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

Authors:

Tsung-Hsien Yu (ericthyu@ntu.edu.tw)
Yu-Chang Hou (houyc0115@mail2000.com.tw)
Kuan-Chia Lin (kuanchia@ntunhs.edu.tw)
Kuo-Piao Chung (kpchung@ntu.edu.tw)

Version: 3
Date: 25 January 2014

Author's response to reviews: see over
Dear Editor-in-Chief, Dr. van Mourik, and Dr. Petherick:
Thank you for your helpful comments and encouragement, which have helped us tremendously in revising the manuscript. We are so sorry for the delayed reply, as we are preparing for a new grant proposal.

As you will see, we have made substantial modifications (see italicised and bold sentences with underlining) to improve this manuscript. The following are our point-to-point responses to your comments and recommendations.

Yours sincerely,

Kuo-Piao Chung, PhD
Professor, Institute of Health Policy and Management
College of Public Health, National Taiwan University

Reviewer: Maaike SM van Mourik

Major compulsory revision
1. The authors do not clearly explain the ‘traditional method’ of surveillance. This is a key component of any research of surveillance methods and should be presented clearly to readers (who performs surveillance, how is it performed, are there any validations etc…?). I would strongly recommend using the STARD guidelines for reporting of diagnostic accuracy studies.

Authors’ reply:
The traditional model mentioned in this study was to identify infection cases based on the ICD-9-CM codes. While the problem is, although surgeons gives the ICD-9-CM surgical site infection code when patients had surgical site infection, hospitals use the discharge diagnosis list to petition for reimbursement. Under this mechanism, ICD-9-CM codes that represent important clinical conditions but are associated with low reimbursement can be discarded from the list (Jhung et al., 2009). Therefore, many studies indicated that the ICD-9-CM codes might not be an appropriate tool to identify surgical site infection/ healthcare associated infection in claim data or administrative data. In this revision, we have already changed the original term into ICD9-based model.

Reference:

In addition, we also used the STARD guidelines to check the information
<table>
<thead>
<tr>
<th>Section and Topic</th>
<th>Item #</th>
<th>Item</th>
<th>On page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE/ABSTRACT/KEYWORDS</td>
<td>1</td>
<td>Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').</td>
<td>1,2</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>2</td>
<td>State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.</td>
<td>6</td>
</tr>
<tr>
<td>METHODS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>3</td>
<td>The study population: The inclusion and exclusion criteria, setting and locations where data were collected.</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?</td>
<td>7</td>
</tr>
<tr>
<td>Test methods</td>
<td>7</td>
<td>The reference standard and its rationale.</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>The number, training and expertise of the persons executing and reading the index tests and the reference standard.</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.</td>
<td>NA</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12</td>
<td>Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Methods for calculating test reproducibility, if done.</td>
<td>10</td>
</tr>
<tr>
<td>RESULTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>14</td>
<td>When study was performed, including beginning and end dates of recruitment.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).</td>
<td></td>
</tr>
<tr>
<td><strong>Test results</strong></td>
<td>17</td>
<td>Time-interval between the index tests and the reference standard, and any treatment administered in between.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>Any adverse events from performing the index tests or the reference standard.</td>
<td></td>
</tr>
<tr>
<td><strong>Estimates</strong></td>
<td>21</td>
<td>Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>How indeterminate results, missing data and outliers of the index tests were handled.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>Estimates of test reproducibility, if done.</td>
<td></td>
</tr>
<tr>
<td><strong>DISCUSSION</strong></td>
<td>25</td>
<td>Discuss the clinical applicability of the study findings.</td>
<td></td>
</tr>
</tbody>
</table>

2. For their ‘identification based on use’ the authors describe three alternative approaches (criteria-based, probabilistic and based on CART) using combinations signs of infection. In the logistic regression model, the authors do not specify how co-linearity between variables was handled, if variable selection was performed.
and what arguments were for selection of the predicted probability cut-off. In the CART model, model reduction (pruning) was allowed, therefore perhaps giving an unfair comparison.

