Reviewer's report

Title: Record linkage of census, mortality and hospitalisation databases to explore ethnic variations in disease: ethical creation of a retrospective cohort study

Version: 1 Date: 19 December 2006

Reviewer: seeromanie harding

Reviewer's report:

General
This is an exciting initiative, a record linkage exercise to determine ethnic differences in events, survival and mortality from MI in Scotland, worthy of publication.

Both papers: Given that a detailed report on these studies is available on the web, I would urge the authors to combine the 2 papers and focus on the potential use of the data.

Paper 2: The South Asian population is considerably younger than the Non-south Asian population (84% of South Asian women under age 45 years compared with 59% of Non South Asian) and the age specific incidence rates and Kaplan-Meier survival curves based on small number of events/deaths. I would suggest moderating the comments to reflect that the ethnic difference could be due to the younger age distribution, present survival curves and days survived (Table 2) for only those aged 55-84 years (the oldest age group is least reliable due to the very small number of South Asians) and mention the need for linkage of events over longer periods to improve the reliability of effect sizes.

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

Paper 1
1. As implied above, this paper could be shortened and merged with paper 2. It would seem appropriate particularly to shorten the bits on computing details avoiding jargon such as 'hashing', 'partitioning' etc.

2. It would interesting to show the ethnic differences in matching at each stage – i.e. on matching CHI to Census, this merged record to morbidity, to mortality etc. – rather than the overall matching percentages by ethnicity.

3. It would be helpful to clarify when and how the matching of the mortality records was done. Were death records linked to the merged record containing both census and morbidity records? Were there ethnic differences in mismatching of these records to the census or to the morbidity or to both?

4. It would be useful for researchers planning to construct similar data sets to elaborate further the statement on ‘strategies a human checker employs to decide whether a pair match is ‘good’. For example, were there more mismatches for women due to changes in names on marriage?

5. The section on ethics could be substantially reduced as this would differ in different settings.

Paper 2
Methods
1. To avoid confusing the reader, suggest that the methods refer only to the data used here - Census 2001, admissions April 2001-2004 and deaths April 2001-2003. Delete reference to discharges 1981-2004 as it seems that the data covering the period 1981-March 2001 were not used?

2. Elaborate on how statistical adjustments were done for the hazard ratios. Refer to comment above re censoring of analyses to avoid the confounding from differences in age distributions.

Tables
1. Table 1 Add 95% Confidence intervals and p-values for differences between rates for South Asians and Non South Asians.
2. Table 2 could be amended to show results of Cox models, showing hazard ratios and number of deaths for those aged 55-84 years, unadjusted and adjusted for deprivation.

3. It is interesting that the gender difference in survival for the South Asians is different from that of the Non-south Asians. Do the authors have any ideas why?

Discussion
1. The authors may find the following paper useful

2. Given the small number of deaths, it is difficult to censor the data adequately but some discussion about the bias from having admissions (2001-2004) and deaths (2001-2003) for different periods would be useful.

3. There are a few sentences in the discussion which need some elaboration/linking to the preceding sentences. For example – Page 8 3rd para, last sentence beginning ‘South Asians are less likely to get…’. Also page 9, 1st para, last sentence beginning with ‘Anatomical differences’.

What next?: Accept after minor essential revisions

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests