism of SMRs per se, but of the use of administrative data as the source for SMRs or quality monitoring in general, on the grounds that these data are not adequate to the task, and alternative 'better' data sources are available – namely 'clinical audit' data – data recorded separately by clinicians for audit purposes.

This is wrong. Studies have found that clinical audit data tend to be under-recorded and therefore of limited value in understanding quality. While there are exceptions, in the main, clinical audit registries remain incomplete. A recent study of NSQIP, the national surgical clinical audit programme in the United States[15], came to this conclusion, as have studies in the UK[16] [17] [18].

---

[15] Krell RW, Hozain A; Kao LS, Dimick JB. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg* 2014. Mar 12. doi: 10.1001/jamasurg.2013.4249.
[16] Aylin P; Bottle A; Majeed A. (2007) Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;334: 1044.
[17] Aylin P, Lees T, Baker S, Prytherch D, Ashley S. Descriptive study comparing routine hospital administrative data with the Vascular Society of Great Britain and Ireland's National Vascular Database. *Eur J Vasc Endovasc Surg* 2007;33:461-465.
[18] Garout M, Tilney HS, Tekkis PP, Aylin P. Comparison of administrative data with the Association of Coloproctology of Great Britain and Ireland (ACPGBI) colorectal cancer database. *Int J Colorectal Dis* 2008 Feb;23(2):155-63.

Other studies have shown that failure to report data to clinical systems may be biased towards under-reporting of poor outcomes.[19] Even relatively small (5%) omissions of data can change appearances substantially if they are selective.

The difficulties with NSQIP identify another problem with clinically collected data; it is extremely expensive to collect. As a result some clinical audit systems, such as NSQIP, use sampling. However sampling techniques are also prone to bias.[10]

Clinical audit systems operated independently of administrative data are expensive and unreliable. While they are useful, they do not offer an alternative to the use of administrative data. They should instead be seen as a valuable additional source of information, ideally linked into administrative data at source and drawn from electronic patient records.

Claim 6: one large teaching hospital has a high mortality rate

The final argument in the *File on Four* programme was more anecdote than evidence. It was the statement that it was surprising that, if HSMRs measured quality, University Hospitals Birmingham NHS Foundation Trust, a teaching hospital with an 'international reputation', should have a high HSMR. The issue here is whether or not HSMRs pass the most basic reality check of appearing to have some relationship to our general understanding of quality. Certainly, it would be a concern if the hospitals that were regarded as the best in the country appeared among the worst performers.

The presenter is right to raise the question. However, the evidence points in the opposite direction. Large teaching hospitals with international reputations have among the lowest adjusted mortality rates in the country. University Hospitals Birmingham NHS Foundation Trust is the only one of England's large teaching hospitals that has a high mortality rate[16]. This difference should be regarded as a signal for more detailed investigation, not as a reason to argue against mortality rates.

[The section above is also published in 'Mortality Measurement: the case in favour' by Roger Taylor (Co-founder and Director of Research, Dr Foster) and Dr Paul Aylin (Clinical Reader in Epidemiology and Public Health, Co-Director of the Dr Foster Unit at Imperial College London), July 2014.

2. Criticism of HSMRs by Dr Mohammed A Mohammed

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'

---

[19] Almoudaris AM, Burns EM, Bottle A, et al. A colorectal perspective on voluntary submission of outcome data to clinical registries. *Br J Surg* 2011;98(1):132-9.
[16] Dr Foster Hospital Guide 2013. Large teaching hospitals are defined as members of the Shelford group of NHS trusts.

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. Using national data, findings by the Dr Foster Unit at Imperial College London in their paper, *Monitoring hospital mortality: A response to the University of Birmingham report on HSMRs* suggest only a weak relationship between coding depth and HSMR[20].

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[21] 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 (R2=0.922) of HSMRs calculated using 30-day in and out of hospital deaths, with HSMRs calculated using just in-hospital deaths.

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' (i.e. 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 (FNOF) 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.

