Reviewer's report

Title: A systematic review of barriers to data sharing in public health

Version: 1  Date: 7 August 2014

Reviewer: Rohit Ramaswamy

Reviewer's report:

The authors of the paper are addressing an important issue. Effective collection and sharing of public health data is important both for national or state policy, and for delivering public health programs with quality. However, the conclusions of the paper could be significantly enhanced with some additional detail about the methods and the results. These are all major compulsory revisions and are enumerated below:

1) The definition of what exactly constitutes public health data should be clarified, perhaps with a table. In the methods section, the authors define public health data as “data that were primarily collected by public health agencies for routine purposes such as disease surveillance or program monitoring without primary intention of research ” but this is so broad that it is difficult to comprehend what exactly is encompassed in this definition. Depending on the program or the disease, or whether the data is collected at the local or national level, there can be different barriers to data sharing. In addition to this definition, the authors also point to emerging data sources, such as "genetics, electronic medical records, mobile technology, climate sciences". Clearly, the field of public health data is vast and complex, and becoming more so, and this complexity might create even more barriers to collection and sharing. Some specific characteristics describing the kinds of data that is the focus of is paper will help to better define its scope.

2) Along the same lines, all data are not equal. Since resources are needed for data collection and sharing, it is unrealistic to expect that all data will be collected and shared. Some discussion on what data is most important for sharing, which should be the primary focus of attempts to address barriers would improve the recommendations of the paper for specific courses of action. The discussion section calls for more research to better understand barriers. While this is important, explicitly prioritizing some key pieces of data that are of critical importance and providing recommendations on how collection and sharing of these pieces can be enhanced will enhance the immediate usability of this paper.

3) In the methods section, the authors state that “an initial list of barriers was extracted”. The authors need to state how this was done. How were these barriers identified? For example, did the paper explicitly need to use the word “barriers” or were definitions of barriers agreed to by the team in advance, other than the definition mentioned in the supplementary document. It would be important to know what constituted the definition of a barrier.
4) Some additional clarification on the reasons for eliminating the 92 papers that made it to the full text eligibility screening would be useful.

5) In Table 1 and in the text of the paper, the authors distinguish between “empirical” and “non-empirical” and define non-empirical as “no or little original data presented”. It is not clear what that means, and whether non-empirical studies (and unpublished studies) were given different weights. For example technical barrier number 2 relies entirely on non-empirical and unpublished sources, while technical barrier number 7 has the benefit of multiple studies with empirical data. Is there a difference in the strength of the assertions made about this barrier 2 compared to barrier 7? If it is possible to develop a composite score for each barrier that takes into account the number of studies in which the barrier is reported, and the relative reliability of the study based on the category (empirical, non-empirical, unpublished), it will provide a way of identifying the barriers that are clearly understood and separating them from the ones that need additional research.

6) Following the recommendations in 5 will also help to make the conclusions from the paper stronger and be of greater use to those who are working currently to reduce the barriers. Rather than a generic call for more research, the paper would benefit from a set of concrete recommendations about how to address the barriers about which knowledge already exists (e.g. barrier 1 or 18) and then calling for research on barriers about which little is known (e.g. barrier 11 or 19).

**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 have no competing interests