Reviewer's report

Title: Distributed data processing for public health surveillance

Version: 3 Date: 23 August 2006

Reviewer: Debjani Das

Reviewer's report:

General

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

none

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Minor Essential Revisions:
1. Reference #2 correct title of journal from “Morbidity and Mortality Weely Report”.

Discretionary Revisions (which the author can choose to ignore)

Discretionary Revisions/Comments:

The articles listed below do describe the system and briefly mention advantages and disadvantages of the system. As an objective reviewer, I would say that the rewritten manuscript does go into greater detail of the system.


Even though the manuscript contrasts the system to a centralized system, the authors state in the background that they will describe the system in greater detail in the manuscript. I would have liked to see more detail given to data transmission issues. In particular, I would have liked to hear about what time of day the files from the sites arrive and what happens when data are missing from one or more sites. It would be interesting to hear how easy/difficult it is to troubleshoot data transmission problems with these offsite data providers. I would have also liked to hear about how much time on average it takes a site from initial contact to start sending data. It wasn’t clear whether NDP personnel visited each site to set up the PHIN-MS service and train IT personnel on how to use the service/server.

One advantage mentioned in the paper was the fact that events for the same patient may be collapsed if they occur within a certain time frame. However, I also was not sure how this was done. Is there a de-duping process in place at the data provider side? On what basis are duplicates identified? In some institutions, emergency department data may be updated with a different chief complaint or zip code or age. This may result in a duplicate record which may be seen as another unique individual visiting the ED.

The line list reports never get transmitted to the datacenter, but I wasn’t sure how they get transferred to the public health authorities. I realize that these reports only get sent to public health authorities as warranted, but the author may want to mention if the line list report gets manually transmitted through your secure system (or if this will be so in the future) or if the manual report transmission is independent of the electronic system.

One thing I did notice is that data are aggregated by zip code. If only one or two patients were reported from
a particular zip code, this could identify the patients.

The authors may want to mention any changes that they are making in the near future to the system or even any further goals they hope to achieve with the system.

The authors may want to note how many current data providers are in the system since it is not mentioned elsewhere.

I would also suggest removing Figure 2 since the content of the screen is presented in tables 2 and 3.

**What next?:** Accept after discretionary revisions

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests