**RESEARCH METHODS & REPORTING**

**Strengths and weaknesses of hospital standardised mortality ratios**

Alex Bottle, Brian Jarman, Paul Aylin

Hospital standardised mortality ratios are fairly easy to produce and, as the example of Mid Staffordshire shows, can help identify hospitals with poor performance. However, they are not without problems.

Hospital standardised mortality ratios (HSMRs) are intended as an overall measure of deaths in hospital, a proportion of which will be preventable. High ratios may thus suggest potential problems with quality of care. Although a growing number of countries are using HSMRs, they are controversial, especially if the figures are made public, as in England and Canada.1,3

The HSMR is complex but cheap and relatively easy to calculate from national or other benchmark data that allow calculation of patients’ predicted risks of death. However, there are a number of methodological challenges in their construction and interpretation, which we discuss below. Although there are other versions of the HSMR, we focus on the Jarman one.4 Full methodological details of its construction in England are given on bmj.com. A few of the finer points that we discuss are specific to English hospital data, but most of the methodological concerns are relevant to HSMRs (or other composite hospital mortality measures) in any developed country.

What is an HSMR?
The HSMR is derived from administrative data commonly used for billing purposes from hospital information systems such as Hospital Episode Statistics in England. It is the ratio of the observed to expected deaths, multiplied by 100, with expected deaths derived from statistical models that adjust for available case mix factors such as age and comorbidity.

<table>
<thead>
<tr>
<th>SUMMARY POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSMRs are intended as an overall measure of in-hospital mortality</td>
</tr>
<tr>
<td>They are used as a screening tool in increasing numbers of countries</td>
</tr>
<tr>
<td>Case mix adjustment, admissions consisting of multiple care episodes and inter-hospital transfers, and patients with multiple admissions complicate their construction</td>
</tr>
</tbody>
</table>

The HSMR is meant as an overall measure of adjusted in-hospital mortality and serves as a screening tool. Some of the deaths in the numerator will be preventable. Thus some of the variation in HSMRs between hospitals will be due to important variation in preventable deaths, although much will be due to other factors. Estimates of the number of preventable in-hospital deaths vary; in the UK it is estimated to be in the thousands.5 Nevertheless, two English hospital groups, Mid Staffordshire NHS Trust and Basildon and Thurrock NHS Trust, had high HSMRs when they were investigated by the national healthcare regulator and found to have substandard care (box).

How they are used
HSMRs are used as a small part of our system for monitoring quality of care.4 Some hospitals also use them as part of quality improvement efforts.7,8 Their boards monitor the HSMR alongside other indicators such as mortality for individual diagnosis groups, infections, and patient experience. In England, Dr Foster Intelligence, a private company and joint venture with the NHS Information Centre, includes HSMRs in its annual Hospital Guide, and the figures are publicly available on the NHS Choices website.

We envision the role of HSMRs as part of a suite of measures for hospitals’ internal use. The figures can be broken down by diagnosis group, and any potential problems investigated by checking the data and analysing processes, often going as far as a case note review.9 This is considered the gold standard method for deciding whether an individual death was preventable but has inherent difficulties such as inter-rater reliability.10

Methodological uncertainties
We have divided the uncertainties into those relating to the numerator, denominator, risk modelling, interpretation, and coding.

Numerator
Most hospital administrative databases capture only deaths that occur in hospital. The choice is then between including all in-hospital deaths or only those that occur within a set number of days since admission. Inclusion of all in-hospital deaths will capture long stay patients, perhaps with more chronic disease or complications from...
Although linking these transfers is desirable, some administrative systems do not allow this, in which case transfers can affect a unit’s estimated performance.²¹ Linkage of hospital admissions to death registrations is possible in some countries, but in England currently incurs a considerable time lag, limiting its utility.

Denominator Hospital administrative data generally use the International Classification of Diseases (ICD) to code diagnostic information, which is typically divided into the primary diagnosis
(main problem treated) and various secondary diagnoses (including comorbidities and complications). The Agency for Healthcare Research and Quality’s Clinical Classification System is one way of grouping ICD codes, and is intended to be clinically meaningful for health services research, but other groupings exist. The HSMR is based on admissions with a primary diagnosis belonging to one of the Clinical Classification System groups that cover a combined total of 80% of in-hospital deaths. In England, 56 of the 259 groups achieve this, but this varies by country. 80% is chosen because of the Pareto principle (80% of the effects come from 20% of the causes). All diagnoses could be included, though, with simplified risk models in groups with few deaths.

