Author's response to reviews

Title: The SAIL Databank: building a national architecture for e-health research and evaluation

Authors:

David V Ford (d.v.ford@swansea.ac.uk)
Kerina H Jones (k.h.jones@swansea.ac.uk)
Jean-philippe Verplancke (p.verplancke@swansea.ac.uk)
Ronan A Lyons (r.a.lyons@swansea.ac.uk)
Gareth John (gareth.john@hsw.wales.nhs.uk)
Ginevra Brown (g.brown@swansea.ac.uk)
Caroline J Brooks (c.brooks@swansea.ac.uk)
Simon Thompson (simon@chi.swan.ac.uk)
Owen Bodger (o.bodger@swansea.ac.uk)
Tony Couch (tony.couch@hsw.wales.nhs.uk)
Ken Leake (ken.leake@hsw.wales.nhs.uk)

Version: 3 Date: 31 July 2009

Author's response to reviews: see over
Dear Editor,

Re: Submission of revised (2nd) paper – The SAIL Databank: building a national architecture for e-health research and evaluation

Please find attached a revised (2nd) copy of a paper entitled: ‘The SAIL Databank: building a national architecture for e-health research and evaluation’. The paper is included in Word (.doc) and there are two separate Word files with Figures 1 and 2. The titles of these additional files are:

Figure1 for national architecture - system
Figure2 for national architecture - split-file

We would once again like to thank the reviewers for taking the time to provide us with thorough and helpful comments, from their expertise in the field, and we believe their reviews have greatly assisted us in improving the paper.

We have noted that the reviewers’ comments this time are minor or discretionary but we have sought to address them nonetheless. But as they are minor or discretionary we sincerely hope that the amendments we have provided will allow you to publish the revised paper swiftly, as we are keen to see the work disseminated.

Reviewer: Fortunato Castillo
Reviewer’s report:
Discretionary Revisions
1. Although the authors provide a reference to a previous publication that describes the record matching process and its efficacy there is significant information that has not been provided. Neither the referred publication nor this manuscript provide details of the total number of records held in the 13 SAIL datasets. The previous publication only examined 3 datasets. It would be informative to the reader to be told the total number of records in each datasets, how many individuals these relate to and how effective the linkage process has proved for the additional datasets.

Sincere apologies for any confusion, but we provided the total number of records in the SAIL databank to date in the revised manuscript (18/6/09). The abstract and the section headed ‘datasets incorporated’ includes that there are over 500 million records in the databank. We have not included exact figures as the number of records is continually increasing. Also, we would like to keep the option of reporting on matching efficacies of other datasets in possible later publications on studies of those datasets.

Reviewer: Mark McGilchrist
Reviewer’s report:
(1) [Minor] on p23 a distributed solution called MILA is described incorrectly. The authors should state without reference to Grid:

MILA (Multi-Institutional Linkage and Authorisation) is a technology-independent methodology for the sharing of distributed data, which can be implemented through any suitable combination of transmission and encryption technologies. Explicit demonstrators of this method are the Scottish Family Health Study (SFHS - www.generationscotland.org/SFHS.htm, a national family-based genetics study) and the Scottish Diabetes Research Network (SDRN - www.sdrn.org.uk). SFHS distributes phenotype, genotype, demographics, identity and routine healthcare data over multiple independent institutions for improved governance (McGilchrist et.al., in preparation).
Thank you for this very helpful information. It has been incorporated into the text about MILA under ‘future work’.

(2) [Discretionary] The authors have responded to my original comments (1)-(7) and I will leave the reader to decide whether they are convincing. The authors might wish to consider the following thoughts when finalising the paper:

(a) It is not obvious that pseudonymisation is any weaker than anonymisation provided appropriate institutional arrangements are made; for example physically separating the keys from the data using separate, independent organisations whose responsibilities are clear. There are many advantages to using pseudonymised data (p4).
The text has been amended to say ‘Other approaches’ rather than ‘Weaker approaches’.

(b) The authors accept that anonymous datasets are open to linkage attacks but believe the proposed arrangements are sufficient protection. The proposal for secure anonymous data transfer to authorised studies (p22) may complicate this, and emphasises that at some point in the future the anonymous datasets may be exposed to other parties. Is the Blue-C supercomputer that processes HIRU’s datasets under HIRU’s control? Is this an example of the datasets being outside the defined jurisdictions? This is an important point to consider for our proposed future work in securely transferring anonymous data to authorised studies. No datasets are outside of HIRU’s control and the internal audit has assessed the security arrangements and control measures of the SAIL system on Blue-C.

(c) Securing confidentiality inevitably involves a multi-faceted solution whatever architecture is chosen (p4). Similar issues arise in each solution, but inevitably there is always more than one way to solve these problems. For example it is not clear (to me) that disclosure control methodologies applied by a central authority are more effective and of lower overall risk than placing researchers under a legal obligation to comply with agreed standards for the use of anonymised datasets and the publication of results.
This is very true, and we are attempting to utilise the best combination of approaches available to us through technology, access, permissions and responsible usage, with shared control and authorisation responsibilities with relevant stakeholders.

(d) Finally, the centralisation of datasets (even anonymous datasets) creates new systemic risks. While extensive governance arrangements have been demonstrated, it cannot be assumed that these arrangements are stable to political and economic changes both inside and outside the collaboration; only experience will tell.
This is also a very valid point. The text of ‘Observations on the SAIL system’ has been amended to reflect this observation (p.20)

Thank you once again for the reviewers’ comments and the editorial guidance. The advice received has been particularly helpful and we hope that the points raised have been addressed to their and your satisfaction. If you have any queries, please contact me and I will endeavour to provide you with any additional information you require.

I look forward to hearing from you shortly and hope this will be sufficient for the paper to be accepted.

Thank you.

Yours sincerely,

Dr Kerina Jones
Research Development Fellow
HIRU

Tel: 01792 602764
Email: k.h.jones@swansea.ac.uk