Emerging Health Threats Journal 2011, **4**:s7. **doi:** 10.3134/ehtj.10.007
© 2011 A Baer *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

EDITORIAL

International Society for Disease Surveillance Conference 2010: Enhancing the Synergy Between Research, Informatics, and Practice in Public Health

A Baer[1], DR Olson[2], and JS Brownstein[3]

[1]*Public Health, Seattle & King County, Seattle, WA, USA;* [2]*International Society for Disease Surveillance, New York, NY, USA and* [3]*Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA*
E-mail: drolson@gmail.com

The International Society for Disease Surveillance (ISDS) was created to improve population health by advancing the field of disease surveillance. In support of this mission, ISDS provides educational and scientific forums where researchers, informaticists, public health practitioners, and healthcare providers can explore and address population health monitoring across institutional and professional boundaries.

As the scope of many journals tends to be focused on only a single discipline, advances in the field of population surveillance have been scattered across the scientific literature. In order to improve opportunities for its diverse membership to publish its work in a more cohesive way, ISDS has sought to create a partnership with a journal that will promote publication of cross-disciplinary approaches to population health monitoring. Also, acknowledging the severe financial pressures in the public health setting, ISDS sought a journal that was both free to publish in and openly accessible. ISDS is proud to announce its new partnership with the Emerging Health Threats Journal (EHTJ). As a first step towards this partnership, ISDS and EHTJ are collaborating to publish the 2010 ISDS Annual Conference abstracts.

The ISDS annual conference brings together researchers and practitioners in public health, epidemiology, health policy, biostatistics and mathematical modeling, informatics, computer science, and related fields focused on biosurveillance and emerging challenges to public health practice. The theme of the ninth annual conference, held on 1 and 2 December in Park City, Utah, was, 'Enhancing the Synergy Between Research, Informatics, and Practice in Public Health.' The conference was preceded by a workshop focused on disease surveillance standards, the interaction between disease surveillance and the International Health Regulations, and an introduction to syndromic surveillance. There were also several satellite meetings and user's groups focused on syndromic surveillance systems, global surveillance, and syndrome standardization.

The conference featured plenary, panel, and parallel sessions covering a wide range of topics addressing the following themes: Analytics/Research Methodologies, Public Health Surveillance/Practice, Applications of Methodologies to New Domains, Informatics Applications/Practice, and Informatics Architectures/Integration/Interoperability. A keynote address on bridging theory, technology, and practice in public health surveillance was given by Dr Samuel Groseclose, Acting Director of the Division of Healthcare Information, Public Health Surveillance Program Office (PHSPO), US Centers for Disease Control & Prevention. A special call for proposals was announced to address the topic of data quality, which was featured as a panel presentation. Other plenary and panel presentations focused on the implications of meaningful use on research and practice, and the use of social networking tools for surveillance. A total of 186 abstracts were submitted to the conference; abstracts accepted for either an oral or poster presentation are published in this supplemental issue of the Journal.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

International Society for Disease Surveillance Conference 2010

Track 1: Analytics / Research Methodologies

Emerging Health Threats Journal 2011, **4**:s8. **doi:** 10.3134/ehtj.10.008
© 2011 T-C Lu and N Abernethy.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Using cross-correlation networks to identify and visualize patterns in disease transmission

T-C Lu and N Abernethy

Biomedical and Health Informatics, School of Medicine, University of Washington, Seattle, WA, USA
E-mail: neila@uw.edu

Objective

Time series of influenza-like illness (ILI) events are often used to depict case rates in different regions. We explore the suitability of network visualization to highlight geographic patterns in this data on the basis of cross-correlation of the time series data.

Introduction

Syndromic surveillance data such as the incidence of influenza-like illness (ILI) is broadly monitored to provide awareness of respiratory disease epidemiology. Diverse algorithms have been employed to find geospatial trends in surveillance data, however, these methods often do not point to a route of transmission. We seek to use correlations between regions in time series data to identify patterns that point to transmission trends and routes. Toward this aim, we employ network analysis to summarize the correlation structure between regions, whereas also providing an interpretation based on infectious disease transmission.

Cross-correlation has been used to quantify associations between climate variables and disease transmission.[1,2] The related method of autocorrelation has been widely used to identify patterns in time series surveillance data.[3] This research seeks to improve interpretation of time series data and shed light on the spatial–temporal transmission of respiratory infections based on cross-correlation of ILI case rates.

Methods

For this pilot study, we analyzed patient visits to health care providers for ILI, collected through the US Outpatient Influenza-like Illness Surveillance Network (ILINet). Aggregate data for the 27-week period from the 35th week of 2009 to the 9th week of 2010 were used. The model involves the 10 Human and Health Services (HHS) regions for which ILI data are publicly available through the CDC. The data consist of the weighted percentage of all patient visits to healthcare providers for ILI reported each week. Additional networks were generated using confirmed cases from the 2009 H1N1 pandemic and city-level data from Google Flu Trends.[4] Using the cross-correlation function to measure maximal correlation and the corresponding latency

Figure 1 Cross-correlation network representing fraction of ILI cases in the 10 HHS regions during the 2009–2010 influenza season. The corresponding latency between time series is represented by directed ties in the graph. Gray (bidirectional) edges represent regions correlated with no time lag.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

between each pair of regions, the resulting matrix was visualized with the R statistical package.

Results

We demonstrate this technique on several data sources. An example network (Figure 1) shows ILI correlations between HHS regions. For clarity, a minimum threshold was used to highlight regions having the most similar rates. Regions four and nine are disconnected from the core network because of weak correlation with other regions. These outliers reflect an early peak (region four) and weak transmission (region nine) of ILI cases in hospital visit data. Other observations from the network are discussed. We demonstrate how this visualization motif might enhance existing tools such as DISTRIBUTE.

Conclusions

Network analysis has a role in several aspects of infectious-disease modeling, including simulation, contact investigation, and sampling. We have proposed a new technique to interpret the correlation between case rates in geographic regions. This method may help epidemiologists to quickly visualize the similarities in case rates between regions. In future work, we plan to validate the ability of these networks to reveal patterns in disease transmission. Comparison of cross-correlation networks to transmission routes such as airline networks or US Census worker flow data may reflect the route of transmission of new outbreaks or diseases such as SARS. The networks may then serve as a basis to evaluate intervention options during outbreak of pandemic influenza or other emerging infectious diseases.

Acknowledgements

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

References

1 Naish S, Hu W, Nicholls N, Mackenzie JS, McMichael AJ, Dale P, *et al*. Weather variability, tides, and Barmah Forest virus disease in the Gladstone region, Australia. *Environ Health Perspect* 2006;**114:**678–83.
2 Zhang WY, Guo WD, Fang LQ, Li CP, Bi P, Glass GE. Climate variability and hemorrhagic fever with renal syndrome transmission in Northeastern China. *Environ Health Perspect* 2010;**118:** 915–20.
3 Espino JU, Hogan WR, Wagner MM. Telephone triage: a timely data source for surveillance of influenza-like diseases. *AMIA Annu Symp Proc* 2003; 215–19.
4 Valdivia A, Lopez-Alcalde J, Vicente M, Pichiule M, Ruiz M, Ordobas M. Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks—results for 2009–10. *Euro Surveillance* 2010;**15**, pii: 19621.

Emerging Health Threats Journal 2011, **4**:s9. **doi:** 10.3134/ehtj.10.009
© 2011 AL Buczak *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Methodology for prediction of outbreaks of diseases of military importance

AL Buczak, PT Koshute, SM Babin, BH Feighner, and SH Lewis

Applied Physics Laboratory, Johns Hopkins University, Baltimore, MD, USA
E-mail: anna.buczak@jhuapl.edu

Objective

This paper addresses the problem of predicting outbreaks of diseases of military importance in a chosen region of the world, one to several months in advance.

Introduction

Traditional public health practice has relied on public health surveillance of disease to detect outbreaks in an effort to mitigate their effects. Often the earlier an outbreak is detected, the greater the mitigation of its effects. The logical extension of this relationship is to predict outbreaks before they occur. A predictive model for an emerging infectious disease would forecast, when and where an outbreak of a given disease will occur, well before its emergence. This is a challenging task and truly predictive models for emerging infectious diseases and is still in their infancy.

Methods

Burnette, *et al.*[1] developed a methodology for identifying infectious diseases of military importance, which identified malaria, bacterial diarrheal diseases, and dengue fever as the top three endemic threats to deployed US forces. Of those three diseases, we believe dengue fever is the most suitable for predictive modeling. Dengue was at one time a significant disease in the United States and concern about its potential domestic re-emergence is growing.[2] Dengue fever is recognized in over 100 countries and there are an estimated 50–100 million cases of dengue fever annually.[2]

Predictive disease modeling attempts to exploit the complicated relationship between disease outbreaks and measurable environmental, biological, ecological, and socio-political variables. Previous studies[3–5] identified several factors associated with dengue outbreaks. The most common include past cases, ambient temperature (daily or weekly mean, min, and max), precipitation (weekly, cumulative), and relative humidity (mean, min). Less commonly used factors include Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index, Southern Oscillation Index, Sea Surface Temperature, socio-political stability, sanitation, altitude, wind, cloudiness, seroprevalence in humans, mosquito infection rate, mosquito biting activity, and public health interventions.

For rigorous disease prediction, all predictor variables need to be collected for the previous time period (for example, month) and be used for prediction of outbreaks during a later time period. This ensures a realistic prediction, that is, one in which the values of all the predictor variables can be obtained before performing prediction for the next time period.

Another necessary step when developing disease prediction methods is their validation. The model needs to be validated on a part of the data set, which were not used in its development. The prediction accuracy should then be described in terms of sensitivity (the proportion of actual outbreaks correctly identified) and specificity (the proportion of non-outbreaks correctly identified).

We are presently in the process of collecting the predictor data, the dengue epidemiological data, and developing dengue prediction models including linear regression and Support Vector Machines.

Conclusions

Effective methodologies to predict outbreaks of diseases of military importance would allow taking the preventive actions early to avert large epidemics. For best results, the researchers must have access to data streams with timely, detailed, and accurate values of predictor variables. Model validation is of paramount importance as health officials may be unlikely to spend resources on mitigation efforts based on model predictions without evidence of accuracy on past outbreaks.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

1 Burnette WN, Hoke CH, Scovill J, Clark K, Abrams J, Kitchen LW, *et al*. Infectious diseases investment decision evaluation algorithm: a quantitative algorithm for prioritization of naturally occurring infectious disease threats to the US military. *Military Medicine* 2008;**173**:174–81.

2 Guzman MG, Kouri G. Dengue: an update. *Lancet* 2002;**2**: 33–42.

3 Lu L, Lin H, Tian L, Yang W, Sun J, Liu Q. Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health* 2009;**9**:395.

4 Hu W, Clements A, Williams G, Tong S. Dengue fever and El Nino/Southern Oscillation in Queensland, Australia: a time series predictive model. *Occup Environ Med* 2010;**67**:307–11.

5 Choudhury Z, Banu S, Islam A. Forecasting dengue incidence in Dhaka, Bangladesh: a time series analysis. *Dengue Bulletin* 2008;**32**:29–36.

Emerging Health Threats Journal 2011, **4**:s10. **doi:** 10.3134/ehtj.10.010
© 2011 H Burkom *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Analytic disease surveillance methodology based on emulation of experienced human monitors

H Burkom, JS Coberly, and SL Lewis

Johns Hopkins University Applied Physics Laboratory, National Security Technology Department, Baltimore, MD, USA
E-mail: howard.burkom@jhuapl.edu

Objective

This presentation gives a method of monitoring surveillance time series on the basis of the human expert preference. The method does not require detailed history for the current series, modeling expertize, or a well-defined data signal. It is designed for application to many data types and without need for a sophisticated environment or historical data analysis.

Introduction

Recently published studies evaluate statistical alerting methods for disease surveillance based on detection of modeled signals in a data background of either authentic historical data or randomized samples. Differences in regional and jurisdictional data, collection and filtering methods, investigation resources, monitoring objectives, and system requirements have hindered acceptance of standard monitoring methodology. The signature of a disease outbreak and the baseline data behavior depend on various factors, including population coverage, quality and timeliness of data, symptomatology, and the care-seeking behavior of the monitored population. For this reason, statistical process control methods based on standard data distributions or stylized signals may not alert as desired. Practical algorithm evaluation and adjustment may be possible by judging algorithm performance according to the preferences of experienced human monitors.

Methods

In this approach, introduced at the 2009 ISDS Conference, consensus alerting preferences of human monitors based on historic data time series were used as a gold standard. The study data were highly aggregated, unlabeled time series from the National Notifiable Disease Surveillance System approved for public release by the US Centers for Disease Control and Prevention. Multiple unmarked time series with known outbreak effects were given to 18 experienced epidemiologist raters in an EXCEL tool. Raters were asked to mark dates of desired alerts for possible outbreaks on 18-weekly time series each spanning 2–3 years. Individual response sets and derived consensus sets were used as 'truth data' for calculation of sensitivities and false alarm rates for the purpose of comparing candidate algorithms. These 136 algorithms included 45 CUSUM variants at each of three threshold levels and the standard CDC Historical Limits method.

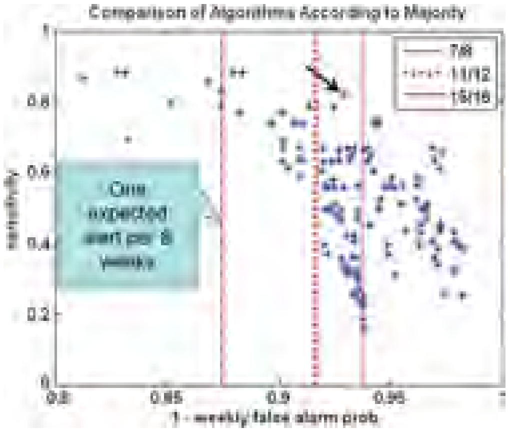


Figure 1 Sensitivity comparison among 136 algorithm variations based on rater majority. Subsequent applications to weekly time series of reportable disease data will also be discussed from (a) a state health department and (b) a resource-limited setting.

Results

An adaptive CUSUM variant with $k = 0.5$ and an 8-week sliding baseline yielded the best sensitivity/false-alarm-rate combination according to the consensus set and to multiple individual sets, as indicated by the arrow in the Figure 1.

Conclusions

On the basis of agreement analysis and the stability of algorithm performance comparisons, the approach is viable and worthy of comparison to conventional approaches based on known data distributions and signal types.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s11. **doi:** 10.3134/ehtj.10.011
© 2011 E Chan *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Forecasting high-priority surveillance regions: a socioeconomic model

E Chan[1], T Choden[1], T Brewer[2], L Madoff[2], M Pollack[2], and J Brownstein[1]

[1]*Children's Hospital Informatics Program, Children's Hospital Boston, Boston, MA, USA; and* [2]*ProMED-Mail, International Society for Infectious Diseases, Brookline, MA, USA*
E-mail: emily.chan2@childrens.harvard.edu

Objective

To evaluate the association between socioeconomic factors and infectious disease outbreaks, to develop a prediction model for where future outbreaks would most likely to occur worldwide and identify priority countries for surveillance capacity building.

Introduction

It has been suggested that changes in various socioeconomic, environmental and biological factors have been drivers of emerging and reemerging infectious diseases,[1,2] although few have assessed these relationships on a global scale.[3] Understanding these associations could help build better forecasting models, and therefore identify high-priority regions for public health and surveillance implementation. Although infectious disease surveillance and research have tended to be concentrated in wealthier, developed countries in North America, Europe and Australia, it is developing countries that have been predicted to be the next hotspots for emerging infectious diseases.[3]

Methods

We constructed negative binomial regression models to analyze the relationship between a set of outbreaks reported by the WHO during 1996–2009 and 60 national socioeconomic variables from the World Bank's World Development Indicators database; a Human Development Index (HDI) reported by the UN Development Programme; and a Democracy Index developed by The Economist. Initial 'univariate' models examined one socioeconomic variable at a time, while controlling for latitude, as parasitic and infectious disease species richness has been associated with latitude.[4] All significant variables ($\alpha = 0.0008$ with Bonferroni correction for multiple comparisons) were then combined in a multivariate model. Variables that were still significant ($\alpha = 0.05$) comprised the final model, which was then used for forecasting using the latest available socioeconomic data for each country.

Results

Birth rate, measles immunization, urban population with access to improved sanitation facilities, life expectancy, infant mortality rate, public health expenditure, net official development assistance (ODA) received, total population, refugee population by country of origin or territory and HDI were significant predictors in the 'univariate' analyses; the final multivariate model, with McFadden's pseudo R^2 of 0. 181, is shown in Table 1 and was used to predict regions at risk for future infectious disease outbreaks (Figure 1).

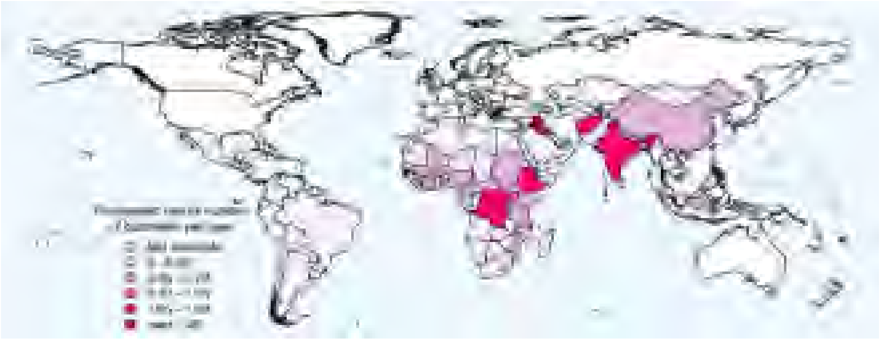


Figure 1 Risk map for future infectious disease outbreaks forecasted by a model based on various socioeconomic factors.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1 Final negative binomial regression model predicting number of outbreaks in each country from socioeconomic variables

Variable	*Model coefficient (95% CI)*
(Intercept)	2.893 (2.404;3.381)
Public health expenditure (Percentage of total health expenditure)	−0.010 (−0.017;−0.003)
Net ODA received (US$)	3.941e-10 (2.171e-10;5.620e-10)
Total population	9.715e-10 (4.421e-10;1.481e-09)
Human development index	−1.501 (−2.363;−0.650)
Average latitude	−0.034 (−0.047;−0.021)

Abbreviations: CI, confidence interval; ODA, official development assistance.

Conclusions

Public health expenditure, net ODA received, total population and HDI were significantly associated with the number of outbreaks in a country, even after controlling for latitude. The final model identified regions in Africa, Asia and the Middle East as high-risk regions for future infectious disease outbreaks. The significance of ODA as a variable in the model may indicate that many of these countries may have limited capacity to support surveillance efforts, and highlights the need for assistance through international collaborations and financial support.

Acknowledgements

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

References

1 Morse SS. Factors and determinants of disease emergence. *Rev Sci Tech* 2004;**23:**443–51.

2 Stephens DS, Moxon ER, Adams J, Altizer S, Antonovics J, Aral S, *et al.* Emerging and reemerging infectious diseases: a multidisciplinary perspective. *Am J Med Sci* 1998;**315:**64–75.

3 Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, *et al.* Global trends in emerging infectious diseases. *Nature* 2008;**451:**990–3.

4 Guernier V, Hochberg ME, Guegan JF. Ecology drives the worldwide distribution of human diseases. *PLoS Biol* 2004;**2:**e141.

Emerging Health Threats Journal 2011, **4**:s13. **doi:** 10.3134/ehtj.10.013
© 2011 JF Conley.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Sample size and spatial cluster detection power

JF Conley

Department of Geology and Geography, West Virginia University, Morgantown, WV, USA
E-mail: Jamison.Conley@mail.wvu.edu

Objective

In syndromic surveillance settings, the use of samples may be unavoidable, as when only a part of the population reports flu-like symptoms to their physician. Taking samples from a complete population weakens the power of spatial cluster detection methods.[1] This research examines the effectiveness of different sampling strategies and sample sizes on the power of cluster detection methods.

Introduction

Prior work demonstrates the extent to which sampling strategies reduce the power to detect clusters.[1] Additionally, the power to detect clusters can vary across space.[2] A third, unexplored, effect is how much the sample size impacts the power of spatial cluster detection methods. This research examines this effect.

Methods

The same six simulated clusters in the Pittsburgh, PA area are used as in [1]. A total of 1000 samples of three different strategies (random, stratified and case–control) were taken. Furthermore, 1000 independent random samples of sizes ranging from 1 to 50% of the complete population were taken. All sample data sets and the complete data sets were analyzed using FleXScan (http://www.niph.go.jp/soshiki/gijutsu/download/flexscan/),[3] which has been proven effective at finding arbitrary shaped clusters.[1,4] The detected clusters were then evaluated with a weighted power statistic that assesses the amount of overlap between the detected cluster and the actual cluster.[4]

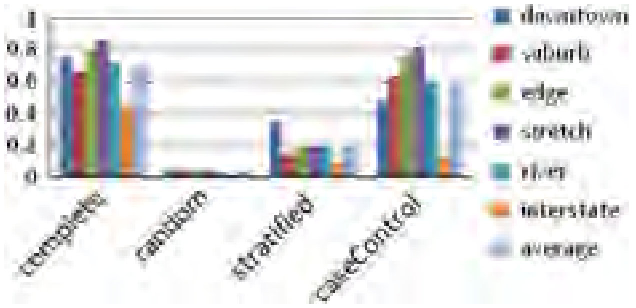


Figure 1 Weighted powers of different sampling strategies for FleXScan across all cluster shapes.

Results

Figure 1 shows the relative performance of the different sampling strategies using FleXScan. Case–control sampling is clearly the best method, followed by stratified sampling and random sampling.

Figure 2 shows the effect of sample size on the weighted power of FleXScan when using random samples. As the chart shows, the weighted power increases rapidly up to a sample size of 15–20% of the complete population, and increases slowly thereafter.

Conclusions

These results demonstrate some of the impacts of the use of samples to detect spatial clusters. First, the method of sampling is important, as case–control sampling is more effective than random sampling and stratified sampling. However, in a syndromic surveillance situation, case–control data may be

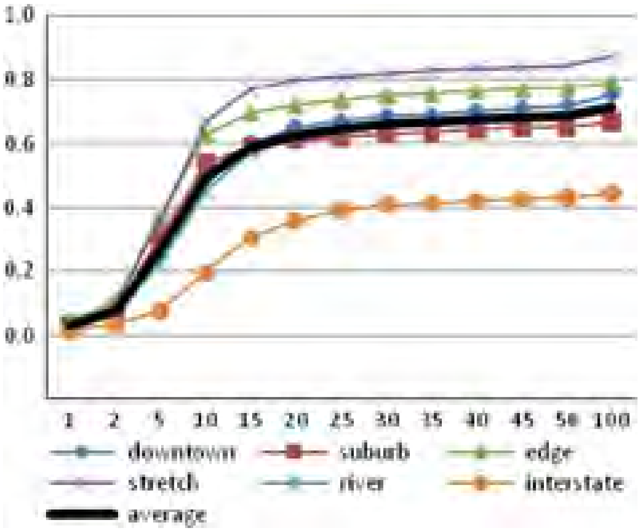


Figure 2 Weighted power versus sample size for FleXScan using random sampling.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

unavailable, and the only available data may be considered as a random sample, such as when relying upon patients to report symptoms to their physicians. In this situation, a random sample is almost as effective at accurately detecting the shape of a disease cluster when the sample is more than 15% of the population.

Acknowledgements

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

References

1 Conley JF. An evaluation of the extended power of cluster detection methods under sampling strategies. *Proc of GeoComputation 2009*, Sydney Aus. http://www.biodiverse.unsw.edu.au/geocomputation/proceedings/PDF/Conley.pdf Accessed Sep 2 2010.
2 Waller LA, Hill EG, Rudd RA. The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations. *Statist Med* 2006;**25**:853–65.
3 Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geog* 2005;**4**:11.
4 Takahashi K, Tango T. An extended power of cluster detection tests. *Statist Med* 2006;**25**:841–52.

Emerging Health Threats Journal 2011, **4**:s126. **doi:** 10.3134/ehtj.10.126
© 2011 S Day *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Spatial cluster detection in schools using school catchment information

S Day, KJ Konty, and C Goranson

New York City Department of Health and Mental Hygiene, New York, NY, USA
E-mail: sday@health.nyc.gov

Objective

To improve cluster detection of influenza-like illness within New York City (NYC) public schools using school health and absenteeism data by characterizing the degree to which schools interact.

Introduction

The H1N1 outbreak in the spring of 2009 in NYC originated in a school in Queens before spreading to others nearby.[1] Active surveillance established epidemiological links between students at the school and new cases at other schools through household connections. Such findings suggest that spatial cluster detection methods should be useful for identifying new influenza outbreaks in school-aged children.

As school-to-school transmission should occur between those with high levels of interaction, existing cluster detection methods can be improved by accurately characterizing these links. We establish a prospective surveillance system that detects outbreaks in NYC schools using a flexible spatial scan statistic (FlexScan), with clusters identified on a network constructed from student interactions.[2]

Methods

We use three data sets provided by the New York City Department of Education: 2008–2009 enrollment data, NYC Automated Student Health Record data, and absentee data that reports the number of students absent per school class on a given day.

The school interaction network can be constructed in several ways including quantifying the association between school catchment areas or enumerating the potential interactions between schools. For the former, two schools with students distributed equivalently across geographic areas would be perfectly associated ($\tau = 1$), whereas schools with no overlap would have none ($\tau = 0$). For the latter, the number of potential interactions between two schools could be summed across geographic areas. One example of this would be the number of household connections between the two schools. For our analyses, we used the number of potential interactions between the two schools by student home zip code. The resulting interaction matrix is then converted to a network by choosing a cutoff value.

The data streams included absenteeism as a percentage of enrollment and total influenza-like illness visits. For each day, the observed values were compared with a baseline period constructed from the previously reported 15 school days. We input the interaction matrix, XY coordinates, and observed/expected data into FlexScan. We ran prospective daily analyses for the school year 2008–2009, excluding high school students. Details of daily clusters were recorded. Here we focus on the 2009 H1N1 outbreak period.

Results

The results for the Spring 2009 H1N1 outbreak show two periods of significant activity: 28 April 2009–30 April 2009 and 13 May 2009–1 June 2009. The former period reported nine clusters (five in Queens, two in Brooklyn, and two in the Bronx). The latter has a total of 42 clusters (23 in Queens, 9 in Brooklyn, 9 in the Bronx, and 1 in Staten Island). During this period, the activity began in Queens, which reported at least one cluster per day, and towards late May spread to Brooklyn and the Bronx. By early June, activity slows and clusters are generally smaller. The clusters themselves differed greatly from those identified using only the circular scan method (Figure 1).

Conclusions

The method presented here improves upon existing approaches by quantifying interactions between students at different schools and is thereby more consistent with the actual process of disease diffusion. Clusters are identified when these interacting schools have increased absenteeism or school nurse visits. The analysis has a couple of limitations: (1) a school must have reported on a given day to be included in that particular analysis and (2) the data only includes public schools. However, with increased coverage both in terms of number of schools and regularity of reporting the system could alleviate these problems.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

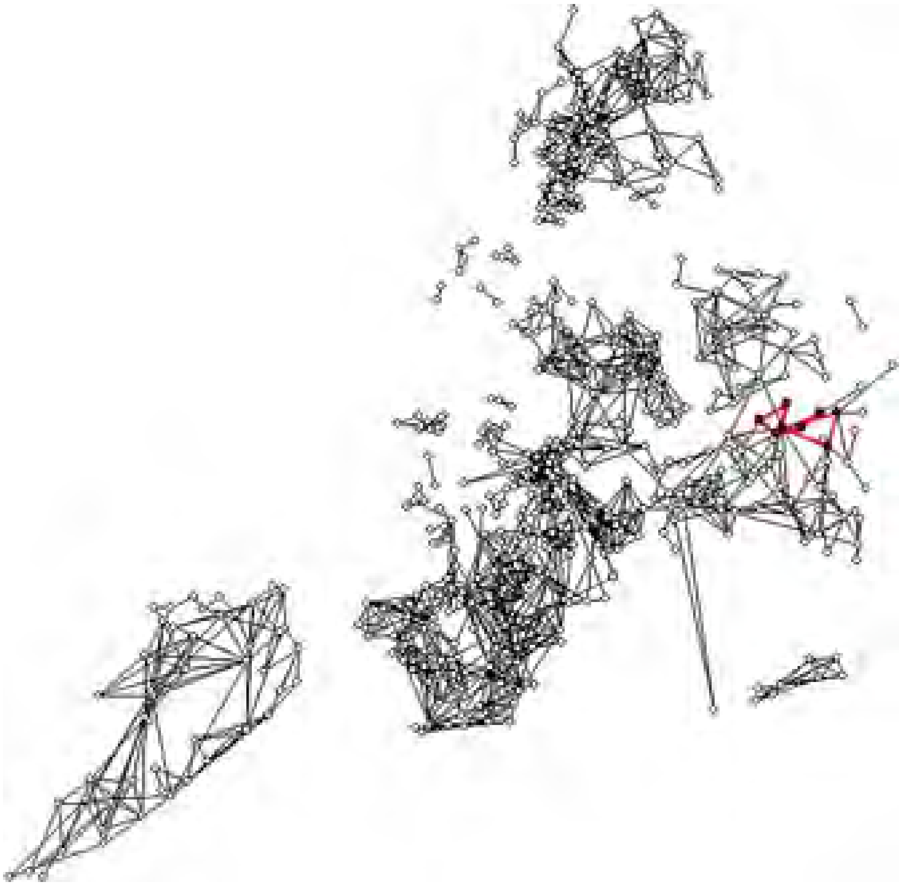


Figure 1 An example of network of elementary schools with connections defined by potential interactions between school catchments, highlighting an active cluster in Queens (4 May 2009).

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 CDC. Swine-origin influenza A (H1N1) virus infections in a school—New York City, April 2009. *MMWR* 2009;**58**:1–3.
2 Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr* 2005;**4**:11.

Emerging Health Threats Journal 2011, **4**:s17. **doi:** 10.3134/ehtj.10.017
© 2011 H Zheng *et al*.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Incorporating seasonality and other long-term trends improves surveillance for acute respiratory infections

H Zheng[1,2], WH Woodall[3], and S DeLisle[1,2]

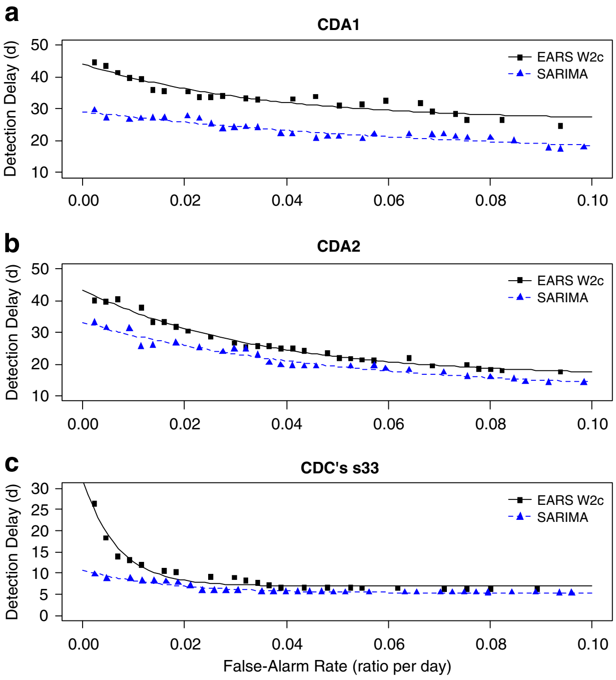
[1] *Veterans Administration Maryland Health Care System, Baltimore, MD, USA;* [2] *Department of Medicine, University of Maryland, Baltimore, MD, USA; and* [3] *Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*
E-mail: sdelisle@umaryland.edu

Introduction

As the electronic medical record (EMR) market matures, long-term time series of EMR-based surveillance data are becoming available. In this work, we hypothesized that statistical aberrancy-detection methods that incorporate seasonality and other long-term data trends reduce the time required to discover an influenza outbreak compared with methods that only consider the most recent past.

Methods

Authentic background time series of daily case counts were created by applying either of two case detection algorithms (CDAs) to EMR entries related to outpatient encounters at the Baltimore VA. The CDAs targeted acute respiratory infections (ARIs), and had the following composition and performance, compared with a manual review of 15,377 records:[1] (1) CDA1, a grouping of ICD-9 diagnostic codes similar to that used by the BioSense surveillance system (sensitivity 63%, positive predictive value (PPV) 16%; (2) CDA2, an optimized grouping of ICD-9 codes combined with the results of a computerized free-text analysis of whole-clinical notes (sensitivity 69%, PPV 54%). We used an age-structured metapopulation influenza epidemic model for Baltimore to inject factitious influenza cases into these backgrounds. From the time of this injection, aberrancy-detection statistics were applied each successive day on paired background + injection vs background-only time series. Each injection-prospective-surveillance cycle was repeated 52 times, each time shifting the injection to a different week of the study period (2003-04). We applied the same study scheme to CDC's s33 synthetic background and injection data sets (http://www.bt.cdc.gov//surveillance/ears/datasets.asp, accessed 20 February 2010). We computed two whole-system benchmarks: (1) the 'Detection Delay', the average time from injection to the first true-positive signal, defined as a statistical alarm originating in the background + injection data set but not present in the background-only data set; (2) the 'False-Alarm Rate' (FAR), defined as the number of unique false-alarms originating in the background-only data set during the study year, divided by 365 days. Detection Delay–FAR pairs were determined empirically over broad ranges of alert thresholds. We compared two aberrancy-detection approaches: (1) CDC's early aberration reporting system (EARS) W2c,[2] which makes predictions using 4 weeks of past data; (2) background-specific seasonal autoregressive integrated moving average (SARIMA) models,[3] which used 8 years of historical data. These methods were implemented for weekdays and for weekend/holidays time series.



OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Results

The figure shows activity monitoring operating characteristic (AMOC) curves for epidemic Detection Delay (*y*-axis, in days) as a function of FAR (*x*-axis). Results using EARS W2c (squares) are compared with those using SARIMA (triangles) for CDA1 (panel 'a'), CDA2 (panel 'b') and for CDC's s33 synthetic data set (panel 'c'). Note that at any given FAR within a practical range (0–10%), the SARIMA methods yield lower detection delays than EARS W2c.

Conclusions

Forecasting approaches that incorporate long-term data trends may improve the performance of surveillance systems, at least for diseases that exhibit strong seasonality.

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 DeLisle S, South B, Anthony JA, Kalp E, Gundlapallli A, Curriero FC, *et al.* Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PLoS ONE* 2010;**5:** e13377.
2 Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, *et al.* Enhancing time-series detection algorithms for automated biosurveillance. *Emerg Infect Dis* 2009;**15:**533–9.
3 Williamson GD, Weatherby Hudson G. A monitoring system for detecting aberrations in public health surveillance reports. *Stat Med* 1999;**18:**3283–98.

Emerging Health Threats Journal 2011, **4**:s18. **doi:** 10.3134/ehtj.10.018

© 2011 H Zheng *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Surveillance for acute respiratory infections: should we include all outpatient visits or focus on urgent care areas?