Authors’ reply:
In this study, we did not perform variables selection. Before data analysis, we consulted a statistician. He suggested that the selection process was unnecessary, because we were not to explore the relationship between surgical site infection and those variables, but to calculate the probability of surgical site infection. As for the cut-off point, we calculated the probability of surgical site infection for each case, and then we conducted the interpolation to find the optimal point which the areas under ROC curve was the maximum. After continuous testing, we found that a cut-off point of 0.022 was associated with the maximum (0.9050) area under ROC curve. In this revision, we provided these information in the last paragraph of page 9. We also provided the performance of parsimonious model of decision tree model(Figure 2/ second paragraph of page 12).

3. In the results, the authors do not specify whether a threshold could be found for model 4 (logistic regression) that gives similar performance to model 5 (in other words, if accepting a sensitivity of 87.5, what are the associated specificity, PPV and NPV). This would allow for a more balanced interpretation of the findings. I would also be very interested to know what variables were included in the logistic regression model and what the overall performance of the model was (eg area under the ROC curve, calibration).

Authors’ reply:
The variables included in logistic regression model were: (1) use of more than 3 types of antibiotics, (2) use of more than 7 DDD (defined daily doses) of antibiotics, (3) use of more than 7 DDD (defined daily doses) of cefazolin, (4) use of second-line antibiotics, (5) length of stay > 21 days, and (6) number of vessels obstructed > 2. We had already provided such information in previous version (the last paragraph of page 9). And, we also mentioned that all criteria were converted into continuous variables. In this revision, we added information of the process and results (AUC) of the logistic regression model. (last paragraph of page 9)

Minor essential revisions
1. In the introduction, the authors discuss several approaches to using data in models for SSI detection. They refer to the probability-based approach, also in comparison to the criteria-based method. I would like to point the author’s attention to a recent review in Clinical infectious Disease comparing exactly these two methods for HAI surveillance (vol. 57 p 85 – 93).

Authors’ reply:
Thank you. Your studies were the benchmark to us. We have already considered and cited it in the introduction and discussion sections. And we also modified the “criteria-based approach” into “the classification algorithms”, and the “probability-based approach” into “the multivariate regression model”

2. In tables 2 and 3, the authors present their performance estimates as percentages. Could the authors somehow present raw number or confidence intervals for their estimates of performance?

Authors’ reply:
We provided the raw number in this revision.

3. In the exclusion criteria paragraph, it is stated that cases of ‘postoperative surgical site infection and mortality’ are excluded. This is unclear to me, isn’t this an important group of cases that one would be interested in finding? Or perhaps I misunderstand what the authors mean.

Authors’ reply:
The statement was incorrect; we copied it from our another manuscript and did not modified it correctly. We are sorry for the negligence. We corrected it in this revision.

4. Figure 1 is very difficult to understand for untrained readers. It requires better layout and a more comprehensive legend.

Authors’ reply:
Figure 1 was redrawn in this revision, and all information on the figure were provided in figure legend (page 22)

5. Table 1: could you present length of stay as a median with interquartile range? This is likely to be a heavily skewed variable.

Authors’ reply:
We provided it in this revision. (Table 1)
6. In your discussion, you may want to expand your statement concerning the lack to generalizability “to hospitals offering different levels of care” to include other payment systems. In order words, do you expect that your findings will be applicable to other countries/payment systems?

Authors’ reply:
Theoretically, we believe the findings of our study can be applied to other levels of hospitals as well as other countries and payment systems. The major reason is the methodological strength of the decision tree/CART model. But, this merits further examination as we don’t have the data from different hospital levels and different countries. In this revision, we made a little modification in this paragraph (the last paragraph of page 16).

