Author’s response to reviews

Title: The most dangerous hospital or the most dangerous equation?

Authors:

Yu-Kang Tu (y.k.tu@leeds.ac.uk)
Mark Gilthorpe (m.s.gilthorpe@leeds.ac.uk)

Version: 2 Date: 30 July 2007

Author’s response to reviews: see over
Dear Editor,

Re: MS1654892127147596: The most dangerous hospital or the most dangerous equation?

Many thanks for forwarding us the three reviewers’ very helpful comments on our manuscript. We have addressed the concerns raised by your reviewers and made extensive revisions according to reviewers’ comments. The length of the manuscript has been substantially expanded by including detailed explanation of the relationship between the variations in mortality rates and hospital sizes. The examples and simulations have been revised and the errors pointed out by the reviewers have been corrected. New data and a new figure have been added to help explain our arguments. Below, we set out referees’ comments in italics and our response in normal text. All changes in the manuscript are highlighted in blue.

Reviewer One’s comments:

Reviewer: Jerod Loeb

General
1. To begin, I was really intrigued by the title of this paper; however, I'm afraid I found rather little that is new in this short rumination on statistical methodology. I should note immediately that I am not a biostatistician. The observation that hospitals with smaller sample sizes (although the authors do not refer to the sample size directly but a surrogate of the sample size, namely the number of hospital beds) have more variability in measure rates is well known.

We agree with Dr Loeb that it has been known for some time now in the literature that small hospitals/Trust have greater variations in their performance indicators. However, to the best of our knowledge, it has not been explicitly explained in the literature regarding the origins of this variation and the role of random processes from an underlying statistical viewpoint. We hope this brings clarity as well as explanation to many less statistically minded researchers – with clarity, perhaps the message will not be forgotten and even transferable to other similar settings.

2. Perhaps more importantly, one might wonder whether the authors should have focused on actual mortality associated with a specific procedure (i.e., CABG as in the hypothetical construct in the paper) and its association with case volume (rather than global mortality and bed size alone).

Our general arguments apply to all performance indicators. Moreover, the intention of our article is to present a statistical viewpoint. Discussion of a specific procedure is an empirical question that is not the focus of our article. However, we now do use empirical data in the Table 1 and Figure 2 to illustrate our arguments.

3. We can only hope that no one appropriately using statistical methodology would make a decision about whether a hospital is a high or low performer based only on the hospital’s overall mortality rank; the standard deviation and confidence interval need to be taken into account as well.
We agree, and we now included discussion and citation of alternative methods such as quality control charts and funnel plots.

4. Even better would be to use a statistical methodology that incorporates the sample size into a measure of the hospital rate, such as hierarchical methods, to determine the statistical significance of the HSMR. If the Dr. Foster report only reported the HSMR with no measure of variation or statistical significance, then the observation that the results could be misinterpreted is valid, but perhaps more appropriate as a letter to the editor rather than a paper.

Funnel plots and quality control charts take sample sizes into account but do not give the rankings. Dr Forster intelligence’s report presents funnel’s plots, but interestingly it also presents league tables.

Specific Comments
1. I found the abstract very confusing. It appears that the authors meant that the true mortality rates are independent of hospital sizes in the second sentence from the end, but this is a very strong assumption. I think the authors meant to point out that the variation is proportional to the square root of the sample size in the fourth sentence, as they assume in the paper (without any justification), but this isn't made clear here.

There was a typing error that has been corrected now, and the abstract has been revised to make it clearer.

2. A 352 bed hospital small would not be considered small by US standards. More appropriately, the actual sample size of the measure should be used to evaluate the potential variability of a hospital's rate, not a surrogate such as hospital beds. Sample sizes of this magnitude would not be expected to vary appreciably, although only if all the hospitals actually had the same rate and the same patient mix (again a strong assumption) would this be expected to misidentify a hospital as being an outlier. The assumption is implicitly made that the patient populations are similar across hospitals. No mention is made of the risk-adjustment methodology that the Dr. Foster report used and its potential influence in identifying poor or good performing hospitals.

