Reviewer's report

Title: Managing Personal Health Information in Distributed Research Network Environments

Version: 2 Date: 16 July 2013

Reviewer: James Boyd

Reviewer's report:

Reviewer's report
-----------------
This manuscript describes a study methodology and associated tools which will help collaborators across health research partner organisations to reduce the risk of data privacy and confidentiality breaches for data sharing projects. Using the methods outlined streamline processes and minimises the transfer of health research data from one organisation to another (i.e. avoiding large-scale data transfer between organisations).

The tools presented in the manuscript are a work plan template and multi-site program checklist. These provide a ‘baseline’ for good practice when managing personal health information and are already used as part of information governance processes and protocols in other fields to ensure data confidentiality and integrity.

- Major Compulsory Revisions

1) The tools described in the manuscript obviously fit into a larger information governance framework to manage the creation, storage, use and deletion of information for collaborative research studies. As these tools will be part of a larger ‘good practice’ governance framework, it would be good to include information on the wider standards, processes and protocols which are being applied and how conflicts or inconsistencies between organisations are addressed (it seems unlikely that data use agreements will be agreed unless organisational process and standards are applied – this must introduce inconsistencies about the processes between organisations).

2) The approach outlines the requirements of the Institutional Review Board in the project approval process but does not provide information of how privacy and ethics approval would be included (e.g. identified through the lead organisations IRB or through separate requirements).

3) Risk assessment must be an important factor in how data use agreements and the study protocols are prepared (especially when combining data from several sources can increase potential for re-identification). It would be good to know how this fits into the process, how the various organisations ensure data is consistently assessed between projects, the parameters involved and if these are
4) Writing consistent programs must be a challenging part of the process (especially as it is carried out by each organisation) – it not clear if standard testing and checking criteria (black-box tests) are applied across organisations.

5) It would be good to discuss how generalisable the tools would be and how easily they can be customised.

- Discretionary Revisions

The manuscript mentions the studies where the tools would be useful i.e. studying rare outcomes, new interventions, and diverse populations across multiple healthcare institutions. The challenge of using personal data from multiple organisations has become a far more frequent research problem. It may be worth identifying other research in the area which identifies similar issues, for example:

- The Commonwealth Government of Australia has defined process for the integration of data across states and portfolios. It identifies integration as one of the most important steps in the process and outlines some of the challenges.

- The Multi-Institutional Linkage and Anonymisation (MILA) model employed by Dundee University for record linkage project has explored robust ways to prepare extracts of data for researchers.

- DataSHIELD – which provides some solutions around performing a pooled analysis of individual-level data without sharing the data.

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

**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