H Zheng[1,2], T Siddiqui[1,2], and S DeLisle[1,2]

[1] *Veterans Administration Maryland Health Care System, Baltimore, MD, USA; and* [2] *Department of Medicine, University of Maryland, Baltimore, MD, USA*
E-mail: sdelisle@umaryland.edu

Introduction

A comprehensive electronic medical record (EMR) represents a rich source of information that can be harnessed for epidemic surveillance. At this time, however, we do not know how EMR-based data elements should be combined to improve the performance of surveillance systems. In a manual EMR review of over 15 000 outpatient encounters, we observed that two-thirds of the cases with an acute respiratory infection (ARI) were seen in the emergency room or other urgent care areas, but that these areas received only 15% of total outpatient visits.[1] Because of this seemingly favorable signal-to-noise ratio, we hypothesized that an ARI surveillance system that focused on urgent visits would outperform one that monitored all outpatient visits.

Methods

Time series of daily casecounts (background) were created by applying one of eight different ARI case detection algorithms (CDAs) to EMR entries related to 'all' or to 'urgent-only' outpatient encounters at the VA Maryland Health Care System. The CDAs were constructed using various combinations of diagnostic codes, medications, vital signs, and/or computerized free-text analyses of whole clinical notes.[1] We used an age-structured metapopulation influenza epidemic model for Baltimore to inject factitious influenza cases into these backgrounds. Injections were discounted by the known sensitivity of each CDA.[1] Injections destined to urgent-only backgrounds were further discounted by 33%, to reflect the proportion of ARI patients who present to routine rather than urgent care areas. From the time of injection, CDC's EARS-W2c statistics[2] were applied on each successive day on paired background + injection vs background-only time series. Each injection-prospective-surveillance cycle was repeated 52 times, each time with the injection shifted to a different week of the 1-year study period (2003–2004). We computed: (1) the 'detection delay', the average time from injection to the first alarm present in the background + injection dataset but absent from the background-only dataset; (2) the 'false alarm rate' (FAR), defined as the number of unique false alarms originating in the background-only dataset during the study year, divided by 365 days; (3) the 'caseload', defined as the total number of cases contained in 1 year of false alarms. To create activity monitoring operating characteristic (AMOC) curves, we empirically determined the corresponding delay-FAR or delay-caseload pairs over a wide range of alarm thresholds.

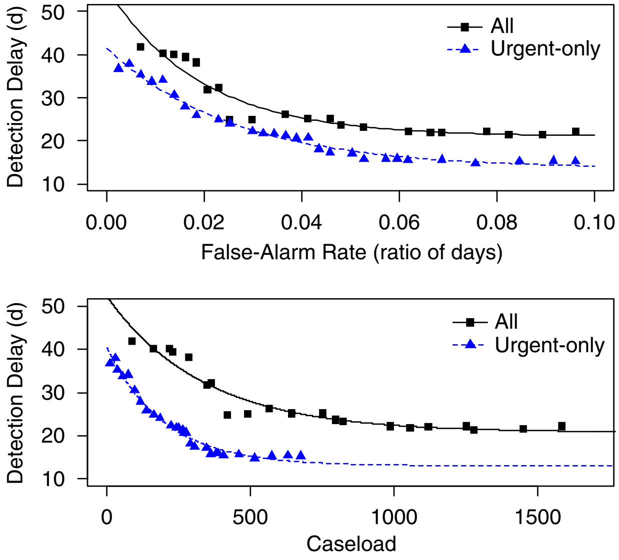


Figure 1 Performance of surveillance systems that focus either on all outpatient visits (squares) or on the subset of these visits that are urgent (triangles). Time to outbreak detection (*y*-axis) is plotted as a function of false-alarm rate (upper panel) or caseload (lower panel).

Results

Figure 1 compares AMOC curves for a representative ARI CDA (ARI-related ICD-9 codes or a new cough suppressant or

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

two non-negated ARI symptoms from our case definition by text analysis), in otherwise identical surveillance systems that included either 'all' outpatient visits (black squares) or urgent-only visits (blue triangles). Note that detection delay (*y*-axis) is lower at any given FAR (upper panel) or caseload (lower panel). These findings were consistent across all eight CDA tested.

Conclusions

Our results suggest that ARI surveillance systems that focus on urgent/emergent care areas outperform systems that monitor all outpatient visits, even if they ignore a significant number of outpatients whose ARI coincides with routine visits.

Acknowledgements

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

References

1 DeLisle S, South B, Anthony JA, Kalp E, Gundlapallli A, Curriero FC, *et al*. Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PLoS ONE* 2010;**5**:e13377.

2 Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, *et al*. Enhancing time-series detection algorithms for automated biosurveillance. *Emerg Infect Dis* 2009;**15**:533–9.

Emerging Health Threats Journal 2011, **4**:s20. **doi:** 10.3134/ehtj.10.020
© 2011 LH Duczmal *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A Voronoi based scan for space-time cluster detection in point event data

LH Duczmal, GJP Moreira, D Burgarelli, RHC Takahashi, FCO Magalhães, and EC Bodevan

Federal University of Belo Horizonte, Minas Gerais, Brazil
E-mail: duczmal@est.ufmg.br

Objective

We describe a method for prospective space-time cluster detection of point event data based on the scan statistic. Our aim is to detect as early as possible the appearance of an emerging cluster of syndromic individuals because of a real outbreak of disease amidst the heterogeneous population at risk.

Introduction

Scan statistics are highly successful for the evaluation of space-time clusters.[1] Recently, concepts from the graph theory were applied to evaluate the set of potential clusters. Wieland *et al.*[2] introduced a graph theoretical method for detecting arbitrarily shaped clusters on the basis of the Euclidean minimum spanning tree of cartogram transformed case locations, which is quite effective, but the cartogram construction step of this algorithm is computationally expensive and complicated.

Methods

The data set consist of locations of entities (cases/controls) often represented as points in a two-dimensional map space or in a three-dimensional space-time domain. Here, cases refer to individuals with a particular disease of interest, and controls refer to individuals with similar characteristics, but do not have the disease. We build the *Voronoi Diagram* of the case-control observed data space-time set. Assign to each point p_i (case or control) the part of the domain closer to p_i than to any other point. In this construction, each cell contains exactly one case or control. Let $C=\{c_i^t\}$ be the subset representing the cases. We define a weighted complete graph $G(C)=(V, E)$ with vertex set $V=\{c_i^t | c_i^t \in C\}$ and edge set $E=\{(c_i^t, c_j^t) | c_i^t, c_j^t \in C, i \neq j\}$. Edges $(u, v) \in E$ have weight defined by the *case-to-case cell-crossing count* (C^5), that is, the number of Voronoi cells between cases u and v. This *Voronoi metric* is used instead of the usual Euclidean metric. After computing the minimum spanning tree (MST) T of the weighted graph $G(C)$, every potential cluster is defined as a connected subgraph of T. Given a set V of n cases, we evaluate only $2n$-1 potential clusters, by sequentially deleting the longest remaining edge of T. We consider the two newly emergent connected components for each step. Figure 1 shows an example of the spatial projection of the MST.

Results

Our C^5 scan was compared numerically with the elliptic version of the popular prospective space-time scan[1,3] and an identical scan (but using the Euclidean metric) according to power of detection, positive predicted value (PPV) and sensitivity. We used artificial datasets with total population

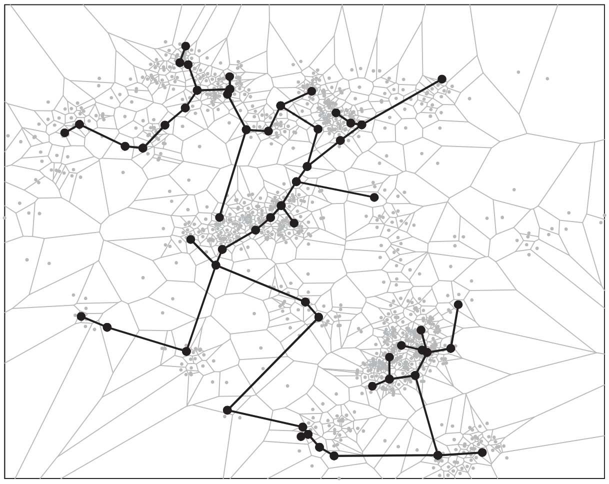


Figure 1 A MST linking cases according to the Voronoi metric.

Table 1 Power, PPV and sensitivity for three space time clusters

Cluster	*Power*		*PPV*		*Sensitivity*	
Model	*C^5*	*Elliptic*	*C^5*	*Elliptic*	*C^5*	*Elliptic*
Cylinder	0.6072	0.4789	0.6763	0.6415	0.6762	0.5447
Cone	0.4540	0.3863	0.6078	0.5822	0.6025	0.4683
L-shape	0.5115	0.3316	0.5980	0.5323	0.6301	0.4530

Abbreviation: PPV, positive predicted value.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

at risk of 1000 individuals, 100 cases, for a period of 10 days. Three alternative hypotheses models of space-time clusters with different shapes were simulated; exactly the same sets were used for all algorithms, with 10 000 Monte Carlo simulations for each model (Table 1).

Conclusions

We have proposed a space-time cluster scan for case event data. For all three-shaped clusters our scan has better performance compared with the prospective elliptic scan[1,3] and is much faster than the density-equalizing Euclidean minimum spanning trees algorithm[2] and the elliptic scan. It also is significantly better than the Euclidean metric MST scan (results not shown).

Acknowledgements

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

References

1 Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *JRSS A* 2001;**164**:61–72.
2 Wieland SC, Brownstein JS, Berger B, Mandl KD. Density-equalizing Euclidean minimum spanning trees for the detection of all disease cluster shapes. *PNAS* 2007;**104**:9404–9.
3 Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med* 2006;**25**:3929–43.

Emerging Health Threats Journal 2011, **4**:s22. **doi:** 10.3134/ehtj.10.022

© 2011 G Fairchild *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Evaluating the performance of two alternative geographic surveillance schemes

G Fairchild, A Segre, P Polgreen, and G Rushton

Computational Epidemiology Research Group, Department of Computer Science, University of Iowa, Iowa City, IA, USA
E-mail: gcfairch@gmail.com

Objective

This paper describes an experiment to evaluate the performance of several alternative surveillance site placement algorithms with respect to the standard influenza-like illness (ILI) surveillance system in Iowa.

Introduction

ILI data is collected by an Influenza Sentinel Provider Surveillance Network at the state (Iowa, USA) level. Historically, the Iowa Department of Public Health (IDPH) has maintained 19 different influenza sentinel surveillance sites. Because participation is voluntary, locations of the sentinel providers may not reflect optimal geographic placement. This study analyzes two different geographic placement algorithms—a maximal coverage model (MCM) and a K-median model.[1] The MCM operates as follows: given a specified radius of coverage for each of the *n* candidate surveillance sites, we greedily choose the *m* sites that result in the highest population coverage. In previous work, we showed that the MCM can be used for site placement.[2] In this paper, we introduce an alternative to the MCM—the K-median model. The K-median model, often called the P-median model in geographic literature, operates by greedily choosing the *m* sites which minimize the sum of the distances from each person in a population to that person's nearest site. In other words, it minimizes the average travel distance for a population.

Methods

This project is split into two phases—a calculation phase and a validation phase. In the first phase, we developed a user-friendly web-based calculator to help public health entities locate sites on their own. In the second phase, we compared and analyzed the placements of sites to show that the two models implemented in the web calculator choose 'better' sites *de novo* for Iowa than the existing 19 sites hand-picked by the IDPH. First, we generated 19 sites *de novo* using each algorithm. Then, we used 8 years of statewide Medicaid billing data to simulate the spread of influenza across the state of Iowa. In total, there were approximately 2 million cases with ILI-related ICD9 codes present in this Medicaid database. We used two different probability functions, an exponentially decreasing function relating to patient distance from site, and the Huff probability model, a model commonly used in geography, to probabilistically measure the likelihood of a case being 'noticed' by a surveillance site (and thus having their visit reported back to the IDPH). The Huff model operates by computing the probability based both on the distance to a site and the 'attractiveness' of a site, which, for this experiment, is the population coverage of that site.

Results

	Existing	*MCM*	*K-median*
Exponential	16.0%	12.0%	19.3%
Huff	32.3%	37.4%	41.3%

The Table above shows the percentage of Medicaid cases detected during the simulation for each model and for each probability function. The 19 sites calculated *de novo* using the K-median model capture 3.3% more cases than the 19 existing sites when using the exponentially decreasing probability function based on distance. Additionally, using the commonly used Huff probability model, the same sites achieve 9% better detection than the existing sites. Although the sites calculated using the MCM detect fewer cases than the existing sites using the distance-based probability function, they capture 5.1% more cases using the Huff model.

Conclusion

We have described a systematic, intuitive and easy to use method by which surveillance sites may be located in a given geographic region. We offer two models for site placement. One model, the maximal coverage model, greedily chooses

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sites which maximize total population coverage. The other model, the K-median model, greedily chooses sites which minimize the average travel distance for a given population. When simulating 8 years of ILI Medicaid cases in Iowa against the chosen sites, we found that we are able to generate sites that outperform the existing sites in Iowa in terms of cases reported when compared using two different probabilistic models for patient visits.

To generate the best results from a disease surveillance network, site placement is of the utmost importance, and the methods we describe help to ensure that sites are located as well as possible.

Acknowledgements

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

References

1 Church RL. Location modeling and GIS. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds). *Geographical Information Systems, Principles, Techniques, Applications and Management*. 2nd edn. New York: Wiley, 1999, pp 293–303.

2 Polgreen P, Chen Z, Segre AM, Harris ML, Pentella MA, Rushton G. Optimizing influenza sentinel surveillance at the state level. *Am J Epidemiol* 2009;**170:**1300–6.

Emerging Health Threats Journal 2011, **4**:s23. **doi:** 10.3134/ehtj.10.023
© 2011 AR Duarte *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Weighted non-connectivity for detection of irregular clusters

AR Duarte[1], SB Silva[2], L Duczmal[2], SJ Ferreira[2], and ALF Cançado[3]

[1]*Mathematics Department, Federal University of Ouro Preto, Ouro Preto, Brazil;* [2]*Statistics Department, Federal University of Minas Gerais, Belo Horizonte, Brazil; and* [3]*Statistics Department, University of Brasilia, Brasilia, Brazil*
E-mail: sabjfn@gmail.com.br

Objective

Irregularly shaped clusters in maps divided into regions are very common in disease surveillance. However, they are difficult to delineate, and usually we notice a loss of power of detection. Several penalty measures for the excessive freedom of shape have been proposed to attack this problem, involving the geometry and graph topology of clusters. We present a novel topological measure that displays better performance in numerical tests.

Introduction

Heuristics to detect irregularly shaped spatial clusters were reviewed recently.[1] The spatial scan statistic is a widely used measure of the strength of clusters.[2] However, other measures may also be useful, such as the geometric compactness penalty,[3] the non-connectivity penalty[4] and other measures based on graph topology and weak links.[5,6] Those penalties $p(z)$ are often coupled with the spatial scan statistic $T(z)$, employing either the multiplicative formula maximization $\max_z T(z) \times p(z)^4$ or a multiobjective optimization procedure $\max_z(T(z),\ p(z))$,[3,6] or even a combination of both approaches.[5] The geometric penalty of a cluster z is defined as the quotient of the area of z by the area of the circle, with the same perimeter as the convex hull of z, thus penalizing more the less rounded clusters. Now, let V and E be the vertices and edges sets, respectively, of the graph $G_z(V,\ E)$ associated with the potential cluster z. The non-connectivity penalty $y(z)$ is a function of the number of edges $e(z)$ and the number of nodes $n(z)$ of $G_z(V,\ E)$, defined as $y(z)=e(z)/3[n(z)-2]$. The less interconnected tree-shaped clusters are the most penalized. However, none of those two measures includes the effect of the individual populations.

Methods

Based on the non-connectivity penalty, we propose the concept of weighted non-connectivity $w(z)$, taking into account the non-uniform distribution of the populations of the regions of the cluster z. Let

$$w(z) = \sum\nolimits_{a \in A} q(a)/3 \left[\sum\nolimits_{v \in V} p(v) - 2 \sum\nolimits_{v \in V} p(v)/n(z)\right]$$

where $p(v)$ is the population of the region associated with the node $v \in V$ and $q(a)$, $a \in E$ is a weight, defined as the average of the populations of the regions linked by the edge a (see Figure 1). The function $w(z)$ is used as the second objective in the multiobjective problem $\max_z(T(z),\ w(z))$, employing a genetic algorithm. The solution is a Pareto set, consisting of all clusters that are not worse on both objectives simultaneously. The statistical significance evaluation is performed for all clusters in the Pareto set through Monte Carlo simulations, with the help of the attainment function concept,[6] determining the best solution.

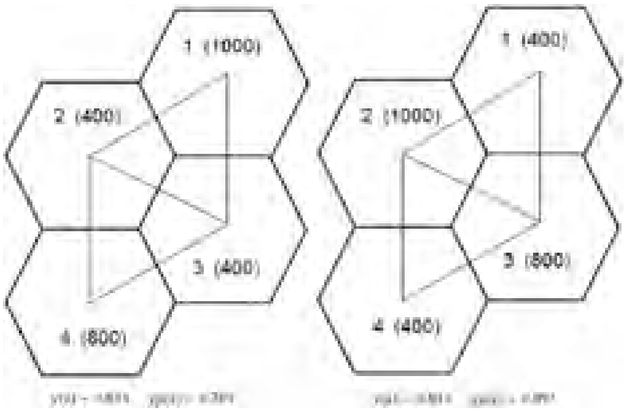


Figure 1 Comparison between measures of non-connectivity and weighted non-connectivity for two hypothetical clusters.

Results

Numerical simulations were conducted to assess the power of detection, sensitivity and positive predictive value. We run the weighted non-connectivity scan for the benchmarks

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in ref. 3 and compare it with the algorithms defined in refs 3–6. The weighted non-connectivity scan presented significantly higher power of detection and sensitivity, and about the same positive predictive value, compared with the other algorithms.

Conclusion

The weighted non-connectivity scan is more efficient when compared with the geometric, non-connectivity and other topological penalty based scans.

Acknowledgements

SJF acknowledges the support by Fapemig, Brazil. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Duczmal L, Duarte AR, Tavares R. Extensions of the scan statistic for the detection and inference of spatial clusters. J. Glaz, V. Pozdnyakov, S. Wallenstein. (Org.). *Scan Statistics: Methods and Applications*. Birkhauser, Boston, MA, 2009;157–82.
2 Kulldorff M. A spatial scan statistic. *Comm Statist Theory Meth* 1997;**26**:1481–96.
3 Duczmal L, Cançado ALF, Takahashi RHC. Geographic delineation of disease clusters through multi-objective optimization. *J Comput Graph Stat* 2008;**17**:243–62.
4 Yiannakoulias N, Rosychuk RJ, Hodgson J. Adaptations for finding irregularly shaped disease clusters. *IJ Health G* 2007;**6**:28.
5 Duarte AR, Duczmal L, Ferreira SJ, Cançado ALF. Internal cohesion and geometric shape of spatial clusters. *Environ Ecol Stat* 2010;**17**:203–29.
6 Cançado ALF, Duczmal L, Duarte AR, Ferreira SJ, Fonseca CM, Gontijo ED. Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters. *Int J Health Geogr* 2010;**9**:55.

Emerging Health Threats Journal 2011, **4**:s132. **doi:** 10.3134/ehtj.10.132
© 2011 RE Gamache and S Grannis.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A model for flu outbreak surveillance that describes the time lag for data reporting from the first presentation of a case to diagnosis

RE Gamache and S Grannis

Regenstrief Institute, Inc., Indianapolis, IN and the Indiana University School of Medicine, Indianapolis, IN, USA
E-mail: rgamache@regenstrief.org

Objective

This poster will present a predictive model to describe the actual number of confirmed cases for an outbreak (H1N1) based on the current number of confirmed cases reported to public health. The model describes the methods used to calculate the number of cases expected in a community based on the lag time in the diagnosis and reporting of these cases to public health departments.

Introduction

Reporting notifiable conditions to public health authorities by health-care providers and laboratories is fundamental to the prevention, control, and monitoring of population-based disease.[1,2] To successfully develop community centered health, public health strives to understand and to manage the diseases in its community. Public health surveillance systems provide the mechanisms for public health professionals to ascertain the true disease burden of the population in their community.[3] The information necessary to determine the disease burden is primarily found in the data generated during clinical care processes.[4]

Methods

Syndromic and clinical data were collected from the various participating providers throughout the state.[5] These data sources include; emergency departments, hospitals, outpatient centers, and other ancillary care data sources. Table 1 lists the ICD9 codes used to establish an influenza diagnosis. An influenza diagnosis is based on the CDC definition. All messages for patients in the Marion County area were processed to identify any individual with a flu diagnosis. For each case, the data were searched for the initial encounter for that patient with the health care system concerning their symptoms. The lag time from the initial encounter and diagnosis was calculated from the difference of these two events.

Results

Figure 1 provides a summary of the lag time information. The median time for a clinical diagnosis to be seen in this system is 5 days. An analysis of the time differences in relationship to the number of syndromic cases provides a ‘near real-time’ estimate of the actual number of influenza cases (seeking care) occurring in the community. These results are compared with the actual information collected and other community-based simulations.[6] The primary difference is that this simulation is based on events that are currently occurring in the population.

Table 1 ICD9 codes used to determine influenza or an influenza-like illness diagnosis

ICD9 code	*Simple description*
487.x	Flu diagnosis
488	H1N1 diagnosis
780.6, 780.6x	Fever diagnosis
786.2	Cough diagnosis
460, 461.x, 462, 463, 464.x, 465.x	Upper respiratory tract infection

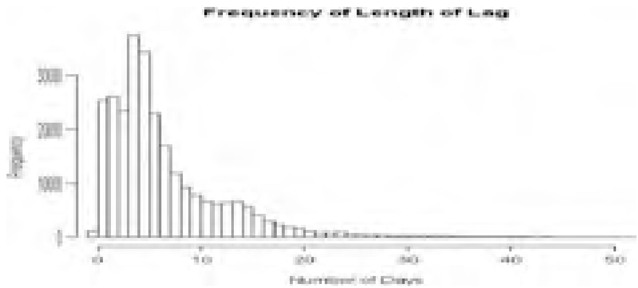


Figure 1 Counts of the number of days after a provider sees a patient until a diagnosis is received through the electronic messaging system. Over 95% of the data was received by day 21.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

Although diagnosis data may not be required to identify a disease outbreak in the community, the additional clinical diagnosis information is beneficial for public health departments particularly for the deployment of specific countermeasures based on the presences of a given infectious agent in the population. A better understanding of the actual number of infected and contagious individuals in a community provides public health departments a more realistic level of the disease burden, which is particularly useful when deploying countermeasures in the community.

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 Chorba TL, Berkelman RL, Safford SK, Gibbs NP, Hull HF. Mandatory reporting of infectious diseases by clinicians. *JAMA* 1989;**262**:3018–26.
2 Committee on Emerging Microbial Threats to Health Division of Health Sciences Policy, D.o.I.H. *Microbial Threats to Health in the United States*, Joshua L, Robert ES and Stanley Jr CO (eds). Institute of Medicine, National Academy Press: Washington, DC, 1992.
3 Germann TC, Kadau K, Longini Jr IM, Macken CA. Mitigation strategies for pandemic influenza in the United States. *Proc of the Natl Acad of Sci* 2006;**103**:5935–40.
4 Overhage JM, Grannis S, McDonald CJ. A Comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. *Am J Public Health* 2008;**98**:344–50.
5 Grannis S, Wade M, Gibson J, Overhage JM. The Indiana Public Health Emergency Surveillance System: ongoing progress early findings and future directions. *AMIA Annu Symp Proc* 2006; 304–8.
6 Halder N, Kelso JK, Milne GJ. Analysis of the effectiveness of interventions used during the 2009 A/H1N1 influenza pandemic. *BMC Public Health* 2010;**10**:168.

Emerging Health Threats Journal 2011, **4**:s25. **doi:** 10.3134/ehtj.10.025
© 2011 SK Greene *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Gastrointestinal disease outbreak detection using multiple data streams from electronic medical records

SK Greene[1], J Huang[2], AM Abrams[1], M Reed[2], R Platt[1], and M Kulldorff[1]

[1]*Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; and* [2]*Kaiser Permanente Division of Research, Oakland, CA, USA*
E-mail: Sharon_Greene@hphc.org

Objective

To identify which syndromic and laboratory-based data streams from electronic medical records (EMRs) may be used to detect gastrointestinal (GI) disease outbreaks in a timely manner.

Introduction

GI disease outbreaks can be focal (for example, restaurant associated), generalized (for example, seasonal rotavirus increases) or intermediate (for example, widely disseminated contaminated commercial products). Health departments (HDs) are commonly notified of focal outbreaks by passive reporting, whereas generalized outbreaks in non-institutional settings are seldom reported as clusters. Intermediate outbreaks are often detected via laboratory testing, which may be subjected to backlogs and delays. Healthcare systems routinely collect in EMRs clinical data related to GI disease, such as ambulatory care diagnoses, that could be exploited for surveillance.[1] Multiple syndromic and laboratory data sources could potentially be used to prospectively detect generalized and intermediate GI disease outbreaks for situational awareness and possible epidemiological investigation.

Methods

Kaiser Permanente Northern California (KPNC) is a large integrated health care delivery system utilizing EMRs. Zip code-specific daily episode counts in 2009 were generated for 22 data streams: 9 streams of ICD-9 codes for upper and lower GI diagnoses in ambulatory care, emergency department, and inpatient settings; four streams of ICD-9 codes in association with an antibiotic prescription; one stream of text strings for hospital admissions; one stream of prescription antidiarrheal medication dispensings; one stream of stool culture tests ordered; and six streams of tests positive for GI pathogens.

We mimicked prospective surveillance of these data streams with univariate cylindrical space–time permutation scan statistic analyses,[2] using only data that would have been available as of each surveillance day. 'Alerts,' that is, cluster(s) unlikely to occur by chance alone in a 1-year period, were evaluated and will be compared with GI outbreaks from preliminary or final reports to the state HD in non-institutional settings affecting any of the 16 countries for which KPNC had $\geqslant$10% population coverage. *Salmonella* and *E. coli* clusters were identified in retrospective analyses, and genetic relatedness of isolates in these clusters will be assessed.

Results

Analyses of 22 EMR data streams yielded 35 unique alerts. Seven streams did not contribute to alerts, including two streams of ICD-9 codes and an antibiotic prescription and five streams of positive tests. In total, 25 GI outbreaks (six laboratory confirmed) were reported to the state HD for the study area in 2009; the median number of recognized cases was 10 (range: 3–280). Spatio-temporal overlap, if any, between alerts in electronic data and known outbreaks will be characterized. Three clusters each of *Salmonella* and *E. coli* were identified in EMR data; the genetic relatedness of isolates in these clusters will be described.

Conclusions

We plan to identify which streams may be most useful for GI outbreak surveillance. Alerts could reflect previously unknown (particularly generalized or intermediate) outbreaks, in contrast with outbreaks reported to the state HD, which generally had small numbers of recognized cases and were highly localized. Prospectively identifying clusters of tests positive for specific GI pathogens may be a promising complement to syndromic surveillance, as isolates in clusters could be prioritized for testing, potentially reducing the time to outbreak detection.

Acknowledgements

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

References

1 Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, Kruskal BA, *et al.* Evaluating real-time syndromic surveillance signals from ambulatory care data in four states. *Public Health Rep* 2010;**125**:111–20.
2 Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005;**2**:e59.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s27. **doi:** 10.3134/ehtj.10.027
© 2011 CL Hayden *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Impact of including physician's prescribing directions on calculations of medication possession ratios

CL Hayden, BC Sauer, and GW Cannon

Clinical Epidemiology, University of Utah; Veterans Affairs Medical Center, Salt Lake City, UT, USA
E-mail: candace.hayden@hsc.utah.edu

Introduction

Medication adherence studies typically use pharmacy-dispensing data to infer drug exposures. These studies often require calculations reflecting the intensity and duration of drug exposure. The typical approach to estimating duration of drug exposure is to use dispensing dates and day supply.[1–3] Often, pharmacy databases have random and/or systematic errors causing improbable calculations.[1] These errors become particularly problematic when estimating medication duration in drugs with complicated dosing schedules. Experts recommending cleaning data or removing erroneous data before analysis,[1] but do not provide instructional guidelines. We developed an algorithmic approach to improve estimation of drug-course duration, dosing and medication possession ratios (MPRs). This study compares estimated MPRs produced by the standard method with MPRs by the algorithmic approach. Methotrexate was chosen as the first drug to implement the algorithm because of its widespread use for rheumatoid arthritis (RA) and for its complexity in dosing schedules.

Methods

The data used in this study were provided by the Pharmacy Benefits Management (PBM) Database for patients enrolled in the Veterans Affairs RA (VARA) Registry. The algorithm was based on clinically feasible weekly doses to calculate our research variables. A course was defined as any number of prescriptions for the same drug for a given individual without a prescribing gap of greater than or equal to 90 days. The prescribed course duration of drug exposure was defined as the sum of the expected durations for each prescription within a course. The average dose prescribed was calculated as the total dose dispensed divided by the prescribed duration. The average dose consumed was the total dose dispensed divided by the observed course duration. The MPR was calculated as the prescribed duration divided by the observed course duration. When calculated doses fell out of the clinically expected range, the algorithm was triggered to flag the course and use the sig interpretation and give a set of alternate calculations. Alternate sig calculations were performed on the whole data set for comparison purposes.

Results

We identified 2127 unique courses of methotrexate in 1034 individuals. Approximately 2% of the prescription courses triggered the algorithm. A paired *t*-test was run on MPRs calculated by both methods on the whole data set. MPR values were significantly lower when calculated without using the sig interpretation (mean difference $= -0.03$, $P = 0.0005$). On the subset of records that triggered the algorithm, the effect was more dramatic with a mean difference was -0.27 with a P-value of <0.0001.

Conclusions

This algorithm provides a systematic approach to error detection and correction in secondary databases. Researchers in need of careful precision and accuracy of drug exposure and compliance may benefit from this algorithmic approach.

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 Peterson AM, Nau DP, Cramer JA, Brenner J, Gwadry-Sridhar F, Nichol M. A checklist for medication compliance and persistence studies using retrospective databases. *Value in Health* 2007;**10**:3–12.
2 Gilmer TP, Dolder CR, Lacro JP, Folsom DP, Lindamer L, Garcia P, *et al.* Adherence to treatment with antipsychotic medcation and heath care costs among Medicaid beneficiaries with schitzophrenia. *Am J Psychiatry* 2004;**161**:692–9.
3 Hoffman L, Enders J, Luo J, Segal R, Pippins J, Kimberlin C. Impact of antidepressant management program on medication adherence. *Am J Manag Car* 2003;**9**:70–80.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s28. **doi:** 10.3134/ehtj.10.028
© 2011 MJ Heaton *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A spatio-temporal absorbing state model for disease and syndromic surveillance

MJ Heaton[1], F Zou[2], D Banks[1], A Karr[2], G Datta[3], J Lynch[4], and F Vera[5]

[1]*Department of Statistical Science, Duke University, Durham, NC, USA;* [2]*National Institute of Statistical Sciences, Research Triangle Park, NC, USA;* [3]*Department of Statistics, University of Georgia, Athens, GA, USA;* [4]*Department of Statistics, University of South Carolina, Columbia, SC, USA; and* [5]*Department of Mathematical Sciences, Clemson University, Clemson, SC, USA*
E-mail: matt@stat.duke.edu

Objective

Syndromic surveillance for new disease outbreaks is an important problem in public health. Many statistical techniques have been devised to address the problem, but none are able to simultaneously achieve important practical goals (good sensitivity and specificity, proper use of domain information, and transparent support to decision-makers). The objective, here, is to improve model-based surveillance methods by (i) detailing the structure of a hierarchical hidden Markov model (HMM) for the surveillance of disease across space and time and (ii) proposing a new, non-separable spatio-temporal autoregressive model.

Introduction

The goal of disease and syndromic surveillance is to monitor and detect aberrations in disease prevalence across space and time. Disease surveillance typically refers to the monitoring of confirmed cases of disease, whereas syndromic surveillance uses syndromes associated with disease to detect aberrations. In either situation, any proper surveillance system should be able to (i) detect, as early as possible, potentially harmful deviations from baseline levels of disease while maintaining low false positive detection rates, (ii) incorporate the spatial and temporal dynamics of a disease system, (iii) be widely applicable to multiple diseases or syndromes, (iv) incorporate covariate information and (v) produce results that are readily interpretable by policy decision makers.

Early approaches to surveillance were primarily computational algorithms. For example, the CUSUM[1] technique and its variants (see, for example, Fricker *et al.*[2]) monitor the cumulative deviation (over time) of disease counts from some baseline rate. A second line of work uses spatial scan statistics, originally proposed by Kulldorff[3] with later extensions given in Walther[4] and Neill *et al.*[5]

Methods

As the data layer for the HMM, let,

$$Y_s(t) \sim P(\mu_s(t) + \delta_s(t)\lambda_s(t))$$

where $Y_s(t)$ represents a count of a disease at location s at time t, $\mu_s(t)$ represents a baseline rate of disease, $\delta_s(t) \in \{0,1\}$ is an indicator as to whether or not the disease is in an epidemic state and $\lambda_s(t)$ represents an added rate of disease because of an epidemic state.

As models for the rates of disease (that is, $\mu_s(t)$ and $\lambda_s(t)$), a novel non-separable spatio-temporal structure is assumed. Furthermore, the indicators, $\delta_s(t)$ are assumed to follow an absorbing state Markov chain, where the state transitions are governed by the number of neighbors in an epidemic state.

Conclusions

The model performs well by correctly classifying states as either epidemic or non-epidemic in both a large simulation study and in an application to influenza/pneumonia fatality data.

Acknowledgements

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

References

1 Page ES. Continuous inspection schemes. *Biometrika* 1954;**41**:100–15.
2 Fricker RD, Hegler BL, Dunfee DA. Comparing syndromic surveillance detection methods: EARS versus a CUSUM-based methodology. *Stat Med* 2008;**27**:3407–29.
3 Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997;**26**:1481–96.
4 Walther G. Optimal and fast detection of spatial clusters with scan statistics. *Ann Stat* 2010;**38**:1010–33.
5 Neill DB, Moore AW, Cooper GF. A Bayesian spatial scan statistic. In: Weiss Y, Scholkopf B, Platt J (eds) *Advances in Neural Information and Processing Systems*, vol 18. MIT Press: Cambridge, MA, USA, 2006, pp 1003–10.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s29. **doi:** 10.3134/ehtj.10.029

© 2011 ATY Ho *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

A prediction market for H5N1 influenza compared with statistical forecasting model

ATY Ho[1], FD Nelson[1], GR Neumann[1], and PM Polgreen[2]

[1]*Department of Economics, University of Iowa, Iowa City, IA, USA; and* [2]*Department of Internal Medicine, University of Iowa, Iowa City, IA, USA*
E-mail: ansontaiyat-ho@uiowa.edu

Objective

The purpose of this study is to compare the results of an H5N1 influenza prediction market model with a standard statistical model.

Introduction

Prediction markets have been successfully used to forecast future events in other fields.[1–4] We adapted this method to provide estimates of the likelihood of H5N1 influenza related events.

Methods

Participants were given educational grants of $100 with which to trade financial contracts whose future values depend on the outcome of selected avian influenza watershed events. These events were on the basis of health-related policy decisions, numbers, and locations of human and animal H5N1 cases. For example, one contract was worth $1.00 if the total number of H5N1 human cases confirmed by WHO was at least 350 by 1/1/2008. After 1/1/2008, it ceased trading and was replaced by a similar contract with a 7/1/2008 target. Traders bought and sold contracts with one another at prices that depended on their beliefs about the likelihood of the underlying event. The resulting prices can be interpreted as the consensus probabilities of event occurrence.