A 352 bed hospital being small is relative to the size of other hospitals in England. In any event, we now use hospital admissions as the surrogate instead of using beds. Many factors influence performance indicators and we have now included a brief discussion of this in the text. However, our key message is that even were all these factors held constant (i.e. adjusted for), variation in mortality rates is still greater for small Trusts due to random variation. These adjustments only affect the actual number of mortality rates for individual hospitals, not the variation in mortality rates across hospitals.
Reviewer Two's comments:

**Reviewer:** Mohammed Amin Mohammed

1. The manuscript falls between two stools – it falls short of being a full paper (because it lacks sufficient data) yet it is too long as a letter. To develop the manuscript into a paper the authors need to carry out an empirical study of the relationship league table ranks and sample size in healthcare. This would make a very interesting paper and will likely involve several case-studies. To develop the manuscript into a letter, the authors need to edit out aspects of the text which are not central to a single message – that sample size influences variability. Arguably the authors intend neither a paper nor a letter – but instead a statistical note.

We now included an analysis of empirical data between Hospital Standardised Mortality Ratio (HMSR) obtained from Dr Foster’s report “How healthy is your hospital” and the number of hospital admissions (used as a proxy for the sizes of NHS Trusts) obtained from Hospital Episode Statistics. We also included more statistical explanation on the relationship between variation in the number of deaths and hospitals/Trusts sizes. Our article can now be read as an ‘educational paper’ with an empirical content.

2. Importantly, the manuscript implies that ranking exercises can be “fixed” to take account of sample size and policy makers can be protected (how? This is not explained) from misinterpreting league tables, where as Deming and others (See Lancet 2001; 357:463-7 ; BMJ 2002; 324: 95-98…..) have clearly argued that ranking has fundamental flaws which cannot be “fixed”. Also, the use of control charts (also known as funnel plots –see Statist. Med. 2005; 24:1185–1202) as a way of protecting misinterpretation and accounting for differences in sample sizes has been omitted from the discussions, despite the inclusion of figure 1.

We agree with Dr Mohammed that it is misleading to use league tables to rank the performance of hospitals, and this is why we wrote this article. We now mention control charts and funnel plots as better alternatives to compare hospital performance. New references are cited in the text.

3. The authors concluding line in the abstract is that this statistical phenomenon needs to be taken account of especially with regards to policy, but don’t specify what this means exactly (even in the discussion).

We now state in the abstract that this statistical phenomenon needs to be taken account of “in the comparison of hospitals/Trusts performance”. A new figure on the relationship between HSMR and the number of hospital admissions has been added, and in Table 1 the number of beds has been replaced by the number of admissions. The concluding paragraphs have been totally revised to make it clear that our focus is on the relationship between the size of hospitals/Trusts and variations in mortality rates.

4. In their hypothetical explanation of why size matters, the authors consider cross-sectional data (ie comparing hospitals) but then introduce a time based dimension to the illustration and then finally revert back to a cross-sectional comparison. This seems inconsistent.
Table 1 has been changed into compare the HSMR in 2005/6 in order to make it consistent.

5. It is not clear why the authors consider +/-2 as the bounds on the observed number of deaths? I wonder if a funnel plot with illustrative data is not better suited to making a point which is easy to understand once you can visualise the data. Also, for an illustrative note, the authors need to explain every step of the calculations – how the +/-2 is derived and how the +/-% are derived. In comparing the two sets of numbers for the notional hospitals it is not explicitly clear how the calculations used to derive the final %+/- differ. This is important if the novice reader is to follow the construction of the argument.

Many thanks for pointing this out. We have extensively revised and expanded this whole section. It is now hopefully clearer that +/- mean plus and minus standard deviations and hence the range is a 95% confidence interval. The numbers are revised and derived from the binomial distribution. A new section is added to explain and illustrate how to use the binomial distribution to obtain the standard error of the mean and hence the 95% confidence interval for variations in the number of deaths and mortality rates.

5. The text related to Figure 1 is confusing, because until now the reader is weaned on the importance of the sample size (and its square root) only to find the “expected” mortality as a proxy for sample size. The link between expected mortality and sample size has not been made satisfactorily. Furthermore the labelling on figure 1 is inadequate in respect of the additional lines.

The new section on the binomial distribution before Figure 1 provides information for readers to understand the Figure 1. Explanation for the upper and lower confidence limits in Figure 1 has now been given in the text.

6. Table 1 is useful and provides empirical support for the message. However would a plot of SMR Rank vs Hospital Size be useful? Furthermore the authors state that smaller hospitals are over-represented in table 1 without explicitly stating their rationale.

