Reviewer’s report

Title: The SAIL databank: linking multiple health and social care datasets

Version: 2 Date: 23 December 2008

Reviewer: Miranda Tromp

Reviewer’s report:

This paper on the SAIL databank is of importance to the field of Health Information Research. The method section has been made much clearer in the revised manuscript. The information on the actual matching process strengthens the paper.

Minor essential revisions

1) I still feel the background section is too elaborate and should be focused more on the specific subject of this paper. In addition, the conclusion paragraph is also quite elaborate and would benefit from making it shorter and more to the point.

2) It is still not clear to me from table 2 how the last line in the table ‘Overall combining valid NHS, DRL & PRL (50%)’ can have more records matched when compared to the former last line ‘Valid NHS plus PRL 50%’. The only difference is the adding of DRL and DRL is exact matching based on the five variables. How can these additional 13 and 3 record pairs not have been included in the PRL 50% matching?

3) Table 2 only mentions the number and percentage of matches records of the specific databases, information on error rate is mentioned in the text. Can the error rate be incorporated in the table?

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.