However, the process of care measures looked at were limited and did not include *C-difficile*, wound infections, bed sores, missed antibiotics, poor fluid control, hospital acquired chest infection rates, suture line leaks, etc. Despite this, in 33 per cent of deaths, they did find 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

---

[20]Aylin, P., Bottle, A. & Jarman, B. 2009. *Monitoring hospital mortality: A response to the University of Birmingham report on HSMRs*.London: Imperial College London. Available at: http://www1.imperial.ac.uk/resources/1133EEEE-5AE7-4E07-9940-CEA0E5F2120D/
[21] Jarmon, B. 1999. 'Explaining differences in English hospital death rates using routinely collected data'. *BMJ*. 318:1515. Available at: http://www.bmj.com/content/318/7197/1515

Quarterly, 2008). One of the other hospitals with a high HSMR, and contributing to the report's case-note reviews, has been Mid Staffordshire Hospital, severely criticised by the Healthcare Commission for its emergency care.

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

The final chapter (and a subsequent paper Mohammed A Mohammed in 2009, 'Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals'[22]) 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.

Appendix 9 of the Francis enquiry is a detailed review of mortality statistics produced by two Harvard academics. It concludes that the Mohammed A Mohammed et al paper, 'Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals'[23], should not be used as an excuse to ignore high HSMRs or write them off as being methodologically flawed. The author's state:

"We are disturbed by the final sentence summarising the author's conclusions: "In other words, quality of care should remain innocent until proven guilty". This is a hospital-centric admonition, but certainly not one that would be acceptable to most patients or to the regulators entrusted with ensuring the quality of their care. We accept that there is no single, perfect mechanism for assessing health care quality. We also agree that every statistical quality monitoring algorithm, including Dr Foster, should be critically examined by experts to determine its validity. However, we believe that in the case of Mid Staffordshire, there were so many different warning flags from different entities, using different approaches, and over multiple time periods, that it would have been completely irresponsible not to aggressively investigate further."

[22] Mohammed, M.A. 2009. 'Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals'. *BMJ*. 338:b780. Available at: http://dx.doi.org/10.1136/bmj.b780

[23] Mohammed, M.A. 2009. 'Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals'. *BMJ*. 338:b780. Available at: http://dx.doi.org/10.1136/bmj.b780

Further reading

- Aylin, P., Bottle, A. "Intelligent Information: A National System for Monitoring Clinical Performance". *HSR: Health Services Research.* February 2008. 43:1, Part I

- Aylin, P., Bottle, A., Majeed, A. "Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models". *BMJ.* May 2007. 334:1044-1044

- Aylin, P., Bottle, A., Jarman, B. "Monitoring hospital mortality: A response to the University of Birmingham report on HSMRs". Imperial College. January 2009. Available at: http://www1.imperial.ac.uk/resources/1133EEEE-5AE7-4E07-9940-CEA0E5F2120D/

- Francis, R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. 2013. London: The Stationary Office. Available at: http://www.midstaffspublicinquiry.com/report

- Healthcare Commission. "Investigation into Mid Staffordshire NHS Foundation Trust". March 2009 Available at: http://www.cqc.org.uk/_db/_documents/Investigation_into_Mid_Staffordshire_NHS_Foundation_Trust.pdf

- House of Commons (Chaired by Robert Francis QC). "Independent Inquiry into care provided by Mid Staffordshire NHS Foundation Trust". January 2005 – March 2009. Available at: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/@ps/documents/digitalasset/dh_113068.pdf

- Cohen, I.B. "Florence Nightingale". *Scientific American.* March 1984. **250:**128-137

- Jarman, B., Gault, S., Alves, B. et al. "Explaining differences in English hospital death rates using routinely collected data". *BMJ.* 1999. 318:1515–20

- Jarman, B., Bottle, A., Aylin, P., Browne, M. ''Monitoring Changes in Hospital Standardised Mortality Ratios.'' *BMJ.* 2005. 330: 329

- Kafetz, A., Bedford, Z., Taylor, R., Rowell, H. eds. *The Hospital Guide: The health of our hospital revealed.* 2008. London: Dr Foster Available at: www.drfosterhealth.co.uk

- Kafetz, A., Bedford, Z., eds.2009. *The Doctor Foster Hospital Guide 2009 - How Safe is Your Hospital?* 2009. London: Dr Foster.Available at: www.drfosterhealth.co.uk

- Nightingale, F. *Notes on hospitals*. 1863. London: Longman

- Spiegelhalter, D. "Funnel plots for institutional comparison". *Quality and Safety in Health Care.* December 2002.11(4):390-1

- Wright, J., Dugdale, B., Hammond, I., Jarman, B., Neary, M., Newton, D., Patterson, C., Russon, L., Stanley, P., Stephens, R., Warren, E.; "Learning from death: a hospital mortality reduction programme". *Journal of the Royal Society of Medicine*. 2006. 99:303–308

Technical Document

HSMR Mortality Indicators Full Methodology

dr foster ®
intelligence

A toolkit on Hospital Standardised Mortality Ratios

Overview

Measures of survival are an important measure of the quality of care provided by hospitals. Florence Nightingale was one of the first people to identify the importance of measuring survival rates and in the 1860s, she highlighted the variation in survival rates for hospitals across London. Today, many clinicians routinely monitor the survival rates in their services, and use them to improve care.

The analyses are derived from routinely collected hospital data. The statistical process control charts have been adjusted to take into account a range of factors that can affect the survival rates, but which are beyond the control of the individual hospital, for example, the age and sex of the patient or whether they have another medical condition.

Methodology

1. Data sources

Mortality indicators are based on the analysis of the most recent 11 years of inpatient and day case records from Hospital Episode Statistics (HES) for the period 2000/01 to 2005/06, and Secondary Uses Service (SUS) for 2006/07 onwards. These are data that are routinely collected within the health service for administrative purposes and not specifically for clinical audit. There may be issues regarding coverage, completeness and accuracy that need to be considered when interpreting the results.

1.1 Data period

Data are extracted for analysis through SUS by the Dr Foster Unit at Imperial College on the 9th of each month.

2. General data processing

2.1 Cleaning

These data are cleaned according to established HES guidelines with one or two minor additions/modifications. More detailed information is available on request.

2.2 Area-level deprivation

The population-weighted quintiles of the Carstairs deprivation score calculated by 2001 Census Output Area are then added to the data by matching on the patient's postcode. More detailed information is available on request.

2.3 Trust mergers

As hospitals merge and services reorganised, provider codes (PROCODE) may change from one year to the next. In order to track hospitals over time, the provider codes need to be unified, i.e. just one code needs to identify each trust throughout.

intelligence dr foster®

3. "Intelligent" data processing

3.1 Linkage

The data are in the form of consultant episodes (the continuous period during which the patient is under the care of one consultant), which need to be linked into admissions (or "spells"). Records are assumed to belong to the same person if they match on NHS Number (NEWNHSNo), or where NHS Number isn't valid - date of birth, sex and postcode (DOB, SEX, HOMEADD).

Only ages within the ranges 1-120 and 7001-7007 (special values to indicate age in months for children less than 1 year) are considered valid. Duplicate records (those with the same combination of NHS Number (or date of birth, sex, postcode where NHS Number is invalid) date of admission and episode number, unfinished episodes, those with missing/invalid ADMIDATE and regular attenders (CLASSPAT=3, 4) are excluded. Some spells have the same date of admission (ADMIDATE) but different dates of discharge (DISDATE). This is not valid unless the patient was discharged and readmitted on the same day: if not, the spell with the earliest DISDATE was arbitrarily taken to be the valid one. Episodes relating to the invalid spell are excluded at this stage. Remaining episodes are sorted by provider, NHS Number (or date of birth, sex, postcode where NHS Number is invalid), date of admission, date of discharge and episode number (PROCODE, NEWNHSNO (or DOB, SEX, HOMEADD), ADMIDATE, DISDATE, EPIORDER). Episodes are not required to be in strict sequence, only in chronological order. For example, if the first one had EPIORDER=01, the second one had EPIORDER=03 and the last one of the same spell had EPIORDER=99, then the three episodes are treated just the same as if they were numbered 01, 02 and 03 (as most multi-episode spells are). However a spell must have at least one episode with EPIORDER=01 otherwise it is considered invalid and excluded. Spells with invalid length of stay (DISDATE < ADMIDATE) are also excluded.

Spells ending in transfer to another NHS hospital are linked together ("superspell"), allowing for a difference between discharge from the first trust and admission to the next trust of up to two days, using ADMIMETH= 81 or DISDEST/ADMISORC values of 49-53 (which refer to NHS providers).

3.2 Diagnosis derivation

We use the 56 diagnostic groups which contribute to approximately 83% of in-hospital deaths in England. All 56 groups are listed in Table 1, and further information on the Clinical Classification System (including the ICD codes making up the groups) is available at http://www.ahrq.gov/data/hcup/icd10usrgd.htm.