Discretionary revisions

1. Abstract: The wording ‘alternative use-based models’ does not have a clear interpretation. Perhaps the authors can rephrase

Authors’ reply:
We changed the “alternative use-based models” to alternative models based on surrogate indicators in this revision

2. The authors refer to their third approach of surveillance as data mining. This word, however, is used so broadly in the current literature that it is interpretation is not unequivocal. Perhaps they can use more specific wording.

Authors’ reply:
Thanks for your comment. We gave several examples of data mining in the last paragraph. In addition, in methods, results and the discussions sections, decision tree model was used to replace data mining.
Reviewer: Emily Petherick

Reviewer's report:

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?

Authors’ reply:

Thanks for your suggestion. We changed the title into “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”

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.

Authors’ reply:

In this study, we used the healthcare associated infection lists which were collected by infection control professionals as the gold data. Infection control professionals followed the Taiwan CDC criteria to identify the SSI cases. The criteria are the same with the U.S. CDC. Although infection control nurses and infection professionals received strict training, however, in some special cases, inconsistencies might exist among hospitals, although such situation was rare. Thanks for your comment; we added a paragraph to discuss this issue (the last paragraph of page 10 / the first paragraph of page 17).

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?

Authors’ reply:

Yes, we meant that all items in the database were aggregated and were without time information. For example, if a patient received 1 ml cafozaolin before surgery, and another 1 ml cafozaolin during surgery, and 1 ml after surgery. It will be presented as
3 ml cefazolin during the hospitalization. We provided an example in the second paragraph of data source section (page 7)

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.

Authors’ reply:
The statement was incorrect; we copied it from our another manuscript and did not modified it correctly. We are sorry for the negligence. We corrected it in this revision.

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.

Authors’ reply:
We don’t have direct evidence to prove that the patients with the number of vessels obstructed >2 have higher risk of surgical site infection. However, according to the literature, the longer the duration of operation, the higher risk of surgical site infection. We also discussed with infectious disease specialists and decided to use this criterion as the proxy of duration of operation. We added this information in the first paragraph of page 9. *(Mangram et al., 1999)*

Reference:
1999;20(4):250-278

6. Page 9, Results. Can you please provide more detail of how you determined cut points using ROC analysis.

Authors’ reply:
We adopted the interpolation to find the optimal cut-off point while the area under curve was the maximum. We also added this information (last paragraph of page 9) in this revision.

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.
Authors’ reply:
It’s the same thing; we have changed “complexity of surgery” to the number of vessels obstructed throughout the manuscript in this revision.

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?
Authors’ reply:
In Taiwan, surgeons can prescribe antibiotics pre-, during and post-operation for preventing infection. Cefazolin is the most common one for prophylaxis use in CABG. However, there was no compulsory regulation on prescription, and surgeons can freely decide the type and dosage of antibiotics. Nevertheless, most of the surgeons usually prescribe the first-line antibiotics (e.g. cefazolin) as prophylactic antibiotic, that’s why we selected the use of second-line antibiotics as a criterion. And we also provided the context information in second paragraph of page 13 in this revision.

9. Do the authors feel that there 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?
Authors’ reply:
Yes, but your question was the limitation of the database. In Taiwan, surgeons can freely prescribe any type and dosage of antibiotics pre-, during and post-operation for preventing infection. However, the national health insurance database does not tell the types and dosage of antibiotic utilization prior to, during and after surgery. In a previous study, Lee et al. selected the antibiotic types changed or a second antibiotic type added midway as one of the criterions; we thought it was a good criterion, while this information was not available in our database. Furthermore, the antibiotic not only can be used for prophylaxis but also for treatment, surgeons not usually used first-line antibiotics for infection prevention, but also for treatment. Second-line antibiotics are usually for treatment. In general, it is not possible for surgeons to prescribe too many kinds of antibiotics, and to prescribe second-line antibiotics for prophylaxis. This study was inspired by Lee et al., hence we selected “more than 3 types of antibiotics and use of second-line antibiotics” as a criterion. (second
paragraph of page 13/ first paragraph of page 14)