Author's response to reviews

Title: Distributed data processing for public health surveillance

Authors:

Ross Lazarus (ross.lazarus@channing.harvard.edu)
Katherine Yih (Katherine_Yih@harvardpilgrim.org)
Richard Platt (richard_platt@channing.harvard.edu)

Version: 4 Date: 13 September 2006

Author's response to reviews: see over
Reviewer 1 (Das)

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

Done, thanks for pointing this out

**Discretionary Revisions (which the author can choose to ignore)**

*Discretionary Revisions/Comments:*

.....

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.

We have added some additional detail and explanation within the limited scope of our debate. It is one issue we have almost no problems with. We use standard CDC supported messaging software.

**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?**

We have tried to clarify this even more. The identifiable data held at each distributed site make this trivial and of course that is the only place where it can be reliably done – so that's how it's done, in the software we distribute using local patient master index identifiers.

**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**

We collect no data from EDs. In general our providers are very good at keeping track of individual insured or covered patients for billing purposes and duplication is probably very rare.
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.

That’s a local decision and is completely up to the provider. We have discussed automating this process but it is not something providers are comfortable with.

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.

See previous revision response to reviewer 1.

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.

Done.

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

Done.