For each spell we assign a diagnosis based on the primary diagnosis in the first episode of care. However, if the primary diagnosis is a vague symptom or sign we look to the second episode (of a multi-episode spell) to derive a diagnosis.

Table 1

CCS group	Description of CCS group	Financial Year 2012/13 C statistics
2	Septicemia (except in labour)	0.792
12	Cancer of oesophagus	0.840
13	Cancer of stomach	0.834
14	Cancer of colon	0.850
15	Cancer of rectum and anus	0.865
17	Cancer of pancreas	0.778
19	Cancer of bronchus, lung	0.784
24	Cancer of breast	0.956
27	Cancer of ovary	0.861
29	Cancer of prostate	0.897
32	Cancer of bladder	0.938
38	Non-Hodgkin's lymphoma	0.845
39	Leukaemias	0.820
42	Secondary malignancies	0.821
43	Malignant neoplasm without specification of site	0.785
55	Fluid and electrolyte disorders	0.796
59	Deficiency and other anaemia	0.793
68	Senility and organic mental disorders	0.684
100	Acute myocardial infarction	0.765
101	Coronary atherosclerosis and other heart disease	0.864
103	Pulmonary heart disease	0.790
106	Cardiac dysrhythmias	0.864
107	Cardiac arrest and ventricular fibrillation	0.698
108	Congestive heart failure, nonhypertensive	0.688
109	Acute cerebrovascular disease	0.743
114	Peripheral and visceral atherosclerosis	0.890
115	Aortic, peripheral, and visceral artery aneurysms	0.859
117	Other circulatory disease	0.814
122	Pneumonia	0.819
125	Acute bronchitis	0.857
127	Chronic obstructive pulmonary disease and bronchiectasis	0.719
129	Aspiration pneumonitis, food/vomitus	0.710
130	Pleurisy, pneumothorax, pulmonary collapse	0.810
131	Respiratory failure, insufficiency, arrest (adult)	0.758

133	Other lower respiratory disease	0.833
134	Other upper respiratory disease	0.898
145	Intestinal obstruction without hernia	0.829
148	Peritonitis and intestinal abscess	0.862
149	Biliary tract disease	0.931
150	Liver disease, alcohol-related	0.705
151	Other liver diseases	0.825
153	Gastrointestinal haemorrhage	0.830
154	Noninfectious gastroenteritis	0.872
155	Other gastrointestinal disorders	0.899
157	Acute and unspecified renal failure	0.737
158	Chronic renal failure	0.888
159	Urinary tract infections	0.798
197	Skin and subcutaneous tissue infections	0.909
199	Chronic ulcer of skin	0.799
224	Other perinatal conditions	0.764
226	Fracture of neck of femur (hip)	0.761
231	Other fractures	0.831
233	Intracranial injury	0.778
237	Complication of device, implant or graft	0.848
245	Syncope	0.798
251	Abdominal pain	0.945

3.3 Outcome derivation

We define our death outcome when the patient dies in hospital at the end of their stay in hospital (superspell). The spell in which death occurs (DISMETH = 4 or 5) may be post-transfer, but deaths are assigned to all the trusts in the superspell.

3.4 Derivation of additional parameters for risk adjustment

Table 2

Parameter	Definition	Excluded if invalid
Admission method	If ADMIMETH = 11,12,13 in last episode of spell with valid ADMIMETH, then "Elective" else "Non-elective"	Yes, if no episodes in spell contain valid ADMIMETH
Age group	Age on admission in 5-year bands (<1 year,1-4,5-9,…90+)	Yes, if no episodes in spell contain valid age on admission
Year of discharge	Financial year of date of discharge at the end of the superspell	Yes, if no episodes in spell have either valid DISDATE or SPELEND="Y" and valid EPIEND
Deprivation quintile	Derived from postcode on the episode in the spell in the diagnosis dominant episode	No

