MORTALITY MEASUREMENT: THE CASE IN FAVOUR.
This report was co-authored by:

ROGER TAYLOR
Co-founder and Director of Research at Dr Foster.

DR PAUL AYLIN
Clinical Reader in Epidemiology and Public Health.
Co-Director of the Dr Foster Unit at Imperial College London.

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.

About the Dr Foster Unit at Imperial College

The Dr Foster Unit at Imperial was established in 2002 and resides within the Department of Primary Care and Public Health in the School of Public Health, part of the Faculty of Medicine at Imperial College London. Its expertise lies in making the most out of routinely collected health data. The NHS has been said to be awash with data but short on information. The unit uses routinely collected clinical and administrative data to examine variations in quality and safety in healthcare. In developing innovative statistical and computational methods for processing large data sets derived from electronic medical records and NHS databases, they have encouraged the use of data in the management and monitoring of healthcare in the UK and internationally. The identification of significant variation in process and outcome measures using validated indicators, is the vital first step in the quality improvement efforts of NHS managers and clinicians. The Unit is funded through a grant from Dr Foster and also receives funding from NIHR through the Patient Safety Translational Research Centre and other research projects.

For more information about Dr Foster or this report, please contact: info@drfoster.com or call 0207 332 8800
Executive Summary

This document has been produced by Roger Taylor at Dr Foster and Dr Paul Aylin at the Dr Foster Unit at Imperial College in response to criticisms of the use of risk-adjusted mortality measures. Our message is that measurement of quality in healthcare is a complex undertaking. No single indicator will ever adequately summarise the safety and quality of an organisation. Recent polarised arguments proposing one measure over another are misleading and unhelpful. We must learn to accept a multidimensional approach to measuring healthcare if we are to have any hope of monitoring safety and improving quality. The key points made by the document are:

1. There is good evidence of a link between risk-adjusted mortality measures and other measures of quality. Claims that mortality measures do not correlate with either process measures of quality or measures of preventable mortality based on case note reviews are based on selective quoting from the literature.

2. Process measures do not provide an alternative to measuring outcomes as a method of understanding quality in healthcare. Both types of metrics need to be used, not least as the value of process metrics is only established by demonstrating that they are contributing to a better outcome.

3. Preventable mortality measures based on case note reviews are not a more reliable measure of quality of care than risk-adjusted mortality measures. They have a number of aspects that make them inherently unreliable. Also, they fail to capture many aspects of quality of care that can lead to patient deaths. None the less, they are a valuable and useful tool for learning and quality improvement.

4. The signal-to-noise ratio of hospital mortality ratios is unknown but there is strong evidence that it is high enough to provide an extremely useful guide to variations in quality of care. Models showing low signal-to-noise ratios are based on assumptions that underestimate the impact of quality on mortality. Furthermore, such estimates ignore the use of multiple metrics and analyses at the level of the organisation, diagnosis and procedure.

5. Gaming of mortality rates is a problem that needs to be addressed by improving data quality controls. Arguing that mortality rates should be abandoned because some organisations have been gaming is equivalent to arguing that financial accounting rules should be abandoned because some companies are fraudulent. It only makes the problem worse.

6. Clinical audit data do not necessarily provide a more reliable basis for understanding quality of care. Much clinical audit data is too incomplete to be reliable. Most audit data also fail to capture the patient journey through the healthcare system by focussing on a single episode of treatment.

7. Healthcare providers with high-mortality ratios should regard them as a warning of potential problems of care. Organisations that dismiss them on the grounds that they do not accord with their experience of the care they provide, risk perpetuating poor quality care.

In the introduction to the document, we suggest that arguments about metrics are in some cases proxy wars for more 'political' disputes about clinical control of data, clinical accountability to management and the role of transparency in healthcare. These are important and legitimate debates, but they should not be allowed to colour arguments about differing approaches to measuring quality. The technical aspects of measuring quality accurately in healthcare is a hard enough problem to fix without the distractions that arise when it gets caught up in ideological arguments.
Mortality measurement: the case in favour

Index

Introduction and observations........................................................................................................................................ 2
Response to arguments made in the media against hospital mortality rates......................................................... 4
  Claim: there is no association between mortality rates and other measures of quality of care......................... 4
  Claim: preventable deaths is a more reliable measure than standardised mortality rates............................ 10
  Claim: the signal is too weak...................................................................................................................................... 12
  Claim: you can manipulate the indicator by manipulating data............................................................................ 13
  Claim: clinical audit data are better.......................................................................................................................... 13
  Claim: one large teaching hospital has a high mortality rate................................................................................ 14
Introduction and observations

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 ratios or mortality rates\(^1\) as a metric.