For each of the contracts based on the total number of human H5N1 cases, we also used a statistical model to assess the performance of our prediction markets. The statistical model used the average number of cases per week since 1/1/2006 to predict the total number of cases on the contract expiration date. Although this statistical model only forecasts the total number of cases, it also possesses the properties of a Poisson distribution, as the parameter of a Poisson process in weekly frequency is the same as the average case count per week in maximum likelihood estimation. Thus, we can estimate the probability of the contract being true, and this probability is directly comparable to the results from the prediction markets.

Statistical models critically depend on reliable and timely data. Thus, to have a more comprehensive comparison of the performance of the prediction markets and the statistical model with different specifications, we also assessed the performance of the statistical model using different lengths of historical data, and compared it with the results of the prediction markets.

Results

In all, we offered 52 pairs of contracts. The most popular contracts were focused on the total number of human cases by a specific date.

Both the prediction markets and the statistical model correctly forecasted the outcomes of the five human-case-number contracts at least a month in advance. At the beginning of the trading period, the prediction markets and the statistical model exhibit some discrepancies in the probability that the specified event would be true. As the time before the target week decreased, the probability estimation of these two methods converged, and the predictive accuracy of both methods increased.

When only current data (past 4 weeks) are used in the statistical model, although the model still successfully predicted the final outcomes, it exhibited volatile predictions of the outcome during the trading period, especially when forecasts are made far in advance. In contrast, the prediction markets provide stable forecasts with considerable accuracy. This suggests that long history of data is needed for a statistical model to forecast reliable H5N1 event outcomes. This is in contrast to prediction markets as a mechanism for aggregating information *per se*.

Conclusion

In summary, prediction markets may provide a flexible and effective way to aggregate both objective and subjective information about H5N1 influenza. The probabilities generated by such a market may help public health officials to plan for the future and coordinate resources. Prediction markets, because of

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

their reliance on surveillance data to define contracts, will not replace any existing surveillance systems. Instead, we propose our prediction market as a supplement to aggregate expert opinions based on existing surveillance information.

Acknowledgements

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

References

1 Smith VL. Constructivist and ecological rationality in economics. *Am Econ Rev* 2003;**93**:465–508.
2 Pennock DM, Lawrence S, Giles CL, Nielsen FA. The real power of artificial markets. *Science* 2001;**291**:987–8.
3 Cowgill B, Wolfers J, Zitzewitz E. Using prediction markets to track information flows: evidence from Google. *Dartmouth College* (2008), http://ww.bocowgill.com/GooglePredictionMarketPaper.pdf.
4 Dvorak P. Best buy taps 'prediction market'. *Wall Street J* 2008. Available at: http://online.wsj.com/article/SB122152452811139909.html.

Emerging Health Threats Journal 2011, **4**:s31. **doi:** 10.3134/ehtj.10.031
© 2011 S Kandula *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A bootstrapping method to improve cohort identification

S Kandula[1], Q Zeng-Treitler[1], L Chen[2], W Salomon[3], and BE Bray[1]

[1]*Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA;* [2]*Scientific Systems Company Inc., Woburn, MA, USA; and* [3]*Clinical Metrics LLC, Poland, ME, USA*
E-mail: sasi.kandula@utah.edu

Objective

ICD-9 codes are commonly used to identify disease cohorts and are often found to be less than adequate. Data available in structured databases—lab test results, medications etc.—can supplement the diagnosis codes. In this study, we describe an automated method that uses these related data items, and no additional manual annotations to more accurately identify patient cohorts.

Introduction

The research reported in this paper is part of a larger effort[1] to achieve better signal-to-noise ratio, hence accuracy, in pharmacovigilance applications. The relatively low frequency of occurrence of adverse drug reactions (ADRs) leads to weak causal relations between the reaction and any measured signal.[2,3] We hypothesize that by grouping related signals, we can enhance detection rate and suppress false alarm rate.

Methods

The proposed method has the following steps:

1. Identify findings related to the diagnosis of interest and calculate the corresponding values for each instance in patient sample, W. Let X_j be the vector of related findings for patient j and D_j the number of his/her encounters with related ICD-9 codes.
2. Identify a training set T, from W, containing positive and negative instances.
3. Set C_j (the class of patient j) to D_j
4. Begin iterative process (superscript denotes iteration number):
    a. Using T, build support vector machine based classification models to obtain non-linear relationship $\hat{C}_j^i = f^i(X_j)$, where $\hat{C}_j^i$ is the model's estimate for C_j.
    b. Apply $f^i(X_j)$ to W and generate histogram over $\hat{C}_j^i$ for all patients in W.
    c. Select a cut-off threshold η^i to separate the positive and negative populations in the histogram (e.g., η^i can be the global minima of the function describing the histogram).
    d. Set $L_j^i = 1$ if $\hat{C}_j^i > \eta^i$ and $L_j^i = 0$ otherwise; here, L_j^i is our estimate of the patient's label and a positive label indicates the patient is positive for the condition.
    e. Compare L_j^i to L_j^{i-1} and compute F^i, the percentage of patients for whom the label has changed in the current iteration.
    f. If $F^i < \delta$ (where δ is an acceptable threshold), return L_j^i and terminate; else, update $\hat{C}_j^{i+1} = [(i-1) \times \hat{C}_j^i + D]/i$ for all j in T.

It can be seen from the above definition that the influence of D_j, the number of ICD-9 codes of the given diagnosis, tapers down as training progresses, while the learned relationship dominates.

Results

We applied the method described to identify diabetes and hyperlipidemia patient cohorts in a Logician database containing structure data for 800,000 patients. Relevant features (Step 1) were identified by consultation with clinicians and domain experts.

For diabetes, the feature set included the number of abnormal hemoglobin A1C tests, number of anti-diabetic medications (insulin, insulin supplements, biguanides, sulphonylureas, alpha-glucosidase inhibitors and so on) and abnormal blood glucose tests. Although only 15,000 patients in the database had diabetes related ICD codes (250.*), the method described here identified 22000 patients as diabetic.

In the case of hyperlipidemia, the number of abnormal lipid panel tests and related medications (HMG-CoA reductase inhibitors, intestinal cholesterol absorption inhibitors and so on) were used as features. The model labeled 76,000 patients as positive for hyperlipidemia which is almost twice the number of patients who could have been identified using ICD codes alone (272.*).

For both conditions, clinician evaluation was conducted on 100 cases. The values for recall and f-measure observed

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

with the bootstrapping algorithm were found to be higher than those observed with ICD-9 codes.

Acknowledgements

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

References

1 Zeng Q, Cimino JJ, Zou KH. Providing concept-oriented views for clinical data using a knowledge-based system: an evaluation. *J Am Med Inform Assoc* 2002;**9**:294–305.
2 Noren GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug-drug interaction surveillance. *Stat Med* 2008;**27**:3057–70.
3 Noren GN, Edwards IR. Modern methods of pharmacovigilance: detecting adverse effects of drugs. *Clin Med (London, England)* 2009;**9**:486–9.

Emerging Health Threats Journal 2011, **4**:s35. **doi:** 10.3134/ehtj.10.035

© 2011 KJ Konty and DR Olson.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

The spatial-temporal pattern of excess influenza visits at the (sub-)urban scale

KJ Konty, and DR Olson

New York City Department of Health and Mental Hygiene, New York, NY, USA
E-mail: kkonty@health.nyc.gov

Objective

To develop a novel method to characterize the spatial-temporal pattern of seasonal influenza and then use this characterization to: (1) inform the spatial cluster detection efforts of syndromic surveillance, (2) explore the relationship of spatial-temporal patterns and covariates and (3) inform conclusions made about the burden of seasonal and pandemic influenza.

Introduction

Quantifying the spatial-temporal diffusion of diseases such as seasonal influenza is difficult at the urban scale for a variety of reasons including the low specificity of the extant data, the heterogenous nature of healthcare seeking behavior and the speed with which diseases spread throughout the city. Nevertheless, the New York City Department of Health and Mental Hygiene's syndromic surveillance system attempts to detect spatial clusters resulting from outbreaks of influenza. The success of such systems is dependent on there being a discernible spatial-temporal pattern of disease at the neighborhood (sub-urban) scale.

We explore ways to extend global methods such as serfling regression that estimate excess burdens during outbreak periods to characterize these patterns. Traditionally, these methods are aggregated at the national or regional scale and are used only to estimate the total burden of a disease outbreak period. Our extension characterizes the spatial-temporal pattern at the neighborhood scale by day. We then compare our characterizations to prospective spatial cluster detection efforts of our syndromic surveillance system and to demographic covariates.

Methods

We use the NYCDOH's syndromic surveillance emergency department system. The data includes daily data from 50 of 55 ED's in NYC representing 95% of all ED visits. An archive of prospective analyses records all significant spatial signals identified by our syndromic surveillance system for Fever/flu, influenza-like illness (ILI) and respiratory syndromes traditionally used in influenza surveillance. We define historic disease periods using WHO collaborating laboratory isolate data in New York City.

NYCDOH currently utilizes serfling regression to estimate excess visits during influenza periods. We develop a novel spatial serfling model that assigns the excess visits to dates, neighborhoods (as defined by zipcodes) and age groups. Spatial poisson regression methods were also explored for this assignment. The resulting excess patterns are consistent with the citywide excess calculations. We then compare the resulting patterns to the archive of historic spatial clusters as well as spatial demographic patterns. Finally, we notice

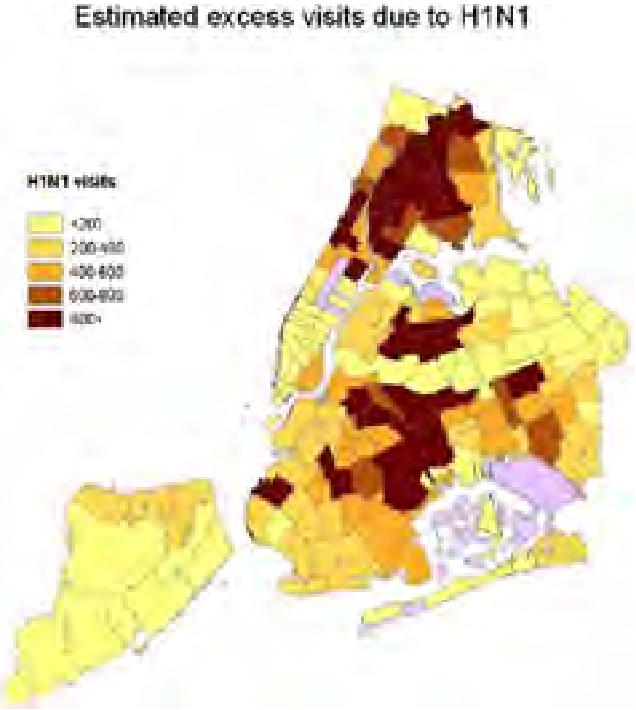


Figure 1 Spatial pattern of excess during H1N1, 2009. The resulting pattern is remarkably similar to past seasonal influenza patterns.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

regularities of the patterns across flu seasons including the 2009 H1N1 season.

Results

Spatial-temporal patterns of excess ILI visits are remarkably consistent and appear to be largely driven by health-seeking behavior. Although spatial clusters are regularly detected by routine syndromic surveillance, their similarity suggests that these patterns may result from shifts in spatial pattern because of demographic characteristics and health-seeking behavior rather than the spatial diffusion of influenza. Only the May 2009 H1N1 pandemic wave shows a clear pattern of spatial diffusion emanating from Queens. However, the spatial pattern of excess for the entire 2009 H1N1 period is similar to the previous seasonal influenza seasons. Figure 1

Conclusions

The spatial serfling model developed here characterizes the spatial-temporal pattern of individual flu seasons. Unlike cluster detection systems, the resulting patterns are global and can be used to explore spatial relationships with demographic covariates. The lack of variation across flu periods suggests that spatial cluster detection for influenza may not be useful at the neighborhood scale. This may be due to the speed of diffusion at the urban scale. The spatial temporal pattern of ED visits during H1N1 was similar to all previous influenza seasons and calls into question some observations made about the uniqueness of the outbreak, most notably that the poor were more likely to be infected—this may well be a trait common to all seasonal influenza epidemics.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s38. **doi:** 10.3134/ehtj.10.038
© 2011 C McCullough *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Where are the data? Accuracy of automated EHR reporting

C McCullough, S Catlett, and S Shih

Primary Care Information Project, New York City Department of Health and Mental Hygiene, New York, NY, USA
E-mail: cmccullo@health.nyc.gov

Objective

Comparison of automated EHR-derived data with manually abstracted patient information on smoking status and cessation intervention.

Introduction

Over 300 independent practices transmit monthly quality reports to a data warehouse using an automated process to summarize patient information into quality measures. All practices have implemented an EHR that captures clinical information to be aggregated for population reporting, and is designed to assist providers by generating point-of-care reminders and simplify ordering and documentation.

Methods

A total of 82 small practices received training on documenting in the EHR to maximize performance on quality measures as part of participation in a pilot pay-for quality program. Retrospective clinical chart reviews were completed for 3278 patients at 46 practices. Chart reviewers collected visit-based information on patients' diagnoses, vitals, laboratory results, smoking status and if they were a smoker, whether the patient received cessation intervention. Chart reviewers also recorded whether the information was documented in the smart form, social history or other locations of the EHR. For the automated calculation of the cessation intervention measure, smokers are identified through a smart form and an intervention is recorded whether a patient receives either a referral to the New York State Fax-to-Quit program (http://www.nysmokefree.com/Pageview.aspx?p = ftq), a prescription for tobacco cessation medication, or counseling within the past year (Table 1).

Table 1 Location of Documentation for Current Smokers

		Smoking status location		
		Smart form	*Other*	*Total*
Cessation	Yes	72	31	103
Intervention	No	129	102	231
Received	Total	201	133	334

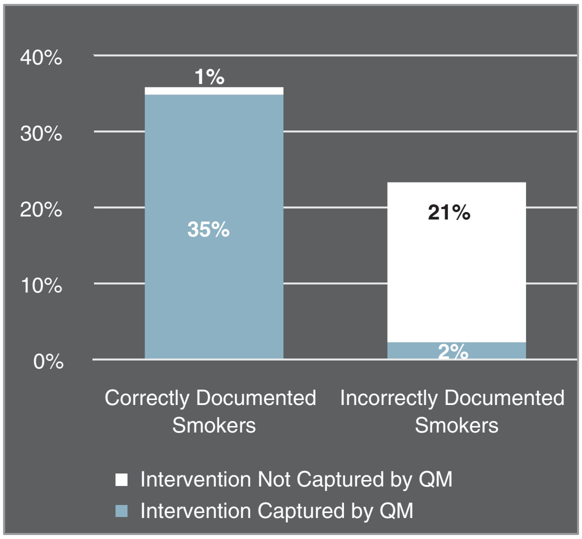


Figure 1 Smokers receiving a cessation intervention.

Results

Providers recorded a smoking status for 85% of 3278 patients, 57% (of 2785) of which were documented in the smart form and captured for automated reporting. For the 1,197 patients not captured for reporting, a majority of the smoking status was recorded in the social history (89%).

The automated quality measure captured 201 (60%) of the 304 current smokers (Figure 1).

Conclusions

Because a limited number of structured fields are used for measure queries, the smoking status and cessation quality measures generally undercount performance. Despite training, providers do not consistently document smoking status as intended by the EHR interface design. As a consequence, smokers not captured by the quality measure are less likely to

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

receive a cessation intervention because EHR-based alerts and reminders are not triggered.

When smoking status was not documented in the smart form, providers consistently used the social history structured field, which is not captured in the current measure query. The social history may be an equally valid place to record the information, and its use may indicate EHR usability and design issues for capturing and reporting on smoking status/cessation.

EHR vendors need to carefully consider provider workflows when designing user interfaces and choosing the structured fields used to generate quality measures. Updates to the EHR software to accurately report on provider delivery of services can be time consuming and disruptive to the practice. Because quality measure specifications are frequently revised, it is also important for vendors to design systems that have the flexibility to incorporate changes with greater ease or be able to update specifications recommended by evidence-based practice guidelines.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s42. **doi:** 10.3134/ehtj.10.042
© 2011 DB Neill *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Fast subset scan for multivariate spatial biosurveillance

DB Neill, E McFowland III, and H Zheng

Event and pattern detection laboratory, Carnegie Mellon University, Pittsburgh, PA, USA
E-mail: neill@cs.cmu.edu

Objective

We extend the recently proposed 'fast subset scan' framework from univariate to multivariate data, enabling computationally efficient detection of irregular space-time clusters even when the numbers of spatial locations and data streams are large. These fast algorithms enable us to perform a detailed empirical comparison of two variants of the multivariate spatial scan statistic, demonstrating the tradeoffs between detection power and characterization accuracy.

Introduction

The spatial scan statistic[1] detects significant spatial clusters of disease by maximizing a likelihood ratio statistic over a large set of spatial regions. Several recent approaches have extended spatial scan to multiple data streams. Burkom[2] aggregates actual and expected counts across streams and applies the univariate scan statistic, thus assuming a constant risk for the affected streams. Kulldorff *et al.*[3] separately apply the univariate statistic to each stream and then aggregate scores across streams, thus assuming independent risks for each affected stream. Neill[4] proposes a 'fast subset scan' approach, which maximizes the scan statistic over proximity-constrained subsets of locations, improving the timeliness of detection for irregularly shaped clusters. In the univariate event detection setting, many commonly used scan statistics satisfy the 'linear-time subset scanning' (LTSS) property, enabling exact and efficient detection of the highest-scoring space-time clusters.[4]

Methods

In the multivariate setting, we wish to search over proximity-constrained subsets of locations and all subsets of the monitored data streams, but an exhaustive search over subsets is computationally infeasible, scaling exponentially with the number of streams and the maximum neighborhood size. We develop computationally efficient algorithms for both the Burkom and Kulldorff multivariate scan approaches. For Burkom's method, we iterate between two steps, optimizing over subsets of streams for the current subset of locations, and optimizing over subsets of locations for the current subset of streams. For Kulldorff's method, we iterate between optimizing over subsets of locations for fixed values of the relative risks for each stream, and recalculating the maximum likelihood risk values for the current subset of locations. Each optimization over subsets can be performed efficiently for any statistic satisfying the LTSS property: we sort the locations (streams) by a priority function, and then consider subsets consisting of the top-k highest priority locations (streams), for $k = 1 \ldots N$. We can prove that one of these will be the highest scoring subset. Both fast algorithms converge to a local maximum of the score function, and our experiments demonstrate that each closely approximates the global maximum with high probability.

Results

We compared fast localized scan (searching over proximity-constrained subsets of locations) and circular scan approaches, for both the Burkom and Kulldorff methods, monitoring multiple streams of real-world Emergency Department data from Allegheny County, PA. Our fast algorithms enable both multivariate scan statistics to be optimized over proximity-constrained subsets of the 97 zip codes and all subsets of the 16 monitored data streams in less than 2 s per day of data, whereas exhaustive search would require hundreds of millions of years. Comparing the Burkom and Kulldorff methods, we find tradeoffs between detection and characterization performance: Kulldorff's method exhibits slight but significant improvements in detection time, whereas Burkom's method more accurately characterizes the affected subset of streams. For both methods, our fast localized scan approach improved timeliness of detection by 1 to 2 days as compared with circular scan, and also increased spatial accuracy (weighted overlap coefficient between true and detected regions) from 70 to 83%. More details of our methods and results are provided in the full paper.[5]

Conclusions

By extending fast subset scan to the multivariate setting, we enable more timely detection of emerging events using

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

multiple data streams, as well as accurate characterization of the affected subset of streams.

Acknowledgements

This work was partially supported by NSF grants IIS-0916345, IIS-0911032, and IIS-0953330. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Kulldorff M. A spatial scan statistic. *Commun Statist: Theory and Methods* 1997;**26:**1481–96.
2 Burkom HS. Biosurveillance applying scan statistics with multiple disparate data sources. *J Urban Health* 2003;**80:**i57–i65.
3 Kulldorff M, Mostashari F, Duczmal L, Yih K, Kleinman K, Platt R. Multivariate spatial scan statistics for disease surveillance. *Stat Med* 2007;**26:**1824–33.
4 Neill DB. Fast and flexible outbreak detection by linear-time subset scanning. *Adv Dis Surveill* 2008;**5:**48.
5 Neill DB, McFowland E, Zheng H. Fast subset scan for multivariate event detection and visualization. *Stat Med* 2011 (in press).

Emerging Health Threats Journal 2011, **4**:s43. **doi:** 10.3134/ehtj.10.043
© 2011 DB Neill and Y Liu.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Generalized fast subset sums for Bayesian detection and visualization

DB Neill, and Y Liu

Event and Pattern Detection Laboratory, Carnegie Mellon University, Pittsburgh, PA, USA
E-mail: neill@cs.cmu.edu

Objective

We propose a new, computationally efficient Bayesian method for detection and visualization of irregularly shaped clusters. This Generalized Fast Subset Sums (GFSS) method extends our recently proposed MBSS and FSS approaches, and substantially improves timeliness and accuracy of event detection.

Introduction

The multivariate Bayesian scan statistic (MBSS)[1] enables timely detection and characterization of emerging events by integrating multiple data streams. MBSS can model and differentiate between multiple event types: it uses Bayes' Theorem to compute the posterior probability that each event type E_k has affected each space-time region S. Results are visualized using a 'posterior probability map' showing the total probability that each location has been affected. Although the original MBSS method assumes a uniform prior over circular regions, and thus loses power to detect elongated and irregular clusters, our Fast Subset Sums (FSS) method[2] assumes a hierarchical prior, which assigns non-zero prior probabilities to every subset of locations, substantially improving detection power and accuracy for irregular regions.

Methods

We propose GFSS, a generalized Bayesian framework, which includes both FSS and the original MBSS method as special cases. As in FSS, we define a hierarchical prior over all 2^N subsets of the N locations. We first choose the center location s_c and size $n \in \{1 \ldots N\}$ uniformly at random. Given the 'neighborhood' Z consisting of s_c and its $n-1$ nearest neighbors, each location $s_i \in Z$ is independently included with probability P, where the parameter P defines the 'sparsity' of the region. FSS assumes a uniform prior over all 2^n subsets of Z, and thus corresponds to GFSS with $P=0.5$, whereas MBSS only considers circular regions, and thus corresponds to $P=1$.

Naïve computation of the posterior probability map using GFSS would require us to compute and sum over an exponential number of region probabilities, which is computationally infeasible for $N>25$. However, we show that, for any value $0<P\leqslant 1$, the posterior probability map can be computed without computing each individual region probability, thus reducing the run time from exponential to polynomial in N. In practice, this means that we can monitor hospital data from 97 Allegheny County zip codes in less than 10 s per day of data using GFSS (Figure 1).

Results

We evaluated the detection power and spatial accuracy of GFSS for 10 values of the sparsity parameter P ranging from 0.1 to 1.0. We tested these methods on 10 differently-shaped semi-synthetic outbreaks (200 injects of each type) injected

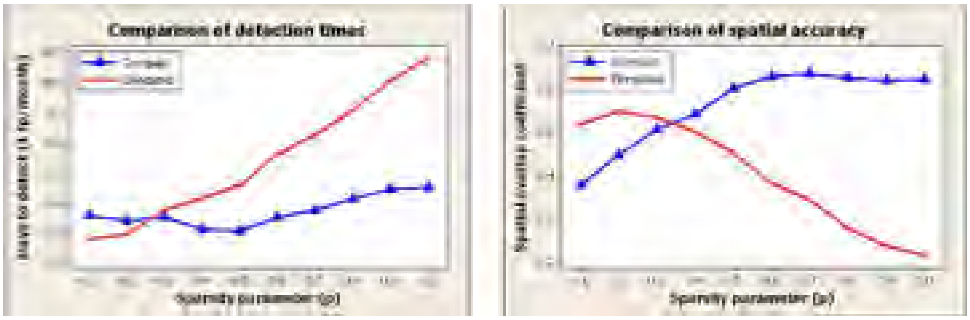


Figure 1 Detection time and spatial accuracy for GFSS, as a function of the sparsity parameter *P*.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

into two streams of real-world Emergency Department data from Allegheny County, PA. Figure 1 shows the average detection time (days to detect at 1 false positive per month) and spatial accuracy (overlap coefficient between true and detected clusters) for each method. Our results show that the optimal value of P depends strongly on the shape of the outbreak: for compact clusters, the original FSS method (GFSS with $P = 0.5$) minimizes detection time, while spatial accuracy was maximized at $P = 0.7$. For highly elongated outbreaks, however, GFSS with $P = 0.2$ achieved substantial improvements as compared with FSS, including 0.8 days faster detection and 10% higher spatial accuracy. These results suggest that incorporating previous information about the density or sparsity of an outbreak can improve detection power. Additionally, GFSS enables us to more accurately distinguish between multiple outbreak types with different sparsity parameters. The optimal value of P for each outbreak type can be learned automatically, using labeled data from outbreaks of that type.

Conclusions

Our results demonstrate that GFSS can dramatically improve event detection and visualization as compared with MBSS and FSS, while still enabling fast, exact computation of the posterior probability map.

Acknowledgements

This work was partially supported by NSF Grants IIS-0916345, IIS-0911032 and IIS-0953330. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Neill DB, Cooper GF. A multivariate Bayesian scan statistic for early event detection and characterization. *Mach Learn* 2010;7:261–82.
2 Neill DB. Fast multivariate Bayesian scan statistics for event detection and visualization. *Stat Med* 2010 (in press).

Emerging Health Threats Journal 2011, **4**:s44. **doi:** 10.3134/ehtj.10.044
© 2011 J Niemi and M Ludkovski.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Optimal sequential management decisions for influenza outbreaks

J Niemi, and M Ludkovski

Department of Statistics and Applied Probability, University of California-Santa Barbara, Santa Barbara, CA, USA
E-mail: niemi@pstat.ucsb.edu

Objective

This abstract highlights a methodology to build optimal management policy maps for use in influenza outbreaks in small populations.

Introduction

Management policies for influenza outbreaks balance the expected morbidity and mortality costs versus the cost of intervention policies under outbreak parameter uncertainty. Previous approaches have not updated parameter estimates as data arrives[1] or have had a limited set of possible intervention policies.[2] We present a methodology for dynamic determination of optimal policies in a stochastic compartmental model with sequentially updated parameter uncertainty that searches the full set of sequential control strategies.

Methods

We model small population influenza outbreaks using a stochastic SIR-model with parameters controlling the S→I and I→R transition rates. The full posterior distribution for the parameters is sequentially updated at each data point, for example, Figure 1.

Isolation and vaccination can be initiated at any time to modify the S→I transition rate and create an S→R path, respectively. The total cost for an outbreak is the sum of the costs for infected individuals as well as ongoing and fixed costs for intervention strategies that depend on when each intervention is initiated. Combining the techniques of dynamic programming and regression Monte Carlo, we use simulation techniques to build an optimal policy map for all possible future outbreak scenarios. Once an outbreak has begun, real-time decisions are made by calculating current parameter and state estimates and then consulting the policy map.

Results

As an illustration of our methodology, we consider the case of a flu outbreak in an English boarding school described in[3] and recently used by Merl D *et al.*[2] As our initial state, we take two infected and 761 susceptible individuals. In Table 1, we compare the cost of our dynamic strategy to simpler strategies over random outbreak realizations.

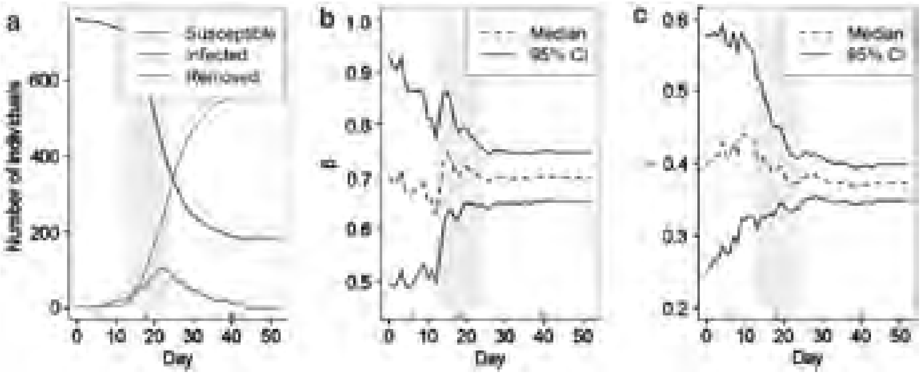


Figure 1 A realization of the stochastic SIR model along with sequential point wise medians and 95% credible intervals for sequential inference on model parameters.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1 Average cost comparison of the dynamic control (Dyn) strategy versus no control (NC), isolation always (Iso), threshold (Thr) and fixed-date (F-D) for various infected cost structures, $c(i) = i + 0.1(\max(i - I, 0))^2$

I	*NC*	*Iso*	*Thr*	*F-D*	*Dyn*
∞	1030	1400	1091	942	876
75	3350	1440	1091	1183	886
50	4950	1515	1091	1378	902
25	7250	1666	1097	1667	931

Conclusion

We have built novel methodology to find optimal policy maps for dynamic risk management of flu outbreaks in a stochastic framework with parameter uncertainty. The methodology accounts both for stochastic interactions between individuals during outbreaks as well as uncertainty about the outbreak parameters that are important for policy makers. The simulation-based control algorithm computes an approximately optimal adaptive and dynamic management strategy, which creates a full policy map across all possible outbreak scenarios.

Acknowledgements

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

References

1 Tanner MW, Sattenspiel L, Ntaimo L. Finding optimal vaccination strategies under parameter uncertainty using stochastic programming. *Math Biosci* 2008;**215**:144–51.
2 Merl D, Johnson LR, Gramacy RB, Mangel M. A statistical framework for the adaptive management of epidemiological interventions. *PLoS ONE* 2009;**4**:e5087.
3 Murray JD. *Mathematical Biology: An Introduction*. Springer: New York, NY, USA, 1993.

Emerging Health Threats Journal 2011, **4**:s45. **doi:** 10.3134/ehtj.10.045
© 2011 Y Ohkusa *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Real-time estimation and prediction for pandemic A/H1N1 (2009) in Japan

Y Ohkusa, T Sugwara, K Taniguchi and N Okabe

Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Tokyo, Japan
E-mail: ohkusa@nih.go.jp

Objective

This paper summarized our effort for real-time estimation of pandemic influenza A/H1N1pdm in Japan.

Introduction

Unfortunately, confirmation and notification of all A/H1N1 (2009) patients in Japan was ceased on 24 July when the cumulative number of patients was about 5000. After that, as all suspected patients are not necessarily confirmed or reported, the only official surveillance was the sentinel surveillance for influenza-like-illness (ILI) patients from 5000 clinics accounting for almost 10% of all clinics and hospitals in Japan. However, because the surveillance results are reported weekly, it tends to lack timeliness. To collect and analyze the information in more timely manner, we, Infectious Disease Surveillance Center, National Institute of Infectious Diseases, developed a full automatic daily reporting system of ILI patients. Using this information, we had estimated *R*v and predict its course in every week.

Methods

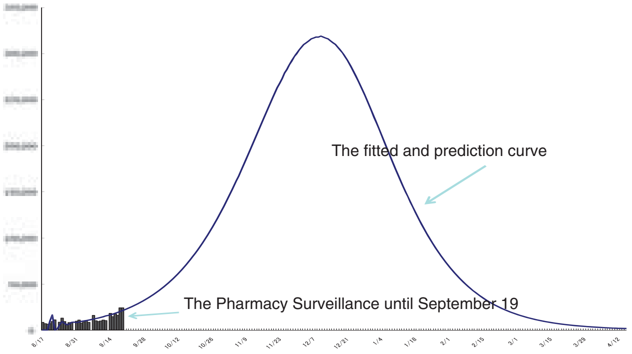
In Japan, prescription of anti-flu drugs, such as oseltamivir and zanamivir, have been automatically monitored daily since April 2009 at about 3350 pharmacies that account for almost 7% of all pharmacies in Japan. By weighing the number of prescription for oseltamivir and zanamivir with the proportion of participating pharmacies, we can estimate the total number of patients in Japan of the previous day. We usually publish it online at 0700 hours daily.

Our estimation and prediction model is a very simple SIR model and parameters used are described in the previous study, that is, natural history, period when the patient is infectious, rate and infectiousness of asymptomatic patients, and withdrawal rate are borrowed from these two previous researches. *R*v is estimated by maximum likelihood, which minimizes the difference between an actual epidemic curve and estimated one, assuming it to have a normal distribution. Taking into the situation in Japan, we can assume that all symptomatic patients visit a doctor to receive oseltamivir or zanamivir on onset date. Thus, we can compare the number of prescription of oseltamivir or zanamivir in the pharmacy surveillance and the number of patients on onset date in the SIR model. Confidence interval (CI) of *R*v is calculated using the second derivative of likelihood as its estimator of variance. We estimated and predicted every week since September 2009.

Results

Bars in the figure show the estimated total number of patients each day from the pharmacy surveillance between August and 19 September, and the smoothed line is a fitted curve by the model. The estimation on 19 September shows estimated *R*v of 1.797, but we did not provide its CI at that time. The figure shows prediction results over the whole course of the pandemic, and its predicted peak would be 10 December.

The estimation on 29 November shows estimated *R*v of 1.72 and its 95% CI was 1.69, 1.75. It is estimated that the peak will be reached on the 23 December with 95% CI of 14 December and 2 January 2010. Then, the number of patients is estimated to reach 227 000 with 95% CI of 193 000 and 262 000 at the peak. The cumulative number of patients over the period will be as high as 17.8 (16.6, 19.0)% of total



OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

population. These estimates are circulated weekly among central and local government officers in charge of pandemic control.

Conclusions

We, Infectious Disease Surveillance Center, National Institute of Infectious Diseases, performed real-time estimation and prediction every week from September till the end of November, and the obtained information was circulated among central and local government officers fighting pandemic. It was a helpful tool to understand the speed and impact of pandemic and to plan for better counter-measure against pandemic.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s46. **doi:** 10.3134/ehtj.10.046
© 2011 A Okhmatovskaia *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Quantifying the potential benefit of early detection for preventing morbidity and mortality: a simulation study of cryptosporidium outbreak

A Okhmatovskaia[1], A Verma[1], C Jauvin[1], B Barbeau[2], R Allard[3], and DL Buckeridge[1]

[1]*Clinical and Health Informatics, McGill University, Montréal, Canada;* [2]*École Polytechnique de Montréal, Montréal, Canada; and* [3]*Direction de Santé Publique de Montréal, Montréal, Canada*
E-mail: okhmatovskaia@gmail.com

Objective

To quantitatively assess the benefit of issuing a boil-water advisory for preventing morbidity and mortality from a waterborne outbreak of cryptosporidiosis.

Introduction

Many studies evaluate the timeliness and accuracy of outbreak detection algorithms used in syndromic surveillance. Of greater interest, however, is defining the outcome associated with improved detection.

In case of a waterborne cryptosporidiosis outbreak, public health interventions are aimed exclusively at preventing new infections, and not at medical treatment of infected individuals. The effectiveness of these interventions in reducing morbidity and mortality will depend on their timeliness, the level of compliance, and the duration of exposure to pathogen.

In this work, we use simulation modeling to examine several scenarios of issuing a boil-water advisory (BWA) as a response to outbreak detection through syndromic surveillance, and quantify the possible benefits of earlier interventions.

Methods

We developed an agent-based model for simulating realistic outbreak signals similar to historical waterborne outbreaks of gastrointestinal disease.[1] The model generates a synthetic population and simulates the spread of pathogenic organisms through water system, individual mobility and water consumption, disease progression in infected individuals, and patterns of healthcare utilization and disease reporting—all within the geographical setting of the Island of Montreal. We also model the effect of BWA on water consumption determined by the rate of compliance with the advisory.

