Reviewer's report

Title: Should policy-makers and managers trust PSI? An empirical validation study of five Patient Safety Indicators in a National Health Service

Version: 1 Date: 25 November 2011

Reviewer: Alex Bottle

Reviewer's report:

General remarks

This study uses appropriate statistical measures to look at inter-hospital variation in rates of five PSIs in Spain, trying to differentiate between random and systematic variation and patient versus hospital variation. There are a large number of inaccuracies with the English that need to be fixed, though these can be corrected at a later stage.

Comments – all would be classed as “Minor Essential Revisions”

Background

P3, para 2: “c) Are measurements precise enough…?” Do you mean in terms of statistical power or amount of noise or something else?

Methods

P5: the term “adjusted cumulative incidences” is used – why are they “cumulative”? The rates seem to be adjusted incidences.

P8: OR>=2 is used as a cut-off. This seems a bit large to me, as several comorbidities in the Charlson and Elixhauser indices, for example, have smaller ORs. Are the results still the same if you simply retain those variables with e.g. p<0.05?

P8: the c statistic is not a measure of goodness of fit, as stated.

P10, data sources: were ALL publicly funded hospitals in Spain included, even specialist ones whose casemix will be very different from the average? I didn’t know there was a central database for hospital admissions in Spain – how does this overcome the very regionalised administration?

P10, data sources: up to 30 secondary diagnoses can be recorded. It would be useful to know the typical (e.g. mean) coding depth, in order to compare internationally, as this will clearly influence the PSI rates.

Results

P12, first line: do you mean the rho rather than c statistic for the % of variance explained by patient factors? It would in fact be interesting to know the c statistics for the risk adjustment models.
P12, bottom: the “underperformers” are mentioned, but it is important to say what % are “good performers”. For instance, poor coding will lead to a hospital having a low PSI rate. Asymmetry would give us another clue to the reliability of the recording.

Discussion
P13, near the bottom: “this distribution is expected to be centred on zero”. If you are referring to the PSI rate, then this will of course not be true, as rates lower than zero are impossible. Do you mean the log of the rates or something else?

P14: the comment on whether variation is attributed to the patient or the hospital in the case where risk adjustment is limited is a good one.

P16, subheading: “detect hospital performing above the expected”. Do you mean “better performing” (i.e. with low rates) or with higher PSI rates? You need to be clear here and elsewhere what “above expected” means.

P17, line 2 (also lower down on p17): “proven sensitive enough to detect hospitals above the expected”. I don’t think one can make this conclusion. Just because a graph has some statistical outliers does not help us estimate the sensitivity of the indicator. For that, we would need to know which hospitals are actual outliers (from some ideal data source) and we don’t. I wouldn’t use the word “sensitivity” for this reason – power might be better (esp in the Conclusion).

P18: if hospitals do not “properly report secondary diagnoses” they will more likely be flagged as good, not bad performers.

Table 1: for the EB figures, it would be helpful to say what a “high” value looks like as most people are unfamiliar with this statistic. The same would be useful for rho (could give a footnote).

Fig 1, Y axis label: I think this should be log of adjusted risk, as it’s not a log scale

Fig 2: this needs axis labels. Also, I wonder if a funnel plot would be more useful as it would show the outliers better. Alternatively, you could superimpose horizontal lines to show the threshold values for “high” and “low” residuals.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests