Reviewer's report

Title: Deaths from heart failure: Using coarsened exact matching to correct cause of death statistics

Version: 1 Date: 20 November 2009

Reviewer: Andrew Moran

Reviewer's report:

General Review Criteria:

1. Is the question posed by the authors new and well defined?
   yes
2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?
   yes
3. Are the data sound and well controlled?
   yes
4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   yes
5. Are the discussion and conclusions well balanced and adequately supported by the data?
   yes
6. Do the title and abstract accurately convey what has been found?
   Yes, but an expanded title is suggested below
7. Is the writing acceptable?
   Well written

General Comments:

Heart failure is a major illness, and sometimes a fatal one, but heart failure’s causal link with a number of sometimes overlapping etiologies [ischemic heart disease, hypertension, infections, valvular disease, cardiomyopathies, endocrine and metabolic disorders, and toxic exposures (e.g. alcohol)] make measurement of heart failure’s impact difficult to measure. As more and more patients survive the primary insult (the etiology) and the world population ages, heart failure is becoming a more common disease. Due to this, and to chronic errors in recording cause of death on death certificates, heart failure is also now commonly listed as a cause of death, but the death certificate includes no underlying cause of death (the etiology or etiologies that preceded and caused
heart failure). This breaks with ICD rules—heart failure by definition is not an ‘underlying cause of death’ in the ICD. The problem is of improperly coded deaths has a big impact on mortality estimates, and even more importantly it confounds public health efforts to address out the conditions that are the root causes of heart failure. The ICD system is also limited in this case because it is best at describing a linear causal chain, and no multiple causality.

Stevens et al. describe an approach called “coarsened exact matching” to assigning the heart-failure-without-an-underlying-cause deaths to proper underlying causes. The method proposed is conceptually simple, and makes use of a number of variables available in death registration data from three nations. The authors address the possibility of confounding by matching, rather than multivariate adjustment, and efforts were made to validate the overall approach in large sub-populations within the mortality data.

The method proposed has the potential to improve the accuracy of heart failure epidemiology estimates as well as improve estimation of the mortality burden of its underlying causes. I support publication of the manuscript, though I must state several issues that concern me but are addressable by the authors. These are as follows:

1) Authors reference an analysis in which a polynomial logistic regression approach to reassigning deaths with diabetes reported as both a multiple cause of death and an underlying cause of death to an ICD-appropriate underlying cause. This analysis was performed for death registries in the US and Mexico. The authors make a compelling argument for the matching method over the logistic regression method—it makes sense that the matching approach allows one to develop a large number of outcomes which would certainly be appropriate in this case as heart failure has so many etiologies/underlying causes. But their argument in favor of matching is theoretical only. Though the authors make a case for the limitations of the logistic method (mainly the need for assumptions of independence of covariates), the reader is not entirely convinced—the problem confronted seems not all that different from the one faced for the Murray diabetes analysis (citation 13) and mortality data from the US and Mexico were again analyzed. It would seem in this case appropriate to at least compare the two methods, and if the presumably simpler, lower information matching approach performed as well it would make the approach more attractive.

2) If the reason that the authors advocate this method is that it will be a better method than other methods to use in nations with less complete vital registration, this was never explicitly stated. If true, it could be a powerful endorsement for more widespread applicability of the coarsened exact matching method.

3) The choice of matching categories appears to have been chosen apriori. It is not clear if the categories were selected based on past experience or common sense. It would be more interesting and data-driven if the matching categories were determined by growing a larger number of matching categories until a predetermined statistical test of heterogeneity found that adding more categories would produce no change in the outcome between categories (i.e., a CART
analysis of categorical data). I have no sense if the number of matching variables or the number of categories per variable are too few or too many.

4) I’m troubled by use of average relative error statistic—was there no criteria or threshold value decided a priori for accepting or rejecting validation of the method in subsets of the death registry? The way the statistic is used now is very qualitative—it would seem that a wide range of differences could be argued for as acceptable or unacceptable after the calculation is made. It would seem that the investigators embrace the base approach, while noting the difference when the autopsy subset is analyzed in Mexico and Brazil. There needs to be a firm justification for rejecting (perhaps as not representative of the entire population?) the autopsy subset derived adjustment since so many readers will consider autopsy the gold standard of ascertainment. On this note, I would expect a lower rate of heart failure listed as underlying cause in the autopsy subset—was this the case? Perhaps the distributions of cause of death shown in Table 1 should be repeated for the validation subsets in an appendix table.

5) Why no discussion of limitations of the ICD system? How could it be structured to better capture multiple causes of death/multiple underlying causes of death? What would improve death certificate reporting? Should heart failure be designated an underlying cause?

6) External validity: How do controls/matches and allocation results match up with differences in the incidence, mortality for the underlying causes listed— at least IHD and HTN in at least US and Mexico?

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

1. At the least, add to argument for the matching method in the Discussion that ideally regression and matching methods would be compared head-to-head—how do they compare with one another (equivalency) and which produces the most consistent results across different populations?

2. Explain better how the matching categories chosen in the Methods. Data driven, based on prior experience?

3. State clearly if the average relative error was meant to be descriptive or the basis of decisions regarding valid or invalid.

- Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. Consider adding a description or small Table 1 with main etiologies of heart failure (ie hypertension, IHD, valvular disease, infections (including Chagas), idiopathic cardiomyopathies, etc), perhaps in the Introduction.

2. While discussing this, under what underlying cause of death listed in the present Table 1 does valvular disease fall? Can you confirm in the text that
Chagas belongs under Cardiomyopathy?

3. Background page 3: “Causes such as heart failure...violate standard protocols” not strong or clear enough, state “in the ICD system, heart failure is not allowed to be an underlying cause of death”

4. Methods first full paragraph on page 7. Assume that authors meant something like:

“US deaths with heart failure listed as the UCD were matched with...2888 deaths with death certificate mention of heart failure but another disease as the UCD”

5. Limitations—suggest mentioning that the 13% of heart failure assigned to diabetes in Mexico may be due to cause of death assignment problems related to diabetes itself.

6. Is the high rate of cardiomyopathy as heart failure UCD in Brazil due to Chagas? Can you check how often Chagas is mentioned on death certificate part 2 of “cardiomyopathy” UCDs in Brazil?

7. “Table 2” mentioned second paragraph of page 7 should be Table 3.

8. typo (backslash) in the first full paragraph, first sentence of page 10

9. Same paragraph, page 10, more attribution to HF in Egypt may not just be “physician culture” but also more diagnostic uncertainty (unable to identify the UCD).

- Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

10. As stated above, the final paragraph of the Discussion, and perhaps a few words in the Conclusion could be devoted to: How the ICD be structured to better capture multiple causes of death/multiple underlying causes of death? What would improve death certificate reporting? Should heart failure be designated an underlying cause?

11. For external validity, brief description in the Discussion of prevalence and/or incidence of hypertension, ischemic heart disease, smoking in each nation (this is available for the U.S. (NHANES), should be for Mexico, but may be unavailable for Brazil. The pattern of more stroke, less IHD mortality in Brazil in Mexico—would expect reflective of more hypertension-related disease.

12. The authors do not state how often one of the heart failure causes is listed in death certificate part 2 (multiple cause) alongside a heart failure UCD. For instance, how often is hypertension listed in a part 2 below IHD as the UCD? Could this information be used to establish some uncertainty bounds around the allocations, and give a hint of overlapping multiple causes of heart failure?

Level of interest: An article of importance in its field
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.