We simulated an outbreak scenario resulting from a 2-week failure of a water treatment plant, similar to that observed in Milwaukee in 1993. We varied the BWA compliance rate, and timing relative to symptom onset (earliest possible time for BWA guided by syndromic surveillance only). Because Milwaukee incident was the largest documented outbreak in terms of duration and pathogen concentration, and might not be the most representative scenario, we also simulated a shorter outbreak with a 72-hour treatment plant failure. In all experiments, we measured the total number of disease cases and deaths.

Results

Figure 1 displays the number of cases prevented by the BWA under a 2-week contamination scenario, as a function of the delay for different compliance levels. There appears to be little benefit in issuing a BWA 5 or more days after the beginning of the outbreak. In the 72-hour scenario, same happens with the delay exceeding 1 day. We can also see that

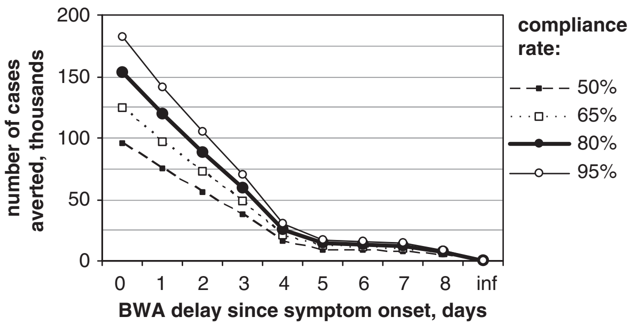


Figure 1 Benefit of BWA as a function of its timeliness and compliance rate.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the benefits of BWA decline faster with time than with deteriorating compliance: issuing a BWA 1 day later is equivalent to about 20% loss in compliance.

Conclusions

We have shown that even for large outbreaks like the Milwaukee outbreak, the time window for effective intervention can be only 5 days wide if BWA relies on detection through human health surveillance, suggesting that using additional sources of information (for example, water quality surveillance) to enable earlier intervention can be beneficial in preventing morbidity and mortality. This finding, however, depends on some additional factors related to outbreak scenario, like daily infection rate (a function of pathogen concentration) and specifics of the water distribution system, and thus should be generalized with care.

Our results also suggest that timeliness of BWA is more critical than high compliance rates. A good decision about issuing a BWA must take into account its cost and the uncertainty of outbreak detection.

Acknowledgements

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

Reference

1 Okhmatovskaia A, Verma AD, Barbeau B, Carriere A, Pasquet R, Buckeridge DL, *et al*. A simulation model of waterborne gastrointestinal disease outbreaks: description and initial evaluation. AMIA Annual Symposium, 2010.

Emerging Health Threats Journal 2011, **4**:s47. **doi:** 10.3134/ehtj.10.047
© 2011 FLP Oliveira *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Non-parametric intensity bounds for the visualization of disease clusters

FLP Oliveira[1,2], LH Duczmal[1], and ALF Cançado[3]

[1]*Statistics Department, Federal University of Minas Gerais, Minas Gerais, Brazil;* [2]*Ibmec, Minas Gerais, Brazil; and* [3]*Statistics Department, University of Brasília, Brasília, Brazil*
E-mail: fernandoluizoliveira@est.ufmg.br

Objective

Given an aggregated-area map with disease cases data, we propose a criterion to measure the plausibility of each area in the map of being part of a possible localized anomaly.

Introduction

Consider the most likely disease cluster produced by any given method, like SaTScan,[1] for the detection and inference of spatial clusters in a map divided into areas; if this cluster is found to be statistically significant, what could be said of the external areas adjacent to the cluster? Do we have enough information to exclude them from a health program of prevention? Do all the areas inside the cluster have the same importance from a practitioner perspective? How to access quantitatively the risk of those regions, given that the information we have (cases count) is also subject to variation in our statistical modeling? A few papers have tackled these questions recently;[2] produces confidence intervals for the risk in every area, which are compared with the risks inside the most likely cluster. There exists a crescent demand of interactive software for the visualization of spatial clusters.[3] A technique was developed[4] to visualize relative risk and statistical significance simultaneously.

Methods

In this work, we assess the problem of finding error bounds for the delineation of spatial clusters in maps of areas with observed populations and number of cases. A given map with the vector of real data (the number of observed cases for each region) shall be considered as one possible realization of the random variable vector with an unknown expected number of cases. Therefore, the process of identification of possible spatial clusters must take into account this source of variation.

In our methodology, we perform *m* Monte Carlo replications of the vector of random variables for fixed rates given the observed number of cases in each area. Then the most likely cluster for each replicated map is detected and the corresponding likelihood values obtained in the *m* replications are ranked. For each area, we determine the maximum-likelihood value among the most likely clusters containing that area. Thus, we obtain the intensity function associated to each area's ranking of its respective likelihood value among the *m* values.

Results

We apply this tool for three different maps for sharply and diffusely delineated clusters. The intensity bounds found by the method reflect the geographic dispersion of the detected clusters, as in Figure 1.

Conclusions

Our technique is able to detect irregularly shaped and multiple clusters, making use of simple tools like the circular scan. Intensity bounds for the delineation of spatial clusters are obtained and indicate the plausibility of each area belonging to the cluster. This tool employs simple mathematical concepts, and interpreting the intensity function is very intuitive in terms of the importance of each region in delineating the possible anomalies of the map of rates.

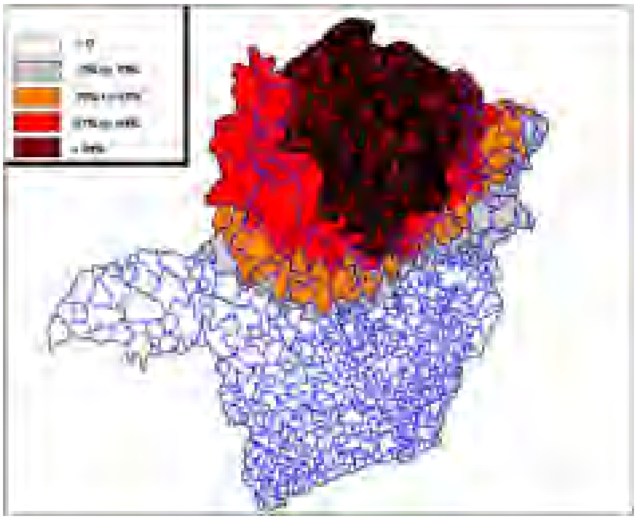


Figure 1 Intensity map quantiles for Chagas' disease data.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Monte Carlo simulation requires an effort similar to the circular scan algorithm, and, therefore, it is quite fast. We hope that this tool should be useful in public health decision making of which areas should be prioritized.

Acknowledgements

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

References

1 Kulldorf M. Information Management Services Inc: SaTScan v7.0: software for the spatial and space-time scan statistics. *Tech Rep* http://www.satscan.org/ 2006.
2 Rosychuk RJ. Identifying geographic areas with high disease rates: when do confidence intervals for rates and a disease cluster detection method agree? *Int J Health Geogr* 2006;**5**:46.
3 Hardisty F, Conley J. Interactive detection of spatial clusters. *Adv Dis Surveillance* 2008;**5**:37.
4 Boscoe FP, McLaughlin C, Schymura MJ, Kielb CL. Visualization of the spatial scan statistic using nested circles. *Health Place* 2003;**9**:273–7.

Emerging Health Threats Journal 2011, **4**:s48. **doi:** 10.3134/ehtj.10.048
© 2011 DR Olson *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Evaluating a standard influenza-like illness syndrome definition across multiple sites in the distribute project: The 'ILI-s' Pilot

DR Olson[1], A Baer[1,2,5], M Paladini[1,3], H Burkom[1,4], D Revere[1,5], I Painter[1,5], WB Lober[1,5], JS Brownstein[1,6] and DL Buckeridge[1,7], for the Distribute Research and Evaluation Working Group

[1]*International Society for Disease Surveillance (ISDS), Boston, MA, USA;* [2]*Public Health-Seattle and King County, Seatlle, WA, USA;* [3]*New York City Department of Health and Mental Hygiene, New York City, NY, USA;* [4]*Johns Hopkins University Applied Physics Laboratory, Baltimore, MD, USA;* [5]*University of Washington, Seattle, WA, USA;* [6]*Harvard University Children's Hospital, Boston, MA, USA; and* [7]*McGill University, Montreal, Quebec, Canada*
E-mail: drolson@gmail.com

Objective

To describe the initial phase of the ISDS Distribute project influenza-like illness (ILI) syndrome standardization ('ILI-s') pilot.

Introduction

The Distribute project began in 2006 as a distributed, syndromic surveillance demonstration project that networked state and local health departments to share aggregate emergency department-based ILI syndrome data.[1] Preliminary work found that local systems often applied syndrome definitions specific to their regions; these definitions were sometimes trusted and understood better than standardized ones because they allowed for regional variations in idiom and coding and were tailored by departments for their own surveillance needs.[2,3] Originally, sites were asked to send whatever syndrome definition they had found most useful for monitoring ILI. Places using multiple definitions were asked to send their broader, higher count syndrome.[4] In 2008, sites were asked to send both a broad syndrome (ILI-broad), and a narrow syndrome (ILI-narrow) specific to ILI.[5]

Methods

Selected Distribute health department sites were sent an email inviting them to participate in the pilot. Sites were told that the pilot was intended to assess their ability to respond to a query in the event of a public health emergency; to evaluate syndrome component characteristics across sites; and obtain objective evidence regarding whether and how to standardize syndromes. Sites were asked to submit historical ILI data using a standardized syndrome definition (ILI-s) made-up of three component sub-syndromes: 'fever and cough', 'fever and sore throat' and 'flu'. SAS code was provided. A table with inclusion and exclusion terms was included for sites to create the syndromes in code other than

Figure 1 Sample region dataset of locally applied ILI and ILI-s time-series visualized on the Distribute restricted website.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

SAS if preferred. Sites were asked to follow the ILI-s concepts and coding as closely as possible, and to document and share local ILI-narrow and ILI-broad code already in use. ILI-s pilot data were collected through the same upload procedure used for routine data submission. After ILI-s submission, data could be visualized through the Distribute restricted website. Comparison of local ILI-narrow, ILI-broad and ILI-s by sub-syndrome was conducted by age group and region (Figure 1).

Results

Of the 12 invited sites, six participated in the ILI-s exercise (several expressed interest but were unable to dedicate the necessary time). Comparison of ILI-s and regionally defined syndromes found considerable variation within and between sites in the relative proportion of each one, the level of signal-to-noise and the age-specific trends. However, inter-regional comparisons were clearer using the standardized syndromes, which had less noise compared with the regionally defined syndromes.

Conclusions

Response to the initial phase of the Distribute query and standardization ILI-s pilot suggests the exercise can be expanded to the larger Distribute project, and evaluation of local ILI and ILI-s can be conducted collaboratively with participating sites. This work will be ongoing within the Distribute community over the coming year.

Acknowledgements

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

References

1 Olson DR, Paladini M, Buehler JW, Mostashari F. *Adv Dis Surv* 2008;**5**:185 www.isdsjournal.org/article/view/3307/2456.
2 Paladini M, Pendarvis J, Murray E, Gunn J, Olson D. *Adv Dis Surv* 2008;**5**:50 www.isdsjournal.org/article/view/3252/2359.
3 Pendarvis J, Murray E, Paladini M, Gunn J, Olson D. *Adv Dis Surv* 2008;**5**:53 www.isdsjournal.org/article/view/3258/2362.
4 Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. *Emerg Infect Dis* 2004;**10**:858–64, www.cdc.gov/ncidod/EID/vol10no5/03-0646.htm.
5 Olson DR, Heffernan RT, Paladini M, Konty K, Weiss D, Mostashari F. *PLoS Med* 2007;**4**:e247 http://dx.doi.org/10.1371/journal.pmed.0040247.

Emerging Health Threats Journal 2011, **4**:s141. **doi:** 10.3134/ehtj.10.141
© 2011 J Paton *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Samos: a community-driven open-access prediction market system

J Paton[1], F Nelson[2], A Segre[1], and P Polgreen[3]

[1]*Department of Computer Science,The University of Iowa, Iowa City, IA, USA;* [2]*Department of Economics, The University of Iowa, Iowa City, IA, USA; and* [3]*Department of Internal Medicine, The University of Iowa, Iowa City, IA, USA*
E-mail: james-paton@uiowa.edu

Objective

This poster presents a software system to provide a community-driven, user-generated, low-overhead, web-based prediction market system called Samos.

Introduction

Prediction markets are a type of futures market in which users trade shares that pay off if the event to which they are connected occurs. They are used to aggregate knowledge on a large scale, as the prices of the various contracts can be interpreted as probabilities of their events. Since 2006, our group has been using prediction markets and testing their utility in predicting the spread and impact of diseases, including seasonal influenza, syphilis, and others on a market called the Iowa Electronic Health Markets (IEhM), found at http://iehm.uiowa.edu. For example, in 2009, a series of markets were run on novel influenza A (H1N1), which showed success in predicting the extent and duration of the outbreak.[1]

We currently plan to move into a new phase of development that will allow the community of users to submit proposals for new prediction markets, which will then be approved by site editors and referees. We call the new system Samos.

Methods

Samos consists of the prediction market engine already in use coupled with a new proposal management system (Proposals) currently in development. Proposals provide a workflow for the submission, construction, and approval of prediction markets that is modeled after the workflow for the submission and approval of journal articles. A user can submit a proposal by first submitting an abstract. Then an editor can decide whether to accept, reject, request revise-and-resubmit, or send the abstract to referees who can then vote accept, reject, or revise-and-resubmit. Once an abstract has been accepted, the user must submit a mock-up of the market, which contains the market's question, further explanation of the market, and the various contracts that can be traded. Proposals provide a browser-based online tool to create mock-ups. This information is used for the eventual presentation of the market to users for trading. Like the abstract, the mock-up can either be accepted, rejected, have revision requested, or sent to referees. Once a mock-up has been accepted, an editor chooses start and end dates for the market. The engine takes over from there, starting and ending the market accordingly. When the market has finished, the user must submit a final write-up, summarizing the market's motivation, question, resolution, and performance. The write-up must also go through the same approval process as the mock-up and abstract. Once the write-up has been approved, it becomes available to users and visitors of the site to view.

The abstract's purpose is to ensure that the market has been designed according to the proper prediction market principles and is noteworthy from a public health perspective. As such, an editor will generally choose one referee from each of the prediction market and public health fields. Abstracts must include information regarding the motivation and objective of the market as well as the final data sources to be used in deciding the winning contract.

Samos also includes a forum, or discussion board, which can be posted to by registered users. The forum will also provide private messaging capabilities so that users can communicate with one another, discuss running markets, and collaborate on prediction market proposals.

Results

We have developed a platform for eliciting ideas for prediction markets from a user community, transforming these ideas into contracts suitable for prediction markets, and encouraging collaboration among users. The platform's workflow system supports an incremental refinement process

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

between editors and the community members who generate new contract ideas. Samos increases the efficiency and versatility of prediction markets.

Conclusions

Prediction markets aggregate expert opinion, tying the price of a contract to the probability of its associated event. Samos provides a method of allowing a community of users to collectively manage a website for running prediction markets.

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.

Reference

1 Polgreen PM, Nelson FD. Using prediction markets to forecast infectious diseases. In: T Kass-Hout and X. Zhang (eds.) *Biosurveillance: Methods and Case Studies*. CRC Press: Boca Raton, FL, 2010, pp 145–60.

Emerging Health Threats Journal 2011, **4**:s49. **doi:** 10.3134/ehtj.10.049
© 2011 J Que and F-C Tsui.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A data simulation model using NRDM pharmaceutical sales counts

Jialan Que, and Fu-Chiang Tsui

RODS Laboratory, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: jiq4@pitt.edu

Objective

This study proposes a simulation model to generate the daily counts of over-the-counter medication sales, such as thermometer sales from all ZIP code areas in a study region that include the areas without retail stores based on the daily sales collected from the ZIP codes with retail stores through the National Retail Data Monitor (NRDM). This simulation allows us to apply NRDM data in addition to other data sources in a multivariate analysis in order to rapidly detect outbreaks.

Introduction

In disease surveillance, an outbreak is often present in more than one data type. If each data type is analyzed separately rather than combined, the statistical power to detect an outbreak may suffer because no single data source captures all the individuals in the outbreak.[1] Researchers, thus, started to take multivariate approaches to syndromic surveillance. The data sources often analyzed include emergency department (ED) data, categorized by chief complaint; over-the-counter (OTC) pharmaceutical sales data collected by the National Retail Data Monitor (NRDM), and some other syndromic data.[1,2]

Methods

Owing to the limit of the existing dataset collected in NRDM, in that it does not have information about home ZIP codes of the patients for each of the product sales, we proposed a data simulation model to allocate the counts of OTC sales in patient's residential ZIP code areas.

To illustrate, we use an example of OTC medication purchases made by the patients living in six ZIP code areas with or without pharmacy stores (Figure 1). The nodes are connected by three types of arrows representing different types of commuting we presume: (1) for people who live in the ZIP code areas with pharmacy stores, they purchase OTC medications from those stores; (2) for people who live in ZIP code areas without stores, they will purchase OTC medications from (a) the adjacent ZIP code areas that have stores (solid arrows) (b) the nearest with-store ZIP code areas if neither their living ZIP codes nor the adjacent have stores (dashed arrows) or (c) their working ZIP code area with stores (doubled arrows).

Our methods consist of three steps. First, we split each non-store node into sub-nodes so that each sub-node only has one arrow going out. In the rightmost graph in Figure 1, s_{ij}^{w} represents the population of work flow between s_i and s_j, which was collected during the 2000 census, and s_{ij} represents the remaining population in area S_i who purchased OTC medication in area S_j, which is computed as proportional to the population of its target node. Second, for each with-store node, the sales counts are then re-allocated to all of its incoming nodes and itself assuming a multinomial

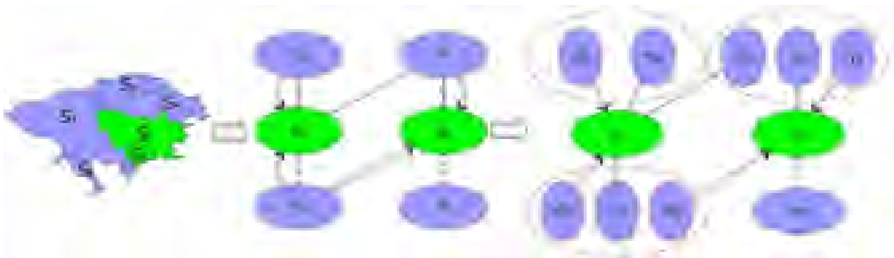


Figure 1 Modeling OTC medication purchases made by the patients living in six ZIP code areas. The leftmost figure shows their geographic relations. The green areas represent the ZIP codes with stores and the blue ones represent the ones without. The middle graph is to illustrate the three types of commuting in between. The right graph shows the sub-nodes after splitting.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 2 The simulated OTC counts in Allegheny County.

distribution. Third, we combine the sub-nodes back into the original node by adding the allocated counts together.

Results

Figure 2 is an example of the simulated counts in Allegheny County, Pennsylvania. The model re-allocated the counts from 53 ZIP code areas with stores (in green) to the remaining 44 ZIP codes without stores.

Summary

We have presented a method to simulate the counts of purchased OTC medications in terms of residential location of the patients. This dataset can be used in multivariate analysis in combination with the syndromic dataset collected during the ED visits of patients in order to improve the power of early outbreak detection.

Acknowledgements

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

References

1 Kulldorff M, Mostashari F, Duczmal L, Katherine Yih W, Kleinman K, Platt R. Multivariate scan statistics for disease surveillance. *Stat Med* 2007;**26:**1824–33.
2 Wagner MM, Tsui FC, Espino J, Hogan W, Hutman J, Hersh J. National Retail Data Monitor for public health surveillance. *Morb Mortal Wkly Rep* 2004;**53** (Suppl): 40–2.

Emerging Health Threats Journal 2011, **4**:s52. **doi:** 10.3134/ehtj.10.052
© 2011 M Sabhnani *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Detection of multiple overlapping anomalous clusters in categorical data

M Sabhnani, A Dubrawski, and J Schneider

Auton Lab, Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA
E-mail: sabhnani@cs.cmu.edu

Objective

We present Disjunctive Anomaly Detection (DAD), a novel algorithm to detect multiple overlapping anomalous clusters in large sets of categorical time series data. We compare performance of DAD and What's Strange About Recent Events (WSARE) on a disease surveillance data from Sri Lanka Ministry of Health.

Introduction

Syndromic surveillance typically involves collecting time-stamped transactional data, such as patient triage or examination records or pharmacy sales. Such records usually span multiple categorical features, such as location, age group, gender, symptoms, chief complaints, drug category and so on. The key analytic objective to identify potential disease clusters in such data observed recently (for example during last one week) as compared with baseline (for example derived from data observed over previous few months). In real world scenarios, a disease outbreak can impact any subset of categorical dimensions and any subset of values along each categorical dimension. As evaluating all possible outbreak hypotheses can be computationally challenging, popular state-of-the-art algorithms either limit the scope of search to exclusively conjunctive definitions[1] or focus only on detecting spatially co-located clusters[2] for disease outbreak detection. Further, it is also common to see multiple disease outbreaks happening simultaneously and affecting overlapping subsets of dimensions and values. Most such algorithms[1,2] focus on finding just one most significant anomalous cluster corresponding to a possible disease outbreak, and ignore the possibility of a concurrent emergence of additional clusters.

Methods

DAD model assumes that there are multiple anomalous clusters in data where each cluster is defined as a conjunction over data dimensions and disjunctions over values along each dimension. The cluster definitions are allowed to overlap across multiple dimensions and values. It is convenient to visualize the data aggregated in a multidimensional cube with as many cells as there are unique conjunctions of all data dimensions. Each cluster spans a sub-tensor in this view of data. It is defined by two factors: location (the sub-tensor), which defines the scope of disease outbreak, and intensity, which defines the disease rate. DAD assumes that effect of overlapping clusters on any cell of the data cube are additive.

During detection, DAD algorithm iteratively adds new clusters to the model and optimizes their distribution along the data cube simultaneously. It alternately fits cluster intensities using non-negative least squares approach, and cluster locations using best subset selection approach. The algorithm uses AIC regularization to control the number of clusters reported by the model.

Results and Conclusions

We evaluated DAD against WSARE on Sri Lanka Weekly Epidemiological Reports[3]. The data stores patient visits spanning 26 regions and 9 diseases reported over 2.5 years. We injected multiple overlapping disease outbreaks in the data and then executed both algorithms to see how well they could be detected. Figure 1 (left) shows the detection accuracy (ROC) of DAD (shown in solid) and WSARE (dotted). Each experiment involved three simultaneous overlapping clusters, and the graph shows average performance over 100 such experiments. Figure 1 (right) shows time-to-detection (AMOC) characteristic. When both

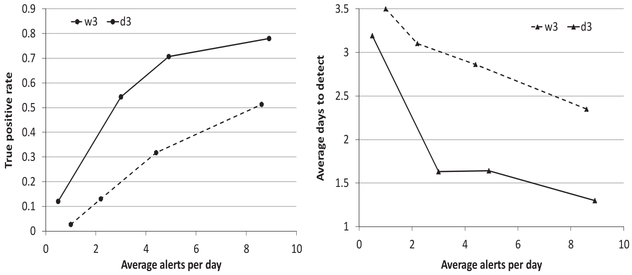


Figure 1 Accuracy (left) and timeliness of detection (right) performance of DAD and WSARE algorithms.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

algorithms are allowed to generate at most three alerts per day, DAD can detect 55% of injected clusters, whereas WSARE can only detect 20%. Also, DAD can detect them in 1.5 days after onset, whereas WSARE takes almost 3 days. We found similar results for evaluations across various injection parameters: number of clusters, size of clusters, and extend of overlap between predicted and injected clusters.

Acknowledgements

This work was supported, in part, by National Science Foundation (Grant 0911032). This paper was presented as an oral presentation at the 2010, International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Wong W, Moore A, Cooper G, Wagner M. What's Strange About Recent Events (WSARE): an algorithm for the early detection of disease outbreaks. *J Mach Learn Res* 2005;**6**:196–98.
2 Neill D, Cooper G. A multivariate Bayesian scan statistic for early event detection and characterization. *Mach Learn J* 2009.
3 Weekly Epidemiological Reports. Sri Lanka Ministry of Health and Nutrition. http://wwwepid.gov.lk/wer.htm.

Emerging Health Threats Journal 2011, **4**:s53. **doi:** 10.3134/ehtj.10.053
© 2011 C Safta *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Characterization of communicable disease epidemics using bayesian inversion

C Safta[1,2], J Ray[1,2], K Cheng[1,2], and D Crary[1,2]

[1]*Sandia National Laboratories, Livermore, CA, USA; and* [2]*Applied Research Associates, Inc., Arlington, VA, USA*
E-mail: csafta@sandia.gov

Objective

We present a statistical method to characterize an epidemic of a communicable disease from a time series of patients exhibiting symptoms. Characterization is defined as estimating an unobserved, time-dependent infection rate and associated parameters that completely define the evolution of an epidemic. The problem is posed as one of Bayesian inference, where parameters are inferred with quantified uncertainty. The method is demonstrated on synthetic and historical epidemic data.

Introduction

The evolution of a communicable disease in a human population is not entirely predictable. However, the spreading process can be assumed to vary smoothly in time. The time-dependent infection process can be linked to observations of the epidemic's evolution by convolving it with a stochastic delay model. In retrospective analyses of epidemics, when the observations are the dates of exhibition of patients' symptoms, the delay is the incubation period. In case of biosurveillance data, the delay is caused by incubation and a (hospital) visit delay, modeled as independent random variables. A model for observational error is also required. The time-dependent infection/spread rate may be inferred from observations by a deconvolution process.[1] The smooth temporal variation of the infection rate allows its representation using a low dimensional parametric model, and the inference may be performed with relatively little data. For large outbreaks, the data may be available early in the epidemic, allowing timely modeling of the outbreak. Short-term forecasts using the model could thereafter be used for medical planning.

Methods

We extend the model by Brookmeyer and Gail, for use with biosurveillance data, by adding a model for visit delay. The model is also augmented, for use in bioterrorism scenarios, with an additive term modeling the existence of a significant number of index cases. We use the model to construct a Bayesian inverse problem for various parameters of epidemiological interest for example, spread rate parameters, index cases etc, and solve it using a Markov Chain Monte Carlo method. This procedure develops posterior distributions for the objects of inference, allowing us to quantify the uncertainty in the estimates. The inference procedure, when using biosurveillance data, can be computationally expensive as parameter estimation involves a double deconvolution. We accelerate this process by developing a surrogate model offline, which is trivially parallelizable. The surrogate model consists of a weighted sum of computationally inexpensive polynomial chaos expansions,[2] allowing the inference to proceed in a timely manner.

Results

We demonstrate our method on data from historical plague outbreaks to study its ability to estimate infection rate properties. The methodology is validated by comparisons with published results. We also validate the inversion procedure with synthetic data from simulated plague bioattacks to gauge the inversion accuracy and computational savings when using the surrogate model. The improvement of the parameter estimates with the availability of data is also explored.

Conclusions

Bayesian inference, using infection rate models, allows a simple way of characterizing epidemics with few observations. In conjunction with surrogate models, it may also be performed with modest computational effort.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010. This work was funded by DoD/DTRA, under contract HDTRA1-09-C-0034. The PI of the program is Ms Karen Cheng.

References

1 Brookmeyer R, Gail M. A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *JASA* 1988;**83**:301–8.
2 Ghanem R, Spanos P. Stochastic Finite Elements, Springer-Verlag: New York, 1991.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s148. **doi:** 10.3134/ehtj.10.148
© 2011 TK Sell *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Using a prediction market to forecast dengue fever activity

TK Sell[1], C Franco[1], ATY Ho[2], and PM Polgreen[2]

[1]*Center for Biosecurity of UPMC, Baltimore, MD, USA; and* [2]*University of Iowa, Iowa City, IA, USA*
E-mail: tksell@upmc-biosecurity.org

Objective

The objective of this project is to use prediction markets to forecast the spread of dengue.

Introduction

Dengue is a mosquito-borne viral disease, and there is considerable evidence that case numbers are rising and geographical distribution of the disease is widening within the United States, and around the world.

The accuracy and reporting frequency of dengue morbidity and mortality information varies geographically, and often is an underestimation of the actual number of dengue infections. As traditional methods of disease surveillance may not accurately capture the true impact of this disease, it is important to gather professional observations and opinions from individuals in the public health, medical, and vector control fields of practice. Prediction markets are one way of supplementing traditional surveillance and quantifying the observations and predictions of professionals in the field.

Prediction markets have been successfully used to forecast future events,[1–4] including future influenza activity.[5] For these markets, we divided the possible outcomes for each question into multiple mutually exclusive contracts to forecast dengue-related events. This differed from many previous prediction markets that offered single sets of yes-no questions and used ‘real’ money in the form of educational grants. However, with more detailed contracts, we were able to generate more refined predictions of dengue activity.

Methods

Participants are given $100 of a valueless currency with which to trade. The future values of these contracts depend on the outcome of selected dengue-related events. On 17 August 2010, we opened contracts that specified (1) the total number of the United States dengue cases in 2010, (2) the percentage increase in clinical dengue in the Americas in 2010, and (3) the number of states that will report locally acquired dengue cases in 2010.

For each question, we offered multiple mutually exclusive contracts. For example, for the number of cases by the end of 2010 question, we provide contracts specified on five possible categories: (1) 150 or less, (2) 151–200, (3) 201–250, (4) 251–300, and (5) 301 or more. Traders buy and sell contracts with one another at prices that depend on their beliefs about the likelihood of the underlying event. These prices can be interpreted as the consensus probabilities of event occurrence.

Results

As of 9 September 2010, 28 active participants were trading on the markets. A total of 333 transactions occurred, with a total of 752 contracts being traded. The current predictions on the total number of the United States dengue cases are plotted in Figure 1. The height of each shaded region at any particular date represents the predicted probability, as of that date, that the corresponding event will occur. For example, on 4 September 2010, market prices indicated a 45 chance of 251–300-US dengue cases—the most likely outcome. Prices of other contracts predict that the percentage increase in the Americas in 2010 will most likely be between 75 and 100 (51 chance), and only 1 state will report locally acquired dengue cases in 2010 (43 chance).

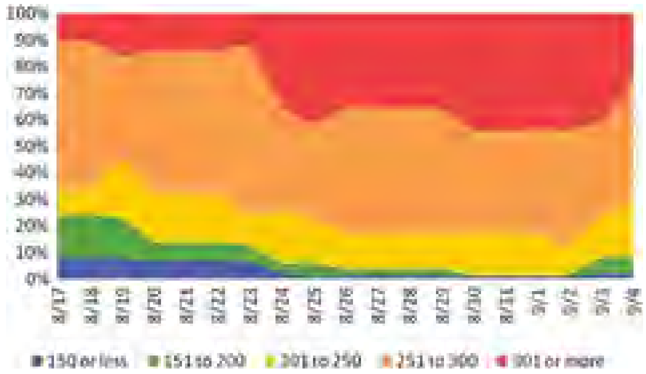


Figure 1 Probability distribution of case count outcomes.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusion

The consensus opinion of each dengue market has reflected changes in dengue activity. This demonstrates the potential of markets as useful tools for disease surveillance through aggregating the experience and knowledge of experts.

Acknowledgements

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

References

1 Smith VL. Constructivist and ecological rationality in economics. *Am Econ Rev* 2003;**93**:465–508.
2 Pennock DM, Lawrence S, Giles CL, Nielsen FA. The real power of artificial markets. *Science* 2001;**291**:987–8.
3 Cowgill B, Wolfers J, Zitzewitz E. Using prediction markets to track information flows: Evidence from Google, Dartmouth College (2008). www.bocowgill.com/GooglePredictionMarketPaper.pdf.
4 Dvorak P. Best buy taps prediction market. Wall Street Journal. 16 September 2008. Available at: http://online.wsj.com/article/SB122152452811139909.html.
5 Polgreen PM, Nelson FD, Neumann GR. Use of prediction markets to forecast infectious diseases activity. *Clin Infect Dis* 2007;**44**:272–9.

Emerging Health Threats Journal 2011, **4**:s55. **doi:** 10.3134/ehtj.10.055
© 2011 N Um *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Data quality in federated disease surveillance: using variability as an indicator of quality

N Um[1], S Visweswaran[1], J Espino[2], and M Wagner[1,2]

[1]*Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; and* [2]*Real-Time Outbreak and Disease Surveillance Laboratory, University of Pittsburgh, Pittsburgh, PA, USA*
E-mail: nau2@pitt.edu

Objective

We developed a novel method for monitoring the quality of data in a federated disease surveillance system, which we define as 'a surveillance system in which a set of organizations that are not owned or controlled by public health provide data.'

Introduction

Most, if not all, disease surveillance systems are federated in the sense that hospitals, doctors' offices, pharmacies are the source of most surveillance data. Although a health department may request or mandate that these organizations report data, we are not aware of any requirements about the method of data collection or audits or other measures of quality control.

Because of the heterogeneity and lack of control over the processes by which the data are generated, data sources in a federated disease surveillance system are black boxes the reliability, completeness, and accuracy of which are not fully understood by the recipient.[1]

In this paper, we use the variance-to-mean ratio (VMR) of daily counts of surveillance events as a metric of data quality. We use thermometer sales data as an example of data from a federated disease surveillance system. We test a hypothesis that removing stores with higher baseline variability from pooled surveillance data will improve the signal-to-noise ratio of thermometer sales for an influenza outbreak.

Methods

We computed the VMR for each of 178 drug stores in Allegheny County, PA, USA. In particular, we computed VMR for a non-influenza period, which we term the 'base period' (BP). We used the 81-day period from 1 June 2009 to 20 August 2009, inclusive.

Before computing VMR, we smoothed daily thermometer sales for each story by applying a 7-day moving average (MA7) to remove day-of-the-week effects.

To determine whether removal of stores with highest VMR improves the ability to detect an influenza outbreak, we systematically removed stores with the highest VMR from the total daily counts summed for the 178 stores in Allegheny County. To determine whether the timeliness of algorithmic detection was affected by the removal of high VMR stores, we used a detection threshold of three s.d. above

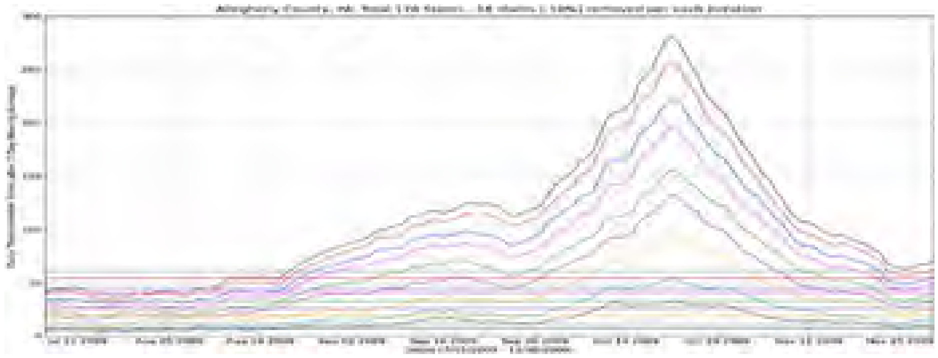


Figure 1 7-day moving average daily thermometer sales TS plot with progressive removal of stores. The uppermost line represents all 178 stores in Allegheny County, PA. Each subsequent TSs represents removal of 18 stores by VMR scores.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the mean of the BP. We also compared the date of peak thermometer sales and the signal-to-noise ratio at peak.

