Reviewer's report

Title: Distributed data processing for public health surveillance

Version: 2 Date: 2 June 2006

Reviewer: David Muscatello

Reviewer's report:

General

Clear and well written. Important because it provides a practical method that health departments can adopt that can address the delicate balance between public health surveillance needs, patient and provider confidentiality, surveillance feedback and adequately supporting public health response.

I'm puzzled that the abstract has Background, Methods, Results and Conclusion, but the main manuscript does not have all these sections. A consistent structure should be used for the manuscript. For example, much of the Discussion section could fit under the heading of Methods.

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

My main concern, particularly for an international audience, is that the health-care delivery and public health context in which the system operates isn't clear. While the paper's intentions are to generally describe how the system operates, the context is important in guiding other potential users of this method as to what challenges they might face in implementing it in their own context. Does it operate in hospital emergency departments/rooms, large ambulatory/primary care clinics, small private medical practices? Are health-care providers required to participate or is it voluntary? I find it hard to imagine that the system would successfully operate in small private medical practices, given the multiplicity of information systems that are likely to be used. Even though clinical providers can easily produce simple files for input into the data aggregator, I'd be surprised if diverse information systems use consistent data definitions for the same data items. Unless those private practices participate in an information system that is centrally resourced through a larger data centre.

Are clinical providers reimbursed for their efforts in changing their information system to provide input into the data aggregator?

Do data transmission failures occur? What resources are required for monitoring and dealing with data transmission failures from participating health-care providers at both the surveillance data centre end and the clinical provider end?

Please state why 'payer type' is important to collect for surveillance (Table 1). I understand that socioeconomic status is an important public health risk factor, but I can't immediately see why it's important for syndromic surveillance.

The aggregate data feed includes counts of 1 patient in some rows (Table 1), and may be potentially identifying. While there are limited other identifying data items, the potential identifiability needs to be discussed as well as why it is still acceptable.

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

Some acronyms are not explained in the main text of the manuscript (GLMM, NDP, SSL).

Figures need acronyms explained in a footnote or the figure itself (eg EMR, EPHI in Figure 1).

The first sentence at the end of page 12 does no make sense "It would be straightforward to would send line
lists..."

Page 12: "recurrence interval" is a term that would be unfamiliar to many readers.

Discretionary Revisions (which the author can choose to ignore)

It would be useful to know whether the various software tools developed for the project are publicly available and if, so, how does one get them.

Another advantage that could be mentioned is that pre-processing of data at the data provider would reduce the data processing time required at the surveillance data centre. This is a limiting factor in real-time surveillance and should be more widely recognised.

Is the confidentiality of the clinical provider a matter that should be discussed? There are likely to be commercial reasons that a provider may be concerned about the release of these data.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory 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