Reviewer's report

Title: Is it possible to identify cases of coronary artery bypass graft postoperative surgical site infection accurately from claim data? A multi-model comparison study over 2005-2008

Version: 3
Date: 24 February 2014

Reviewer: Maaike SM van Mourik

Reviewer's report:

Review of revised manuscript: “Is it possible to identify cases of coronary artery bypass graft surgical site infection accurately from claim data? A multi-model comparison study”

The authors have made several changes to their manuscript to improve readability and quality of reporting – several of my questions, however, remain and I have listed them below.

Major compulsory revisions:

1. The authors use the terms ‘traditional surveillance’ in a way that may cause unnecessary confusion. I fear that many readers will think of traditional surveillance as surveillance by manual chart review, hence I would recommend to state in the introduction that all approaches (ICD-9CM model and the alternative approaches) will be compared to surveillance by infection control personnel (your gold or reference standard) and that the relative performance of the ICD-9CM approach will be compared to other methods and refrain from referring to ICD-9 based surveillance as traditional surveillance. I feel that the authors also make the confusion between the traditional approach and the ICD9 approach on p.11

2. The statement on page nine referring to the interpretation of the logistic regression model is, to my knowledge, not entirely accurate. The area under the curve is determined by all the points on the ROC curve, not just a single point. I would therefore state the area under the curve (with confidence intervals) in the results section for both model development and validation, and refer to selection of the cut-off point as based on maximizing sensitivity and specificity (or any other criteria the authors may have used).

3. Furthermore, please explicitly state the variables included in the logistic regression model and (reasons for) any variable selection. Also it is unclear to me how the use of more than 7DDD of cefazolin and antibiotics could have been made continuous. Was the number of DDD included instead of the binary indicator? Also it seems that some of the antibiotic use variables will overlap, e.g. DDD of cefazolin and all antibiotics, how was this handled? Please clarify the variable specifications in more detail, perhaps in a supplementary table.

4. Results, p11: please state in the first paragraph the number of SSIs identified by infection control.
5. As I also suggested in my previous review, I would be very interested to see what the positive predictive value would the logistic regression model would have been when selecting different cut-off points. For example selecting the cut-off where sensitivity is equal to model 5 (87.5), with the best possible PPV. This will allow for a fair head-to-head comparison of the methods; this is currently not the case and thus with the way the information is currently presented I would not support all the conclusions drawn by the authors.

Minor essential revisions

1. Please change ‘multivariate model’ into multivariable model’ as this is a more accurate description.

2. I would suggest paraphrasing the citations to references [12] instead of directly quoting.

3. Given that manual surveillance is not perfect I would rephrase ‘gold standard’ into ‘reference standard’. This is a general recommendation for diagnostic accuracy studies.

4. Please clarify: model 5 is the model with or without pruning?

5. Expanding the figure legends for the CART models may help inexperienced readers understand the use of the model.

6. The authors should present the estimates (including the intercept and standard errors) of the multivariable approach to allow for reproducibility of the results, perhaps in a supplemental data file.

7. Wording issue p 15: the authors refer to the PPV as the ‘worst performance indicator’. This sentence is open for two interpretations a) PPV is not a useful measure of performance or b) some models have a very low PPV. I think the authors mean the latter but I would rephrase the statement to avoid confusion.

**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