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: 3 Date: 31 July 2006

Author's response to reviews:

Editor

BMC Public Health

Dear Sir,

With thanks to the two reviewers, we submit a revised version of our MS which takes into account all of the helpful suggestions made. We hope that the MS is now suitably improved and strengthened. A copy of the revised MS marked up with all changes is attached as BMCPH_distribdataproc27July2006_showing_changes.pdf

Our responses to each point made by each reviewer are set out in a separate pdf file (BMC_1389358455984688_response_june22.pdf) whose text is set out below.

Yours truly

(Ross Lazarus, for the authors)

============================================================

BMC Public Health MS 1389358455984688

Distributed data processing for public health surveillance

Author response to Reviewers
Background: We have added material making the advantages in terms of establishing collaboration with data custodians, of transmitting and centrally collating only HIPAA compliant, de-identified data even more clear.

Disadvantages to researchers: We have tried to make the fundamental point that we adopted this strategy in order to gain cooperation with data custodians, despite the disadvantages pointed out in the Discussion for researchers, even more obvious.

Methods: The Editor might have already explained to the Reviewer that no Methods section is stipulated in the guidelines for authors. Please see http://www.biomedcentral.com/bmcpublichealth/ifora/?txt_jou_id=2024&txt_mst_id=2010

Results: As in Methods above.

Discussion: We have added explanatory material on the use of CDC mandated standards in CDC funded research.

On page 11, we wrote:

"Current capacity of the NDP to respond to public health needs

When unexpectedly high counts of particular syndromes are detected in geographically defined areas, the data center automatically generates electronic alerts, which are automatically routed to appropriate public health authorities. For example, in Massachusetts, electronic messages are automatically sent to the Massachusetts Alert Network where they are forwarded to the appropriate personnel for follow up."

Given this reasonably explicit text, the Reviewer is mistaken in asserting that "Currently the system is only being used to alert medical providers or others at the participating health plans". We welcome specific suggestions from the Reviewer on making the point even more transparently obvious, but we believe that the original text should be adequate for all but the most cursory readers.

Timeliness: We have added material on timeliness although most of this is related to implementation details within the relevant health department, well outside our primary purpose of provoking debate about distributed processing. As now discussed, on average, alerts are transmitted to the HAN within minutes of being generated. Subsequent communication to subscribers within the HAN is as close to real-time as each subscriber requires, within the limits of current technical capacity - for example, a subscriber can choose to be alerted by a telephone call with the alert message text read announced by a text to speech translator, at their home telephone or via a cell phone depending on the time of day. The HAN offers fallback delivery methods including email and fax if the preferred delivery mechanism fails to elicit an acknowledgement of receipt.
Recommendation: We are not aware of any prior publications describing or debating the merits and disadvantages of a distributed BT or Public Health surveillance strategy, or of other systems offering configurable, near real-time potential event alerting for public health authorities, that we have described. We must assume that an objective reviewer would only have made the assertion that the MS contains little of novelty or value, on the basis of refereed publications already covering all of the material addressed in our MS, so we ask the Reviewer to make those available, since we clearly must have missed them all.

If these are not available, we will assume that the lack of evidence of prior art, the new insights gained from our comments in response to the review, and the changes we have made to the MS, would likely result in a less dismissive assessment by any fair and objective Reviewer.

Reviewer 2 (Muscatello)

We thank the reviewer for the thoughtful and thorough review. We have adopted all of the constructive and helpful suggestions and fixed the errors.

General: The abstract has been reformatted to the headings recommended in the guidelines.

Major compulsory revisions:

Context: We have tried to provide context where this may be unfamiliar to non-North American readers. However, as now stated in the MS, we believe that although implementation details are highly context specific, the idea of using a distributed approach with preliminary aggregation to decrease inadvertent PHI disclosure risk, is a generalisable model addressing an important problem in most developed countries.

Applicability: We have clarified the applicability of the approach. We focused on relatively large practices since they tend to have adequate counts for the GLMM models, appropriate EMR systems operating, and IT infrastructure support. In principle the design is applicable to existing electronic medical record systems using ICD9 codes for encounters with spare disk and CPU capacity, and an outgoing internet connection, provided that the vendor is willing to provide a module to extract relevant data in the required file formats. Small counts in each cell may make the statistical inference from a single small practice unhelpful, but we believe that combining counts from multiple small EMR systems might overcome this scale problem. We have expanded the applicability section to reflect these useful points.

Funding: Funding arrangements to cover costs for participating providers were part of our CDC grant, and this point has been added to the discussion.

Transmission errors: Data transmission failures are relatively rare in our experience, and PHIN-MS performs automated periodic retransmission until successful message delivery is acknowledged. We have had a web-based system to monitor and review data flow since the second year of the project, which is now briefly described.
Payer type: Payer type is related to specifics of local reimbursement and insurance arrangements. We do not distinguish payer type in data leaving the provider so it is only available internally or to public health authorities during an investigation when it is of potential importance for local context. For example, a cluster amongst Veterans Administration payer type patients might provide a clue about a potential collective exposure. To simplify the paper, we have removed it from the Figure since it is not central to our argument.

Identifiability: We have added comments about this problem.

Minor Essential Revisions:

Acronyms are now all spelled out on first use

Figures now have explanations for acronyms

Syntax: The offending sentence has been repaired.

Recurrence interval: The term "recurrence interval" is no longer in the text

Discretionary Revisions:

Source code: A statement about source code availability has been added.

Remote processing time: We do not agree that this is sufficiently advantageous to advocate - the time to process even a large number of encounters is relatively small, so the cost of doing this centrally is small - having the data at a central location is far more convenient and we would always prefer to use that traditional approach because it is so much more flexible. We advocate the distributed approach because it may be the only feasible way to convince data custodians to participate - as was the case for our system.

Commercial confidentiality for data custodians: this is an issue we would frankly prefer not to raise. Our commercial partners are very mindful of this, but choose to participate, being well aware of the risks.