Results

The mean VMR for the 178 stores was 1.16 with maximum of 2.73 and minimum of 0.76. Figure 1 shows that the shape of the plot of county-wide sales of thermometers remained as high-variability stores were removed in 10% tranches. The detection date and peak date were unchanged (August 24 ± 1 day and October 21, 2009 ± 1 day, respectively) through the progressive removal process. The signal-to-noise ratio, measured as number of s.d.s above the BP mean on the 'peak' was 45 s.d. at 0% removal, 41 s.d. at −30%, 33 s.d. at −60%, and 21 s.d. at −80%.

Conclusions

There was significant difference in the VMR for sales of thermometers by different stores. However, removal of those stores from a surveillance system did not improve the ability of a typical surveillance algorithm to detect the 2009 influenza epidemic and the signal to noise ratio at the peak of the epidemic was not improved by the removal of the stores with more baseline variability.

Acknowledgements

This research was supported by a grant from the Lockheed-Martin Corporation. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

Reference

1 Wagner MM. Methods for evaluating surveillance data. *Handbook of Biosurveillance*, Chapter 21, p 313–20, 2006.

Emerging Health Threats Journal 2011, **4**:s56. **doi:** 10.3134/ehtj.10.056
© 2011 L van Asten *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Estimating the number of deaths attributable to nine common infectious pathogens adjusted for seasonality and temperature

L van Asten, C van den Wijngaard, W van Pelt, J van de Kassteele, A Meijer, W van der Hoek, M Kretzschmar, and M Koopmans

Centre for Infectious Disease Control Netherlands, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
E-mail: Liselotte.van.asten@rivm.nl

Objective

Mortality exhibits clear seasonality mainly caused by an increase in deaths in the elderly in winter. As there may be substantial hidden mortality for a number of common pathogens,[1,2] we estimated the number of elderly deaths attributable to common seasonal viruses and bacteria for which robust weekly laboratory surveillance data were available.

Introduction

Accurately assigning causes or contributing causes to deaths remains a universal challenge, especially in the elderly with underlying disease. Cause of death statistics commonly record the underlying cause of death, and influenza deaths in winter are often attributed to underlying circulatory disorders. Estimating the number of deaths attributable to influenza is, therefore, usually performed using statistical models. These regression models (usually linear or poisson regression are applied) are flexible and can be built to incorporate trends in addition to influenza virus activity such as surveillance data on other viruses, bacteria, pure seasonal trends and temperature trends.

Methods

On weekly time series (1999–2007), we used Poisson regression models (which included linear and periodic components) to characterize the association of total death counts with trends in common seasonal viruses and bacteria (influenza A and influenza B, RSV, parainfluenza, enterovirus, rotavirus, norovirus, campylobacter and salmonella) adjusted for extreme outdoor temperatures and stratified by age. Model coefficients were used to calculate the numbers of deaths attributable to the various included pathogens.

Results

With increasing age, the number of pathogens (all viruses) significantly associated with mortality also increased: influenza A, RSV, influenza B and parainfluenza were significant contributors in individuals aged 75–85, and additionally norovirus in those aged 85 years and older. Adjusting for temperature decreased parameter magnitudes. Influenza attributable mortality was directly associated with mortality (that is, not lagged), while other viruses were associated with deaths 1–4 weeks later. Our most conservative models attributed 6.9% of all deaths in those aged 85 years and older to multiple winter viruses. The attributed proportion decreased with decreasing age (75–85 years: 3.6%; 65–75years: 1.5%; 55–65years: 1.1%), also because with decreasing age less viruses were significantly associated with death (but always including influenza A for which the estimated attributable overall mortality ranged from 2% in the eldest down to 0.6% in the youngest age group of 55–65 years) (Figure 1).

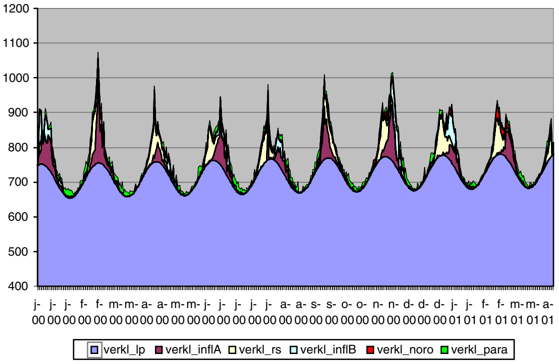


Figure 1 Modeled weekly deaths attributable to respiratory and gastrointestinal pathogens in individuals aged 85 years and older (stacked and adjusted for high temperature).

Conclusions

In the population of 55 years and older, the number of common viruses that were associated with overall mortality increased with increasing age to include: influenza A, respiratory syncytial virus, influenza B, norovirus and parainfluenza.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Together, these pathogens were associated with 6.9% of all deaths in the individuals aged 85 years and older.

Acknowledgements

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

References

1 Pitman RJ, Melegaro A, Gelb D, Siddiqui MR, Gay NJ, Edmunds WJ. Assessing the burden of influenza and other respiratory infections in England and Wales. *J Infect* 2007;**54:** 530–8.
2 Markov PV, Crowcroft NS. Modelling the unidentified mortality burden from thirteen infectious pathogenic microorganisms in infants. *Epidemiol Infect* 2007;**135:**17–26.

Emerging Health Threats Journal 2011, **4**:s57. **doi:** 10.3134/ehtj.10.057
© 2011 R Villamarín *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Estimating the incidence of influenza cases that present to emergency departments

R Villamarín, G Cooper, F-C Tsui, M Wagner, and J Espino

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: ricardov@pitt.edu

Objective

We developed a model that predicts the incidence of influenza (flu) cases that present to Emergency Departments (ED) in a given region based on thermometer sales (TS).

Introduction

Our laboratory previously established the value of over-the-counter (OTC) sales data for the early detection of disease outbreaks.[1] We found that TS increased significantly and early during influenza season. Recently, the 2009 H1N1 outbreak has highlighted the need for developing methods that not only detect an outbreak but also estimate incidence so that public-health decision makers can allocate appropriate resources in response to an outbreak. Although a few studies.[2] have tried to estimate the H1N1 incidence in the 2009 outbreak, these were done months afterward and were based on data that are either not easy to collect or not available in a timely fashion (for example, surveys or confirmed laboratory cases).

Here, we explore the hypothesis that OTC sales data can also be used for predicting a disease activity. Towards that end, we developed a model to predict the number of ED flu cases in a region based on TS. We obtain sales information from the National Retail Data Monitor (NRDM) project. NRDM collects daily sales data of 18 OTC categories across the US.[1]

Methods

To create the model, we obtained (1) the number of thermometers sold per day from 1 May to 31 December 2009 in Allegheny County (AC), Pennsylvania (PA), and (2) estimates of the number of daily cases that presented with flu to monitored EDs in AC during the same period. These estimates were produced by our 'Bayesian case detector (BCD)' algorithm, which estimates from an ED report the probability that a given ED patient has flu. We developed a linear regression model with zero time lag to predict flu cases solely from TS; let S and I denote the slope and intercept parameters of the model, respectively.

Presently, the BCD only collects information from a fraction of all EDs in AC ($A = 0.44$), and TS data only cover a fraction of all sales from OTC retail stores in AC ($B = 0.91$). To apply our model to other regions in the US, we used A and B to adjust the regression parameters. Moreover, we also must account for differences in population between AC ($\sim$1.2 million) and the region to which the model is being applied. We obtain the following formula, which assumes that TS and ED visits are proportional to population size (pop) and that our regression model holds:

$$\mathrm{EDFlu(region)} = \mathrm{TS(region)S}(B/A) + (I/A)(\mathrm{pop(region)/pop(AC)})$$

Given the lack of data with which to validate our model in other regions of PA or the country, we used counts of constitutional symptoms (referred to as constitutional counts, CC) of people who come to EDs as a rough surrogate of ED flu cases. We had such CCs and TS for all counties in PA

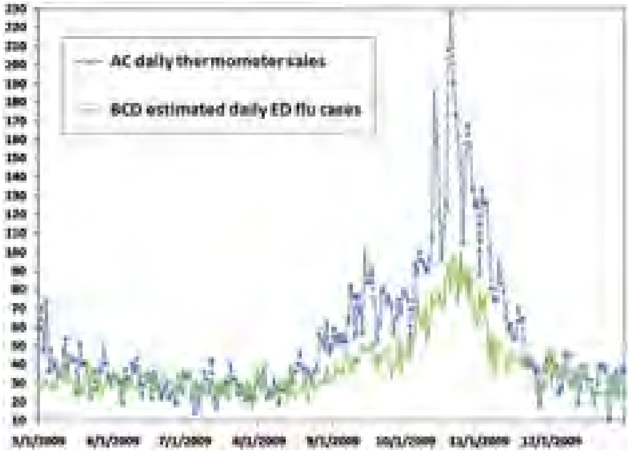


Figure 1 Plots of TS and estimated ED flu cases in AC (1 May–31 December 2009).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in 2009. We trained a linear model to predict CCs for a subset of counties in PA based on those counties' TS in 2009. We applied that model to predict the CCs in another subset of PA counties from their TS. We repeated this experiment 10 times, each time randomly selecting a different subset of counties for training and testing. (Figure 1).

Results

Least squares linear regression applied to TS and counts of ED flu cases (from BCD) produced $S=0.39$ (95% CI: 0.37, 0.42) and $I=17.6$ (95% CI: 15.9, 19.2) with R^2 of 0.79 and a cross-correlation (0 time lag) of 0.90. The regressions to predict CCs also had high R^2 values (mean = 0.81) and high correlations (mean = 0.89) between the predictions and actual CCs.

Conclusions

The ability of our regression on TS to predict flu cases (as measured by R^2) shows that a linear relationship between those quantities fits well. Moreover, the regressions done to predict CCs support the validity of predicting outcomes that are related to flu ED visits based on TS. Our results suggest that the use of TS, which are available nationwide daily, has potential to be helpful in estimating ED flu activity.

Acknowledgements

The project was funded by CDC grants P01 HK000086 and 1U38 HK000063-01, NIH grant R01 LM009132, and NSF grant IIS-0911032. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Hogan WR, Wagner MM. Sales of over-the-counter healthcare products. *Handbook of Biosurveillance*, 1st edn. 2006, p 321.
2 Presanis AM, De Angelis D, Hagy A, Reed C, Riley S, Cooper BS, *et al.* The severity of pandemic H1N1 influenza in the U.S., from April to July 2009: a Bayesian analysis. *PLoS Med* 2009;**6**:e1000207.

Emerging Health Threats Journal 2011, **4**:s158. **doi:** 10.3134/ehtj.10.158
© 2011 AP Wilper *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Temporo-spatial surveillance of influenza-like illness: preliminary results from the Idaho infectious disease reporting network

AP Wilper[1,2], WG Weppner[1,2], K Elgethun[3], J Mitten[1], J Tiumalu[1], and D Stevens[1,2]

[1]*Boise VA Medical Center, Boise, ID, USA;* [2]*University of Washington School of Medicine, Seattle, WA, USA; and* [3]*Boise State University, Boise, ID, USA*
E-mail: andrew.wilper@va.gov

Objective

The objective of this study is to describe initial efforts to establish a real-time syndromic surveillance of influenza-like illness (ILI) in Idaho, using data from the Veterans Administration electronic medical record (Computerized Patient Record System (CPRS)).

Introduction

Current influenza-like illness (ILI) monitoring in Idaho is based on syndromic surveillance using laboratory data, combined with periodic person-to-person reports collected by Idaho state workers. This system relies on voluntary reporting.

Electronic medical records offer a method of obtaining data in an automated fashion. CPRS captures real-time visit information, vital signs, ICD-9, pharmacy, and lab data. The electronic medical record surveillance has been utilized for syndromic surveillance on a regional level.[1] Funds supporting expansion of electronic medical records offer increased ability for use in biosurveillance.[2] The addition of temporo-spatial modeling may improve identification of clusters of cases.[3] This abstract reviews our efforts to develop a real-time system of identifying ILI in Idaho using Veterans Administration data and temporo-spatial techniques.

Methods

The Boise Veterans Affairs Medical Center provides care to over 20,000 veterans living in Idaho with clinics in Boise, Caldwell, Twin Falls, Salmon, Idaho, Burns, and Oregon. Using retrospective data from the Veterans Integrated Service Network (VISN 20) data warehouse for the 2008–2009 influenza season, we identified ILI cases from these clinics using ICD-9 codes[4] collected as weekly counts. Duplicates and incompletes were removed, zip code was extracted, and clusters less than five per zip code were suppressed. We used SaTScan program v9.1.0 (SaTScan is a trademark of Martin Kulldorff. The SaTScan software was developed under the auspices of Martin Kuldorff, the National Cancer Institute and Farzad Mostashari at the New York City Department of Health and Mental Hygiene. Available from http://www.satscan.org) for cluster analysis, with Monte Carlo simulation for an expected incidence based on distribution of sample over time and space; geographic extent of cluster was not limited. We used ArcMap 10 (ArcGIS 10 is a product of Esri, Redlands, CA, 2010) for visualization on the basis of the United States Census Map data. The Veterans Administration Puget Sound Institutional Review Board (IRB) approved this study.

Results

We identified one primary and one significant secondary cluster ($P \leqslant 0.05$) of ILI (Figure 1). The relative risk was 7.8 for ILI in the primary cluster identified in southeast Idaho over a month-long period from 17 July 2009 to 20 August 2009. The secondary cluster in west–central Idaho occurred over a shorter two-week period in January. These results were shared with Idaho Public Health District directors, who confirmed the existence of an ILI cluster in southeast Idaho, the site of the primary cluster.

Limitations

We sampled a small percent of the state population; women and children are underrepresented.

Conclusions

Retrospective data obtained from VA electronic health records appear to be useful in locating ILI outbreaks in space and time. Further work is needed to evaluate the ability of our system to identify outbreaks in real time.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

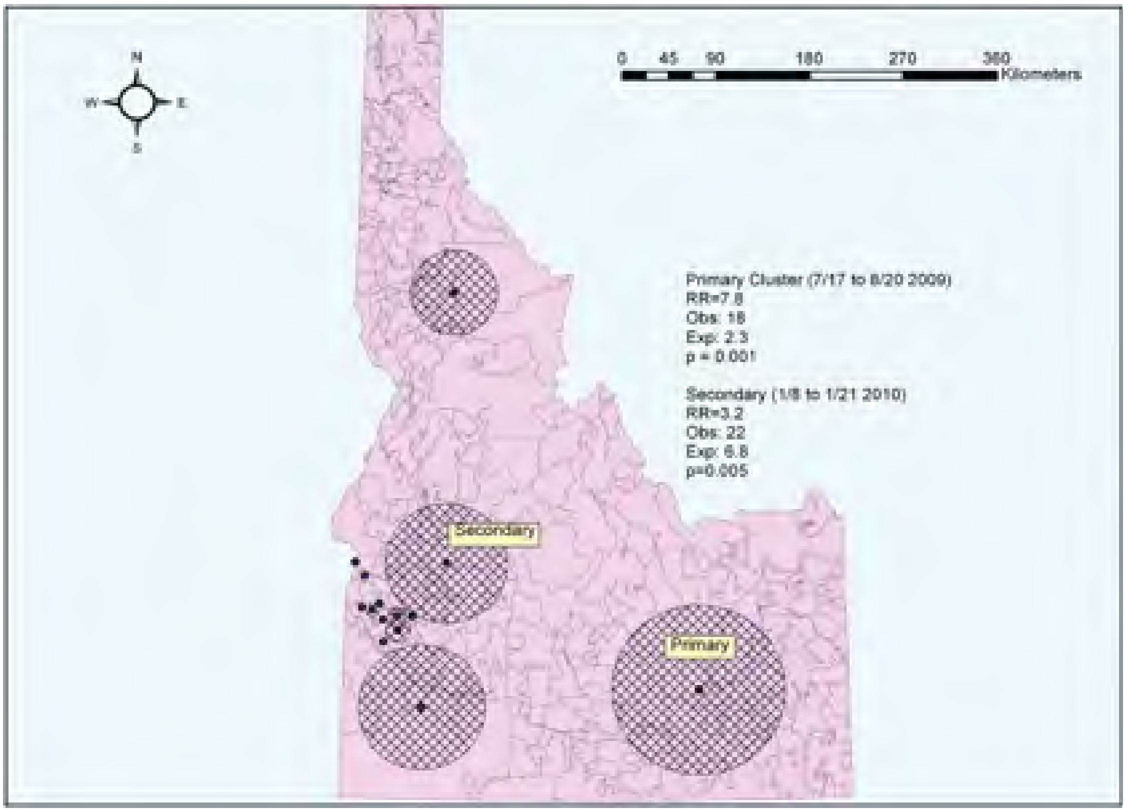


Figure 1 Primary and secondary cluster of ILI in the state of Idaho, 2008–2009 flu season.

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 Lombardo JS, Burkom H, Pavlin J. II ESSENCE and the framework for evaluating syndromic surveillance systems. *MMWR* 2004;**53** (suppl): 159–65.
2 Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. *N Eng J Med* 2010;**363:**501–4 (www.NEJM.org).
3 Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, Kruskal BA, *et al.* Evaluating real-time syndromic surveillance signals from ambulatory care data in four states. *Public Health Rep* 2010;**125:**111–25.
4 Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenza-like illness by International Classification of Diseases, ninth revision. *Emerg Infect Dis* 2007;**13:**207–16.

Emerging Health Threats Journal 2011, **4**:s59. **doi:** 10.3134/ehtj.10.059

© 2011 K Zinszer *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

The influence of address errors on detecting outbreaks of campylobacteriosis

K Zinszer[1,2], K Charland[1,2], C Jauvin[2], L Bedard[3], R Allard[1,3], and DL Buckeridge[1,2,3]

[1]*Department of Epidemiology & Biostatistics, McGill University, Montreal, Quebec, Canada;* [2]*Clinical and Health Informatics Research Group, McGill University, Montreal, Quebec, Canada; and* [3]*Direction de Santé Publique, Agence de la Santé et des services Sociaux de Montreal, Montreal, Quebec, Canada*

E-mail: kate.zinszer@mail.mcgill.ca

Objective

To examine, via a simulation study, the potential impact of residential address errors on the identification of a point source outbreak of campylobacter.

Introduction

Mandatory notification to public health of priority communicable diseases (CDs) is a cornerstone of disease prevention and control programs. Increasingly, the addresses of CD cases are used for spatial monitoring and cluster detection and public health may direct interventions based on the results of routine spatial surveillance. There has been little assessment of the quality of addresses in surveillance data and the impact of address errors on public health practice.

We launched a pilot study at the Montreal Public Health Department (DSP), wherein our objective was to determine the prevalence of address errors in the CD surveillance data. We identified address errors in 25% of all reported cases of communicable diseases from 1995 to 2008. We also demonstrated that address errors could bias routine public health analyses by inappropriately flagging regions as having a high or low disease incidence, with the potential of triggering misguided outbreak investigations or interventions. The final step in our analysis was to determine the impact of address errors on the spatial associations of campylobacter cases in a simulated point source outbreak.

Methods

We based our simulations on campylobacter case records extracted from the Montreal department of public health for the period 1995–2008. We compiled area-level counts of campylobacter cases for each of Montreal's 102 subregions and applied an address verification algorithm to determine the validity of the primary address and to identify an alternative, valid address. An address error was defined as the street number, street name, and/or postal code being identified by the algorithm as incorrect. We then re-computed the number of cases for each subregion.

For the outbreak scenario, we added an additional 21 cases to the caseload that occurred during July and August 2008, simulating a point source outbreak of a contaminated food product. Using Bayesian hierarchical models, we estimated subregional-level risk ratios (RRs) based on the original addresses and then again after correcting the address errors. We increased the prevalence of address errors by 5% increments, to a maximum of 40% address error. High-risk regions were identified by using a cutoff probability of 0.8 for the posterior proportion and a reference threshold of 1.2.

Results

We found significant subregional campylobacteriosis rate differences after correcting for address error (Figure 1).

The outbreak simulation also indicated significant differences in high-risk subregions after correcting for address errors and when increasing the prevalence of address errors.

Conclusions

Our study demonstrates that the positional shifting of cases due to address errors can produce diverging results in spatial

Figure 1 Map of campylobacteriosis rate differences (original address versus corrected address) per 10,000 per subregion in Montreal.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

analyses. The magnitude of address errors in our simulation was reasonable, given the rates of actual address errors we found in our pilot study. Positional errors have the potential to impact cluster detection and related public health interventions.

Acknowledgements

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

International Society for Disease Surveillance Conference 2010

Track 2: Public Health Surveillance

42. The utility of inpatient data obtained from regional health information organizations for pneumonia and influenza surveillance
G Johnson, X Jin, J-H Chen, H-G Chang, and D Morse

44. Florida's ESSENCE systemFfrom syndromic surveillance to routine epidemiologic analysis across syndromic and non-syndromic data sources
A Kite-Powell, J Hamilton, R Wojcik, W Loschen, and R Hopkins

46. Comparison of Distribute and ILINet for national influenza surveillance
K Kniss, S Benoit, A McIntyre, D Buckeridge, J Brownstein, L Brammer, and L Finelli

48. Emergency department diagnosis code data for surveillance of vaccine adverse events: comparison with the national vaccine adverse event reporting system
E Lamb, H Vaughan-Batten, NJM Dailey, J-M Maillard, L Johns, AT Fleischauer, Z Moore, and M Davies

50. Meaningful use and public health surveillance: to travel fast or far?
LA Lenert

52. Surveillance for nationally notifiable infectious conditions using ICD-9-CM diagnosis codes in the VA ESSENCE biosurveillance system
C Lucero, P Schirmer, G Oda, and M Holodniy

54. Using laboratory and medical encounter records to identify impact of prophylaxis policies on group A Streptococcus in three recruit training facilities
TM Luse, K Shannon, and G Nowak

56. EpiSentry^{TM}: simulation-based threat identification, response management and decision support for outbreaks and pandemics
DW Mazeroski and T Arthur-Hartranft

58. Quantifying the relationship between influenza-related emergency department visits and hospital admissions in BioSense
AF McIntyre, L Finelli, HS Burkom, SR Benoit, KL Kniss, and TA Kass-Hout

60. Evaluation of the NATO Disease Surveillance System in Kosovo in 2010
J-B Meynard, H Chaudet, L Pellegrin, Y Moreno, O Deboosere, P Mornand, K Pennin, S Stein, B Queyriaux, H-U Holterm, A Zuendorf, J Kocik, R Migliani, X Deparis, and G Texier

62. Creating a fast and flexible syndromic surveillance reporting system
J Morin, A Ali, D Ferguson, RG Gutierrez, A Riley, N Vinson, S McClinton, C McDermaid, K Ali, J Martin, and RF Davies

64. Comparison of respiratory, febrile and influenza-like illness syndromes to detect laboratory-reported H1N1 and RSV, Influenza Season 2009–10, New York City
E Westheimer, TQ Nguyen, M Paladini, D Weiss, and S Balter

66. Early Aberration Reporting System (EARS) update: present and future
G Rainisch, and K Hayashi

68. Description of the quality of public health case reports received at a local health department and potential impact on workflow
D Rajeev, A Price, K Schnirel, M Hill, C Staes, and I Risk

70. Asthma patterns in Boston emergency department visits for children age five and under
M Rennick, J Gunn, M Donovan, and J Salvia

72. Evaluating the use of syndromic surveillance for the detection of influenza-like illness in Salt Lake County, Utah
J Robertson and M Hill

74. School absenteeism and emergency department ILI rates in King County, WA 2003–2009
C Rodriguez, A Baer, J Baseman, I Painter, D Yanez, T Koepsell, and JS Duchin

76. Impact of alternate diagnoses on the accuracy of influenza-like illness case definition used for H1N1 screening in the emergency department
D Rumoro, J Silva, S Shah, J Bayram, M Hallock, G Gibbs, M Waddell, and K Thomas

78. Access to and use of syndromic surveillance information at the local health department level
E Samoff, A Ising, A Waller, E Sickbert-Bennett, M Davis, M Park, SW Haas, L DiBiase, and PDM MacDonald

80. Tracking H1N1 vaccine doses administered using CRA system
B Nichols, S Waite, H Hill, L Han-Lee, S Sapkota, and J Tropper

81. User requirements for a user-centered design (UCD) redesign of a public health surveillance system: BioSense
REN Sembajwe, TR Farris, OL Rojas-Smith, BL Massoudi, AK Banger, and A Casoglos

83. Using Twitter to estimate H1N1 influenza activity
A Signorini, PM Polgreen, and AM Segre

84. How did specific specialties and new sentinel providers affect ILINet data post 2009 A/H1N1 recognition
O Simwale

85. Use of a public health working group to coordinate multi-jurisdiction response to bioterrorism surveillance signals and influenza outbreaks
AJ Siniscalchi

87. A comparison of syndromic surveillance chief complaint data and discharge data in a pediatric hospital system during 2009 H1N1
D Peace, W Smith, A Reeves, D Little, K Soetebier, and C Drenzek

Emerging Health Threats Journal 2011, **4**:s60. **doi:** 10.3134/ehtj.10.060
© 2011 DG Cochrane and JR Allegra.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

The use of patient temperature data for biosurveillance in the emergency department

DG Cochrane, and JR Allegra

New Jersey Research Foundation, Morristown Memorial Hospital Residency in Emergency Medicine, Morristown, NJ, USA
E-mail: johnalle@verizon.net

Introduction

Biosurveillance systems commonly use emergency department (ED) patient chief complaint data (CC) for surveillance of influenza-like illness (ILI). Daily volumes are tracked using a computerized patient CC classifier for fever (CC Fever) to identify febrile patients. Limitations in this method have led to efforts to identify other sources of ED data. At many EDs the triage nurse measures the patient's temperature on arrival and records it in the electronic medical record (EMR). This makes it possible to directly identify patients who meet the CDC temperature criteria for ILI: temperature greater than 100 degrees F (T > 100F).

Objective

To evaluate whether a classifier based on temperature > 100F would perform similarly to CC Fever and might identify additional patients.

Methods

Design: Retrospective review of EMR data. *Setting*: Five EDs in New York and New Jersey from 10-1-07 to 5-31-09. *Protocol*: For patients with a measured temperature, we determined daily volumes for CC Fever and for T > 100 and compared them using linear regression analysis and visual inspection of the time-series graphs. We also determined the additional

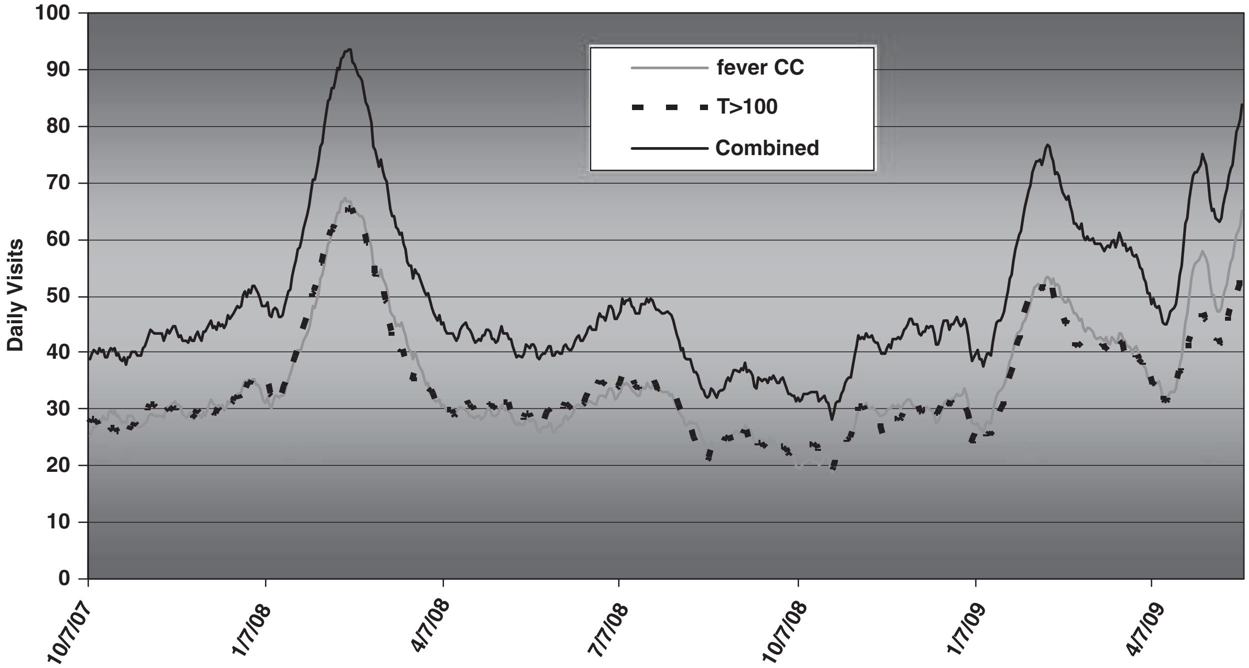


Figure 1 14-day moving average of visits for fever CC, T > 100 and combined fever CC and T > 100 classifiers.

This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

number of visits that were identified by adding T > 100 to CC Fever.

Results

There were a total of 352 742 ED visits, of which 265 099 (75%) had a measured temperature. Within that group, the total volumes were 21 411 for CC fever and 20 876 for T > 100. The correlation coefficient for the daily volumes was $r^2 = 0.94$, $P < 0.001$. Visual inspection of the time-series graphs (see Figure 1) showed a close match of seasonal peaks. The T > 100 classifier identified an additional 8959 patients not identified by the CC fever classifier alone (42% increase).

Conclusions

The classifier for measured temperature identified seasonal fever peaks similarly to the classifier based on CC fever and identified many additional patients. Further study is needed to determine the potential advantage of using measured temperature in ED surveillance for ILI.

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.

Emerging Health Threats Journal 2011, **4**:s61. **doi:** 10.3134/ehtj.10.061
© 2011 S Aslam *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Maryland ESSENCE expansion to incorporate prescription medication data

S Aslam, I Ajit, S Adams, and Z Faigen

Office of Preparedness and Response, Maryland Department of Health and Mental Hygiene, Baltimore, MD, USA
E-mail: saslam@dhmh.state.md.us

Objective

The purpose of this paper is to describe Maryland's process of enhancing its Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) by incorporating additional data sources such as prescription medication data.

Introduction

Maryland has a powerful syndromic surveillance system, ESSENCE, which is used for the early detection of disease outbreaks, suspicious patterns of illness, and public health emergencies. ESSENCE incorporates traditional and non-traditional health indicators from multiple data sources (emergency department chief complaints, over-the-counter (OTC) medication sales, and poison control center data).

Initially, 15 (30%) acute care hospitals in the National Capital Region and Baltimore Metro Region of the state were sending emergency department (ED) data to ESSENCE. DHMH began planning several years ago to increase the number of hospitals reporting to ESSENCE.

In 2007, Maryland's Governor introduced a homeland security initiative that outlined 12 Core Goals for A Prepared Maryland. One of core goals was to improve biosurveillance and in 2009, Maryland successfully incorporated 100% (45) acute-care hospitals into ESSENCE. Maryland continues to enhance and improve ESSENCE by incorporating additional data sources such as prescription medication data.

Methods

First, Maryland contacted the agency which developed ESSENCE, to learn about what data and software developments needs to occur in order to add another data source into ESSENCE system. Next, Maryland started on efforts to obtain the prescription medication data.

Currently, there is only one state that collects prescription medication data for biosurveillance. Maryland contacted the state to learn more about how this process is setup in their state.

With the help of Board of Pharmacy, Maryland started to work with chain pharmacies that were already sending OTC data to ESSENCE to incorporate prescription medication and an additional pharmacy that was participating in the State's Antiviral purchase program. Although pursing efforts to obtain data from pharmacies, Maryland learned that Center for Disease Control and Prevention (CDC) already receives prescription medication data from a third party vendor. Maryland contacted CDC and learned that CDC has data, which Maryland is seeking and is willing to share it with State.

Table 1 List of surveillance medications

Amino-penicillins
Tetra-cyclines
Quinolones/Fluoroquinolones
Neuraminidase Inhibitors
M2 Inhibitors

Results

Currently, Maryland is the in the process of working with its partners to incorporate the prescription medication data into ESSENCE. Maryland will have a fully automated system that creates alerts and allows near real-time surveillance on antivirals and prescription medications used to treat category A agents (see Table 1 above).

Conclusion

Maryland will be one of the first States to incorporate prescription medication data into its syndromic surveillance system. Addition of prescription medication into ESSENCE will make Maryland's syndromic surveillance more robust and enhance its capability for early detection of outbreaks and public health emergencies. ESSENCE will continue to be a tool used for situational awareness and inform decision makers in Maryland. Lastly, the process of incorporation prescription medication has provided an opportunity for Maryland to collaborate with other partners and CDC.

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.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s62. **doi:** 10.3134/ehtj.10.062

© 2011 S Aslam *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Utilities of Maryland's syndromic surveillance system: indentifying threats, case investigation and situational awareness

S Aslam, I Ajit, S Adams, and Z Faigen

Office of Preparedness and Response, Maryland Department of Health and Mental Hygiene, Baltimore, MD, USA
E-mail: saslam@dhmh.state.md.us

Objective

The purpose of this paper is to describe how Maryland's syndromic surveillance system, electronic surveillance system for the early notification of community-based epidemics (ESSENCE), has many utilities including identifying threats, case investigation and situational awareness.

Introduction

Maryland's electronic surveillance system for the early notification of community-based epidemics (ESSENCE) data includes emergency department visits from all acute care hospitals, over-the-counter medication sales and poison control data that cover all jurisdictions in Maryland.

Maryland Department of Health and Mental Hygiene (DHMH) uses ESSENCE daily for the early detection of public health emergencies. DHMH also utilizes ESSENCE for other purposes including situational awareness during high security events, assistance with outbreak investigation and for the H1N1 pandemic.

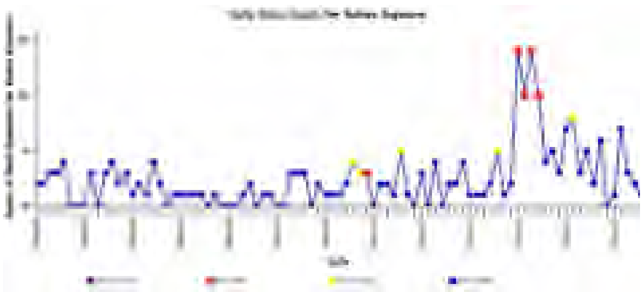
Methods

ESSENCE data are reviewed closely on a daily basis looking for clusters indicating a public health threat. However, during the 2008 nationwide *Salmonella* outbreak, a query was developed to search for patients with chief complaints related to *Salmonella* or tomatoes from April to June 2008.

During the initial H1N1 outbreak, DHMH used ESSENCE as a tool to enhance case finding efforts. In addition, throughout the H1N1 response, ESSENCE was also used to monitor levels of ILI across the state. Daily reports were provided to the Governor's office and state executives to enhance situational awareness and inform decision-making.

In 2009, ESSENCE was used for situational awareness during the Presidential inauguration. The following list of queries was used during the inaugural period: symptoms for hypothermia, injury, dehydration, influenza like illness, meningitis and Center of Disease Control and prevention category A diseases (anthrax, botulism, plague, small pox, tularemia and viral hemorrhagic fever).

Most recently in 2010, ESSENCE was queried to identify the number of patients that reported to Maryland hospitals for prophylaxis after being exposed to a rabid calf.



Results

On numerous occasions during the daily review, ESSENCE has identified clusters of meningitis. A statistically significant increase in ILI activity was observed during the H1N1 pandemic in spring and fall of 2009. Monitoring of the over-the-counter medication sales allows for the early prediction of an increase in ILI activity. During the 2009 Presidential inauguration, ESSENCE identified a statistically significant increase in the number of hypothermia cases. During the 2008 *Salmonella* outbreak, ESSENCE identified a case. This case was detected in ESSENCE 1 week earlier than its report date to the state health department and about 3 weeks earlier than when it was identified as part of the outbreak. Lastly, ESSENCE was utilized to monitor the number of cases that reported to Maryland hospitals for prophylaxis after being exposed to a rabid calf.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusion

ESSENCE is a vital tool for DHMH that provides a timely and ongoing situational awareness for policy decision makers. ESSENCE is a useful tool during public health events such as outbreaks and pandemic.

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.

Emerging Health Threats Journal 2011, **4**:s63. **doi:** 10.3134/ehtj.10.063

© 2011 T Azarian *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Syndrome evaluation of individuals reporting emergency department utilization for notifiable disease and conditions

T Azarian[1], A Kite-Powell[2], and S Zaheer[1]

[1]*Duval County Health Department, Jacksonville, FL, USA; and* [2]*Bureau of Epidemiology, Tallahassee, FL, USA*
E-mail: Taj_Azarian@doh.state.fl.us

Objective

This paper characterizes emergency department (ED) utilization among individuals diagnosed and reported with notifiable diseases and conditions (NDC). Furthermore, it evaluates the subsequent assignment of syndromic surveillance (SS) syndromes based on the patient's chief complaint (CC) during their ED visit.

Introduction

The Duval County Health Department (DCHD) serves a community of over one million people in Jacksonville, FL, USA. Each year, DCHD Epidemiology Program reports an average of 1133 (4-year average) NDC with the exception of STD/HIV, TB, and viral hepatitis. Within Duval County, emergency medical care is provided by eight local hospitals, including one pediatric facility and a level-1 trauma center. These facilities contribute SS CC data to the Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE) of Florida.

Historically, evaluations of SS systems have used ICD-9 diagnoses as the gold standard to determine predictive values.[1] However, limited studies have surveyed reports of NDC to identify related ED visits and subsequent CC-based syndrome categorization. These data may provide public health investigators insight into health seeking behaviors, interpretation of SS signals, and prevalence of NDC within ED data.

Methods

Between March and August 2010, DCHD epidemiology investigators collected ED visitation history from reported cases of NDC. ED usage during their course of illness was documented. An ESSENCE review was then conducted to match cases reporting ED visitation. To reduce the burden on hospital staff, cases were matched by ED, date of visit, age, zip code, sex, and race, instead of ESSENCE unique patient identifier. Case investigation data was then abstracted from Merlin, NDC reporting registry of Florida. ESSENCE syndromes were compared with appropriate NDC-specific syndromes determined by CDC ICD-9-based syndrome definitions.[2] Prevalence of notifiable diseases within ED SS data was evaluated. Comparisons between populations using ED services and those not reporting ED visitation were made.

Results

From 11 March 2010 to 20 August 2010, 486 NDC were reported by DCHD. In all, 61 cases were not directly interviewed. Of the remaining 425, 53.2% (226) reported visiting an ED during the course of their illness. Of these, 198 (87.6%) were diagnosed through the ED or during admission. The remaining 28 were diagnosed during subsequent visits to clinics or primary care providers. In all, 200 cases were able to be matched to the corresponding ED visits in ESSENCE. In addition, 13 cases reported two ED visits and 1 reported three visits for their illnesses; comprising a total of 215 visits associated with NDC.

Regarding syndrome classification, 58.14% (125/215) of syndromes were correctly matched. In addition, 14.8% (32/215) of visits were assigned to the 'other' category and not identified appropriately as a syndrome based on the reported NDC. The rash syndrome category had the highest proportion of matches. The gastrointestinal illness (GI) category had the highest frequency of related ED visits. The GI category also had the highest proportion of CC misclassified as a single syndrome; 30/154 (19.5%) classified in the fever syndrome. This was most prominent among children <2 years old.

Conclusions

This study provides an evaluation of syndrome classification among individuals with reported NDC. In addition, this review revealed limitations for the use of SS data to provide trends for NDC. For instance, children <2 years old are

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

non-verbal and as a result, signs such as fever and rash can predominate CC of the caregiver. This was demonstrated by the misclassification of reported enteric illnesses within the fever syndrome among children <2 years old. Individuals may also present to EDs for unrelated illnesses and later be diagnosed with a NDC. Also, the seasonality of enteric NDC may have resulted in higher frequencies of GI visits during the study period. Overall, collection and evaluation of ED visitation histories during NDC case interviews has improved the local interpretation of SS signals and provided context for syndromic and NDC surveillance.

Acknowledgements

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

References

1 Lombardo J, Burkom H, Pavlin J. Centers for Disease Control and Prevention (CDC). II ESSENCE and the framework for evaluating syndromic surveillance systems. *MMWR Morb Mortal Wkly Rep* 2004;**53** (Suppl): 159–65.
2 DHHS and CDC, syndrome definitions for diseases associated with critical bioterrorism-associated agents, 2003, http://www.bt.cdc.gov/surveillance/syndromedef/.

Emerging Health Threats Journal 2011, **4**:s121. **doi:** 10.3134/ehtj.10.121
© 2011 R Borroto *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Tracking emergency department (ED) patients with gastrointestinal symptoms during a Norovirus epidemic in Miami-Dade

R Borroto, J Suarez, E Rico, E O'Connell, A Llau, L Londono, P Mann, and G Zhang

Office of Epidemiology, Disease Control and Immunization Service (EDC-IS), Miami-Dade County Health Department (MDCHD), Miami, FL, USA
E-mail: Rene_Borroto-Ponce@doh.state.fl.us

Objective

To demonstrate how the Epidemiology, Disease Control and Immunization Service (EDC-IS) at the Miami-Dade County Health Department used ESSENCE to track gastrointestinal symptoms during a Norovirus epidemic.

Introduction

From January to March 2010, thirteen outbreaks of Norovirus infection were reported to the EDC-IS, up from four outbreaks in the entire 2008 and same number during 2009. Individual cases of Norovirus are not reportable in the State of Florida. That makes it difficult to track the onset, rise, peak, and fading of epidemics of this disease.

Methods

A countywide query with the string nausea, or vomit, or diarrhea, or gastroenteritis is performed in ESSENCE on a daily basis. The incidence of enteric illnesses other than Norovirus was analyzed by comparing the count of the period January–March 2010 to the mean of the same months from 2000 through 2009. The reportable diseases included in the analysis were of bacterial origin (shigellosis, *Escherichia coli* infection, campylobacteriosis, and salmonellosis) and parasitic nature (cyclosporiasis, cryptosporidiosis, and giardiasis), as notified in the Merlin database of reportable diseases of the State of Florida. No outbreak was notified during the January–March 2010 period among non-reportable enteric diseases (that is, caused by *Bacillus cereus*, *Clostridium perfringens*, *Plesiomonas shigelloides*, *Aeromonas*, and *Staphylococcus*). As for non-reportable rotavirus infection, the monthly count of diagnosed children aged 0–4 years from January 2003 to April 2010 at the largest pediatric hospital in Miami-Dade was analyzed.

Results

There had been a spike of emergency department (ED) patients right after Christmas in late December 2009 (see Figure 1), but, unlike the spikes of January–March 2010, it

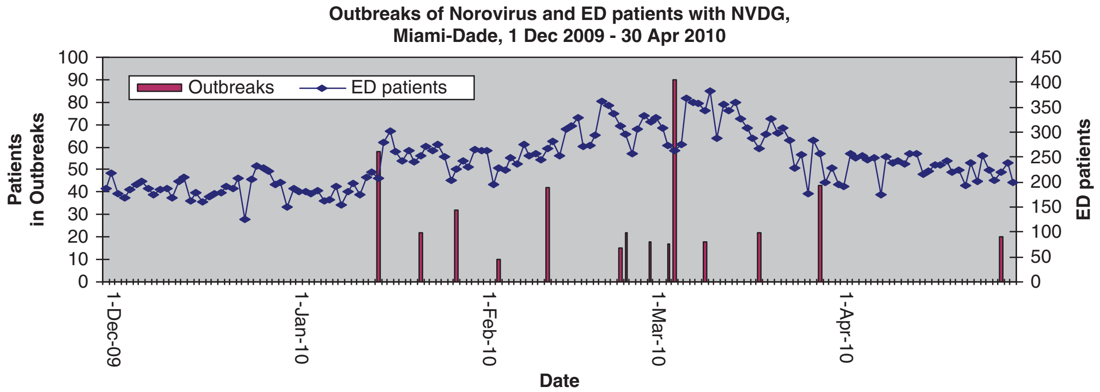


Figure 1 Norovirus outbreaks and ED patients with gastrointestinal symptoms.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

occurred only among adults, and spared children. This spurt in December returned to background levels after 3 days, unlike the spikes of January–March 2010. Moreover, no outbreak of Norovirus was reported in December 2009. The first outbreak of 2010 had its onset on Wednesday, 13 January, whereas the first significant increase in ED patients occurred from 16 (Saturday) to 18 (Monday) January, during the Dr Martin Luther King Jr holiday weekend, when 3 consecutive days of red flags were posted among the all-age group. The next leg up in the uptrend of the count of ED patients came right after St Valentine Day, when four consecutive flags were posted from 15 through 18 February. The count of ED patients peaked in early March 2010, when the number of Norovirus outbreaks also peaked. The only reportable enteric disease that showed a peak in March 2010 was campylobacteriosis, mostly because of changes in the case definition, rather than an actual increase in incidence. Rotavirus infection did not appear to make a contribution to the increase in the count of ED patients.

Conclusions

ESSENCE can help tracking epidemics of Norovirus infection, although a careful analysis of the temporal patterns of the incidence of other enteric illnesses should be ascertained.

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.

Emerging Health Threats Journal 2011, **4**:s65. **doi:** 10.3134/ehtj.10.065
© 2011 H Burkom *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A collaboration to enhance detection of disease outbreaks clustered by time of patient arrival

H Burkom[1], L Ramac-Thomas[1], R Arvizu[1], C Lee[1], W Loschen[1], R Wojcik[1], and A Kite-Powell[2]

[1]*Johns Hopkins University Applied Physics Laboratory, National Security Technology Department, Baltimore, MA, USA; and* [2]*Florida Department of Health, Bureau of Epidemiology, Tallahassee, FL, USA*
E-mail: howard.burkom@jhuapl.edu

Objective

This presentation discusses the approach and results of collaboration to enable a solution of a hospital time-of-arrival (TOA) monitoring problem in syndromic surveillance applied to public health data at the hospital level for county monitoring.

Introduction

One objective of public health surveillance is detecting disease outbreaks by looking for changes in the disease occurrence, so that control measures can be implemented and the spread of disease minimized. For this purpose, the Florida Department of Health (FDOH) employs the Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE). The current problem was spawned by a labor-intensive process at the FDOH: authentic outbreaks were detected by epidemiologists inspecting ESSENCE time series and derived event lists. The corresponding records indicated that patients arrived at an ED within a short interval, often less than 30 minutes. TOA task was to develop and automate a capability to detect events with clustered patient arrival times at the hospital level for a list of subsyndrome categories of concern to the monitoring counties.

Methods

Factors that drove the choice of the statistical alerting algorithm for the TOA surveillance were: user objectives and processing performance needs, the existing ESSENCE operational framework, and the limitation to time series and covariates that could be extracted quickly with ESSENCE queries.

The chosen approach was to monitor each of the many time-series cells, in which cells could be stratified by several variables. Following analysis of the influence of selected variables in historical data, a cell was defined as a hospital/syndrome/time-of-day combination (Figure 1). For the majority of these cells, a Poisson distribution test was applied, and a negative binomial test was applied for cells whose time-series counts were overdispersed. The method adjusts for daily patterns and for levels of activity that vary widely across hospitals.

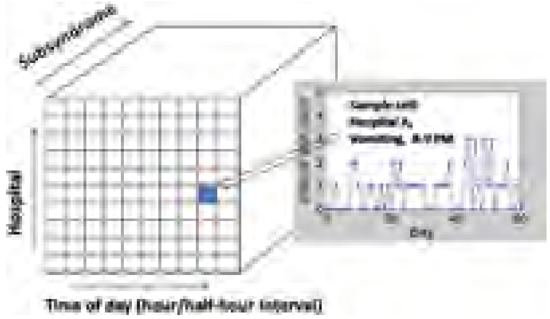


Figure 1 Schematic depicting the temporal cell-based monitoring method.

Results

This method was tested and validated on ED visit count data covering over 6 years and aggregated at 30-minute intervals for 51 subsyndromes from each of 134 hospitals covering 34 Florida counties.

An alerting threshold of $P = 0.0001$ gave detection of 18 of the 20 known signals in the data, with manageable alert counts for the most common subsyndromes in the largest hospitals. This method is implemented in an operational ESSENCE website module.

Conclusion

The TOA application resulted from (a) a well-defined problem and associated constraints; (b) involvement of all stakeholders for the duration of the project, including a state-level epidemiologist domain expert; and (c) data streams sufficient for a useful solution. The approach and implementation were the result of a give-and-take process informed by data analysis, with features that were informative for the efficiency of future collaborative efforts.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s66. **doi:** 10.3134/ehtj.10.066
© 2011 BF Buss *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Influenza mortality surveillance using a statewide electronic death registration system — Nebraska, 2009–2010

BF Buss[1,2], R Williams[1], P Pillai[1,2], and T Safranek[1]

[1]*Nebraska Department of Health and Human Service, Lincoln NE, USA; and* [2]*Centers for Disease Control and Prevention, CDC Office of Public Health Preparedness, Atlanta, GA, USA*
E-mail: bryan.buss@nebraska.gov

Objective

This report describes use and evaluation of a near real time, novel electronic influenza mortality surveillance system developed in Nebraska.

Introduction

Public health surveillance using death data is critical for tracking the impact of diseases such as influenza. However, utility of such systems is compromised by delayed reporting, particularly when it is paper based. In Nebraska, funeral directors are encouraged to initiate death certificates electronically by an electronic death registration system (EDRS). Although paper-based or mixed (electronic followed by paper) registration is still accepted statewide, EDRS usage has gradually increased over time. Fact of death (FOD) data that includes time and place of death, and a deceased person's identity are usually recorded by a funeral director. Cause of death data in the medical portion are provided by physicians or medical examiners at a later date. FOD data entered into EDRS are immediately available, whereas paper-based data must first be mailed to vital records whereupon staff enter it into EDRS. Although implemented in 2006, epidemiology surveillance staff did not have real-time access to EDRS data until early 2009, when a collaboration was formed between the Office of vital records and the Office of epidemiology within the Nebraska Department of Health and Human Services. Daily electronic access by surveillance staff to death certificate data was established enabling the conductance of public health death surveillance.

Methods

Timeliness of electronically-filed versus paper-based FOD data was evaluated from 19 July 2009 to 1 May 2010. Additionally, sensitivity and specificity of an influenza-specific death surveillance system using EDRS were compared with active surveillance using laboratory and provider reports also conducted from 19 July to 1 May 2010. For active surveillance, an influenza-associated death was defined as a death, with laboratory confirmation or clinical impression of influenza upon chart review. For surveillance by EDRS, funeral directors were instructed during this time period to query decedent's next-of-kin whether the death 'was in any way associated with influenza.' Responses including yes, no, or unknown were entered into a new field, created specifically in the FOD section in EDRS for purposes of influenza death surveillance. Deaths reportedly associated with influenza were investigated to ascertain cases.

Results

Of 11,598 deaths, filed in Nebraska during 19 July 2009 to 1 May 2010, registration for a total of 7,354 (63.4%) were initiated electronically by EDRS, for which FOD notice and data were available at a median of 10 days sooner than paper-based data (Table 1). Among the deaths registered by EDRS, 31 (0.4%) were reportedly influenza associated. The remaining 7,323 (99.6%) were classified as 'no,' 'unknown,' or had missing data. Active surveillance identified 21 influenza-associated deaths that met the case definition. Of these, 16 (76.2%) were initiated electronically, and were included in the analysis; approximately half being reported by funeral directors as influenza associated

Table 1 Time interval in days from date of death to date when death registration was initiated

Initiation method	N	*Mean (days)*	*Median (days)*	*Range (days)*
Electronic	7354	2.0	1.0	0–113
Paper based	4244	13.8	11.0	1–221

This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 2 Influenza-associated death notification, with EDRS, compared with active surveillance

Surveillance by EDRS	*Active surveillance*		*Total*
	Case	*Not a case*	
Case	7	24	31
Not a case	9[a]	7314	7323
	16	7338	7354

Abbreviation: EDRS, electronic death registration system. [a]Five 'no,' three 'unknown,' and one missing. Sensitivity = 43.8% (95% confidence interval, 19.4–68.1). Specificity = 99.7% (95% confidence interval, 99.5–99.8). Positive predictive value = 22.6% (95% confidence interval, 7.9–37.3). Negative predictive value = 99.9% (95% confidence interval, 99.8–100.0).

(sensitivity, 43.8%). Overall, few false positives as a proportion of all deaths (specificity, 99.7%) were reported (Table 2).

Conclusion

Nebraska's EDRS substantially enhances timeliness of FOD data availability. Although sensitivity and positive predictive values were limited, disease-specific death notification by funeral directors by EDRS holds promise as an early detection tool for emergent public health problems such as influenza-associated mortality. Additional training and refinement are needed to improve utility. Further study is needed to evaluate its utility during future influenza seasons when heightened public awareness attributable to 2009–2010 pandemic H1N1 has waned, and for mortality surveillance applied to other public health conditions.

Acknowledgements

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

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Likewise, the authors' findings and conclusions do not necessarily represent the views of the Nebraska Department of Health and Human Services.

Emerging Health Threats Journal 2011, **4**:s122. **doi:** 10.3134/ehtj.10.122
© 2011 BF Buss and TJ Safranek.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Population-based surveillance for methicillin-resistant *Staphylococcus aureus* infections among high school football and wrestling participants—Nebraska, 2008–2010

BF Buss[1,2], and TJ Safranek[1]

[1]*Nebraska Department of Health and Human Service, Lincoln, Nebraska, USA; and* [2]*Centers for Disease Control and Prevention, Atlanta, Georgia, USA*
E-mail: bryan.buss@nebraska.gov

Objectives

The Nebraska Department of Heath and Human Services (NDHHS) Office of Epidemiology conducted ongoing surveillance to monitor statewide incidence of physician-diagnosed methicillin-resistant *Staphylococcus aureus* (MRSA) infections among high school football and wrestling participants during school years 2008–09 and 2009–10.

Introduction

MRSA, which had been known primarily as a cause of healthcare-associated infections, increasingly has been recognized as a cause of skin infections among persons of all ages who have little or no contact with healthcare settings. Such infections are the most common cause of skin and soft-tissue infections among patients presenting to emergency departments.[1] Community-associated MRSA infections among high school, college, and professional athletes, and outbreaks within teams have established otherwise healthy athletes as a population at risk.[2–5] To monitor the occurrence of such infections among Nebraska high school athletes, the NDHHS Office of Epidemiology first initiated population-based surveillance to determine incidence during the 2006–07 and 2007–08 school years. Findings of this surveillance established MRSA as an emerging cause of infections among Nebraska high school athletes, particularly football and wrestling participants; from 2006–07 to 2007–08, physician-diagnosed MRSA incidence per 10,000 wrestlers increased from 19.6 to 60.1, and incidence per 10,000 football players increased from 5.0 to 25.1.[6] In response to this finding, surveillance conducted at regular intervals has since been continued to monitor trends.

Methods

To facilitate ongoing surveillance, a list of official contacts in each of Nebraska's 312 high schools has been maintained and regularly updated. To foster collaboration and enhance response rates, MRSA prevention and control resources and surveillance findings are regularly provided to the targeted officials in each school.

During both the 2008–09 and 2009–10 school years, we developed four Internet-based questionnaires to survey delegated officials in all 312 Nebraska high schools, following the completion of the respective football and wrestling seasons. Each school's administration has delegated responsibility for completing the survey to an appropriate official associated with that school's athletic programs and having knowledge of procedures for prevention, identification, and control of infections among athletes. We collected number of participants for both football and wrestling in each school and number of athletes with physician-diagnosed MRSA infections to calculate statewide attack rates per 10,000 participants.

Results

Of 312 Nebraska schools, 239 (76.6%) and 251 (80.4%) responded to 2008–09 and 2009–10 football surveys, respectively. Of respondents, 97.5% (233/239) and 98.4% (247/251) had football programs of which 6.4% (15/233) and 3.6% (9/247) reported $\geqslant 1$ physician-diagnosed MRSA-affected participant. Similarly, of 312 schools, 177 (56.7%), and 251 (80.4%) responded to 2008–09 and 2009–10 wrestling surveys, respectively. Of respondents, 78.0% (138/177) and 75.3% (189/251) had wrestling programs of which 7.2% (10/138) and 7.4% (14/189) reported $\geqslant 1$ physician-diagnosed MRSA-affected participant. Incidence per 10,000 football players decreased from 19.0 to 11.5, and incidence per 10,000 wrestlers decreased from 60.8 to 50.0 from 2008–09 to 2009–10, respectively.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

Estimated incidence of physician-diagnosed MRSA infections among Nebraska high school football and wrestling participants decreased substantially during 2009–10 compared with 2008–09, demonstrating that the apparent increasing incidence has not continued. By maintaining updated contact information of at least one official in each Nebraska high school and communicating regularly, we have continued surveillance with high rates of participation to monitor MRSA incidence among high school athletes over four consecutive school years.

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 Moran GJ, Krishnadasan A, Gorwitz RJ, Fosheim GE, McDougal LK, Carey RB, *et al*. Methicillin-resistant S. aureus infections among patients in the emergency department. *N Engl J Med* 2006;**355:** 666–74.

2 Lindenmayer JM, Schoenfeld S, O'Grady R, Carney JK. Methicillin-resistant Staphylococcus aureus in a high school wrestling team and the surrounding community. *Arch Int Med* 1998;**158:**895–9.

3 Romano R, Lu D, Holtom P. Outbreak of community-acquired methicillin-resistant Staphylococcus aureus skin infections among a collegiate football team. *J Athl Train* 2006;**41**(2): 141–5.

4 Nguyen DM, Mascola L, Brancoft E. Recurring methicillin-resistant Staphylococcus aureus infections in a football team. *Emerg Infect Dis*. 2005;**11**(4): 526–32.

5 Kazakova SV, Hageman JC, Matava M, Srinivasan A, Phelan L, Garfinkel B. A clone of methicillin-resistant Staphylococcus aureus among professional football players. *N Eng J Med* 2005;**352:**468–75.

6 Buss BF, Mueller SW, Theis M, Keyser A, Safranek TJ. Population-based estimates of Methicillin-Resistant Staphylococcus Areus (MRSA) infections among high school athletes—Nebraska, 2006–2008. *J Sch Nurs* 2009;**25**(4): 282–91.

Emerging Health Threats Journal 2011, **4**:s70. **doi:** 10.3134/ehtj.10.070
© 2011 N Close *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Influenza surveillance using inpatient health information exchange data

N Close[1], PF Bugni[2], and K Lofy[1]

[1]*Communicable Disease Epidemiology, Washington State Department of Health, Shoreline, WA, USA; and* [2]*Center for Public Health Informatics, University of Washington, Seattle, WA, USA*
E-mail: natasha.close@doh.wa.gov

Objective

To evaluate the timeliness, completeness, and representativeness of influenza hospitalization data from an inpatient health information exchange.

Introduction

During the 2009 H1N1 influenza pandemic, the Washington State Department of Health (DOH) temporarily made lab-confirmed influenza hospitalizations and deaths reportable. As reporting influenza hospitalizations is resource intensive for hospitals, electronic sources of inpatient influenza surveillance data are being explored. A large Health Information Exchange (WA-HIE) currently sends DOH the following data elements on patients admitted to 14 hospitals throughout eastern Washington: hospital, admission date, age, gender, patient zip code, chief complaint, final diagnoses, discharge disposition, and unique identifiers. WA-HIE inpatient data may be valuable for monitoring influenza activity, influenza morbidity, and the basic epidemiology of hospitalized influenza cases in Washington.

Methods

Timeliness and completeness of records received through the WA-HIE were evaluated. Basic characteristics and time trends of WA-HIE influenza hospitalizations (that is, records with ICD9 487–488 listed in the final diagnoses) between April 27, 2009–May 15, 2010 were compared with lab-confirmed influenza hospitalizations of eastern Washington residents reported through the WA notifiable condition reporting system (PHIMS). Overlap between the WA-HIE and PHIMS influenza cases was assessed by conducting a record-level comparison with a subset of WA-HIE and PHIMS influenza cases.

Results

For 90% of admissions into WA-HIE hospitals, the first record is transmitted 1 day after admission (range 0–26 days). One or more data elements were missing from 34% of records 28 days after admission. Of these incomplete records, 96% were missing the final diagnoses.

Data for influenza hospitalizations from the WA-HIE do not differ significantly from data collected through PHIMS with regard to sex, age, and mortality (Table 1). The time series of influenza-related hospital admissions from the WA-HIE and PHIMS are highly correlated ($r = 0.97$, Figure 1).

Table 1 Characteristics of influenza cases detected through WA-HIE and reported through PHIMS

Case characteristic	*WA-HIE* (N = *375*) N *(%)*	*PHIMS* (N = *489*) N *(%)*	P[a]
Sex			
Male	154 (41.1)	199 (40.7)	0.91
Female	221 (58.9)	290 (59.3)	0.91
Age group (years)			
0–4	78 (20.8)	116 (23.7)	0.32
5–24	83 (22.1)	114 (23.3)	0.69
25–49	94 (25.1)	127 (26.0)	0.76
50–64	77 (20.5)	85 (17.4)	0.24
65+	43 (11.5)	47 (9.6)	0.38
Deaths	12 (3.2)	22 (4.5)	0.33

[a]*P*-value calculated using the χ^2-test.

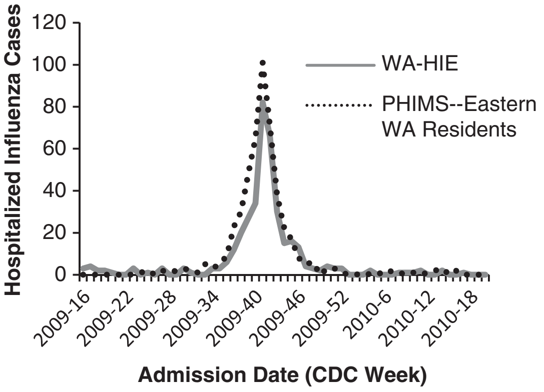


Figure 1 Time series of influenza cases detected through WA-HIE and reported through PHIMS.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Of 305 WA-HIE influenza hospitalizations during October–December 2009, 136 (44.6%) matched to a case reported in PHIMS.

Conclusions

WA-HIE data is an automated, relatively timely source of hospitalization information. Despite a lack of complete overlap between records reported through PHIMS and the WA-HIE, the WA-HIE data are representative of hospitalized influenza cases reported through PHIMS with respect to basic case characteristics and time trends.

Future directions

A prospective evaluation of hospitalized influenza cases is planned. In addition, the WA-HIE data elements are being expanding to allow for more in-depth data evaluation, detection of lab-confirmed influenza cases, assessment of illness severity, and evaluation of risk factors.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s71. **doi:** 10.3134/ehtj.10.071
© 2011 JS Coberly *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Biosurveillance applications for resource-limited settings: open ESSENCE and ESSENCE desktop edition

JS Coberly[1], RA Wojcik[1], JF Skora[1], AM Poku[1], ZC Whitley[1], TC Campbell[1], CJ Hodanics[1], SL Lewis[1], and DL Blazes[2]

[1]*The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; and* [2]*Division of GEIS Operations, Armed Forces Health Surveillance Center, Silver Spring, MD, USA*
E-mail: jacqueline.coberly@jhuapl.edu

Objective

This paper describes the development and early implementation of two freely available electronic biosurveillance software applications: Open ESSENCE (OE), and ESSENCE Desktop Edition (EDE).

Introduction

More than a decade ago, in collaboration with the U.S. Department of Defense, the Johns Hopkins University Applied Physics Laboratory (JHU/APL) developed the Electronic Surveillance System for the Early Notification of Community-based Epidemics (Enterprise ESSENCE),[1] which is currently used by federal, state and local health authorities in the US. As emerging infections will most likely originate outside of the US (for example, SARS) the application of electronic biosurveillance is increasingly important in resource limited areas.[2] In addition, such systems help governments respond to the recently modified International Health Regulations.[3] Leveraging the experience gained in the development of Enterprise ESSENCE, JHU/APL has developed two freely available electronic biosurveillance systems suitable for use in resource-limited areas: OE and EDE.

Methods

Biosurveillance is needed at all levels of public health, the local health clinic, as well as city, regional and national health authorities. The availability of computers and trained personnel varies considerably among these entities. OE and EDE were designed with this disparity in mind. OE is a web-based application designed to run on a system with internet or intranet access. It provides web-based data entry and data visualization to multiple users. EDE, on the other hand, is a stand-alone application designed to run on a single desktop or laptop computer. Both systems have much of the functionality found in the original Enterprise ESSENCE. They produce time series for the health data being analyzed, using both proprietary and open source alerting algorithms. Each also produces case-level data lists, pie and bar charts of any available variable, and alert or prevalence maps if appropriate geographical shape files are available. In addition, all of the user interface components in OE are internationalized. Addition of a language table created by the user allows OE to function in any language. Although EDE is currently only available with an English-user interface, it can read and present data in English, as well as in character-based languages, such as Thai. The most notable difference between Enterprise ESSENCE, OE and EDE is the data repositories. Enterprise ESSENCE has an established database schema and the data must conform to its schema. OE and EDE were designed to be more flexible. OE connects and adapts to the user's database and schema and EDE connects to databases and text files in various user formats.

Results

OE is currently being piloted in Peru by the Peruvian Navy and Army in conjunction with the U.S. Naval Medical Research Center Detachment (NMRCD). It was installed in 2009 and evaluation is ongoing. EDE is being piloted by the Cebu City Health Office in the Republic of the Philippines in conjunction with the Armed Forces Research Institute of Medical Sciences Virology Research Unit (PAVRU). Version 2.0 of EDE was installed early in 2010. An ongoing Simple Message Service (SMS) fever surveillance program will be used to pilot the application.

Conclusions

OE and EDE are freely available electronic biosurveillance applications, designed to facilitate surveillance at all levels of public health infrastructure. OE is web-based, and requires internet access for functionality, whereas EDE is a stand-alone desktop application designed for local users and emergency situations. Both applications provide an electronic surveillance capability, which can help countries conform to IHR 2005 regulations.

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.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

1 Lombardo JS, Ross DA. Disease surveillance: a public health priority. In: Lombardo JS, Buckeridge DL (eds). *Disease surveillance: a public health informatics approach*. John Wiley & Sons, Inc: Hoboken, 2007.
2 Heymann D. The international response to the 2003 SARS outbreak. *Philos Trans R Soc Lond B Biol Sci* 2004;**359:**1127–9.
3 Rodier G. New rules on international public health security. *Bull World Health Organ* 2007;**85:**428–30.

Emerging Health Threats Journal 2011, **4**:s75. **doi:** 10.3134/ehtj.10.075
© 2011 K Desy *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Relationship between neighborhood poverty and emergency department utilization for fever/flu syndrome

K Desy, K Purviance, M Cartter, and JL Hadler

Connecticut Department of Public Health, Epidemiology and Emerging Infections Program, Hartford, CT, USA
E-mail: kristen.desy@ct.gov

Objective

To describe the relationship between neighborhood poverty and emergency department visits for fever/flu syndrome illnesses reported through the Connecticut Hospital Emergency Department Syndromic Surveillance (HEDSS) system.

Introduction

The HEDSS system was implemented in 2004 to monitor disease activity.[1] In all, 18 of 32 emergency departments (ED) and 1 urgent care clinic provide data. Chief complaints are routinely categorized into eight syndromes. The fever/flu syndrome is used for early detection and monitoring of influenza in the community.[2]

Area-based measures, such as zip code, enable linkage to area-based socioeconomic census data. Neighborhood poverty, defined as the percentage of persons living below the federal poverty level in a geographic area, predicts a wide range of disease outcomes.[3]

Methods

HEDSS data were categorized into 6-week periods of maximal influenza activity: baseline (9/06-10/18/2008), 2008–2009 seasonal, spring 2009 H1N1 and fall 2009 H1N1. Neighborhood poverty was categorized as low ($<5\%$ below federal poverty level), medium–low (5–$<10\%$), medium–high (10–$<20\%$), and high ($\geqslant 20\%$) based on zipcode. Chi-square for trend was calculated for age and poverty-stratified percentages of total ED visits for fever/flu for each time period. The percentage increase over baseline in ED visits for fever/flu was calculated by neighborhood poverty level (Figure 1).

Results

There were a total of 4051 fever/flu visits at baseline, 8369 during the seasonal peak, 11 426 during the 2009 spring H1N1 peak, and 14 057 during the fall 2009 H1N1 peak. During each of these periods, the percentage of ED visits that were fever/flu increased with increasing poverty level ($P<0.0001$, χ^2 for trend). The strength of the association was similar at baseline and for each influenza period and for each age group except $\geqslant 65$ years for which no relationship was seen between poverty and ED utilization for fever/flu.

The percentage increase in ED visits for fever/flu compared with baseline was similar for all poverty levels and age groups in each influenza season, with the exception of 5–17 year olds who had significantly higher rates of ED utilization with increasing poverty level during the spring 2009 H1N1 wave ($P<0.0001$).

Conclusions

There is a strong and consistent association between neighborhood poverty and ED visits for fever/flu compared with all other syndromes combined. The strength of association is similar for fever/flu illnesses at baseline, during seasonal flu, and during each of the H1N1 waves. The percentage increase in ED visits for fever/flu during each flu season was similar for all poverty levels. If visiting the ED makes one more likely to be admitted as an inpatient than not visiting the ED, this could partially explain the increased incidence of hospitalization seen in persons in higher poverty neighborhoods. Efforts to reduce ED visits for

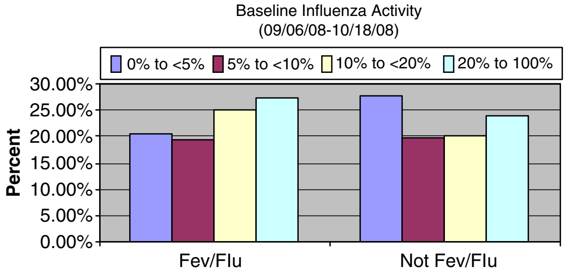


Figure 1 At baseline, ED utilization for fever/flu syndrome was associated with increasing neighborhood poverty. This was not true of non-fever/flu visits. Similar patterns in ED utilization were observed during the other influenza activity periods.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

fever/flu (for example, vaccination, education) should be targeted to higher poverty neighborhoods. Additional study is needed to determine why persons from higher-poverty neighborhoods seem to be more likely to visit the ED for fever/flu syndrome.

Acknowledgements

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

References

1 Dembek Z, Carley K, Hadler J. Guidelines for constructing a statewide hospital syndromic surveillance network. *MMWR* 2005;**54** (Suppl): 21–4.
2 Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. *EID* 2004;**10:**858–64.
3 Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures—The Public Health Disparities Geocoding Project. *Am J Public Health* 2003;**93:**1655–71.