It could be argued that palliative care and not-for-resuscitation patients (some US but not UK data can identify the latter) should be excluded, providing that this was based on intention to treat on admission. However, if the HSMRs are used for judgment (by the regulator or in a pay for performance scheme) this creates the potential for gaming.

Another approach would seek diagnoses for which it is recognised that mortality is one of the most useful markers of quality of care, ideally with documented variations between hospitals. The Agency for Healthcare Research and Quality produced a patient safety indicator, death in diagnosis related groups with low mortality (<0.5%). Death in patients with these conditions would be considered unusual and hence these might represent preventable deaths. This indicator will of course have a small numerator, hampering inter-hospital comparisons, but would be practical for clinical audit. However, this indicator would exclude frail and elderly people, who are most vulnerable to deficiencies in care.

A further consideration is whether to count patients or admissions. Administrative databases count admissions or discharges. In England, the basic units are “finished consultant episodes” (time spent under the care of a given senior doctor), which need to be linked to form admissions. In 2008-9, 15.5% of the overall total and 28.7% of HSMR admissions had more than one consultant episode. As each episode can contain different diagnostic information, this raises the question of which to use. In some hospitals, the first episode can be short, covering a preadmission or observation ward. The primary diagnosis may be simply “chest pain” or “abdominal pain.” Diagnoses in subsequent episodes may represent complications rather than the reason for admission, and we would argue the diagnosis recorded in earlier episodes is preferable for monitoring. However, none of these episode diagnoses may equal the cause of death, which may also be of interest.

HSMRs use the first episode (or second if the first has only a symptom code as its primary diagnosis). An extension of this multiphase phenomenon is the admission consisting of one or more hospital transfers (called a superspell). In the UK the ability to capture and link all the transfers varies regionally, resulting in the “loss” of an unknown number of deaths, which also varies by diagnosis. With a patient based measure, however, assuming a suitable patient identifier exists in the dataset, we also need to decide which admission for chronic obstructive pulmonary disease, for example, to include for each patient. Options include the first, last, or randomly chosen admission. However, since each admission is an opportunity for the hospital to save the patient’s life, it could be argued that all admissions should be included. HSMRs count all admissions, partly for practical reasons, with an adjustment for each patient’s number of unplanned admissions within the previous 12 months. This clearly imperfect method tries to take some account of hard to quantify factors such as disease severity and admission thresholds. Another option would be to exclude chronic conditions liable to repeat admissions. Limiting the set of diagnoses to first time events—for example, acute myocardial infarction and stroke—would minimise this problem but be less inclusive.

As well as varying admission thresholds, the definition of what constitutes an admission as opposed to an emergency department attendance or time spent in an assessment unit may change or differ between hospitals or countries. The proportion of patients in England admitted and discharged on the same day has shown a large, steady increase, from 5.9% of all inpatients in 1996-7 to 15.4% in 2008-9. We could therefore exclude these records as not “proper” admissions, but poor emergency care can result in death and therefore any deaths in these admissions should not be excluded. HSMRs currently include all unplanned as well as planned inpatient admissions.

Risk modelling
The expected deaths in an HSMR are derived from sets of logistic regression models, one for each diagnosis group, that include available case mix factors. An ideal risk adjustment model would capture all important patient factors and fully adjust for them. Two commonly used proxies for health status are socioeconomic deprivation and age. These can be problematic if, for example, the deprivation score captures factors related to life expectancy such as smoking and diet better in some areas of the country than in others. Adjusting for deprivation may also remove some of the effect of quality of care if one hospital is less able to deal with the needs of its disadvantaged patients than another hospital. A few studies have combined patient administrative data with information from laboratory systems and found that a few variables such as serum creatinine can improve the model fit, but unfortunately doing this often remains technologically difficult.

The surgery a patient had can give useful risk information not captured elsewhere, but adjustment for procedure is not straightforward and is not done with HSMRs. The degree to which the surgery reflects the surgeon’s choice (which relates to quality of care and should not be adjusted for) will vary and may be hard to ascertain. We would not recommend using systems of grouping patients for billing purposes such as diagnosis related groups or healthcare resource groups in the risk models, as they are based on the treatment given and also include complications, which again partly relate to quality. Furthermore, redo or revision procedures often involve greater risk, but if they were as a result of difficulties during the original procedure, adjusting for the revision will obscure this quality element.