This paper reviews these arguments and explains why they are flawed. But before we come on to the scientific and statistical questions about the accuracy of different approaches to measurement, we should first consider some of the broader concerns that have been raised.

First, is too much emphasis put on mortality rates, particularly in the media, when they are just one indicator amongst many? This can happen – partly because of the lack of other metrics to complete our understanding. Too often metrics, such as mortality rates or patient satisfaction rates, are looked at in isolation. They should always be considered in context if they are to be understood. We would not claim that mortality rates are more important than other indicators but equally we would not agree that they should be ignored in favour of other metrics.

Secondly, there is the worry that imperfect indicators will unfairly penalise hospitals, wrongly identifying some organisations as worse than they truly are, and others as better. We know that this will inevitably occur with any metric from any data source. However, this does not justify ignoring the information. If data can increase our ability to identify poor and high quality care so that overall, patients receive better care, we must use them even if we know false positives and false negatives will occur. At the same time, it is important to always acknowledge the defects of the information we work with.

Thirdly, there are concerns about who controls the data. One of the arguments against some mortality measures is that they are derived from administrative data when clinical audit data is more reliable. Later in this document we demonstrate that this is often not true. However, there is another difference between clinical audit and administrative data – who controls these data. Administrative data are controlled by managers, audit data by clinicians.

This divide will not survive the move to electronic health records. EHRs will provide a consistent underlying data source that will inform not just doctors and patients in decisions about care, but also, managers and researchers in decisions about the effectiveness of systems and treatments. The idea of separate islands of ‘management’ and ‘clinical’ data will become irrelevant.

Finally, there is the question of whether measurement of processes is better than the measurement of mortality or outcomes, and should we therefore focus on the former? The issue here is less about the technical merits of one measure over another – it is about the experiences of people working with different types of data. People find outcome measures frustrating because it is not clear what causes variation and it is hard to determine what action people should take in response to poor outcome measures. In contrast, process measures are easier to interpret and act

\(^1\) Mortality measurement can be expressed in two ways. Mortality ratios are the ratio between the observed mortality and the expected mortality rate. By tradition this is multiplied by one hundred and is expressed as a value usually between 50 and 150. Mortality rates are the percentage of patients who die. This can also be expressed as an adjusted percentage to reflect the expected mortality rate. The terms are often used interchangeably.
on. However, it is wrong to conclude from this that it is better to focus on process measures alone. No system of process metrics will fully capture quality. The selection of limited sets of process metrics can encourage gaming rather than genuine quality improvement, and unmeasured parts of a process then suffer while improvements are made to those aspects that are being measured.

Defining quality purely in terms of process creates bureaucratised mechanisms of quality control that prevent innovation and improvement. The most successful approaches to quality improvement combine outcome measures and process measures in order to identify those aspects of the process that are important to the outcome, and those that make no difference and could be avoided.

Overleaf we summarise the technical arguments against the use of mortality rates and outline the problems with these arguments. Our aim is not to argue that measurement of mortality data is better than other measures; only that they are complementary to our understanding of quality. We must consider the totality of the evidence (mortality rates, other outcome and process measures, case note audit data and inspections) together when judging a hospital’s safety and quality of care.

We also recognise that debates about different approaches to measurement are sometimes proxies for arguments about other issues such as the role of transparency, accountability and professional autonomy in healthcare. These are important matters. We hope that coming to a common understanding on these difficult issues will help to resolve some of the disagreements over the best methods of measuring quality.
Response to arguments made in the media against hospital mortality rates

1. **Claim: 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:

- risk-adjusted mortality ratios, in which the number of patient deaths is compared with the number we would expect from the patients being treated;
- preventable 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:


The table below comments on each of these:

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of participants</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best 1994</td>
<td>222</td>
<td>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.</td>
</tr>
<tr>
<td>Gibbs 2001</td>
<td>739</td>
<td>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”. Preventable mortality was one of these measures but the relationship was not significant (P = 0.13).</td>
</tr>
</tbody>
</table>
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.’


