Author's response to reviews

Title: Improved hospital-level risk adjustment for surveillance of healthcare-associated bloodstream infections: a retrospective cohort study

Authors:

Edward NC Tong (edwardtong@hotmail.com)
Archie CA Clements (a.clements@sph.uq.edu.au)
Michelle A Haynes (m.haynes@uq.edu.au)
Mark A Jones (mark_a_jones@health.qld.gov.au)
Anthony P Morton (amor5444@bigpond.net.au)
Michael Whitby (whitbym@health.qld.gov.au)

Version: 2 Date: 9 June 2009

Author's response to reviews:

Study title: Improved hospital-level risk adjustment for surveillance of healthcare-associated bloodstream infections: a retrospective cohort study


We thank the reviewers for their insightful comments and recommendations for our study. We have made major revisions to the manuscript and address in detail each of the reviewers’ comments below.

The reviewers have requested more results and the manuscript has been expanded to include one new figure, one new table and an extension of an existing table. The hospitals have been de-identified in the results as was done in the first submission, but in the current resubmission the de-identification was done with a different order of numbering to be consistent across the graphs (Figure 1, 2A, 2B) and tables (Table 6). The ID numbers may appear different to the first submission but they identify the same sample of hospitals.

Reviewer 1

1. The background and methods of the abstract has been revised to improve its quality.

2. Descriptive summary statistics have been added to the results section. A new figure of longitudinal plots of overall BSI infections per month stratified by sample of nine hospitals has been included. A new opening paragraph to the results introduces the unadjusted infection rates. An additional table of summary statistics for average infection rate per month and average number of patient days per month, stratified by hospital level (rather than individual hospital, to conserve space) have been included for all infection groups.
3. A new paragraph which describes the dataset in more detail has been added to the method section. We now cite a paper with a detailed description of the CHRISP surveillance methods and include additional details in the text. The CHRISP database has been very complete. The BSI dataset comprises monthly infection submissions at a hospital level. Of the 21 hospitals in the database, 18 hospitals had complete observations with no missing interim data (i.e. missing data in some months). Of the three hospitals with missing interim data, one hospital departed the CHRISP network after submitting some data from 2002 to 2005. The 18 hospitals with complete data did not all begin submitting data at the same time point. CHRISP initiated data collection from the 10 largest public hospitals in the state in February 2001. In May 2002, an additional 11 smaller general hospitals were included. Standard case definitions of BSI, adapted from the NNIS definitions, were used in all hospitals.

4. We now acknowledge in the discussion that “…hospital services remain an indirect indicator of patient case-mix. Use of service-specific infection rates (which were not available in the current study) or attribute data of individual patients (also not available for the general patient population) would facilitate a more accurate and robust approach to risk adjustment.” The definitions of service types are standard and well described. Rather than adding a table or detailed textual definition of different service types, which would not be permissible given the word limitations of the journal, we have added a paragraph in the discussion interpreting the results with respect to high-risk services.

5. Exclusion of the ICU service from the analysis was a result of strong collinearity between ICUs and other services, particularly infectious diseases. In other words many hospitals offered ICUs and the distribution of this service among hospitals was very similar to that of infectious diseases. Inclusion of ICU would have invalidated interpretation of the model coefficients of collinear services and we now explain this in detail in the methods section. While we understand the reviewer’s reservations about not including ICUs we feel that inclusion would adversely affect the integrity of the analysis.

6. The second paragraph of the discussion has been moved to the results with the implications of the shrinkage plots interpreted in the discussion. The inter-hospital variation did not decrease because of a large model and regression to the mean. The inter-hospital variation was calculated using the crude original levels and the reclassified levels respectively. The large regression model of services and subsequent risk scoring was used to identify how best to reclassify/reallocate the level of the hospital based on their infection risk via the risk score. The hospital-specific risk adjustment for OBSI and STAPH essentially involves using the predicted/fitted values from the regression model to estimate the number of infections per month, after supplying the patient day denominators for the months. It is a direct method of finding the expected number of monthly infections but it requires that enough infections are collected per month to be accurate. The validity of this direct method was shown in the optimal goodness-of-fit statistics (e.g. Harrell’s c-index, concordance correlation). Since the goodness-of-fit for the IVD-STAPH and MRSA outcomes were not optimal
(i.e. the infections in those outcomes were rare on a monthly basis), the direct estimation method through regression was not ideal. Instead, for those two outcomes, hospital-level risk adjustment with the risk scoring approach was recommended to reclassify the level of the hospital, to then estimate the expected number of infections at the hospital’s level/group classification. The inter-hospital variation was used as a validity check on whether the reclassification improved the allocation of risk by explaining more of the unaccounted variation in the data.

Minor essential revision: We rigorously and comprehensively applied standard methods for variable selection and are uncertain what additional statistical criteria the reviewer is referring to when he asks about checking for confounding. During variable selection, the non-significant services had large p-values such that use of significance levels of 5% or 10% both resulted in the same medical services remaining in the final models. This result, in conjunction with the validation and goodness-of-fit statistics, has provided confidence that the significant medical services are legitimate predictors of the outcome.

We do not agree that the analysis of use only to ourselves – we strongly contend (in the discussion) that, while the models might only be generalizable to our own hospital population, the methods provide a more objective and refined approach to hospital-level risk adjustment than the status quo approach (crude risk-adjustment based on the hospital level) and a similar approach could be used by other health services where individual patient data from the general patient population are not available for risk-adjustment purposes.