dr foster®
intelligence

Diagnosis subgroup	Based on official CCS sub-groups within each CCS group	n/a
Sex	Derived from the episode with the first valid value (1 or 2) of SEX, going backwards from the end of the spell.	Yes, if no episodes in spell contain valid SEX
Comorbidity (Charlson score)	The CHARLSON score for a spell is calculated as the sum of the scores for each of the conditions (see Appendix A) in the diagnosis-dominant episode (a condition can only be counted once in a spell). This score is capped at 50. We have expanded the coding definition of some conditions such that more patients are identified as having those conditions. Only secondary diagnoses (DIAG2-DIAG14) are now considered. There is now greater variation in weights between conditions and the Charlson index (the sum of the weights) is treated as a continuous variable (limited to the range 0-50) for the purposes of risk adjustment.	n/a
Emergency admissions in previous 12 months (admicount12)	Calculated as the number of superspells in the previous 365 days for the same patient (using the general pseudonymised patient identifier). This includes the current spell, if it is an emergency admission. Possible values are 0, 1, 2, 3+	n/a
Palliative care	If any episode in any of the spells in the superspell has treatment function code 315 or contains Z515 in any of the diagnosis fields, then "Palliative" else "Non-palliative"	n/a
Month of admission		n/a
Source of admission	Home - Initial spell and ADMISORC=19 Transfer (Acute) - Linked transfer from another acute NHS provider Other place - Initial spell and all other ADMISORC values Transfer (Non-acute) - Linked transfer from another non-acute NHS provider Transfer (Unknown) - ADMIMETH=81 or ADMISORC=49-53 but no previous spell found Birth - Initial spell and ADMIMETH=82-83 Transfer (Internal) - Linked transfer from this NHS provider Unknown (non-transfer) - ADMISORC=98-99	No

intelligence dr foster®

4. Risks

4.1 Denominator

We exclude day cases (spells where CLASSPAT = 2 in first episode) from our risk models and where there is more than one spell with the same diagnostic group (CCS) in a superspell, we include only the first occurring spell.

4.2 Logistic regression models

For each diagnosis group (CCS) we derive predicted probabilities for inpatient in-hospital mortality by fitting logistic regression models using the glm and step functions from the stats package in R 2.15.3[24]. We apply R's backwards stepwise elimination procedure for variable selection, which starts with a model including all the selected explanatory variables and then automatically removes the variables until it finds the model with lowest value of AIC.

We use the variables defined in Table 2 as our predictors. We recategorise three variables – age group, deprivation and number of previous admissions – depending on the absolute number of events, so that each category contains at least 10 events. Starting from the first (lowest) category, we combine it with the next lowest category if it contains fewer than 10 events and continue combining until that total has been reached. We then inspect the next highest category and repeat the process as necessary. If the last category is left with fewer than 10 events then it is combined with the second last category as one group.

4.3 Estimate of risk

The risk estimate (R) for each inpatient is calculated from the table of log odds produced by the risk modelling process as follows:

R = exp(sum(logodds)) / (1+exp(sum(logodds)))

For day cases, R=0.

Risk estimates for data in years after the last year included in the risk model are calculated using the log odds value for the last year in the model.

4.4 Quality of risk model (the 'C statistic')

The success of case-mix adjustment for accurately predicting the outcome (discrimination) was evaluated using the area under the receiver operating characteristic curve (c statistic). The c statistic is the probability of assigning a greater risk of death to a randomly selected patient who died compared with a randomly selected patient who survived. A value of 0.5 suggests that the model is no better than random chance in predicting death. A value of 1.0 suggests perfect discrimination. In general, values less than 0.7 are considered to show poor discrimination, values of 0.7-0.8 can be described as reasonable and values above 0.8 suggest good discrimination. These c-statistics are given in Table 1.

---

[24] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

dr foster®
intelligence

5. Calculation of HSMR

For all of the 56 diagnosis groups, the observed deaths is the number of deaths to have occurred following admission (as recorded in HES/SUS) in each NHS Trust during the specified time period.

The expected number of deaths in each analysis is the sum of the estimated risks of death. The ratio is calculated by dividing the actual number of deaths by the expected number and multiplying the figure by 100. It is expressed as a relative risk, where a risk rating of 100 represents the national average. If the trust has an HSMR of 100, that means that the number of patients who died was exactly as it would be expected taking into account the standardisation factors. An HSMR above 100 means more patients died than would be expected; one below 100 means that fewer than expected died.

Control limits tell us the range of values which are consistent with random or chance variation. Data points falling within the control limits are consistent with random or chance variation and are said to display 'common-cause variation'. Data points falling above the upper 99.8% Poisson control limit are said to be significantly 'higher than expected', data points falling below the lower 99.8% Poisson control limit are said to be significantly 'lower than expected'.