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 ‘preventable deaths’ measures and standardised mortality rates.2 As with earlier studies, the paper acknowledges that it looks at too

---

few patients to be able to establish a connection between preventable hospital deaths identified through case note review 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 “preventable 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 preventable 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.3 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 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 8 of the report.

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.

---

3 Pitches D, Mohammed MA, Lilford J. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. BMC Health Services Research 2007;7:91
doi:10.1186/1472-6963-7-91
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).4 The same has been found in the UK.5

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

---


As discussed on page 6, 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’

‘There is some evidence that targeting hospitals with consistently high death rates over periods longer than one year may better identify potential quality problems’

‘Low-mortality hospitals start aspirin within six hours of presentation more often than intermediate or high-mortality hospitals’

‘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’

‘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)’

‘There was evidence for a relationship between some process variables and outcomes at hospital discharge, but the relationships were generally weak’

Teaching hospitals had a higher ICU death rate – ‘More interventions do not necessarily improve outcomes in critically ill patients’

‘The results of this project suggest that there are substantial opportunities to improve the quality of care for CABG in Oklahoma’

‘Ratings of overall quality of care did not differ significantly between patients from hospitals with higher and lower than expected mortality and morbidity’
Box B: **List of recent papers (published since 2006) showing an association between mortality and other quality metrics** (as discussed on page 5)


2. Claim: preventable 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 a ‘preventable deaths’ measure is inherently more reliable (a gold standard). A ‘preventable 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 a ‘preventable deaths’ measure will fail to identify important aspects of quality of care.

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

Most studies attempting to identify whether or not a death was preventable find only moderate levels of agreement between reviewers. 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 preventable deaths show wide variations in the levels of deaths regarded as preventable.

The higher rates of reliability are achieved by setting strict criteria for the reviewers as to what should be considered preventable. 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 preventable death were listed, but there were still events that clearly qualified as avoidable deaths which did not fall within these defined categories.

2. Evidence of bias in assessment of preventability 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 preventable 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 a preventable death. This is counter-intuitive because these patients are the most susceptible to errors or omissions in treatment. As Hogan et al. 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 preventable 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 preventable deaths.

---

3. Limitations of ‘avoidability’ as a concept

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

As an example, Brooke et al.\textsuperscript{10} 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 preventable, simply because they did not receive peri-operative β-blockers.

Delays in treatment for a heart attack increase the likelihood of death. McNamara et al.\textsuperscript{11} 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 with 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 ‘preventable’ deaths from a case note review, but the change in the process of care resulted in more patients surviving\textsuperscript{12}.

For these reasons it is wrong to assume that any lack of correlation between risk-adjusted mortality rates and case-note reviews of preventability 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 preventability of death.


\textsuperscript{12} Presentation from Peter Glennon, Consultant Cardiologist University Hospitals Coventry and Warwickshire NHS Trust. Global Comparators Conference, Boston 2012.
3. **Claim: 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 preventable deaths to account for the degree of variation in HSMRs. As Professor Black put it: 'only' 5% of deaths are preventable 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 \(^{13}\) 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 preventable mortality that, as discussed above, will underestimate the level of mortality attributable to variations in the quality of care. In effect, the 'low PPV' argument is simply restating the argument that 'preventable 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 ratios 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 ratios. As Don Berwick put it:

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


4. **Claim: 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 ratio 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.

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, 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 measures 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.

5. **Claim: 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 14, came to this conclusion, as have studies in the UK 15 16 17.

Other studies have shown that failure to report data to clinical systems may be biased towards under-reporting of poor outcomes.18 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.

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

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 ratios in the country. University Hospitals Birmingham NHS Foundation Trust is the only one of England’s large teaching hospitals that has a high mortality ratio 19. This difference should be regarded as a signal for more detailed investigation, not as a reason to argue against mortality measures.

---


19 Dr Foster Hospital Guide 2013. Large teaching hospitals are defined as members of the Shelford group of NHS trusts.