Reviewer's report

**Title:** The Procedural Index for Mortality Risk (PIMR): an index calculated using administrative data to quantify the independent influence of procedures on risk of hospital death (a cohort study)

**Version:** 1  **Date:** 8 June 2011

**Reviewer:** Alex Bottle

**Reviewer's report:**

The aim is to derive an index of procedures that predicts mortality using Canadian admin data for one hospital. The centre is a regional centre for trauma etc and its casemix may not be typical of others. The index affected only 16% of included patients and transfers were not included. The study has a number of limitations other than those given. The use of the index would be limited to Canada given the coding system used. The index is compared with KP-IRAM, whose use is confined to North America and seems to use physiological information that is often not included in admin databases in other countries. In its current form, it pays insufficient attention to treatment choices and complications. This is a pity, as sound adjustment using procedure info, particularly if transferable to other countries, would be useful for admin data risk models.

**MAJOR COMPULSORY REVISIONS**

**General key points**

(1) Some procedures were (re)classified as emergency if they could not be done electively, e.g. cardiac resus, though these procedures would presumably be complications and therefore perhaps related to quality of care as well as or even instead of patient status. This relates to the purpose of this index, which is currently not well described in this article. If the aim is to compare hospitals' risk-adjusted mortality for quality improvement or for regulatory purposes, then one should not adjust for the quality of care or treatment choices made in the index admission. This is why it would be inappropriate to adjust for DRG as it includes sets of complications and sets of treatments. Whilst sometimes the procedure will reflect only illness severity, other times it will not. Nor should one adjust for factors that are closely resemble the outcome – the inclusion of heart resus concerns me given that cardiac arrest is often followed very soon after by death. There are several other ops in Table 2 that come under this category. This issue needs discussion.

**Selection of procedures**

(2) Being unfamiliar with the system, I don’t know if using the first five characters of the op codes is a valid way of grouping the large of codes and alternatives should be mentioned.

(3) Some filtering was then done as the sample size was modest (4000 deaths).
“First, we only included procedures that were conducted on the day of the principal procedure.” It isn’t clear at this stage of the article what this means – if you are adjusting for procedure then which procedure? How does your list relate to the principal op? Later on you say that later ops were ignored, so it would be better to say that here.

(4) A good example of the difference between admission and procedural urgency is given in the discussion (angioplasty for AMI following elective THR). Are the dates of each procedure known in the data set, as in England for example? If so, then it would be clear that the angioplasty occurred to deal with a complication of the original elective op.

(5) (Also see comment under General key points) Did you consider excluding any procedure codes because they relate more strongly to treatment choices than to illness severity?

Model fitting

(6) Why did you need to include urgency of admission in the models after including IRAM when IRAM already includes this key variable?

(7) A stepwise selection process was used to pick the final model. This is known to be problematic – did you consider e.g. bootstrapping or test the sensitivity of your approach (e.g. p cut-off of 0.05) to other options?

Results

(8) To me, a drop of 1 in the HL statistic cannot be called an improvement and it would be more realistic to say that calibration was unchanged (also in the abstract). This study, with only one hospital, cannot address this, but it would be useful to know the degree of over or underestimation of risk at a typical hospital. The c statistic is the most commonly used one and is worth reporting, but it is inappropriate for comparing nested models, which you sort of hint at in the discussion but don’t say. The theoretical maximum is not 1 for all models (see e.g. N Cook, Circulation 2007;115:928-935). For linear models, the r-squared is typically given and it would be useful to know the McFadden r-squared for these models (this is the version that is suitable for logistic regression).

MINOR ESSENTIAL REVISIONS

(9) If IRAM did as well with COPS as Elixhauser, why not just stick to COPS?

(10) In the introduction, the second para of the methods might better belong to the intro.

(11) Please replace the word multivariate with ‘multiple’ as, strictly speaking, the former refers to multiple Y variables, not multiple covariates.

(12) Table 5 is unnecessary as its contents are given in the text and Table 4 could be put in the text too perhaps.

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

**Quality of written English:** Acceptable

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

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