The distinction between control limits and confidence intervals is important; although they are very similar in construction and the difference between the two is subtle. Control limits are used because they offer hypothesis tests whereas (strictly speaking) confidence intervals do not. Control limits come from the Poisson distribution and are calculated using an exact method using visual basic routines made available by John C Pezzullo[25]. For further information, please read David Spiegelhalter's informative paper[26]. The Eastern Region Public Health Observatory also has a large resource of relevant information and tools available online[27].

---

[25]Available at: http://statpages.org/
[26]Spiegelhalter, D.J. 2005. 'Funnel plots for comparing institutional performance'.*Stat Med*.24(8):1185-202. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15568194
[27] Available at:: www.erpho.org.uk

6. Relevant publications

- Aylin, P., Bottle, A. Are hospital league tables calculated correctly? A commentary. *Public Health*. 2007. 121, ISSN:0033-3506, 905-906

- Aylin, P., Bottle, A., Majeed, A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ.* 2007. 334: 1044-9

- Aylin, P., Lees, T., Baker, S., Prytherch, D., Ashley, S. Descriptive study comparing routine hospital administrative data with the Vascular Society of Great Britain and Ireland's National Vascular Database. *Eur J Vasc Endovasc Surg*. 2007. 33:461-465

- Bottle, A., Aylin, P., Majeed, A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. *JR Soc Med*. 2006. 99:406-414

- Bottle, A., Aylin, P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. *J Clin Epidemiol.* 2011. 64(12):1426-33

- Bottle, A., Aylin, P. Intelligent Information: a national system for monitoring clinical performance. *Health Services Research.* 2008. 43:10-31

- Bottle, A., Aylin, P. Mortality associated with delay in operation after hip fracture: observational study. *BMJ.* 2006. 332:947-951

- Bottle, A., Jarman, B., Aylin, P. Hospital Standardised Mortality Ratios: sensitivity analyses on the impact of coding. *Health Serv Res*. 2011. 46(6pt1):1741-61

- Bottle, A., Jarman, B., Aylin, P. Hospital Standardised Mortality Ratios: Strengths and Weaknesses. *BMJ.* 2011. 342: c7116 (online first)

- Jarman, B., Gault, S., Alves, B., Hider, A., Dolan, S., Cook, A., Hurwitz, B., Iezzoni, L.I. Explaining Differences in English Hospital Death Rates Using Routinely Collected Data. *BMJ* 1999. 318:1515-1520

- Jen, M.H., Bottle, A., Kirkwood, G., Johnston, R., Aylin, P. The performance of automated casemix adjustment regression model building methods in a health outcome prediction setting. *Health Care Manag Sci.* 2011.14:267-278

- Spiegelhalter, D. Funnel plots for institutional comparison. *Quality and Safety in Health Care.* 2002. 11(4):390-1

- Spiegelhalter, D.J. Funnel plots for comparing institutional performance. *Stats Med.* 2005. 30;24(8):1185-202

- Sundararajan, V. et al. New ICD-10 version of the Charlson Comorbidity Index predicted in-hospital mortality. *Journal of Clinical Epidemiology.* 2004. 57: 1288–1294

dr foster®
intelligence

Appendix A. Charlson comorbidity conditions

The original Charlson weights were derived about 25 years ago in the USA. We have updated them and calibrated them on recent English data because of differences in coding practice and hospital patient population characteristics, and changes in mortality linked to comorbid conditions over time. For example, HIV previously had the highest weight of all the conditions accounted for, but due to the fall in mortality in patients with HIV over the last 25 years, this weight no longer accurately reflects the risk associated with it. We had advice from some clinical coders on current English coding practice and, where possible, also assessed the consistency of comorbidity recording among admissions for the same patient. As a result:-

- We have expanded the coding definition of some conditions such that more patients are identified as having those conditions.

- Only secondary diagnoses (DIAG2-DIAG14) are considered.

- There is greater variation in weights between conditions and the Charlson index (the sum of the weights) is treated as a continuous variable (limited to the range 0-50) for the purposes of risk adjustment.

Condition No.	Condition Name	New Coding	Weight
1	Acute myocardial infarction	I21, I22, I23, I252, I258	5
2	Cerebral vascular accident	G450, G451, G452, G454, G458, G459, G46, I60-I69	11
3	Congestive heart failure	I50	13
4	Connective tissue disorder	M05, M060, M063, M069, M32, M332, M34, M353	4
5	Dementia	F00, F01, F02, F03, F051	14
6	Diabetes	E101, E105, E106, E108, E109, E111, E115, E116, E118, E119, E131, E131, E136, E138, E139, E141, E145, E146, E148, E149	3
7	Liver disease	K702, K703, K717, K73, K74	8
8	Peptic ulcer	K25, K26, K27, K28	9
9	Peripheral vascular disease	I71, I739, I790, R02, Z958, Z959	6
10	Pulmonary disease	J40-J47, J60-J76	4
11	Cancer	C00-C76, C80-C97	8
12	Diabetes complications	E102, E103, E104, E107, E112, E113, E114, E117, E132, E133, E134, E137, E142, E143, E144, E147	-1
13	Paraplegia	G041, G81, G820, G821, G822	1
14	Renal disease	I12, I13, N01, N03, N052-N056, N072-N074, N18, N19, N25	10
15	Metastatic cancer	C77, C78, C79	14
16	Severe liver disease	K721, K729, K766, K767	18
17	HIV	B20, B21, B22, B23, B24	2

dr foster®
intelligence

Appendix B. Formula for calculating Carstairs

The formula was used to calculate the Carstairs index for each Output Area (OA) derived from 2001 census within the UK (England, Wales, Scotland, and Northern Ireland). The figures were normalised for the UK as a whole. Output areas were allocated into quintiles based on resident population, giving an equal total population in each quintile.

Unemployment variable: table KS0009b – Economic activity – males

unemp: KS09b0005   -> unemployed males over 16

unempd: KS09b0001   -> males over 16

No car variable: table UV062 – Cars or vans

nocar: UV0620002 -> households without a car or van

nocard: UV0620001 -> all households

Overcrowding variable: table UV058 – Persons per room

overcrow: UV0580004 + UV0580005 -> households with over 1.0 persons per room

overcrod: UV0580001   -> all households

Low social class variable: table UV050 – Approximated social grade

lowclass: UV0500005 + UV0500006 -> number of persons in a grade D or E classified household

lowclasd: UV0500001   -> all people

Calculation:

* delete all the records with no household residents (Census unit without people)

* compute proportions:

   unempp = (unemp/unempd) *100

   nocarp = (nocar/nocard) *100

   overcrop = (overcrow/overcrod) *100

   lowclasp = (lowclass/lowclasd) *100

* compute z-values for unempp, nocarp, overcrop, lowclasp

* Carstairs Index = sum of z-values for the four variables

[Note: The multiplication by 100 is not necessary as it cancels out in the normalisation process]

Further information on the Clinical Classification System (including the ICD codes making up the groups) is available at http://www.ahrq.gov/data/hcup/icd10usrgd.htm.

About Dr Foster

Dr Foster works with healthcare organisations to achieve sustainable improvements in their performance through better use of data. We believe that comparative information about the quality and efficiency of care, based on open methodologies and multiple data sources, drives improvement in healthcare. We are the leading provider of healthcare variation analysis and clinical benchmarking solutions worldwide: our specialist teams share their world class expertise with healthcare leaders in more than 40 pre-eminent academic hospitals in ten countries along with over 70% of English acute hospital trusts. Operating in England, continental Europe, the United States and Australia we work with a wide range of organisations and clients, including healthcare providers and commissioners, clinicians and managers. Our products, people and services provide actionable insights and inform decision-making by leadership teams.

It works closely with the Dr Foster Unit at Imperial College London and all its methodologies are published in full. Its data analysts and academic experts lead the international data comparison project Global Comparators, which drives best practice using healthcare data insight, across the world.

The Dr Foster Unit at Imperial College London has developed pioneering methodologies that enable fast, accurate identification of potential problems in clinical performance – and areas of high achievement.

Dr Foster works to a code of conduct that prohibits political bias and requires it to act in the public interest. The code is monitored by the Ethics Committee, an independent body whose role is to ensure that Dr Foster meets the standards set out in its Code of Conduct. In the event of a complaint, the committee reviews Dr Foster's published material and can adjudicate in any complaints or disputes.

For more information visit: www.drfoster.co.uk or contact us on, 020 7332 8800.