Emerging Health Threats Journal 2011, **4**:s127. **doi:** 10.3134/ehtj.10.127
© 2011 H Dhotre *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

A comparison of the fever–flu syndrome category with the SC ILINet surveillance system in South Carolina: 2009–2010 influenza season

H Dhotre, C Springs, D Drociuk, and YE Cheng

Division of Acute Disease Epidemiology (DADE), SC Department of Health and Environmental Control (SC DHEC), Columbia,SC, USA
E-mail: dhotrehc@dhec.sc.gov

Objective

This paper compares the South Carolina Aberration Alerting Network's (SCAAN) hospital-based fever–flu syndrome category with the South Carolina Outpatient Influenza-like Illness Network (ILINet) provider surveillance system. This is the first comparison of South Carolina's syndromic surveillance SCAAN data with ILINet data since SCAAN's deployment.

Introduction

SCAAN is a collaborative network of syndromic systems within South Carolina. Currently, SCAAN contains the following data sources: SC Hospital Emergency Department (ED) chief-complaint data, Poison Control Center call data, Over-the-Counter pharmaceutical sales surveillance, and CDC's BioSense biosurveillance system. ILINet is collaboration between the Centers for Disease Control, state health departments and health care providers. ILINet is one of several components of SC's influenza surveillance.[1,2]

Methods

The inclusion criteria for this study were any individuals who visited a hospital facility or an ILI provider in South Carolina from 25 April 2009 to 26 June 2010.

As of July 2010, a total of 14 hospital facilities are enrolled in the SCAAN system. There is at least one hospital facility reporting ED data in five of the eight public health regions of the state.

Seventy South Carolina providers were enrolled in ILINet during the 2009–2010 season. Provider enrollment encompassed all eight DHEC public health regions.

The fever–flu syndrome category and SC ILI surveillance share the same definition: fever (⩾100°F) and cough and/or sore throat (in the absence of a known cause other than influenza).

The weekly fever–flu percentage was calculated by dividing the total number of ED visits with a fever–flu chief-complaint seen in all hospitals by the total number of ED visits seen by all hospitals.

The weekly state ILI percentage was calculated by dividing the total number of patients seen with ILI by the total number of patients seen.

The Pearson's correlation coefficient test was performed in SAS v.9.2 to assess the strength of association between the fever–flu syndromic surveillance and ILI surveillance. Additional analysis will be conducted to compare both surveillance methods geographically.

Results

Figure 1 shows the weekly fever–flu and ILI percentage from 25 April 2009 through 26 June 2010. There was an initial increase in ILI-related visits to the hospital ED near the end of April 2009. This was also the same time period the H1N1 novel influenza media releases were occurring nationally and statewide.

A sharp increase in the number of visits related to ILI was evident around the beginning of September 2009 through the end of October 2009. During this time period, the mean for fever–flu and ILI percentage was 9.29%

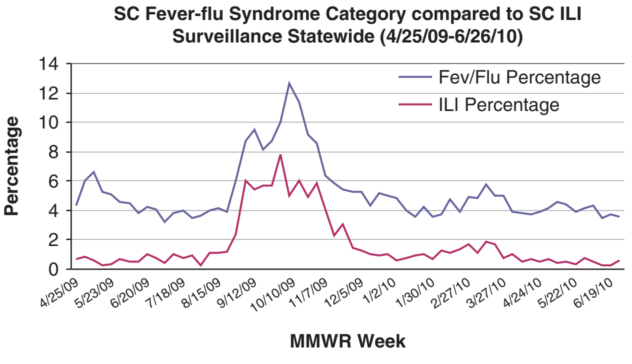


Figure 1 South Carolina's weekly fever–flu and ILI percentage from 25 April 2009 to 26 June 2010.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(range: 6.05–12.67%) and 5.48% (range: 2.40–7.77%), respectively. A peak in the ILI percentage occurred slightly before the peak in fever–flu percentage. Excluding the peak period (September–October), the mean for fever–flu and ILI percentage was 4.44% (range: 3.22–6.66%) and 0.96% (range: 0.24–4.1%), respectively. There was a high correlation ($r = 0.891$) between fever–flu and ILI percentages. Additional correlation analysis will be conducted to account for the geographic distribution of the two data sources (hospital EDs and provider clinics).

Conclusions

On the basis of these findings, the SCAAN fever–flu syndrome category offers an additional surveillance tool to the existing ILI surveillance system. It is useful to understand the population of SC residents who visit the hospital ED versus a private provider clinic for ILI-related issues. A phasing-out of ILI surveillance for the more reliable hospital ED fever–flu surveillance (daily automated analysis) may be a consideration once more hospitals join the SCAAN system.

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 Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenza like illness by International Classification of Diseases, Ninth Revision. *Emerg Infect Dis* 2007;**2**:207 (http://www.cdc.gov/EID/content/13/2/207).
2 The Centers for Disease Control and Prevention. The Early Aberration Reporting System. http://emergency.cdc.gov/surveillance/ears/.

Emerging Health Threats Journal 2011, **4**:s78. **doi:** 10.3134/ehtj.10.078
© 2011 JR Egger *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Can all-cause school absenteeism be used to optimize the timing of school closure in response to pandemic influenza?

JR Egger[1,2,3], AG Hoen[1,2,3], E Wilson[1,2,3], JS Brownstein[1,2,3], DL Buckeridge[1,2,3], and KJ Konty[1,2,3]

[1]*New York City Department of Health and Mental Hygiene, New York, NY, USA;* [2]*Children's Hospital Boston and Harvard Medical School, Boston, MA, USA; and* [3]*McGill Clinical and Health Informatics, McGill University, Montreal, QC, Canada*
E-mail: jegger@health.nyc.gov

Objective

To determine the optimal pattern in school-specific all-cause absenteeism for use in informing school closure related to pandemic influenza.

Introduction

School closure has long been proposed as a non-pharmaceutical intervention in reducing the transmission of pandemic influenza.[1] Children are thought to have high transmission potential because of their low immunity to circulating influenza viruses and high contact rates. In the wake of pandemic (H1N1) 2009, simulation studies suggest that early and sustained school closure might be effective at reducing community-wide transmission of influenza.[2,3] Determining when to close schools once an outbreak occurs has been difficult. Influenza-related absentee data from Japan were previously used to develop an algorithm to predict an outbreak of influenza-related absenteeism.[4] However, the cause of absenteeism is frequently unavailable, making application of this model difficult in certain settings. For this study, we aimed to evaluate the potential for adapting the Japanese algorithm for use with all-cause absenteeism, using data on the rate of influenza-related nurse visits in corresponding schools to validate our findings.

Methods

Data consisted of the daily count of all-cause absenteeism and school nurse visits for fever/flu syndrome in 1206 public schools in New York City from 6 September 2005 to 26 June 2009. A model-adjusted school-specific z-score was calculated for both absenteeism and ILI for each day from 5 September 2006 to 28 June 2009. The z-score was calculated by dividing the model residual by the school-specific standard deviation of the outcome. Both models adjusted for day of week, whether the preceding or proceeding day was a holiday, school type (elem., middle, high), school day and sine and cosine terms to account for seasonality.

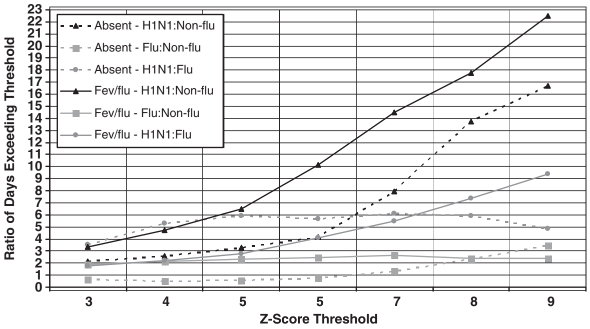


Figure 1 The ratio of the proportion of days exceeding various z-score thresholds, comparing the seasonal, pH1N1, and non-flu periods.

Results

The ratio of the proportion of z-scores higher than a given threshold for the non-flu period, seasonal flu period and the pH1N1 2009 period (April 20–June 19, 2009) are shown in the Figure 1. As expected, the ratio of high fever/flu z-scores during the pH1N1 period to non-flu periods is greater than one. This pattern is similar for high absentee z-scores during the pH1N1 period compared with the non-flu period; however, the ratio of absentee z-scores during the seasonal flu period compared with the non-flu period is below one for z-scores of six or less. This trend reverses itself for z-scores greater than six. In fact, for both absenteeism and fever/flu, the ratio tends to increase with an increasing threshold cutoff.

Conclusions

Preliminary results indicate that for moderately high absenteeism (z-score <6) there is little variation between seasonal and non-seasonal flu periods. However, high absenteeism (z-score >6) occurs more frequently during periods of flu transmission, suggesting that a high absentee threshold may be needed to predict a flu outbreak. Further work will involve determining the optimal temporal pattern

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and threshold levels in absenteeism that best predicts both high absenteeism and fever/flu visits. Results of this work will help determine the relative utility of tracking all-cause school absenteeism for use in school-based influenza surveillance.

Acknowledgements

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

References

1 Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, *et al*. Closure of schools during an influenza pandemic. *Lancet Infect Dis* 2009;**9**:473–81.
2 Halder N, Kelso JK, Milne GJ. Analysis of the effectiveness of interventions used during the 2009 A/H1N1 influenza pandemic. *BMC Public Health* 2010;**10**:168.
3 Lee BY, Brown ST, Cooley P, Potter MA, Wheaton WD, Burke DS, *et al*. Simulating school closure strategies to mitigate an influenza epidemic. *J Public Health Manag Pract* 2010;**16**(3):252–61.
4 Sasaki A, Hoen AG, Ozonoff A, Suzuki H, Tanabe N, Seki N, *et al*. Evidence-based tool for triggering school closures during influenza outbreaks Japan. *Emerg Infect Dis* 2009;**15**:1841–3.

Emerging Health Threats Journal 2011, **4**:s130. **doi:** 10.3134/ehtj.10.130
© 2011 C Faustino *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Evaluating the utility of school absenteeism during the 2009–2010 influenza season in Los Angeles County

C Faustino, P Araki, E Kajita, M Jones, and B Hwang

Los Angeles Department of Public Health (LACDPH), Acute Communicable Disease Control Unit (ACDC), Automated Disease Surveillance Section (ADSS), Los Angeles, CA, USA
E-mail: cfaustino@ph.lacounty.gov

Objective

To evaluate the utility of school absenteeism surveillance data in Los Angeles County during the 2009–2010 influenza season.

Introduction

It has been postulated that school absenteeism, a non-traditional surveillance data source, may allow for early detection of disease outbreaks, particularly among school-aged children who may not seek emergency medical attention. Although a New York City-based study showed moderate utility of school absenteeism in biosurveillance,[1] no study to date has been reported on school absenteeism in Los Angeles County, which contains the second largest school district in the US.[2]

Methods

School absenteeism data on school attendance are negative-based and completed by teachers via the online student information system; once per day for elementary schools, once per period for middle/high schools. Any final corrections to attendance are made at the end of the school day through an administrative portal. School absenteeism data are received by LACDPH in near real-time on a biweekly basis via SFTP.

For the purposes of this study, school absenteeism data were available from 140 elementary, middle, and high schools, during

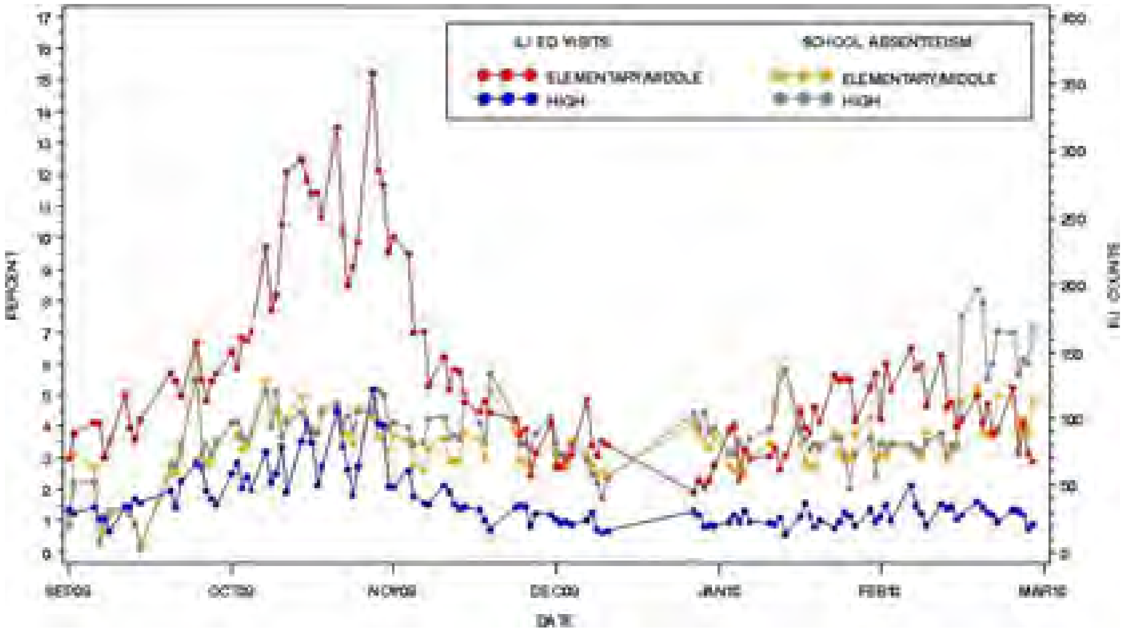


Figure 1 Percent School Absenteeism and ILI ED visits by School Age Groups in Los Angeles County.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the period from 1 September 2009 through 28 February 2010. A retrospective analysis was conducted to examine the correlations between school absenteeism and two other ADSS influenza surveillance systems: (1) influenza-like illness emergency department (ILI ED) visits by ages 5–17 and (2) over-the-counter cough/cold medications and thermometer sales.

Results

During the study period, mean percent absenteeism was 3.28 % for elementary/middle schools and 3.55 % for high schools ($p = 0.06$). Figure 1 shows the number of ILI ED visits, which peaked in early November for both elementary/middle and high school groups. In contrast, however, school absenteeism peaked in late September/early October for both groups. Compared with ILI ED visits in elementary/middle school-aged (5–13 years) and high school-aged (14–17 years) children, school absenteeism showed a weak correlation ($r = 0.38$, $p < 0.001$; $r = 0.33$, $p < 0.001$, respectively). Also, neither cough/cold medication sales nor thermometer sales showed strong correlations with either elementary/middle or high school absenteeism (data not shown).

Conclusions

We found no significant positive correlations of school absenteeism in elementary/middle school and high school groups, with either ILI ED visits or over-the-counter medication sales in Los Angeles County during the 2009–2010 influenza season. Interpreting medical outcomes/trends from a non-traditional source like school absenteeism is challenging, and continued investigation is warranted before determining its role in biosurveillance. Moreover, school absenteeism could serve to help assess the need of school closures during countywide disease outbreaks. However, further assessment of current data capture methods and quality of school absenteeism data within Los Angeles County are necessary to evaluate the utility of absenteeism in early event detection.

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 Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection. New York City. *BMC Public Health* 2005;**5**:105.

2 Snyder TD, Dillow SA, Hoffman CM. *Digest of Education Statistics 2008 (NCES 2009–020)*. National Center for Education Statistics, Institute of Education Sciences, US Department of Education: Washington, DC, 2009, pp 130–5. http://nces.ed.gov/pubs2009/2009020.pdf (retrieved 13 October 2010).

Emerging Health Threats Journal 2011, **4**:s131. **doi:** 10.3134/ehtj.10.131
© 2011 A Fowlkes *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Comparison of the US outpatient influenza-like illness surveillance network and the Google Flu Trends from 2008 to 2010

A Fowlkes[1], M Mohebbi[2], S Cook[2], C Conrad[2], L Brammer[1], and L Finelli[1]

[1]*Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA; and* [2]*Google.org, New York City, NY, USA*
E-mail: ahl4@cdc.gov

Objective

This paper compares estimates of influenza-like illness (ILI) activity with estimates from the Centers for Disease Control's ILINet from October 2008 through March 2010.

Introduction

http://Google.org developed a regression model that used the volume of influenza-related search queries best correlated with the proportion of outpatient visits related to ILI model to estimate the level of ILI activity.[1] For calibration, the model used ILINet data from October 2003 to 2009, which report weekly ILI activity as the percentage of patient visits to health care providers for ILI from the total number patient visits for the week. Estimates of ILI in 121 cities were added in January 2010.

Methods

For the 2008–2009 and 2009–2010 influenza seasons, we used ILINet data to compare the weekly percentage of outpatient ILI visits with the web query-based model estimates of ILI from Google Flu Trends. Data from Google Flu Trends are publicly available through the website. The time period for evaluation was parsed according to the emergence of pandemic influenza A (H1N1) virus (pH1N1): 'pre-pandemic' was defined as October 2008 through March 2009, 'wave one' as April through July 2009, and 'wave two'

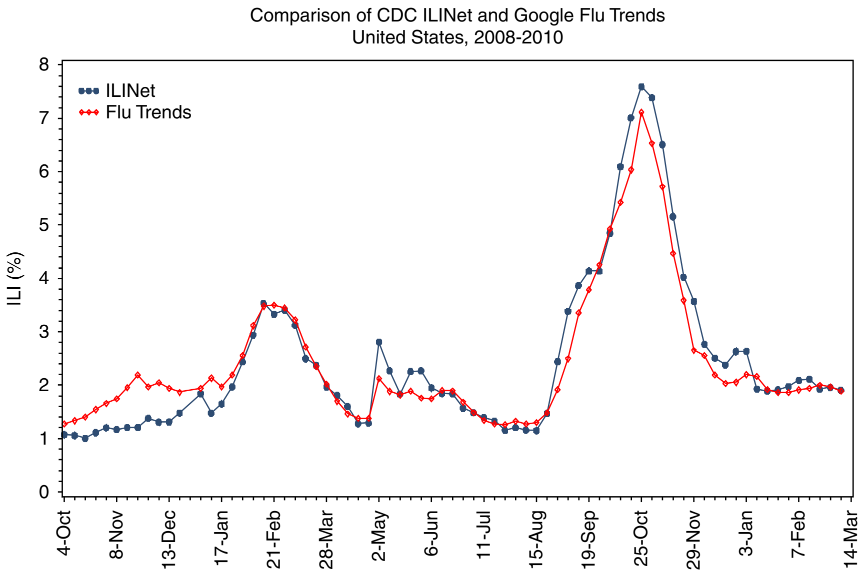


Figure 1 Comparison of Google Flu Trends estimates of influenza-like illness (red) with ILINet estimates from October 2008 to March 2010.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

as August 2009 through April 2010. A correlation coefficient was calculated for the nation, US Department of Health and Human Resources-defined regions, and states, and for 47 major metropolitan areas. Correlations below zero were displayed as zero.

Results

From October 2008 through April 2010, all national and regional ILI percentage estimates reported by ILINet and Google Flu Trends correlated highly (Figure 1). During each of the study time periods, the median correlation was 0.93 for national and regional comparisons. In the state-specific comparisons, the median correlation during the pre-pandemic season was 0.82, including two states with zero correlation (range 0–0.94). During wave one the median state correlation was 0.65, including four states with zero correlation (range 0–0.99). During wave two the median state correlation was 0.94 (range 0.68–0.99). In the major metropolitan areas, the median city correlation was 0.73 (range 0–0.95) during the pre-pandemic period, 0.48 (range 0–0.93) during wave one, and 0.89 (range 0.18–0.99) during wave two.

Conclusions

During periods of widespread influenza circulation, the percentage of ILI-related outpatient visits determined by ILINet and estimated by Google Flu Trends correlated well for national, regional, state and many major metropolitan-area levels. After the introduction of pH1N1, but before widespread circulation, state and metropolitan area estimates of ILI activity were less correlated. Google's experimental estimates of ILI activity in major metropolitan areas correlated well with provider-based surveillance data, but are less well correlated than the more robust national and regional data.

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.

Reference

1 Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;**457**:1012–14.

Emerging Health Threats Journal 2011, **4**:s83. **doi:** 10.3134/ehtj.10.083

© 2011 LH Gren *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Using influenza rapid test positivity as an early indicator for the onset of seasonal influenza

LH Gren[1], SC Alder[1], EA Joy[1], JL Lyon[1], CA Porucznik[1], and CJ Staes[2]

[1]*Departments of Family and Preventive Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA; and* [2]*Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA*
E-mail: lisa.gren@hsc.utah.edu

Objective

The objective of this study is to describe the use of point-of-care rapid influenza testing in an outpatient, setting for the identification of the onset of influenza in the community.

Introduction

Influenza causes significant morbidity and mortality, with attendant costs of roughly $10 billion for treatment and up to $77 billion in indirect costs annually.[1–3] The Centers for Disease Control and Prevention conducts annual influenza surveillance, and includes measures of inpatient and outpatient influenza-related activity, disease severity, and geographic spread.[4] However, inherent lags in the current methods used for data collection and transmission result in a one to two weeks delay in notification of an outbreak via the Centers for Disease Control and Prevention's website.[5] Early notification might facilitate clinical decision-making when patients present with acute respiratory infection during the early stages of the influenza outbreak.

In the United States, the influenza surveillance season typically begins in October and continues through May. The Utah Health Research Network has participated in Centers for Disease Control and Prevention's influenza surveillance since 2002, collecting data on outpatient visits for influenza-like illness (ILI, defined as fever of 100F or higher with either cough or sore throat). Over time, Utah Health Research Network has moved from data collection by hand to automated data collection that extracts information from discrete fields in patients' electronic health records.

We used statistical process control to generate surveillance graphs of ILI and positive rapid influenza tests, using data available electronically from the electronic health records.

Methods

Rapid influenza tests were introduced into Utah Health Research Network in December 2003. We extracted data from the clinical data warehouse for influenza seasons 2004–2005 through 2007–2008. ILI patients were those who: (1) had a measured temperature at the visit of at least 100F, and (2) had a 'reason for visit' field that contained either cough or sore throat. Regardless of ILI status, any patient who had a positive influenza rapid test was also identified.

The daily proportion of patients with ILI or with a positive rapid influenza test was graphed using statistical process control charts. Using the Western electric rules, we defined both the epidemic onset of the influenza outbreak and an early alert signal. We compared the presumed notification date for the early alert signal with the usual notification system in Utah that relies on data collected for Centers for Disease Control and Prevention's influenza surveillance to determine the timeliness of rapid test positivity surveillance.

Results

The peaks for ILI and rapid test positivity coincided in time. An early alert was identified in all four seasons using rapid test positivity, although no early alert was clearly identified for ILI in any season. The early alert for rapid test positivity occurred a median of 16 days before the epidemic onset. Assuming dissemination by email within two days of the alert, Utah Health Research Network clinicians could be notified of the start of the influenza outbreak about nine days earlier than ILI surveillance alerts posted to the Utah Department of Health website.

Conclusions

Influenza rapid test positivity provided an early alert, 16 days before the epidemic onset, and earlier than traditional surveillance for ILI. This early awareness could guide clinician decisions regarding diagnostic testing, prescription of antivirals, and counseling their patients on preventive behaviors.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

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

References

1 Thompson WW, Shay DK, Weintraub E, Brammer L, Cox N, Anderson LJ, *et al.* Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* 2003;**289:**179–86.

2 Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, *et al.* Influenza-associated hospitalizations in the United States. *JAMA* 2004;**292:**1333–40.

3 Molinari NA, Ortega-Sanches IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, *et al.* The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* 2007;**25:**5086–96.

4 Centers for Disease Control and Prevention. Flu activity and surveillance. http://www.cdc.gov/flu/weekly/fluactivity.htm.

5 Gren LH. Outpatient influenza surveillance: bridging the gap between clinical medicine and public health. [dissertation]. [Salt Lake City]: University of Utah; 2010. 88p.

Emerging Health Threats Journal 2011, **4**:s84. **doi:** 10.3134/ehtj.10.084
© 2011 A Grigoryan *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Creating a shared epidemiologic vocabulary: lessons from the former Soviet Union

A Grigoryan[1], C Clarke[2], T Chumachenko[3], and B Smoak[2]

[1]*Office of Surveillance, Epidemiology and Laboratory Services, Centers for Disease Control and Prevention, Atlanta, GA, USA;* [2]*Division of Preventive Medicine, Walter Reed Army Institute of Research, Silver Spring, MD, USA; and* [3]*Kharkiv National Medical University, Kharkov, Ukraine*
E-mail: AGrigoryan@cdc.gov

Objective

The objective of this study is to describe the development of a multilingual lexicon of epidemiology, which is needed for improved communication in public health surveillance internationally.

Introduction

As part of the United States Department of Defense strategy to counter biological threats, the Defense Threat Reduction Agency's biological threat reduction program is enhancing the capabilities of countries in the former Soviet Union (FSU) to detect, diagnose, and report endemic and epidemic, man-made or natural cases of especially dangerous pathogens. During these engagements, it is noted that Western-trained and Soviet-trained epidemiologists have difficulty, beyond that of simple translation, in exchanging ideas.

Before 1991, infectious disease surveillance in the FSU was centrally planned in Moscow. The methodologies of infectious disease surveillance and data analysis have remained almost unaltered since this time in most nations of the FSU. Vlassov[1] describes that FSU physicians and other specialists are not taught epidemiology as it is understood in the West. The Soviet public health system and the scientific discipline of epidemiology developed independently of that of other nations. Consequently, many fundamental Soviet terms and concepts lack simple correlates in English and other languages outside the Soviet sphere; the same is true when attempting to translate from English to Russian and other languages of the FSU. Systematic review of the differences in FSU and Western epidemiologic concepts and terminology is therefore needed for international public health efforts, such as disease surveillance, compliance with International Health Regulations 2005, pandemic preparedness, and response to biological terrorism. A multi-language reference in the form of a dictionary would greatly improve mutual comprehension among epidemiologists in the West and the FSU.

Methods

Following an extensive search of the Russian and English literature by a working group of Western and FSU epidemiologists, we created a matrix containing English and Russian definitions of key epidemiologic terms found in FSU and Western epidemiology manuals and dictionaries, such as *A Dictionary of Epidemiology*[2] and other sources. Particular emphasis was placed on terms relating to infectious disease surveillance, analysis of surveillance data, and outbreak investigation. In order to compare the definitions of each term and to elucidate small or large differences in usage, all definitions were translated into English and Russian so that the definitions could be compared side by side and discussed by the working group.

Results

Three hundred and fifty-four terms from 13 English and 12 Russian sources were chosen for inclusion. Review of the definitions showed that many terms used in biosurveillance and infectious disease public health practice by FSU and Western-trained epidemiologists are used differently, and some concepts are lacking altogether in the Russian or English literature.

Conclusions

Epidemiologic concepts and definitions significantly differed in the FSU and Western literature. To improve biosurveillance and international collaboration, recognition of these differences must occur, and where possible, agreement on standard definitions. Where agreement is not possible, precisely worded definitions illuminating differences in meaning and usage are essential. Implications for biosurveillance and public health practice will be discussed in the presentation and paper. This bilingual dictionary is expected to form the basis for trilingual dictionaries (Russian, English, and other languages

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of the FSU, such as Armenian, Azeri, Georgian and Ukrainian).

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.

The views expressed are those of the authors and should not be construed to represent the positions of the Department of the Army or Department of Defense.

References

1 Vlassov V. Is there epidemiology in Russia? *J Epidemiol Community Health* 2000;**54**:740–4.
2 Last JD. *A Dictionary of Epidemiology*, 4th edn. Oxford University Press: NY, 2001.

Emerging Health Threats Journal 2011, **4**:s85. **doi:** 10.3134/ehtj.10.085

© 2011 AG Hoen *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Effect of expanded recommendations for pediatric seasonal influenza vaccination: an international comparison

AG Hoen[1,2], DL Buckeridge[2], KML Charland[1,2], KD Mandl[1], and JS Brownstein[1]

[1]*Children's Hospital Boston and Harvard Medical School, Boston, MA, USA; and* [2]*McGill Clinical and Health Informatics, McGill University, Montreal, Quebec, Canada*
E-mail: anne.gatewood-hoen@childrens.harvard.edu

Objective

The objective of this study is to estimate the effect of expanding recommendations for routine seasonal influenza vaccination to include 24–59-month-old children.

Introduction

Seasonal influenza epidemics are responsible for over 200,000 hospitalizations in the United States per year, and 39,000 of them are in children.[1,2] In the United States, the Advisory Committee on Immunization Practices guides immunization practices, including influenza vaccination, with recommendations revised on an annual basis. For the 2006–2007 flu season, the Advisory Committee on Immunization Practices recommendations for influenza vaccination began including healthy children aged 24–59 months (two to four years), a shift that added 10.6 million children to the target group.[3]

Canada has a parallel federal organization, the National Advisory Committee on Immunization, which is responsible for guiding the use of vaccines. Recommendations made by the National Advisory Committee on Immunization and the Advisory Committee on Immunization Practices around seasonal influenza vaccination was concordant until the 2006–2007 season. Starting in the 2010–2011 season, the National Advisory Committee on Immunization has further expanded its recommendations to additional pediatric age groups by including two- to four-year-olds for targeted seasonal influenza vaccination.

We took advantage of this divergence in policy between two neighboring countries with similar annual seasonal influenza epidemics to try to understand the effects of the policy change in the United States to expand influenza vaccination coverage to other pediatric populations.

Methods

We analyzed emergency department (ED) data captured by Automated Epidemiologic Geotemporal Integrated Surveillance System (AEGIS), the population health monitoring system for the Massachusetts Department of Public Health (South Street, MA, USA), from Children's Hospital Boston (Boston, MA, USA) for the 2000–2001 to 2008–2009 influenza seasons. Comparable data was extracted from the electronic records of Montreal Children's Hospital (Montreal, Quebec, Canada) to make an international comparison of rates of ED visits for influenza-like illness (ILI) following the United States vaccination policy change. Adjusting for virological factors that impact variation in influenza epidemic intensity, seasonal trends, and all-cause ED utilization, we looked for early evidence of an effect of this policy change by evaluating the post-policy change response of ILI-related ED visits to Children's Hospital Boston, using Montreal Children's Hospital as a control, for four different pediatric age groups.

Results

We observed a 34% decline in the rate of ILI-related ED visits among two- to four-year-olds for Children's Hospital Boston relative to Montreal Children's Hospital following the 2006 policy change (RR ratio 0.66; 95% confidence interval: 0.59–0.75) (Table 1). We also observed smaller, marginally significant declines of 11–17% for the other age groups studied (Table 1).

Conclusions

Preschool-aged children have been identified as important to the household and community spread of influenza, due in

Table 1 Post-/pre-immunization policy change ratios of ILI-related ED visits to Children's Hospital Boston relative to Montreal Children's Hospital for four different pediatric age groups

Age group	*Adjusted rate ratio (95% CI)*	P
0–1	0.89 (0.81, 0.99)	0.03
2–4	0.66 (0.59, 0.75)	<0.001
5–9	0.83 (0.70, 0.98)	0.03
10–18	0.83 (0.70, 0.98)	0.03

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

part to infection rates averaging 25–43% per season, higher than any other age group.[4] We observed a significant decline in ILI-related ED visits by two- to four-year-old children at Children's Hospital Boston relative to a control hospital, Montreal Children's Hospital, following the 2006–2007 divergence in pediatric influenza immunization policy. Interestingly, we also saw smaller decline in rates among other pediatric age groups, a finding that is consistent with previous evidence that vaccinating preschool-aged children also reduces population-wide morbidity due to influenza.[5]

Acknowledgements

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

References

1 Henrickson KJ, Hoover S, Kehl KS, Hua W. National disease burden of respiratory viruses detected in children by polymerase chain reaction. *Pediatr Infect Dis J* 2004;**23:**S11–18.
2 Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, *et al.* Influenza-associated hospitalizations in the United States. *JAMA* 2004;**292:**1333–40.
3 Advisory Committee on Immunization Practices Smith NM, Bresee JS, Shay DK, Uyeki TM, Cox NJ, Strikas RA. Prevention and Control of Influenza: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep* 2006;**55:**1–42.
4 Fox JP, Cooney MK, Hall CE, Foy HM. Influenza virus infections in Seattle families, 1975–1979. II. Pattern of infection in invaded households and relation of age and prior antibody to occurrence of infection and related illness. *Am J Epidemiol* 1982;**116:**228–42.
5 Brownstein JS, Kleinman KP, Mandl KD. Identifying pediatric age groups for influenza vaccination using a real-time regional surveillance system. *Am J Epidemiol* 2005;**162:**686–93.

Emerging Health Threats Journal 2011, **4**:s86. **doi:** 10.3134/ehtj.10.086
© 2011 RA Hora *et al*.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Role of continuous monitoring on performance of surveillance systems in military populations in resource-limited settings

RA Hora[1], DM Vera[1], ZJ Silvera[2], and JM Montgomery[1]

[1]*US Naval Medical Research Center Detachment in Lima, Callao, Peru; and* [2]*Directorate of Health, Peru Air Force, Lima, Peru*
E-mail: ricardo.hora@med.navy.mil

Objective

The objective of this paper is to describe the effect of close monitoring on performance of the electronic disease surveillance system of the Peru Air Force.

Introduction

Timeliness of information has a key role in disease reporting, and may be easily impaired by several factors affecting data entry and utilization.[1] Regarding data entry, previous studies have shown that monitoring strategies, such as telephone reminders and supervision visits ensure reporting timeliness.[2] Likewise, limited reporting infrastructure may prevent adequate reporting and effective data utilization.[3,4] The Peruvian Air Force, in collaboration with the US Naval Medical Research Center Detachment in Lima, Peru, implemented in 2009 an electronic disease surveillance system with the objective of establishing near real-time baseline estimates of disease trends, and detecting disease outbreaks in a timely manner. This system has proven to perform well, although reporting sites vary in their reporting infrastructure. Therefore, we attempted to test the effect of a lack of monitoring on the performance of reporting sites, and explore the influence of other factors potentially affecting timeliness.

Methods

We analyzed the reporting performance of participating units at each epidemiological week (EW) from January to August 2010. We selected a six-week period when monitoring was to be voluntarily withheld for the purpose of our study objectives. Monitoring was resumed fully after six weeks. We compared the report-on-time rates of this six-week period of no reporting with those of the previous six-week period during which monitoring was conducted normally. Additionally, we contacted each key personnel at participating sites and asked them to complete an on-line anonymous questionnaire (www.surveygizmo.com) to explore other factors affecting performance, such as end-user's perception of the system, access to reporting media, and overall workload at establishments.

Results

From January to August 2010, all 31 active sites participated in the study. The report-on-time rate dropped from 100% (EW 17–22) to 42% (EW 23–29) (Mann–Whitney, $P<0.01$). (Figure 1) Our exploratory analysis of other factors affecting timeliness showed that of the 37 (100%) participants who completed the questionnaire, 30 (81%) were female. The median of age was 35 years (range, 22–55). Most participants had less than five years of service ($n=14$, 38%); and 65% of participants were nurses and/or technicians. The overall preferred medium of reporting was the internet (65%), although access to telephone was greater than access to the internet (60 versus 40%, respectively). The majority of participants showed satisfaction with being part of the system (95%), and considered reporting to be as important as their normal routine duties within the site (95%). In addition, 84% of participants considered that reporting did not prevent them from carrying out their normal duties.

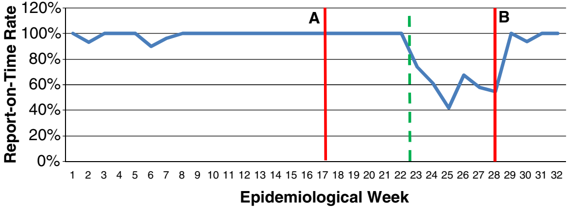


Figure 1 ROT rates during 2010. Red lines, A and B, point at the start and end dates of the study period, respectively. The green dotted line indicates the start of the six-week period at which monitoring was withheld.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

Despite availability of reporting infrastructure, and high acceptability of the system among stakeholders, the need for monitoring of reporting activities remains essential in ensuring timely reporting rates. Other factors possibly explaining this contradiction need to be further characterized.

