

Author's response to reviews

Title: Use of the Social Security Administration Death Master File for ascertainment of mortality status

Authors:

Enrique F Schisterman (Schistee@mail.nih.gov)
Brian W Whitcomb (bwhitcom@epi.umaryland.edu)

Version: 2 **Date:** 4 Dec 2003

PDF covering letter

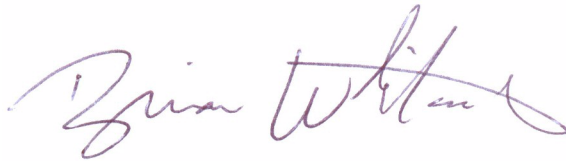
December 3, 2003
Emmanuela Gakidou
Associate Editor
Population Health Metrics

Dear Dr. Emmanuela Gakidou,

Enclosed please find a revised version of our paper entitled "*Use of the Social Security Administration Death Master File for ascertainment of mortality status*" by Enrique F. Schisterman and Brian W. Whitcomb. We have incorporated the suggestions made by the two reviewers. Please find a point-by-point response to the reviewer's comments. Both referees have made very good points that have very much improved our paper, and we extend our thanks to them.

We hope that the paper is now acceptable for publication.

Sincerely,

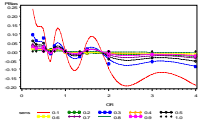
A handwritten signature in purple ink that reads "Brian Whitcomb". The signature is fluid and cursive, with the first name "Brian" and last name "Whitcomb" clearly legible.

Brian Whitcomb
Predoctoral Fellow
NICHD / National Institute of Health

Response to review from Dr *Ajay Tandon*

- 1. The paper would definitely benefit from more examples showing that indeed such Internet-based sources are routinely used for assessing vital statistics such as mortality rates.**
 - Examples of utilization of Internet-based sources have been included in page 3, paragraph 1.
- 2. Furthermore, the paper would benefit from speculating on the extent to which erroneous estimates of mortality rates might bias research based on such Internet-based sources.**
 - We have included a paragraph in the discussion to warn researchers of the potential biases due to misclassification of mortality status. Although such results are well known and extensively studied, we also included for this reviewer a simulation study that evaluates the effects of misclassification of mortality status on the odds ratio (from 0.2-4) by different levels of sensitivity. As shown in Figure 1, we found between a 5%-20% bias of the odds ratio when comparing 2 levels of exposure and assuming different levels of sensitivity (0.1-1). These simulations can be extended to more complicated situations where the interactive effects are accounted for as well. Although these findings are worthy of note, we believe they are beyond the scope of this paper.
- 3. Also, some discussion would be helpful as to what one could do (if anything) to adjust mortality rates in light of such a sample selection bias. There is a big literature on this issue and some reference should be made to some of the methodological approaches that are out there to adjusting for such biases.**
 - We have included in the discussion section methods to correct for misclassification of outcome as well as exposure.
- 4. One final point relates to the generalizability of the findings of the authors. They themselves acknowledge that their study might have limited generalizability because of the sample they use. This is potentially the biggest weakness of their paper. I would recommend that the authors expand their study to include other samples so as to enhance the generalizability of their results. If not, the paper will remain limited in its scope and applicability.**
 - We respectfully disagree with this reviewer's conclusion that the study findings are limited in scope and applicability due to the sample. As in most studies, evaluation of findings in other populations is not entirely feasible for common causes like budget. To be able to generalize the findings in a study of this nature, an investigator would need to evaluate the sensitivity of the SSDI in a representative random sample of the US population. Moreover, despite the use of the convenience sample for this study, we have no reason to suspect that estimates of overall sensitivity or sensitivity as a function of study variables would be grossly different than

population values. However, in the absence of such a sample, sensitivity analyses to account for possible bias are recommended; new methods for sensitivity analysis are available and their use is strongly recommended. We have included such a recommendation to the readers in the discussion section.



