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: 3 Date: 21 April 2010

Author's response to reviews: see over
Dear Dr. Norton,

I am pleased to submit further 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 questions the accuracy of the expansion of the weighting process to ethnic identifiers, as these were not included in the weighting scheme.

   We acknowledge that ethnicity was not specifically used in determining subjects’ weights. Instead, the weights are used primarily to account for the sampling frame, and to account for non-response and other potential selection and sampling biases. However, the subjects who participate in the CCHS are selected at random from the population, so we see no reason to suspect that the use of these weights is inappropriate when studying ethnicity, just as when they are used in analyses stratifying on any of the dozens of other factors measured from the CCHS. Indeed, as per reference 13, Statistics Canada requires the use of weights for analyses of CCHS data to ensure that results are representative of the population, not just the survey sample. This has been clarified on page 6.

2. The reviewer asks whether the sensitivity and PPV were calculated using the weighted population counts rather than the actual population counts.

   As indicated in the first paragraph of page 7, all calculations involving the CCHS data used the sample weights. In fact, as noted above, weighting is required for analyses of these data. Comment #4 in the previous response letter was referring only to data from the RPDB and the Census, not to the CCHS.

3. The reviewer notes that the surname list is selective and affected by “selection bias”.

   We completely agree with the reviewer: the surname lists generate cohorts that are virtually certain to belong to the minority group under consideration, but these cohorts are not representative of the entire population of people belonging to those minority groups. As we noted in the Discussion on page 9, the South Asian list is composed primarily of Hindu, Sikh and Sri Lankan surnames. Surnames from other religious or cultural groups are not included, and the cohorts built using these lists would systematically exclude these people, who may have different biological, behavioural or clinical risks. The same issue arises, although probably to a much less clinically important degree, with the Chinese surname list.

I look forward to hearing from you.

Warm regards,

Baiju Shah