Reviewer's report

Title: Use of hierarchical models to evaluate performance of cardiac surgery centres in the Italian CABG outcome study

Version: 1 Date: 14 March 2007

Reviewer: Patrick S Romano

Reviewer's report:

General

This is an interesting manuscript describing the impact of using hierarchical models instead of traditional logistic models to evaluate cardiac surgery performance in Italy. It offers little new information, but it is definitely a useful contribution to the literature because it extends previous findings that were based largely on data from New York State. I herein ask the authors to improve their paper by fully describing their methods and results for outlier identification, and by better comparing and contrasting their results with those of previous studies (including some that they failed to cite).

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Page, paragraph

5,2 The “conventional stepwise method with a cross validation procedure” should be more fully described. Since no interactions entered the final model, the pre-specified “interaction hypotheses” should also be described.

5,3 Please clarify how you identified hospitals with RAMRs that were significantly higher or lower than the statewide rate. Was this the “model-based” approach given by Hosmer and Lemeshow (Stat Med 1995;14:2161-72), assuming that model coefficients are fixed (Scenario One)? Or was it based on the exact method described by Luft and Brown (HSR 1993;28:419-39)? Or perhaps a bootstrapping approach?

6,end How did you estimate 95% confidence intervals for the group-level residuals? Does citation 16 offer a specific approach to this problem that is consistent with the hierarchical multilevel random-effects model? I have seen three ways of approaching this problem in the applied literature that I scan:

1. Glance et al. (Intensive Care Med 2003;29:2223-9) argue that “since the random effects component of the hierarchical model accounts for between-hospital variability, only the fixed-effects coefficients were used to calculate the expected mortality rate. By using only the fixed-effects portion of the hierarchical model to estimate patient mortality, the impact of hospital quality on patient outcome is removed. In other words, the fixed-effects portion of the model estimates the probability of death for a patient treated at an 'average-quality hospital.'” They used the patient-level probabilities of death to estimate O/E ratios, then SMRs, and then bootstrapped to generate CIs for those SMRs.

2. Most other authors have estimated provider offsets directly within the random-effects hierarchical model. For example, note Models 5 and 8 in DeLong’s paper on CABG mortality (Stat Med 1997;16:2645-64). They use PROC GLIMMIX to estimate these parameters in a logistic model, and then exponentiate them to estimate the ODDS of mortality due to a specific hospital (relative to the average hospitals). Burgess et al. (J Health Econ 2000;19:291-309) seem to use a similar approach, but with a Poisson hierarchical model.

3. Normand et al. (JASA 1997;92:803-14) describe a variety of methods for using the mixed-effects model to estimate hospital performance statistics, including a z score for the standardized difference between observed and expected mortality.

8,2 The list of factors associated with mortality seems incomplete. What about age and diabetes, for instance? Please either list all of the significant risk factors or clarify that you are only mentioned a few of particular interest.

9,3 After explaining in the Methods how you identify outlier hospitals, using each analytic approach independently, you fail to present this information in the Results. How many hospitals were designated as outliers using each approach? To what extent do these two sets of outlier hospitals overlap? The rank
correlation could be quite high, but with relatively little overlap of designated outliers. (This comparison appears briefly in the Abstract, but it is undesirable to include results in the Abstract that are not presented in the text.)

9.5 You do not adequately summarize the existing literature in this area. In addition to the papers cited above, you should review Goldstein and Spiegelhalter (J R Stat Soc A 1996; 159:385-443), Christiansen and Morris (Ann Intern Med 1997; 127:764-8), Shahian et al. (Ann Thorac Surg 2001;72:2155-68), and Austin et al. (Med Decis Making 2003; 23:526-39). Several of these studies (in addition to ref. 17) have directly compared hierarchical with single-level models. Please set your own paper better in this context. Please compare and contrast your results with those from other researchers, as appropriate.

10.5 Your ICC estimate may not be directly comparable with Hannan’s, because Hannan was assessing the PREDICTIVE performance of models by comparing their results with hospital performance two years hence. By contrast, you seem to be using the same data period. What are the implications, if any, of this difference?

Discretionary Revisions (which the author can choose to ignore)

10.2 Should read “to make the same choices,” not “to do the same choices.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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’