Other factors outside the control of the hospital but not the healthcare system include the provision of community services and the proportion of deaths occurring in the community. Jarman et al originally found factors such as the number of general practitioners and NHS facilities per head of population to be statistically significant (although not necessarily important) explanatory factors of in-hospital mortality.4
Investigations into a high HSMR may show problems that lie beyond the hospital.

In-hospital case fatality rates have been falling in England for several years, partly reflecting a fall in total mortality (5.4% reduction in females and 4.0% in males in 2009 compared with 2008) but also potentially due to an increase in total admissions (inflating the denominator) and coding practice changes. The HSMR risk models therefore include adjustment for financial year, meaning that hospitals are compared with the national average for the relevant year. This recalibration is done annually, but in the second half of the year, the continuing fall in mortality means that hospitals will typically have apparently falling HSMRs. More frequent recalibration than annual is possible given adequate resources but, for less common diagnoses, can be impeded by small numbers.

All hospitals are used to derive the predicted risks, but an alternative would be to exclude “atypical” trusts, as is done in the Netherlands, for example. Hospitals could be defined as “atypical” according to case mix, data quality, or even performance. Around 94% of the admissions and deaths used in HSMRs are to acute non-specialist trusts, so the numerical difference is likely to be small.

**Interpretation**

HSMRs are typically published annually, but this may not be the most appropriate timeframe for monitoring. Quarterly and monthly figures will be timelier and may detect changes that an annual summary may miss, but are more subject to chance fluctuations, which can be considerable. A hospital will not have the same number of admissions in each diagnostic group every month, and case mix adjustment is more successful in some groups than others. Consequently, temporal variation in patient mix can affect the HSMR.

Alternatively, hospitals may want to track the progress of their HSMR over time after implementing improvement efforts rather than compare themselves with the contemporaneous national average. To do this, they base all their expected deaths on some previous year so their performance is relative to that year instead of the current one. This also reduces the interpretation problems with diagnostic coding varying between hospitals.

“Unacceptable” or outlying performance needs to be defined. With nearly 150 acute non-specialist hospital trusts in England and many more in other countries, the type I error rate with 95% confidence intervals is not negligible. Funnel plots with 99.8% control limits are therefore increasingly common. With simple random variation in the HSMRs, just 0.2% of hospitals with average rates would be expected to lie outside the control limits just by chance. However, 59 (40%) out of 149 English HSMRs lay outside the 99.8% control limits in 2008-9. If the intention is to detect outliers, the control limits may be widened to adjust for the extra variation.

Unfortunately, it is impossible to tell from the plot whether the greater than random variation is due to signal (differences in quality of care) or noise.

**Coping**

The accuracy of diagnostic coding is clearly vital and is the direct responsibility of the hospital administration. UK clinical coders are checked regularly by the Audit Commission. In 2001 Campbell et al investigated the accuracy of hospital data in the UK through systematic review of studies comparing routinely collected data with case note review. Median coding accuracy rates for primary diagnostic codes were 91% in England or Wales and 82% in Scottish studies. These figures will have improved since that study but still vary by hospital.

Patients transferred from another hospital can be very ill, but it can be hard to capture their high risk using administrative data. We now adjust for “source of admission,” which includes from home or other hospitals, though anecdotal evidence suggests that this field is not well coded.

Secondary diagnosis coding in hospital data includes comorbidities and is likely to vary between hospitals more than primary diagnosis coding, though in elderly patients with multiple problems deciding which should be the primary can be hard. The possibility of distinguishing conditions present on admission from complications developed during the admission exists only in some countries. Several comorbidity indices (the HSMR uses Charlson, for example) have been developed for administrative systems that lack flags for conditions present at admission and try to include only chronic conditions that could not be developed during the hospital stay. Adjustment for comorbidity in risk models is advisable but can incur measurement error. However, the discovery that a hospital’s HSMR is artificially high because of poor coding can be a spur to improve recording, which may have other benefits for that hospital from reimbursement systems. Some specific theoretical concerns have been raised about some of the information used in our risk model, including comorbidities, but it is important to see how influential they are in practice. England now has a list of diagnoses that are mandatory to code for on patient records. The list includes many of the diagnoses used in our comorbidity index, and this should improve coding consistency.

The Dr Foster Unit is affiliated with the Centre for Patient Safety and Service Quality at Imperial College Healthcare NHS Trust, which is funded by the National Institute of Health Research. We are grateful for support from the NIHR Biomedical Research Centre funding scheme. We also thank Liz Robb and Nick Flatt for their helpful comments.