Many thanks for this suggestion. In Table 1, we now use hospital admissions of NHS Trusts in 2005/6 as a proxy for the sizes of the 152 NHS Trusts, for which HSMR is listed in Dr Forster’s report instead of our using the number of beds. To justify our conclusion on the over-presentation of small Trusts, we create Figure 2, using data from Dr Foster’s report and hospital admissions from Hospital Episode Statistics.

Minor Essential Revisions
1. Abstract. The line “From a statistical point of view….” does not make sense – is there a typing error here?

Sorry, there was a typing error that has been corrected now.
2. The authors say that a “close examination of the information” reveals a pattern consistent with a well known statistical phenomenon, but they present no result under the abstract.

At the end of that sentence, we now add “the variations in the performance indicators are greater in small hospitals and smaller in large hospitals”.

3. “We do not know the underlying processes operating in each hospital that may affect a hospital’s mortality rate; for instance, the number of deaths in a hospital fluctuates more for larger hospitals than for smaller hospitals. “

The above sentence appears to contradict the thrust of the paper, and in any case, may be removed without losing the key message.

It has been removed now.

4. The next sentence simply states the fact that the variance is proportional to the square root of the sample size, but the explanation is left to later. This jump in the text is not helpful.

The whole section has been revised to make it easier to follow.

5. The word hospitals should not be interchanged with Trust.

We are more careful about the wording now, though for the benefit of the international audience, emphasis on Trusts (as in the UK) may be more confusing than the emphasis on hospitals, so the distinction is made up front that Trusts manage hospitals and we use hospitals or ‘hospital Trusts’ where appropriate.

Reviewer Three’s comments:

Reviewer: Robert Gibberd

1. This paper discusses the problem of reporting hospital mortality rates when not taking into account random variation. The main point in the paper is to emphasise that the random variation of rates is proportional to the inverse of the square root of the sample size. This is fairly well known to the journal readers, and as such does not provide any new insights. In particular, there are papers that attempt to show how reporting can be improved (eg Gibberd reports on the UK mortality data and how it could be reported), by either using 95% confidence intervals, or Bayesian shrinkage estimates.

As de Moivre’s gave his famous formula 300 years, we do not in our article claim any originality in showing the inverse association between random variations and the square root of sample sizes. We are aware of considerable debate and discussion surrounding the use of league tables as performance indicators, using for instance mortality rates to compare the performance of hospitals/Trusts. We are also aware of control charts and funnel plots and that these have been proposed as alternatives methods. References to this effect are now cited in the text. However, judging from the use of league tables in Dr Foster intelligence’s report “How healthy is your hospital” written by prominent UK
academics, we believe that it sadly remains very necessary to remind many readers of this journal, and indeed the whole research community at large, that there is this persistent issue of the inverse association between random variation and the square root of sample size – many researchers repeatedly overlook this statistical phenomenon. Our article has been extensively revised and expanded to address our intentions more clearly whilst reflecting briefly on the wider issues associated with league tables.

2. The paper also gives examples that are misleading: page 4 assumes that with expected number of deaths, assume that the variation is 10 +/- 2. However, the standard deviation for a Poisson process with mean 10 is square root of 10, or over 3. Thus, 95% of observations would be 10 +/- 6.

Many thanks for pointing this out. We have revised the whole example and simulation. The standard deviation is now derived from the binomial distribution.

3. Table 1 gives the better and poorer hospitals and the number of beds, but the average number of the remaining hospitals should be reported, in order to see that the smaller units have the extreme values. Plotting all the data instead of the simulation as in Figure 1 would be more relevant.

A new figure (Figure 2) now shows the relationship between Hospital Standardised Mortality Ratio (HSMR) from Dr Foster Intelligence’s report and hospital admissions in NHS Trusts (used as a proxy for the patients treated) from Hospital Episode Statistics. It is obvious that the variation in HSMR is much greater in small Trusts than in large Trusts.

We hope that our replies and the revisions to the manuscript will meet with the approval of your reviewers. Both authors have approved the final content of the enclosed manuscript.

Thank you for considering our article for publication in your Journal.

Yours sincerely,

Yu-Kang Tu                              Mark S Gilthorpe