Reviewer's report

Title: Novel use of three administrative datasets to establish a cohort for environmental health research

Version: 2 Date: 8 October 2014

Reviewer: Stephen Haslett

Reviewer's report:

Review of
Novel use of three administrative datasets to establish a cohort for environmental health research

Lucy F Telfar Barnard1§, Michael G Baker1, Simon Hales1, Philippa Howden-Chapman

The paper looks at what is a very important topic, how best to link records in databases.

It provides results from record linkage of three New Zealand datasets used to create a cohort for environmental health research. As such it documents creation of what is potentially a very useful New Zealand resource for public health and environmental research. It also includes a useful set of historical references related to practical issues of data linkage in New Zealand.

The paper, while clearly of considerable practical importance for future New Zealand public health and environmental research, is more a report of results than a documentation, discussion or use of a range of suitable statistical methodologies for record linkage. It is also not focused on how a range of existing methods for record linkage might have been used to improve matching rates. Consequently, the paper provides only one general reference, namely Jutte, Roos and Brownell (2011).

There are other references on data matching, deterministic and probabilistic record linkage, and machine learning, that would be useful for a broader understanding of the methodological issues that underpin the practical implementation. These include the primer by Hertzog, Scheuren and Winkler (2007) which contains a useful and straightforward bridge between the various academic disciplines and professions involved in data linkage. There is also New Zealand based methodological research, e.g. Blakely and Salmond, (2002). Recent more technical references include Vatsalan,, Christen, O'Keefe and Verykios (2014).

In summary, the paper provides a very useful summary of aspects of the combined database created. Documentation of integration of the methodological and practical aspects of data linkage would have made a useful addition to the
References:


Dr Stephen Haslett
Professor - Statistics
Institute of Fundamental Sciences
Massey University
New Zealand