Localized Cluster Detection Applied to Joint and Separate Military and Veteran Subpopulations

Howard Burkom*[1], Yevgeniy Elbert[1], Carla Winston[2], Julie Pavlin[3], Cynthia Lucero-Obusan[2] and Mark Holodniy[2]

[1]Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA; [2]Office of Public Health Surveillance Research, Veterans Health Administration, Palo Alto, CA, USA; [3]Armed Forces Health Surveillance Center, Silver Spring, MD, USA

Objective

We examined the utility of combining surveillance data from the Departments of Defense (DoD) and Veterans Affairs (VA) for spatial cluster detection.

Introduction

The Joint VA/DoD BioSurveillance System for Emerging Biological Threats project seeks to improve situational awareness of the health of VA/DoD populations by combining their respective data. Each system uses a version of the Electronic Surveillance System for Early Notification of Community-Based Epidemics (ESSENCE); a combined version is being tested.

The current effort investigated combining the datasets for disease cluster detection. We compared results of retrospective cluster detection studies using both separate and joined data. — Does combining datasets worsen the rate of background cluster determination?

— Does combining mask clusters detected on the separate datasets?

— Does combining find clusters that the separate datasets alone would miss?

Methods

Cluster determination runs were done with a spatial scan statistics implementation previously verified [1] by comparison with SaTScan software [2] using DoD data from the Biosense system.

Input data files were extracted from a repository of outpatient records from both DoD and VA facilities covering 4 years beginning Jan. 1, 2007. This repository includes over 37 million DoD records and over 86 million VA records. Input files were matrices of daily Influenza-like-Illness (ILI) or gastrointestinal (GI) visit counts. Matrix rows were consecutive days, columns were patient residence zip codes, and entry (i, j) was the number of visits on day i from with zip code j. These files were made for DoD data, VA data, and combined data.

For assessing the alerting burden from combining datasets, three sets of runs were executed using data from three regions, Baltimore/Washington D.C. (dominated by DoD data), Los Angeles (mainly VA data), and Tampa (representation of both). For each region, sets of 1672 daily runs were executed for ILI and GI syndrome data. Lastly, focused runs were done to investigate known outbreaks in New York (GI, Jan-Mar 2010), San Diego (ILI, Dec 2007-Apr 2008 and Fall 2009), and New Jersey (GI, Jan-Mar 2010).

Results

Combining the data sources increased the rate of significant cluster alerting by a manageable 1-10% across run sets. Some clusters found only when the data were combined persisted over several days and may have indicated small events not reported in either system; however, we were unable to validate minor events that may have occurred in past years.

Retrospective looks at known outbreaks were successful in that clustering evidence found in separate DoD and VA runs persisted when data sets were combined. For the New York run, a West Point outbreak was seen in repeated clusters of combined data, beginning days before the event report. However, clustering did not consistently produce alerts before outbreak report dates. In the New Jersey DoD runs, repeated clusters indicated a 10-week GI outbreak at Fort Dix; adding VA data that dominated the record counts gave the same clusters with no added cases, so the DoD event was probably self-contained. The San Diego runs were aimed at detecting unusually severe influenza epidemics in February 2008 and in the fall of 2009, and numerous clusters were found but did not enhance regional disease tracking.

Conclusions

From the analysis, combining DoD and VA data enhances cluster detection capability without loss of sensitivity to events isolated in either population and with manageable effect on the customary alert rate. For cluster detection, there may be many geographic regions where a health monitor in one of the systems would benefit from combined data. More detailed outbreak information is needed to quantify the timeliness/sensitivity advantages of combining datasets. In events examined, clustering itself yielded an occasional but not consistent timeliness advantage.

Keywords

ESSENCE; Department of Defense; scan statistics; cluster detection; Veterans Administration

References

[1] Xing J, Burkom H, Moniz L, Edgerton J, Leuze M, and Tokars J. Evaluation of sliding baseline methods for spatial estimation for cluster detection in the biosurveillance system, International Journal of Health Geographics 2009, 8:45

[2] SaTScan: Software for the spatial, temporal, and space-time scan statistics. www.satscan.org (last accessed 20Aug2012)

*Howard Burkom
E-mail: howard.burkom@jhuapl.edu