Author's response to reviews

Title: Surname lists to identify South Asian and Chinese ethnicity from secondary data in Ontario, Canada: a validation study

Authors:

Baiju R Shah (baiju.shah@ices.on.ca)
Maria Chiu (maria.chiu@ices.on.ca)
Shubarna Amin (shubarna.amin@gmail.com)
Meera Ramani (meera.ramani@utoronto.ca)
Sharon Sadry (s.sadry@utoronto.ca)
Jack V Tu (tu@ices.on.ca)

Version: 2 Date: 5 April 2010

Author's response to reviews: see over
Melissa Norton  
Editor-in-Chief, *BMC Medical Research Methodology*

Dear Dr. Norton,

I am pleased to submit our revisions to the manuscript entitled “Surname lists to identify South Asian and Chinese ethnicity from secondary data in Ontario, Canada: a validation study”, manuscript number 1466319145365545. We have addressed the concerns raised during the review of the manuscript. Below, I have summarized the reviewer’s and the editorial comments, along with our responses and revisions:

**REVIEWER #1**

1. **The reviewer requests further details about the weighting scheme and how it was used in the study.**

   The weighting scheme accounts for the three sampling frames of the survey. It also accounts for other potential biases introduced by sampling, including non-response and refusal. Since it is a telephone survey, it also accounts for sampling errors occurring because of households with multiple telephone lines or those without telephones. Finally, it calibrates the sampling to population age and sex estimates at the health region level. Although ethnicity is not specifically included in the weighting scheme, respondents are selected randomly, so there is no reason to suspect that the weighting scheme would systematically under- or over-represent any ethnic group compared to their population prevalence, particularly since the survey was administered in 24 languages. We have added this information to the manuscript on page 6.

   The study cohort was derived from the CCHS respondents. The sampling weights were used to weight each individual from the cohort in the calculations of sensitivity, specificity, PPV and NPV. This is indicated on page 7.

2. **The reviewer would like to see a table with numbers.**

   We have revised Table 1 to show the numbers as requested.

3. **The reviewer asks if we have done two linkages.**

   We first linked the surname lists with the nominal RPDB by surname to create an ethnic identification file (as described under “Administrative data sources” on page 5). We then linked this file with the CCHS survey responses by encrypted health card number to create the dataset for validation (as described under “Validation” on pages 6–7). We could not link the surname lists directly with the CCHS responses, as the only personal identifier included in this dataset is the health card number.

4. **The reviewer asks whether the South Asians identified in the RPDB represent only 50.2% of all South Asians in Ontario.**

   The RPDB includes both current Ontario residents and former Ontario residents who were eligible for a health card at some time since 1991, but have since moved or died. This has been added to page 5. Thus, considering the sensitivity of the list, the 500,807 people identified as South Asian represent 50.2 percent of all current or former Ontario residents with South Asian origins. According to the latest Census, there were 794,170 South Asians in Ontario in 2006.
5. The reviewer questions whether the surnames lists can be used for population-based studies, and asks that examples of their utility be included in the abstract.

We respectfully disagree. While we acknowledge in the manuscript that the low sensitivities of the surname lists make them unsuitable for population-based studies of ethnicity prevalence within disease cohorts (similar to the types of the studies the reviewer has published in the past), their high PPVs make them suitable for population-based studies of disease prevalence within ethnicity cohorts. Although these ethnicity cohorts would not include the entire population of people from that ethnic group, members of the cohort would be virtually certain to belong to the ethnic group under study. These cohorts could then be used reliably for epidemiologic or health service research studies of people from that ethnic group. We have already published several abstracts that use these surname lists to examine diabetes and cardiovascular disease in Chinese and South Asian populations:


At the reviewer’s suggestion, we have added a short description of the types of studies for which these surname lists would be suitable to the end of the abstract, and these studies are already discussed in the Discussion on page 8.

EDITORIAL REQUESTS

1. The editors ask whether the data sets used are publicly available and, if not, whether appropriate permissions were obtained.

The data for the study are not publicly available, but were provided to the Institute for Clinical Evaluative Sciences by the Ontario Ministry of Health and Long-Term Care under a research agreement. ICES is permitted to hold, link and analyze these data for research purposes as a named “prescribed entity” in Ontario’s health information privacy law, the Health Information Protection Act. We have added this information to the manuscript on page 7.

I look forward to hearing from you.

Warm regards,

Baiju Shah