Reviewer's report

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

Version: 1 Date: 27 June 2007

Reviewer: Mohammed Amin Mohammed

Reviewer's report:

General
The authors present an interesting, short paper, which highlights the (historically well known) influence of sample size in the ranking of hospitals, arguing that smaller hospitals are more likely to appear at extremes of a league table because of increased variability (due to smaller sample sizes).

Whilst I agree with the authors that this phenomenon, although well established is not well known and its reiteration (like that of regression to the mean) is useful and worthwhile, there are some important limitations in the paper which require attention.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

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. As later comments show, there are issues which need to be addressed before the note is sufficiently clear (especially to non-statistical readers).

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.

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).
Why sizes matters

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.

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.

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.

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.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

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

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.

“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.

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 word hospitals should not be interchanged with Trust.

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that i have no competing interests