Reviewer 2

- The first paragraph of the discussion has been amended to provide a more suitable opening for the target audience.
- The title has been changed to include “hospital-level”.
- The sentence has been restructured to include ‘aims’.
- The reviewer is correct in her assumption. The abstract’s conclusion has been restructured to improve clarity.
- A sentence has now been added to the discussion section to raise the possibility of bias.
- Case-wise deletion was employed in the analysis. The missing data arose from 2001 to 2003 by hospitals which had not joined CHRISP in those earlier years. Since data submission was voluntary not all 18 hospitals started contributing data to CHRISP in 2001. By 2003, the CHRISP network expanded and all 18 hospitals were contributing data. A sentence has been added to the method section to explain why the missing data arose.
- The regression coefficient results are generalizable to Queensland hospitals only as the data used was collected within Queensland but the hospital-level risk
adjustment methodology is widely applicable for other parts of the world where data is not available for patient-level risk adjustment. Additional detail on generalizability has been added to the discussion and conclusion section.

• The paragraph relating to the Bayesian shrinkage plot has been moved to the results section and the reviewer's specific recommendations for additional paragraphs in the results and discussion have been included in the manuscript.

Reviewer 3

In reference to item 2:

• We have added interpretation of the high risk services in the discussion, which will be of interest to clinical readers. Many technical details have been removed from the methods. Much of the language has been revised to be more understandable to a non-statistical audience.

In reference to the sub-points of item 4:

• The reviewer is correct in her assumption. We have added such a definition to the method section.

• The paragraph describing the CHRISP dataset and collection methods has been expanded in the method section with more detail as required.

• There is some possibility of bias introduced as a result of excluding the three level II hospitals. This has been acknowledged in the discussion. The missing data in the training dataset of 18 hospitals have been described in more detail in the statistical analysis section of the methods. The missing data (11.3%) arose because hospitals joined the CHRISP network in two different time periods - in February 2001 and May 2002. Hence not all 18 hospitals participated in the study for the same length of time. However once these hospitals joined the network, their data submissions were complete.

• A description of the risk score has been included in the methods with further comments in the results. The comments on the shrinkage plot figures have been moved to the results.

In reference to item 5:

• As described above, the technical aspects of the methods have been reduced and interpretation of the models for a clinical audience has been expanded. Our methods are likely to lead to a lower probability of spurious penalization of a hospital for high infection rates as evidenced by one hospital actually being demonstrated to be "in control" using our reclassification approach when that same hospital had been identified as a high-incidence hospital using the less objectively defined status quo approach.

Reviewer 4

• Reallocation of hospital level for improved risk adjustment is now better explained.
• We acknowledge the reviewer’s concern of circularity in the nature of this study. However, the principle of now widely implemented statistical process control (control charts) does essentially involve assessing the distribution of the outcome being monitored before applying suitable 2 and 3 sigma control limits for identifying outlier observations. The observations that lie outside the control limits are deemed to be non-conforming. The principle of SPC is not unlike the approach used in the present risk-adjustment study. More detailed justification is given at the end of this document (*).

• As noted above, more details of the risk reallocation have been included in the results and the discussion. In particular, the distribution of significant medical services found by the MRSA model is now included in Table 6 with their associated coefficients and MRSA risk scores. The discussion delves into more detail on how the risk reallocation has shifted hospital 1 into a higher level due its offering of high risk medical services.

• Suitable descriptions on Bayesian shrinkage plots and 2/3 sigma control limits have now been added to the results section prior to discussing the plots.

• The method section now includes a sentence that states the infection data are available by aggregated hospital-level only.

• The authors did not intend for non-significant services to be included in the models because we wanted genuine coefficients estimated with sufficient precision to develop the fitted values and risk scores and to avoid the danger of spurious associations and data overfitting. The models in the study used a significance level of 5%, however models developed from a significance level of 10% also produced the same medical services for all outcomes. This is because the non-significant services had large p-values. In short, inclusion of non-significant services in the models was not appropriate.

• We are confident that the MRSA risk score is robust because the model validated well on an out-of-time sample with the ROC analysis and Hosmer-Lemeshow test. Furthermore, after reclassification of the levels using the risk score, there was a reduction in the between hospital variation, providing a more homogenous group of infection rates within each level.

• A brief non-technical description of the risk score has been included in the statistical analysis subsection of the method section. The Framingham reference was used as it had a similar concept of risk scoring (not because we explicitly used an identical risk scoring approach) and it is appropriate to acknowledge this work. The results of the MRSA risk scores in Table 6 now also includes the distribution of medical services and their associated coefficients, in order that readers may add up the coefficients themselves to see how the risk scores are created.

*The goal of risk adjustment is to make a fairer comparison of BSI infection rates among Queensland hospitals. It would not be ideal to use a training sample from for example, a dataset of UK or US hospitals because their healthcare systems
are vastly different with different patient case mixes. The CHRISP Queensland dataset does not have enough hospitals for which a split sample on hospitals can be conducted without introducing bias and sacrificing precision on the regression coefficient estimates. There are only five level I hospitals (large tertiary teaching hospitals) in Queensland, which should all be included in training the coefficients, given the large volume of data they contribute and the varied medical services they offer. Given the lack of hospitals, the split sample validation has used an out-of-time sample to validate the models which is likely the most robust method of checking the models. In an ideal world, BSI risk-adjustment would be analysed on patient-level datasets which have enough observations for a truly independent validation sample but at present, the data collection procedures have not allowed it and yet we still have the responsibility of monitoring the performance of BSI rates in hospitals.