Reviewer’s report

Title: Selection bias: neighbourhood controls and controls selected from those presenting to a health unit in a case control study of efficacy of BCG revaccination

Version: 1 Date: 14 November 2006

Reviewer: Joanna Stewart

Reviewer’s report:

General

--------------------------------------------------------------------------------------
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
methods
1. It is not clear what the effect of the time delay was. Please could this be elucidated. Was the age matching done on the age that the controls were at selection and the cases at time of illness? If the controls were selected 2 years after the cases (I assume the first set of controls were selected at the same time as the cases occurred?) then you would expect the neighbourhood controls to be born 2 years later but the text says they were one year later.

Results

2. As the data is matched, while it is informative to present the raw data as in table 1 it is not valid to produce the p values which, although not explained in the table, look as if they are from the chi square test for a difference in the proportions in the 2 groups. Instead the test of the adjusted odds ratio produced from the matched analysis, as in table 2 but preferably also adjusting for the other potential confounders listed in table 5 should be used. This result could be included in Table 1 rather than having a separate table. This combined analysis would also be preferable to the separate analyses on each variable as is presented in table 5. A global test could be done to see if the inclusion of these confounders could be demonstrated to alter the association of group with vaccination status.

3. At what age were children usually being vaccinated? There is comment in the results that age is associated with vaccination in both groups but in fact the odds ratios for vaccination across age are very different in the 2 groups. Is this because of a decline in vaccination rate in the 7-9 year olds over the 2 years and the movement of the 7-9 year olds into the 10-14 year olds by the time the neighbourhood controls were selected? Presumably including year of birth additionally to matching on age attempted to overcome this but the influence of age is not linear (see OR’s). Also there is confounding with group – all later birthdates will be controls, all earlier cases. This change in the odds ratios I believe needs to be investigated and discussed further in the paper as, as it stands, one interpretation is that the change in vaccination rates from the time the cases occurred to the time the neighbourhood controls were selected could be rendering the controls non comparable with the cases, despite the authors dismissing this in the discussion. At the least the discussion needs to acknowledge that because of this time difference for neighbourhood controls, and because of the different population of people attending a clinic, neither set of controls is in fact valid and the difference in estimates formed using them exemplifies that if you do not choose controls which are in fact a random sample of the population from which the cases came the data is uninterpretable. There is also no mention in the discussion of other biases which could have been introduced by the time difference caused by mobility of families and changes in the structures of communities over time.

4. The abstract and the discussion should be less emphatic that differences in the groups are attributable in their entirety to the bias caused by HU attendance giving more opportunity for vaccination. While this will obviously be part of the reason, the fact that the socio economic and demographic variables which were recorded did not explain differences can not be interpreted as ruling out other unmeasured factors that could act as a confounder between group and vaccination status, including the time factor which renders the neighbourhood controls different from both the HU controls and the cases.

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

5. Typo's – p6 line 5 middle paragraph – same, not some, 9th line, same paragraph – than not that. Last sentence this paragraph 'related to'
Methods – 2nd sentence 'given to'
Results – the sentences referring to table 2 and table 5 need rewording.
Discussion – last sentence first paragraph – 'mainly' misspelt.
Table 1 the * does not appear in the table

Discretionary Revisions (which the author can choose to ignore)

6. Results – the adjusted VE for neighbourhood controls changes quite a lot from the crude one – this is correct is it, and its direction is correct?

7. Although the abstract states that the results of the RCT are used as a comparison, the results of this study are not given in this paper. Although it does not necessarily follow that the fact that a study produces the same results as the RCT means the controls are not biased, if the RCT is to be used as a gold standard it would be informative to quote it.

8. While there is mention that if cases are all registered with a HU then a random sample of controls registered with the same HU's would be appropriate this point could be made clearer, and be followed by the point that it is not appropriate to select from a subgroup which may not be representative, such as those who have attended clinics. In the present case of clinic attendees it is very obvious that there is a reason why the way in which they differ from a sample from the appropriate population would be likely to bias the result. However any sample which is not taken from the population from which the cases come runs this danger, irrespective of whether or not it is obvious. Therefore it would be good to use these examples as illustration of the problem of incorrect controls concluding that if a case control study is to be conducted the method of control selection must always aim to produce a random sample of the population from which the cases are drawn.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

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