Reviewer's report

Title: PHSkb: A Knowledgebase to Support Notifiable Disease Surveillance

Version: 1 Date: 28 March 2005

Reviewer: Antoine Flahault

Reviewer's report:

Summary

The main objective of this work was to improve sensitivity, timeliness and quality of notifiable disease surveillance through interoperable computer knowledge database facilitating electronic data exchange. The authors propose a friendly user tool to be used in a context where controlled vocabulary is needed for data aggregation in public health surveillance systems. Public Health Surveillance knowledgebase (PHSkb) is expected to facilitate automated case detection and surveillance decision support, by linking, understanding, and inferring reportable conditions from clinical, laboratory and public health surveillance information systems. When routine notifiable disease surveillance often suffers from incomplete reporting and poor timeliness, it is assumed that electronic data interchange may improve it. Several sources of medical information, such as laboratory findings, coded discharge diagnoses or pharmacy dispensing data could be used to create tables that link this information to notifiable diseases under surveillance. Systematized nomenclature of medicine (SNOMED) and logical observation identifier names and codes (LOINC) were chosen by the authors as vocabulary standards for integrating these tables into the Public Health Surveillance knowledgebase (PHSkb).

General

This paper provides an interesting model for notifiable diseases, however it has not demonstrated that PHSkb will really contribute to improve sensitivity, timeliness and quality of notifiable disease surveillance as it is initially presented. This should be more clearly stated. It is however a first step in a better understanding of the domain using ontology development. Using ontology development and knowledge representation is an original approach in public health informatics. This approach has been more extensively used in clinical decision making or health information retrieval.

Authors identify root classes (373 reportable diseases, 174 infectious agents, vector, 48 substances, 3 procedures, and 58 individual or population findings) in the ontology and their relations to one another. Instances of the major classes were classified hierarchically with SNOMED CT which allows to harmonize term variations between the 52 US jurisdictions currently reporting notifiables diseases to the CDC.

It seems clear that without a hierarchical representation, it may be difficult to integrate data from various jurisdictions, because different jurisdictions have a high level heterogeneity. In addition - but that remains to be proved - this tool may be used as an inference engine to identify reportable events from medical observations. That would change the historical paradigm of the passive nature of notifiable disease surveillance towards a more active system of reporting. However, as the authors themselves quoted « further advances are needed in the area of natural language processing and automated methods for converting text data to electronic vocabulary standards ».

To date, content of PHSkb is not validated with state surveillance partners. There may be a long way before this tool is used on an ongoing basis, since it is still not distributed in the USA jurisdictions.
**What next?:** Accept after minor essential 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