Reviewer’s report

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

Version: 2
Date: 21 November 2013

Reviewer: Emily Petherick

Reviewer’s report:

The question posed by authors appears well defined and the paper is on the whole well written although would benefit from further clarification.

Major compulsory revisions

1. Page 6, Data sources
   Under your heading data sources I think it would be helpful to add the time period over which these SSIs were detected, e.g. is this over the years 2005-2008 and up to one year after CABG surgery?

2. The authors provide very little detail of the gold standard used to determine CABG SSIs in medical centres A & B. Is this only performed whilst patients are in hospital or does it also continue post discharge and if so what is covered post discharge. Is the gold standard data based on patient self report or verification by health professionals etc. Is there an accepted standard that clinicians use to determine (eg CDC criteria) or is it based on clinical opinion etc. Are there any potential limitations of the gold standard of which readers should be aware.

3. Page 7
   The authors state that data relating to treatment items were ‘cumulative and without time-related information’. I am not sure I understand what the authors mean by this statement. Does this mean they do not have the dates treatments were provided or information on the duration of treatment?

4. Page 7, Exclusion criteria
   Can the authors provide more details of how they identified postoperative SSIs for exclusion. I am unclear how these are different to post discharge SSIs which were included can you please provide more information so that this is clearer.

5. Page 8, Section on SSI case identification based on use
   Can the authors describe how the number of vessels obstructed is a criteria for the definition of SSI to be met? Is there a reference you could add to evidence your inclusion of >2 vessels being indicative of an SSI.

6. Page 9, Results
   Can you please provide more detail of how you determined cut points using ROC analysis.
7. Page 10, Results
In the results the authors refer to the mean level of complexity of surgery but in table 1 of the results this seems to refer to the number of vessels obstructed, is this the same thing or something different? Can you please use the same terminology throughout the manuscript.

General comments
8. It would be useful to have more contextual information to help understand if the methods are appropriate. Firstly does Taiwan have a policy of prophylaxis for persons undergoing CABG in the two hospitals where the study was undertaken and in Taiwan more widely, and how might this impact upon the results?

9. Do the authors feel that their model is compromised by the lack of knowledge of the timing of the dosage of antibiotics? Is it routine to receive antibiotics prior to CABG surgery in Taiwan, how likely is it that this misclassification of prophylactic vs postoperative antibiotic usage has reduced model accuracy? What bias may this result in?

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests