Reviewer's report

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

Version: 5 Date: 16 February 2015

Reviewer: Rosa Gini

Reviewer's report:

this very interesting paper has considerably improved readability and usefulness for a non-new zealander audience.

==============
Minor revisions
==============

Abstract

The acronym "NZDep2006" should not be used or should be spelled out.

Background

"There is a growing literature on the use of administrative health datasets for public health research[5-8], and some on the availability of non-health administrative data[9, 10]. However, there are few descriptions of experiences in combining non-health datasets with health data."

In fact, some of the data sources used in [5-8] are not administrative, rather clinical or registries or national surveys, and some in [9,10] are indeed health-related data sources, some of which are clearly linked with non-health data sources. A more precise sentence would be

"There is a growing literature on the use of health-related[5-8] data sources for public health research, sometimes linked with other types of data sources [9, 10]. Experiences in combining non-health datasets with health data at a national level are still uncommon."

Methods

In the subsection 'data field' the variable 'modified total ethnicity' of the Census is mentioned. Is the paragraph after this list explaining the content of the variable?
If yes, this should be made explicit and more clear, like "The 'modified total ethnicity' variable has categories 'Maori', 'Pacific', 'Asian', 'non Maori, non Pacific, non Asian' (non-MPA). Each subject can be categorised in more than one of the first three categories. The group non-MPA is predominantly of European ethnicity." (if i well understand)

Discussion

==========

The acronym NZHIS should be spelled out

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