Emerging Health Threats Journal 2011, 4:s125. doi: 10.3134/ehtj.10.125

www.eht-journal.org

ABSTRACT

Real-time surveillance of influenza/pneumonia deaths: new strategies using grid computing and natural language processing

K Davis[1], C Staes[1], R Price[2], J Duncan[3], S Igo[1,2], and JC Facelli[1,2]

[1]Department of Biomedical Informatics, Salt Lake City, UT, USA; [2]Center for High Performance Computing, University of Utah, Salt Lake City, UT, USA; and [3]Utah Department of Health, Salt Lake City, UT, USA
E-mail: kailah.davis@utah.edu

Objective

This poster presents the rationale for designing a real-time surveillance system, based on mortality data, using grid and natural language processing tools that will address the current barrier that coded death certificate data not being available for several months. To develop a public health tool that delivers a timely surveillance system for influenza and pneumonia, we integrated death certificates from the Utah Department of Health, analytical grid services, and natural language processing tools to monitor levels of mortality. This example demonstrates how local, state, and national authorities can automate their influenza and pneumonia surveillance system, and expand the number of reporting cities.

Introduction

Surveillance of deaths due to influenza and pneumonia using death records has the potential to be a relatively inexpensive[1] and quick approach to tracking and detecting influenza and respiratory illness outbreaks; however, presently such a system does not exist because of the time delays in processing death records: in Utah, as is similar elsewhere in the United States, coded death certificate data are typically not available for at least 1–3 months after the date of death, and coded national vital statistics data are not available until after 2–3 years.[2]

Advanced informatics methods and tools can be applied to address the problems described above. Grid architecture is a promising methodology for the public health domain because it may combine and analyze unrelated data existing from independent domains, and provide computational and analytic resources on demand.[3] Natural language processing tool can be used to automate the coding of 'cause of death' free text and allows mapping to other standardized codes. Coded information about causes of death is needed to integrate death certificate information into real-time surveillance tools in order to trigger alarms for public health response.

Our objective is to describe a new model for influenza surveillance using NLP tools and grid-enabled data and analytic services.

Methods and preliminary results

The Utah Department of Health has made available de-identified death records from 1989 to 2008. Analytic methods have been developed using R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria to mimic analysis performed by CDC to detect outbreaks. The available data were divided into fifteen years that were used as baseline data and five years that were used as to simulate a real-time data feed. The historic baseline data were used to establish baseline trends and computed threshold values, which were compared with counts to trigger automatic alarms with the simulated real-time feed.

Future work

A grid-based analytic service will be developed using the analytical tool described above. Also, we will explore the feasibility of creating a grid version of the National Library of Medicine's natural language processing tool, MetaMap(http://mmtx.nlm.nih.gov/), which can be used to access the simulated real-time feed of the death records and provide coded information from the 'cause of death' field. Data grid services may be developed to simulate death records located in different administrative domains. For example, access to death records in Utah and neighboring states to identify regional outbreaks.

Conclusions

We expect that the integration of electronic death records with grid-enabled analytic and NLP tools will result in timely and flexible applications for public health surveillance. This

new non-traditional public health surveillance system will allow aggregation across jurisdictions, dynamic monitoring of diseases based on regional or national threats, and allow resource-limited public health agencies to access and share advanced informatics tools.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

1 Dubrow R, Sestito JP, Lalich NR, Burnett CA, Salg JA. Death certificate-based occupational mortality surveillance in the United States. *Am J Ind Med* 1987;**11**:329–42.
2 Center for Disease Control and Prevention. *Death Certificate (document on the Internet)*. Center for Disease Control and Prevention: Georgia, 2006. update 2006 Aug 26, cited 2010 Fed 09]. Available from: http://www.cdc.gov/NCIPC/pub-res/nvdrs-coding/Death-Certificate.pdf.
3 Staes CJ, Xu W, LeFevre SD, Price RC, Narus SP, Gundlapalli A, *et al*. A case for using grid architecture for state public health informatics: the Utah perspective. *BMC Med Inform Decis Mak* 2009;**9**:32.