Reviewer's report

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

Version: 1 Date: 19 October 2008

Reviewer: George Vassilacopoulos

Reviewer's report:

This is an interesting paper concerned with the effective linkage of disparate patient records (or datasets) for health-related research studies, whilst adhering to the principles of good information governance. In this context, the paper describes the SAIL (secure anonymised information linkage) databank whose data are both anonymised and encrypted and are linked using either deterministic or probabilistic record linkage methods.

The research questions posed in the paper are explicitly and well defined. They concern (a) examining the accuracy of using valid NHS number in routine data as a unique identifier and (b) assessing the effect of varying the acceptable threshold in probabilistic matching in different datasets. An SQL-based matching algorithm, called MACRAL, was developed to enable record-linkage. This algorithm was extensively tested using a large variety of datasets. The results obtained were encouraging.

I believe that the manuscript adheres to the relevant standards for reporting and data deposition while the discussion and conclusions well balanced and adequately supported by the data.

Acknowledgement of relevant work has been made to a satisfactory level while the title and abstract accurately convey the essence of the research effort described in the paper.

As a conclusion, I believe that the paper is of satisfactory standard and addresses an important subject, that of record-linkage, which is essential for using electronic patient records in health-related research such as in medical or epidemiological studies.