**Contributors:** AB and PA conceived the study, AB performed all analyses and wrote the first draft. All authors contributed in the revision of the manuscript. AB is the guarantor.

**Competing interests:** All authors have completed the unified competing interest form at www.icmje.org/ico_disclosure.pdf (available on request from the corresponding author) and declare no support from any organisation for the submitted work.

The Dr Foster Unit is funded by the National Institute of Health Research. We are grateful for support from the NIHR Biomedical Research Centre funding scheme. We also thank Liz Robb and Nick Flatt for their helpful comments.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

When I use a word

Formularies and pharmacopoeias

The British National Formulary (BNF), the publication of whose 60th issue we have just celebrated,¹ and the BNF for Children (BNFC), now six issues old, are shining examples of what modern reference books should be. Which is not surprising—“formulary” comes from the Indo-European root MERBH, meaning to shine (as in Morpho, a genus of nymphalid butterflies with shiny blue wings, so named in 1807 by the Danish zoologist Johan Christian Fabricius).

Because what shines is easily seen, MERBH also meant to appear or take shape. The Greek god Morpheus, after whom morphine is named, conjured shapes in men’s minds while they slept. Aphrodite is supposedly named from aphros, the foam of the sea, whence she was born, as Botticelli (The Birth of Venus, 1485) and Titian (Venus Anadyomene, 1525), among others, depicted. But “Aphro” also looks like Morpho metamorphised, and Morpho is what the Spartans called shapely Aphrodite.

Metathesis certainly morphed the Greek “morphe” into the Latin “forma” (appearance) and its diminutive, “formula” (shape). Formosus meant full of shape and therefore beautiful; hence “[Insula] Formosa,” the original Portuguese name for Taiwan.

“Formula” has many meanings in English, but as a pharmaceutical term it was defined by John Kersey, in his 1706 edition of Edward Phillips’ dictionary The New World of English Words, as “a Physician’s Prescription or Bill appointing Medicines to be prepared by an Apothecary.” In other words, a formula describes how a medicinal product should be shaped. And a formulary lists such formulae and describes the formulations (pills, potions, and powders) in which the apothecary compounds them.

The word “formulary” was first used to describe Charles Thomas Haden’s translation of François Magendie’s Formulaires pour la préparation et l’emploi de plusieurs nouveaux médicaments, tels que la noix vomique, la morphine, l’acide prussique, la strychnine, la vénérinaire, les acaulis des quinquinaux, l’iode etc, more prosaically published in 1823 as Formulary, for the preparation and mode of employing several new remedies.

Older than “formulary” is its exact synonym, “pharmacopoeia,” from φαρμακοποιείον (pharmacopoeia, literally “drug-making”) found in the post-classical (Hellenistic) dialect called Koine. Instances in mediaeval Latin include the titles of books such as Pharmaco-Pinax, medicamentorum omnium, quae hodie ad publica medentium munia in officinis extant by Anutius Foesius (Basel, 1561).

“Pharmacopoeia” entered English at the start of the 17th century, as did “pharmacopinax.” “Pinaix,” the Greek word for a writing tablet, appears in several Latin titles, such as Pinax iconici us antiquorum ac variorum in sepolitur ritum by Lillius Gregorius Giraldus (Lyons, 1556), Pinax theatri botanici by Caspar Bauhin (Basel, 1623), and Pinax rerum naturalium Britannicarum, by Christopher Merrett (London, 1666). However, the only English pharmacopoeial instance of which I am aware is Pharmaco-Pinax, or a Table and Table of the Preys of all Usual Medicaments, Simple and composed, contained in D. Gordon’s Apothecaries and Chymicall Shop, published in Aberdeen in 1625. A tablet of tablets, you might say.

Terms such as “dispensatory” and “receipt-book” (that is, recipe book) were also introduced in the 16th and 17th centuries, but the earliest English word for a medical treatise that described treatments was the Anglo-Saxon “laecbeboc.” And laecbeckos contained laecedoms, leechdoms, remedies or prescriptions. The best known, Bald’s Laecceboc, was probably compiled in the early 10th century, soon after the death of King Alfred.² Perhaps we should rename the BNF the British National Leech Book.

Jeff Aronson clinical pharmacologist, Oxford
jeffrey.aronson@dirinpharm.co.uk


Cite this as: BMJ 2011;342:d34