Response to review from Dr *Emmanuella Gakidou*

- 1. The paper seems to assume knowledge from the reader of the data sources involved; the authors should assume that readers outside the USA will not be familiar with the data sources they refer to.**

Information regarding data sources referred to has been included (see page 3, Background, line 6, line 16; page 7, Discussion, line 17).

- 2. The authors state “the SSA Death Master File is comprised of descendants with social security numbers whose death as reported to the SSA.” It would be interesting to know what percent of the US population have social security numbers and what percent of total deaths get reported to the SSA and whether there is any socio-demographic pattern in the SSA Death Master File, like those seen in the results.**

Further explanation about the composition of the U.S. Social Security program has been provided (see page 7, Discussion, line 17). Reasons for reports of death to the SSA, given on page 7, Discussion, line 11, have been further explicated.

- 3. The authors seem to have access to hospital admission records that include the Social Security Number – is that common for researchers in the US? The authors state that previous research linking to the SSA Death Master File without the SSN reported much lower linkage rates for women than men. This implies that linkage to the SSDI should really only be done when the SSN is available (particularly for women) – do the authors agree with this? What are the confidentiality issues involved in getting access to data with the SSN?**

A discussion of the availability of social security numbers to researchers has been included on page 8, Discussion, line 14.

- 4. How was the actual linkage conducted? Which statistical/ database software was used? What exactly were the linkage criteria and the importance weights assigned to each of the variables? In the literature, some researchers use commercial software such as Automatch (e.g. the New Zealand Census Mortality Study) while others develop their own matching algorithms (e.g. StatsCanada). It would be interesting to have more information in the “Methods” section as to how the actual matching was performed in this study and how inconsistencies were dealt with. It is not clear if each record was done individually or if the entire sample was linked at the same time using a statistical procedure. Is the SSDI a downloadable database or does one have to search for each record individually? If this is the case, then this would limit the potential to use the SSDI in studies with large sample sizes.**

The searches were conducted individually and matches determined according to criteria established in previous investigators' work. While the database we utilized is not downloadable, a staff programmer developed a simple program to perform batch searches. The preceding has been described in the second paragraph of the Methods section, page 4, line 22 and page 5, line 3.

- 5. One of the subgroups that displayed lower linkage rates in this study is foreign-born decedents. What percent of the US population are foreign born and what percent of the present study – how big of a concern is this for future studies?**

Demographic information concerning the proportion of the U.S. population that immigrants comprise, as well as in this study population, has been provided (see page 9, Discussion, line 13). The significance of this information for investigators is also noted (page 11, Conclusion, line 4).

- 6. Similar with younger respondents – the authors state that to be included in the SSA one has to have achieved 10 years of work in the US (p.7, 2nd para). This implies that the SSDI is unlikely to ever be a good source of ascertaining mortality for younger populations. Should the SSDI only be used for studies that focus on older adults?**

Some discussion of the impact of age of the study population on the utility of the database has been added (page 9, Discussion, line 12). However, we have not speculated upon whether this source of mortality information will prove useful to other researchers. As noted on page 9, line 16, our study indicated high sensitivity for detection of the oldest decedents and something of a threshold effect for other age groups—the decrease in sensitivity was similar for the first, second and third age quartiles. As shown in Table 3, the age range of the youngest age quartile was broad (from 41 to 70), and we have not extrapolated to individuals younger than those included in this study, but rather displayed estimates for sensitivity as a function of the age quartiles we observed.

- 7. This paper is very interesting and would be well suited for a journal like PHM, after revisions. For the paper to be appreciated by a broad readership more information should be provided on each of the data sources and eligibility criteria for them (e.g. who is eligible for a US SSN, what is the difference between the SSDI and the NDI, how do deaths get reported to the SSA, etc?).**

See response to point numbers one and two.

- 8. Also, more details on the exact methods of the linkage would be very interesting for future investigators wishing to pursue similar analyses. Finally, it would be very interesting to place and compare the results from the current study in the broader literature of record linkage studies.**

As searches for this study were carried out individually, without the aid of linkage algorithms, and according to matching procedures established by previous researchers, we have not expanded upon this study in such a context.