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 Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health* 2004;**4**:29.
2 Huaman MA, Araujo-Castillo RV, Soto G, Neyra JM, Quispe JA, Fernandez MF, *et al.* Impact of two interventions on timeliness and data quality […]. *BMC Med Inform Decis Mak* 2009; **9**:16.
3 Free MJ. Achieving appropriate design and widespread use of health care technologies in the developing world […]. *International Journal of Gynecology & Obstetrics* 2004;**85** (Supplement 1): S3–13.
4 WHO. A safer future: global public health security in the 21st century. In: Hopkins D (ed). *The World Health Report 2007*. World Health Organization: Geneva, Switzerland, 2007.

Emerging Health Threats Journal 2011, **4**:s87. **doi:** 10.3134/ehtj.10.087
© 2011 TA Kass-Hout *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

BioSense program redesign

TA Kass-Hout[1], BL Massoudi[2], L Rojas-Smith[2], SC Kaydos-Daniels[2], JS Brownstein[3], DL Buckeridge[4], and J Buehler[1]

[1]*Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA;* [2]*RTI International, Durham, NC, USA;* [3]*Epidemico, Boston, MA, USA; and* [4]*Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada*
E-mail: tik2@cdc.gov

Objective

The objective of this study is that the Centers for Disease Control and Prevention (CDC) will update the International Society for Disease Surveillance community on the latest activities for the BioSense program redesign (Centers for Disease Control and Prevention, USA).[1]

Introduction

The BioSense program's mission is to support and improve public health surveillance infrastructure and human capacity required to monitor (with minimal lag) critical population health indicators of the scope and severity of acute health threats to the public health; and support national, state, and local responses to those threats. This mission is consistent with the 2006 Pandemic All Hazards Preparedness Act (PAHPA), and 2007 Homeland Security Presidential Directive (HSPD-21), both of which call for regional and nationwide public health situational awareness, through an interoperable network of systems, built on existing state and local situational awareness capability.

Methods

With input and guidance from our local, state, and federal surveillance partners, the new Office of Surveillance, Epidemiology and Laboratory Services (OSELS/CDC) is redesigning the BioSense program. The goal of the redesign is a new BioSense that coordinates and links existing surveillance systems to enable rapid and enhanced interchange of information.

Results

BioSense retains its original purpose of early event (or threat) detection and characterization but will expand its utility for: (1) raising public health situation awareness, (2) improving routine public health practice, (3) improving health outcomes and public health, and (4) monitoring health care quality. BioSense, as an all-hazards and timely electronic surveillance system, will provide a regional (that is, multi-state) and national view of multiple health outcomes and syndromes. By integrating local and state-level data into a cohesive 'picture,' BioSense will improve its utility for state and local users.

The user requirements gathering process identified gaps in biosurveillance practices and systems that BioSense can directly address in the redesign, which will result in more effective and timely public health surveillance at the local, state, and national levels (Figure 1).

Conclusions

Activities for the BioSense program redesign process include: (1) an assessment of the business and work flow needs at the local, state, and federal levels; (2) creation and support of existing communities of practice, (3) update and realignment of the BioSense Strategic Plan to complement and

Figure 1 The BioSense Redesign Collaboration Site (https://sites.google.com/site/biosenseredesign).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

strengthen existing surveillance systems and meet its enhanced and broadened mission; and (4) establishment of a governance structure that will allow public health stakeholders to drive the new direction for BioSense. The new charter and governance structure is supported by a joint steering committee (or the Technical Expert Panel) with a balanced representation of local and state stakeholders and CDC. Under this new governance structure, CDC and its partners will work collaboratively to resolve problems and establish design requirements as set forth by the public health community.

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.

Reference

1 Kass-Hout TA. BioSense Program: Going Forward. [Internet] *Eighth Annual International Society for Disease Surveillance (ISDS) Conference*, Miami, FL, USA, 2009. Available from: http://syndromic.org/conference/2009/presentations/plenaries/KassHout.pdf.

Emerging Health Threats Journal 2011, **4**:s88. **doi:** 10.3134/ehtj.10.088

© 2011 G Johnson *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

The utility of inpatient data obtained from regional health information organizations for pneumonia and influenza surveillance

G Johnson[1], X Jin[2], J-H Chen[2], H-G Chang[2], and D Morse[1]

[1]*Office of Science, New York State Department of Health, Albany, NY, USA; and* [2]*Division of Epidemiology, New York State Department of Health, Albany, NY, USA*
E-mail: gsb03@health.state.ny.us

Objective

The objective of this paper is to assess the potential utility of inpatient minimum biosurveillance data set data obtained from Regional Health Information Organizations (RHIOs) for pneumonia and influenza (P and I) surveillance.

Introduction

Hospital discharge data received by public health agencies has a reporting lag time of greater than six months. This data is often used retrospectively to conduct surveillance to assess severity of illness and outcome, and for evaluating performance of public health surveillance systems.

With the emergence of Health Information Exchanges and RHIOs, inpatient data can be available to public health in near real-time. However, there currently are no established public health practices or information systems for conducting routine surveillance in the inpatient setting.

Through a contract with the Centers for Disease Control and Prevention, New York State Department of Health initiated the development of a statewide public–health Health Information Exchanges with New York RHIOs. Daily minimum biosurveillance data set data-exchange implementation, and evaluation efforts were focused on one RHIO (RHIO A) and one participating hospital system composed of five acute-care facilities.

Methods

Inpatient data for visit dates between 1 June and 29 October 2009 were selected for analysis. The lag times between the availability of patient diagnosis data at RHIO A and the corresponding admission and discharge dates were calculated. Data obtained from RHIO A were compared with electronic hospital discharge data Statewide Planning and Research Cooperative System (SPARCS) admission diagnosis, primary discharge diagnosis, and 14 secondary diagnoses to assess the completeness and concordance of RHIO A data. Patients were matched by facility, date of visit, gender, month and year of birth, and zip code of residence. To evaluate the utility of daily RHIO A data for inpatient P and I surveillance, text and International Classification of Diseases, Ninth Revision (ICD-9) code filters for P and I were applied to RHIO A diagnosis and reason-for-admission fields and compared with an ICD-9 code filter applied to the SPARCS diagnoses.

Results

Reason for visit was 12% complete within one day, 68% within seven days, and 98% within 30 days of admission. Diagnosis was 15% complete within one day, 73% within seven days, and 96% within 30 days of discharge.

There were 26,514 matched patients, who accounted for 99% of the SPARCS patients and 94% of RHIO A patients. Diagnosis code was present for 24,897 (94%) RHIO A patients. Of the remaining 1617 patients, 1536 had at least one RHIO A free-text diagnosis and 81 had no RHIO A diagnosis. Unfiltered RHIO A–ICD-9 diagnosis was an exact match to one or more SPARCS diagnoses for 23,380 (94%) patients.

When the P and I filters were applied to the diagnoses fields, a total of 1685 (6.4%) patients met the criteria. Of those, 1601 (95%) had a RHIO A diagnosis code that matched at least one SPARCS–ICD-9 code, 29 (1.7%) had an uncoded RHIO A diagnosis that met the text filter criteria and at least one SPARCS diagnosis that met the ICD-9 filter criteria, 55 (3.4%) had non-matching RHIO A and SPARCS diagnoses.

When the P and I filters were applied to RHIO A reason-for-admission, 214 (0.8%) met the criteria. A total of 200 (93.5%) patients matched at least one SPARCS–ICD-9 filtered diagnosis. Fourteen (6.5%) patients did not have a matching ICD-9 filtered SPARCS diagnosis.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

Data for almost all patients reported through SPARCS was available through RHIO A and in a much more timely fashion. Both P and I filtered and unfiltered coded RHIO A diagnosis and SPARCS–ICD-9 diagnosis data matched relatively well. Adding P and I text filters to RHIO A diagnosis did not significantly improve pneumonia or influenza case ascertainment. It was not possible to assess the validity of the non-matching diagnoses in any data set. The number of patients meeting the filter criteria in was significantly lower when using RHIO A reason-for-admission compared with using RHIO A diagnosis data. Additional analysis is underway to assess the timeliness and concordance of P and I outcome. Inpatient data received through RHIOs could be a valuable resource to conduct P and I surveillance in the inpatient setting.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s89. **doi:** 10.3134/ehtj.10.089

© 2011 A Kite-Powell *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Florida's ESSENCE system—from syndromic surveillance to routine epidemiologic analysis across syndromic and non-syndromic data sources

A Kite-Powell[1], J Hamilton[1], R Wojcik[2], W Loschen[2], and R Hopkins[1]

[1] *Florida Department of Health, Bureau of Epidemiology, Tallahassee, FL, USA; and* [2] *Johns Hopkins University Applied Physics Laboratory, Columbia, MD, USA*
E-mail: Aaron_Kite-Powell@doh.state.fl.us

Objective

Use of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) in Florida has evolved from early event detection based on emergency department (ED) chief complaints to routine descriptive epidemiologic analysis, data visualization, and reporting across four different data sources, using and building on tools originally developed for syndromic surveillance.

Introduction

Federal laws and national directives have focused attention on the development of more robust biosurveillance systems intended to detect events of public health interest in a timelier manner. Presidential Decision Directive 21 calls for integrated biosurveillance data, enhanced clinician awareness, and an epidemiologic surveillance system with sufficient flexibility to tailor analyses to new syndromes and emerging diseases. In 2007, a statewide syndromic surveillance system (ESSENCE) was implemented and hospitals were recruited to participate. Experience with ESSENCE in the context of the ED data analysis, visualization, and reporting prompted the exploration of integrating new data sources into ESSENCE and new analyses specific to these new data. The purpose of the ESSENCE system is now to provide an intuitive environment for state and local epidemiologists to conduct routine descriptive epidemiologic analysis, to monitor morbidity and mortality trends over time and space and across multiple data sources, thereby providing information that can assist with making decisions on how to improve population health.

Methods

The Florida ESSENCE system is a multi-tiered, automated, and secured (https) web-based application. Microsoft SQL (One Microsoft Way, Redmond, WA, USA) databases are used in data ingestion, statistical analysis, and web databases. System development is primarily in Java (Oracle Corporation, Redwood Shores, CA, USA). It can be accessed via username and password by state-level epidemiologists, county health department epidemiologists, and staff at hospitals that participate. Enhancements to the system are developed through close collaboration with the Johns Hopkins University Applied Physics Laboratory, paying close attention to feedback from state and local epidemiologists in Florida.

Results

The system in Florida now includes four different data sources, each with its own module: (1) ED record data from 145 hospitals (updated once daily); (2) Merlin reportable disease case data from the Merlin system (updated once an hour); (3) Florida Poison Information Network consultation data (updated every 20 min); and (4) Florida Office of Vital Statistics death data (updated once daily). All records have individual identifiers removed.

Users can produce time-series graphs, maps, and tables by a number of possible stratification variables per data source. Several data sources can be combined in the same view. Time-series analysis can be conducted on daily, weekly, monthly, quarterly, and annual time resolutions. Statewide pneumonia and influenza mortality from the Office of Vital Statistics death data feed to ESSENCE is displayed as weekly counts and a weekly proportional mortality, and uses a Serfling regression model for analysis. Death certificates can be analyzed on the basis of analysis of literal text with a five-week delay and on assigned International Classification of Diseases and Related Health Problems 10th revision (ICD-10) codes with a 12-week delay. ED data are analyzed by sub-day time intervals to

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

identify clusters by the time patients are registered at the ED by hospital and sub-syndrome category. The Merlin reportable disease data are now processed through ESSENCE's spatial scan statistic to assist with detecting spatial clusters of public health significance.

Conclusions

A number of barriers often exist for users attempting to access public health data. Data are sometimes siloed within organizations, and often very specialized skill sets are needed to manipulate and analyze data in statistical and mapping software programs. The integration of four data sources into the Florida ESSENCE system creates one location, where users can go to access data and create epidemiologic reports and 'data dashboards' that describe the morbidity and mortality for a given region in Florida. This functionality has allowed for more timely and routine analysis of data that was previously not available, or required significant effort to access, and has helped close the public health surveillance loop with our partners in the clinical community.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s135. **doi:** 10.3134/ehtj.10.135
© 2011 K Kniss *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Comparison of Distribute and ILINet for national influenza surveillance

K Kniss[1,2], S Benoit[2], A McIntyre[2], D Buckeridge[3], J Brownstein[3], L Brammer[2], and L Finelli[2]

[1]*Atlanta Research and Education Foundation, Atlanta, GA, USA;* [2]*Centers for Disease Control and Prevention, Influenza Division, Atlanta, GA, USA; and* [3]*International Society for Disease Surveillance, Boston, MA, USA*
E-mail: krk9@cdc.gov

Objective

To compare ILI data reported to the Distribute surveillance project to data from an existing influenza surveillance system, the US Outpatient Influenza-like Illness Network (ILINet).

Introduction

During the spring of 2009, a public health emergency was declared in response to the emergence of the 2009 Influenza A (H1N1) virus. Owing to the response, timely data were needed to improve situational awareness and to inform public health officials. Traditional influenza surveillance is time-consuming and resource intensive, and electronic data sources are often more timely and resource saving. Collaboration began between the Centers for Disease Control and Prevention (CDC), the International Society for Disease Surveillance (ISDS), and the Public Health Informatics Institute (PHII) to expand syndromic Emergency Department (ED) surveillance through the Distribute project.

Distribute collects aggregate, daily or weekly reports of influenza-like illness (ILI) and total patient visits to EDs from participating health jurisdictions, stratified by age group and other variables. Additional variables included the three digit zip code of the patient's residence as well as the disposition and temperature, however not all jurisdictions collect these variables.[1] Distribute data are typically extracted from ED-based electronic health data systems. The ILI definition is determined by the participating jurisdiction that can be a city, county, or state. At the time of analysis, the network consisted of 33 jurisdictions.

Because ILI data reported to Distribute had not been systematically compared with data reported through other surveillance systems, CDC planned an evaluation of the Distribute data, which included a comparison to ILINet.

ILINet is a collaborative effort between the CDC, local and state health departments and primary health care providers. The network currently consists of approximately 3000 healthcare providers in all 50 states, Chicago, the District of Columbia, New York City, and the US Virgin Islands. Enrolled providers send CDC weekly reports via internet or fax that consist of the total number of patients seen for any reason and the number of those patients with ILI by age group. ILI is defined as fever (temperature of $\geqslant 100^\circ$F ($37.8\,^\circ$C)) and a cough and/or sore throat in the absence of a known cause other than influenza.[2]

Methods

Data from Distribute and ILINet were obtained from 2 September 2007 through 31 October 2009 for all sites where data were available. The weekly percent of visits because of ILI were calculated for each system. For state-based Distribute jurisdictions comparisons were made using ILINet data from the same state. For local jurisdictions, a geographic area was defined using the appropriate city or county boundaries. ILINet facilities that fell into the defined boundaries were used for comparison. Comparisons were made using correlation coefficients as well as visually.

Results

For state-based jurisdictions the correlation coefficients ranged from 0.64 to 0.96, with a mean of 0.83. For local jurisdictions, the correlation coefficients ranged from 0.38 to 0.91 with a mean of 0.76. For all state jurisdictions and all but one local jurisdiction, peaks in ILI tracked similarly in both systems For areas known to have the same EDs contributing data to both systems, ED data were removed from the ILINet data, and comparisons were made again. Correlations among these sites changed very little when the ED data were removed from the ILINet data.

Conclusions

Overall, the correlation between ILI data reported to the two systems was high. These results show that similar ILI data can be obtained using either electronic ED-based syndromic surveillance or traditional, provider-

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

based syndromic surveillance, and could be potentially useful in tracking and describing influenza activity at the national level. Differences in data collected through the two systems could be because of differences in the populations under surveillance, or potentially because of differences in illness seen in emergency departments compared with those seen in traditional primary care. Periodic evaluation of these systems should be conducted. Efforts to expand the Distribute network to a national network should be made.

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 Distribute background at: http://www.isdsdistribute.org/faq.php.
2 Influenza Surveillance Background: http://www.cdc.gov/flu/weekly/fluactivity.htm.

Emerging Health Threats Journal 2011, **4**:s90. **doi:** 10.3134/ehtj.10.090

© 2011 E Lamb *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Emergency department diagnosis code data for surveillance of vaccine adverse events: comparison with the national vaccine adverse event reporting system

E Lamb[1], H Vaughan-Batten[1], NJM Dailey[1,2], J-M Maillard[1], L Johns[1], AT Fleischauer[1,2], Z Moore[1], and M Davies[1]

[1]*Division of Public Health, North Carolina DHHS, Raleigh, NC, USA; and* [2]*Center for Disease Control and Prevention, Atlanta, GA, USA*
E-mail: emilie.lamb@dhhs.nc.gov

Objective

The objectives of this study were: (1) to compare trends in vaccine adverse events identified through emergency department (ED) diagnosis codes and reports from the Vaccine Adverse Event Reporting System (VAERS), and (2) to determine whether 2009 H1N1 vaccine adverse events identified through VAERS could also be identified using ED diagnosis codes.

Introduction

Nationally, vaccine safety is monitored through several systems including Vaccine Adverse Event Reporting System (VAERS), a passive reporting system designed to detect potential vaccine safety concerns.[1] Healthcare providers are encouraged to report adverse events after vaccination to VAERS, whether or not they believe that the vaccine caused the adverse event.[1] The 2009 Pandemic H1N1 influenza vaccine became available in the United States in October 2009. By January 2010, Center for Disease Control and Prevention (Atlanta, GA, USA) estimated that 61 million persons across the United States had received the vaccine.[2] As of January 2010, an estimated 28% of the North Carolina population greater than or equal to six months of age had been vaccinated against 2009 H1N1.[3]

Methods

Vaccine adverse events among North Carolina residents with symptom onset during 1 January 2008–31 December 2009 were identified using VAERS reports and emergency department (ED) diagnosis codes. The ED diagnosis codes for bacterial and other vaccines causing adverse effects in therapeutic use (ICD-9-CM codes E948–E949.9) were obtained from the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) that collects data from 99% of EDs statewide.[4] We used Pearson's correlation coefficient to compare trends in the weekly number of VAERS vaccine adverse event reports with events identified using ED diagnosis codes. We identified adverse events from 2009 H1N1 vaccination during 1 October 2009–31 January 2010 using VAERS reports, and attempted to match reports that indicated that the patient had visited either an ED, or physician's office to ED visits with diagnosis codes possibly related to influenza vaccination (E949.6, E949.7, and E949.9). Events were matched by age, sex, date of birth, county of residence, and vaccine administration.

Results

We identified 1793 vaccine adverse event reports using VAERS and 782 events through ED diagnosis codes among North Carolina residents with symptom onset or visit date during 1 January 2008–31 December 2009. We detected a moderate temporal correlation between vaccine adverse events identified from ED diagnosis codes and VAERS reports ($r = 0.47413$) Figure 1. Of two hundred and seventy two 2009 H1N1 vaccine adverse event reports sent to VAERS regarding North Carolina residents with onset during 1 October 2009–31 January 2010, 100 indicated that the patient visited the ED or a physician's office. Of these, only 8% could be matched to cases identified by ED diagnosis codes.

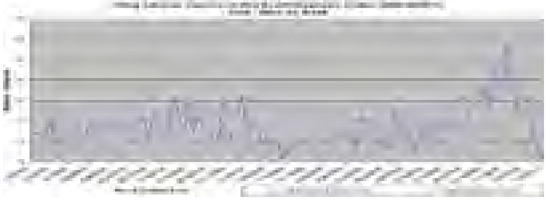


Figure 1 Weekly number of VAERS reports and NC DETECT–ED visits with symptom onset or visit dates during 1 January 2008–31 December 2009.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusion

Temporal trends in ED visits for vaccine adverse events correlate moderately well with trends in VAERS reports. However, the small number of 2009 H1N1 vaccine adverse event, VAERS reports, identified by ED diagnosis codes indicate that different patient populations or types of events are captured by these systems. Further prospective study is required to determine if ED diagnosis code surveillance could prove useful for monitoring trends in vaccine adverse events.

Acknowledgements

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

References

1 Varricchio F, Iskander J, Destefano F, Ball R, Pless R, Braun MM, *et al*. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. *Pediatr Infect Dis J* 2004;**23**(4): 287–9.

2 CDC. Interim Results: Influenza A (H1N1) 2009 Monovalent Vaccination Coverage—United States, October–December 2009. *MMWR* 2010;**59** (No. 2): 44–8.

3 CDC. Interim Results: State-Specific Influenza A (H1N1) 2009 Monovalent Vaccination Coverage—United States, October 2009–January 2010. *MMWR* 2010;**59** (No. 12): 363–638.

4 Waller A, Hakenewerth A, Tintinalli J, Ising A. North Carolina Emergency Department Data, January 1, 2007-December 31, 2007. *N C Med J* 2010;**71**(1): 15–25.

Emerging Health Threats Journal 2011, **4**:s92. **doi:** 10.3134/ehtj.10.092

© 2011 LA Lenert.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Meaningful use and public health surveillance: to travel fast or far?

LA Lenert

Department of Medicine and Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA
E-mail: leslie.lenert@hsc.utah.edu

Objective

The objective of this paper is to review the limitations of current approaches to linkage of public health through meaningful use reporting requirements and to explore alternatives based on integration of public health data reporting requirements, with clinical quality improvement reporting requirements.

Introduction

There is an ancient African proverb that states, 'If you want to travel fast, travel alone—if you want to travel far, travel together.' This paper examines the issue of whether public health can and should 'go it alone' in efforts for creating linkages between clinical care systems and the public health sector, as part of meaningful use requirements. 'Going it alone' in this circumstances refers to whether public health should seek to require data flows, through meaningful use requirements, that meet its work flow needs but do not add value to clinical work flows. An alternative would be to look for synergies between public health goals and the goals of the clinical care system, which public health could exploit to achieve its ends through collaborative means.

Methods

Efforts to create meaningful use requirements are reviewed through the lens of social competition between public health interests seeking more data and more access, healthcare providers seeking to minimize costs and to prevent exposure from loss of confidentiality, and federal regulators seeking workable accommodations that move the state of the programs forward, balancing the needs of society.

Results

Although numerous proposals for submission of data to public health entities were discussed during a series of meetings held by the National Committee of Vital Statistics in Spring of 2009, only two types of data sharing were specified under the final rule for healthcare providers—sharing with immunization registries or submission of syndromic data to health departments, where such capacity exists. Rules for hospitals allow a choice between any of the three tasks: the former two, plus the option of submitting reportable laboratory results to public agencies. The result is an unfavorable one for public health agencies. Agencies need to provide the infrastructure to support all three types of submission, but providers may chose the type of data most convenient for them to submit. Further, because data types submitted by providers are self-selected, the value of the data of public health is diminished. Although the rule allows states to 'require more' data submission, incentives will be paid to all providers who submit one type of data, regardless of any additional requirements (that is, no enforcement.) The conclusion: public health lost this first round of negotiations on integration. It has the most requirements, received only minimal funding from HITECH to support its infrastructure needs to receive data, and has only limited ability to influence the choice that providers make in the type of data they submit through state legislation.

Consider, in contrast, how efforts to enhance population health through quality reporting fared. By 2112, the regulations will require denominator-based reporting on a core set of quality measures pertaining to blood pressure control, smoking cessation, and obesity (or on alternate core measures that include childhood immunization rates). There is also an additional list of 38 other potential quality measures that can be selected for reporting—these are substantiative requirements and everyone must participate.

What if public health chose to align its push for data from meaningful use with these quality measures? For example, smoking cessation quality indicators could be transformed into data on the prevalence of smoking in the practice and on the incidence of cessation interventions? Might not this data, which would use the same mechanisms and infrastructure used to produce quality reporting, be as valuable for public health, from a policy perspective, as data feeds to traditional surveillance systems? The value of the data could

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

be further enhanced by use of geocoding strategies, such as Geographic Interoperable Patient Summary Exchange formatting, which would allow combination of data across practices and views of the health of small areas and into health disparities.

Conclusions

In addition to the well-known (and experienced) financial limitations that public health faces, in social systems where public health interests compete with those of other sectors, there are limits on the political capital that public has as well. Applying that capital in a way that is synergistic, with other interests in the healthcare sector, may produce better long-term results than going against the interests of the sector. Given the alignment between quality of care measures and population level surveillance of chronic diseases, public health may travel farther 'together' alone in pursuit of traditional surveillance measures.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s97. **doi:** 10.3134/ehtj.10.097
© 2011 C Lucero *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Surveillance for nationally notifiable infectious conditions using ICD-9-CM diagnosis codes in the VA ESSENCE biosurveillance system

C Lucero, P Schirmer, G Oda, and M Holodniy

Department of Veterans Affairs, Office of Public Health Surveillance and Research, Palo Alto, CA, USA
E-mail: cynthia.lucero@va.gov

Objective

To determine the utility of ICD-9-CM diagnosis codes in the Veterans Affairs (VA) Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) for detection and public health surveillance of nationally notifiable infectious conditions in veteran patients.

Introduction

ESSENCE obtains electronic data from 153 VA Medical Centers plus outpatient clinics in all 50 states, American Samoa, Guam, Philippines, Puerto Rico, and U.S. Virgin Islands. Currently, there is no centralized VA reporting requirement for nationally notifiable infectious conditions detected in VA facilities. Surveillance and reporting of cases to local public health authorities are performed manually by VA Infection Preventionists (IP) and other clinicians. In this analysis, we examined positive predictive value (PPV) of ICD-9-CM diagnosis codes in VA ESSENCE to determine the utility of this system in electronic detection of reportable conditions in VA.

Methods

VA ESSENCE analyzes ICD-9-CM diagnosis codes and demographic data from outpatient and emergency department visits at all VA facilities.[1] For this review, visits with an ICD-9-CM code(s) for a reportable communicable disease during 2009 were selected. For diseases in which 10 or fewer potential cases were identified, we expanded the date range and selected all potential cases from July 2005 to May 2010. For diseases in which more than 100 potential cases were identified, a minimum of 20% of records were selected for review. Laboratory and clinical data from electronic medical records were reviewed by a VA epidemiologist for concordance and case classification based on the most recent CDC/CSTE case definitions.[2] PPV for each disease was calculated as number of patients who met case definition criteria divided by total number assigned an ICD-9-CM diagnosis code for the disease. For some diseases, numerators included both confirmed and probable/ suspected cases based on CDC definitions.

Results

Disease-specific PPV proportions ranged from 0 to 72% with the overall proportion 25.7% (224/871). Of the 30 reportable diseases reviewed thus far, 13 diseases had PPVs of 0%. Of the 30 diseases, 15 had PPVs above 10% and 4 diseases (cryptosporidiosis, listeriosis, spotted fever rickettsiosis, and salmonellosis) had PPVs greater than 50%. Reasons that visits did not meet CDC case definitions included: (1) Patient had a history of the disease but was not acutely infected; (2) Miscoding for vaccine administration or vaccine reaction; (3) Miscoding of similar sounding disease; (4) Patient was initially diagnosed or treated at another facility, so testing results not available in VA records; (5) Diagnosis was in the differential or a clinical diagnosis only with no confirmatory testing performed; (6) Miscoding for prophylaxis given or exposure to a disease but patient was not acutely ill; (7) Miscoding for antibody titer evaluations.

Conclusions

PPVs for the majority of reportable communicable diseases captured by current VA ESSENCE were low, and therefore ESSENCE is of limited value in detecting new cases. Our results are consistent with a similar analysis performed in North Carolina.[3] For a few diseases, ESSENCE had relatively high PPVs (greater than 50%). For these diseases, VA ESSENCE may be an added tool for more timely and possibly for additional case detection beyond what is currently captured through manual surveillance. Future work will evaluate system sensitivity by comparing reportable diseases captured by VA ESSENCE with those identified by IP manual surveillance as a gold standard. Efforts to enhance PPV by combining ICD-9-CM codes with additional data elements (for example, laboratory results) are also underway. Accurate, automated, electronic reporting of reportable diseases to VA Office of Public Health will be

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

important in understanding disease trends within Veteran populations, ensuring timely reporting and facilitating information sharing with local, state, and federal public health partners.

Acknowledgements

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

References

1 Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, *et al.* A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics. *J Urban Health* 2003;**80** (2 Suppl 1): i32–42.
2 CDC. Nationally Notifiable Infectious Conditions, United States 2010. Available at: http://www.cdc.gov/ncphi/disss/nndss/phs/infdis2010.htm.
3 Sickbert-Bennett EE, Bischoff T, Cleve W, Long R. Utility of ICD-9-CM Diagnosis Codes for Public Health Surveillance. *Advances in Disease Surveillance* 2008;**5**:62.

Emerging Health Threats Journal 2011, **4**:s98. **doi:** 10.3134/ehtj.10.098

© 2011 TM Luse *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Using laboratory and medical encounter records to identify impact of prophylaxis policies on group A Streptococcus in three recruit training facilities

TM Luse, K Shannon, and G Nowak

EpiData Center Division, Navy and Marine Corps Public Health Center, Portsmouth, VA, USA
E-mail: tina.luse@med.navy.mil

Objective

To compare trends of group A beta-hemolytic Streptococcus among recruits before and after changes in prophylaxis implementation using electronic laboratory and medical encounter records.

Introduction

Group A beta-hemolytic Streptococcus (GABHS) has caused outbreaks in recruit training environments, where it leads to significant morbidity and, on occasion, has been linked to deaths.[1,2] Streptococcal surveillance has long been a part of military recruit public health activities. All Navy and Marine Corps training sites are required to track and record positive throat cultures and rapid tests on weekly basis. The Navy and Marine Corps have used bicillin prophylaxis as an effective control measure against GABHS outbreaks at recruit training sites.[3] Though streptococcal control program policies vary by site, a minimum prophylaxis protocol is required and mass prophylax is indicated when local GABHS rates exceed a specific threshold. Before July 2007, prophylaxis upon initial entry was required between October and March, and when the local rate exceeded 10 cases per 1000 recruits. In July 2007, the Navy instituted a policy of mass prophylaxis upon initial entry throughout the year. Evaluation of GABHS cases before and after implementation of the new policy, including overall rates, identification of outbreaks, and inpatient results will help enhance the Navy's ability to evaluate threshold levels, provide systematic/standardized monitoring across the three recruit sites, and inform prophylaxis and monitoring strategies.

Methods

Positive GABHS laboratory results were identified from the Health Level 7 chemistry and microbiology databases from 1 May 2004 through 31 December 2009 for all Navy and Marine Corps recruits. Laboratory results included cultures and rapid strep tests. Positive laboratory test results were matched on specimen collection dates to inpatient and outpatient clinical encounter records using ICD-9 codes. The top diagnoses among GABHS cases were identified. Using these most frequent diagnoses associated with strep positive labs, inpatient and outpatient encounter records during the study timeframe were identified for all recruits with GABHS-like illness, regardless of laboratory test result. Weekly GABHS rates were calculated for laboratory positive cases and GABHS-like diagnosis for each recruit training site. Weekly trends were compared with prophylaxis regimens for each training site.

Results

There were 5576 laboratory-identified cases of GABHS in Navy and Marine Corps recruits between May 2004 and December 2009. Almost half of all cases (47%, $n = 2596$) occurred at one of the three recruit training sites. Comparisons of illness rates before and after implementation of the new policy showed a substantial decrease in rates at all three training facilities. In 2008 and 2009, rates of illness each week rarely exceeded two cases per 1000 individuals and never exceeded four cases per 1000 individuals. Large peaks in cases were identified often before 2007, but have not occurred since 2008. A total of 200,914 outpatient records were identified with GABHS-like illnesses. Close to 60 percent of all outpatient encounters occurred at one of the three training sites ($n = 118{,}043$). Volume of cases was substantially lower after implementation of prophylaxis policies.

Conclusion

Identifying cases of GABHS using electronic laboratory and medical encounter records can assist in determining whether prophylaxis policies are effective. Evaluation of the data has shown a substantial decrease in rates of cases among recruits after the implementation of the new policy in 2007.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

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

References

1 Crum NF, Hale BR, Judd SE, Lim ML, Wallace MR. A case series of group A Streptococcus necrotizing fasciitis among military trainees. *Mil Med* 2004;**169**:373–5.
2 Centers for Disease Control and Prevention. Outbreak of group A streptococcal pneumonia among marine corps recruits—California, November 1—December 20, 2002.
3 Russell KL. *Recruit Medicine*. Borden Institute, Walter Reed Medical Center: US Government Printing Office, Washington, DC, 2006.

Emerging Health Threats Journal 2011, **4**:s139. **doi:** 10.3134/ehtj.10.139
© 2011 DW Mazeroski and T Arthur-Hartranft.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

EpiSentry™: simulation-based threat identification, response management and decision support for outbreaks and pandemics

DW Mazeroski, and T Arthur-Hartranft

Lockheed Martin Corporation, IS&GS, Lansdale, PA, USA
E-mail: david.w.mazeroski@lmco.com

Objective

Traditional real time surveillance systems such as RODS and ESSENCE have focused on the task of threat detection; however, experience with the use of these systems in pandemic and disaster response settings suggests that a more common application is threat characterization and response management. This paper describes EpiSentry™: a novel second generation real-time surveillance software system under development at Lockheed Martin that uses simulation to aid in threat characterization, response management and to provide decision support for disease outbreaks or bio-terror events.

Introduction

The 2009 H1N1 novel flu pandemic demonstrates how a rapidly spreading, contagious illness can affect the world's population in multiple ways including health, economics, education, transportation, and national security. Pandemic disease and the threat of bio-terrorism are prompting the need for a system that integrates disparate data, makes optimal use of the breadth of available health-related analysis and predictive models, and provides timely guidance to decision makers at multiple levels of responsibility.

Methods

Lockheed Martin Corporation is currently developing the EpiSentry™ rapid decision environment. EpiSentry™ integrates epidemiological situational assessment and predictive analysis tools with a decision support system to provide guidance on pandemic and bio-terror investigations (Figure 1).

The project is using an agile development methodology to address the technical challenges of integration of surveillance, simulation and planning capabilities in a series of progressively more advanced software solutions. Challenges include:

- Ability to extrapolate relevant information from non-traditional data sources before data from traditional surveillance streams may even be available.
- Estimation (from both traditional and non-traditional surveillance data) of population-level parameters, with modeling of disease spread and impacts of countermeasures.
- Development of workflow tools for collaborative investigation and response planning.

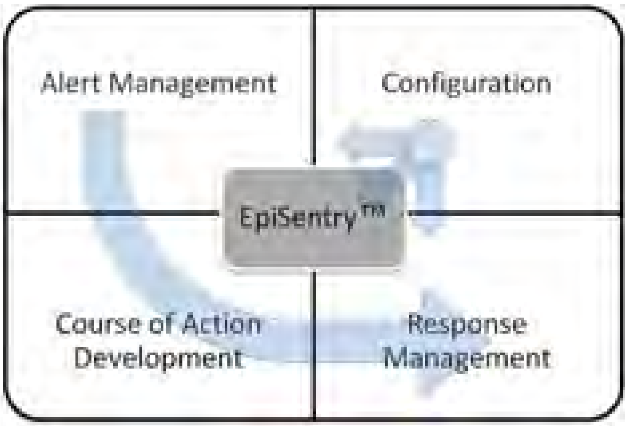


Figure 1 The EpiSentry™ framework provides an integrated configurable alert management, course of action development, and response management capability.

As part of its solution, EpiSentry™ employs geospatial visualization and data fusion techniques, integrates open source monitoring and trending solutions, develops an enhanced disease mobility module, and leverages a decision analytical framework to provide the user with evidence projections of the risks and benefits of different action plans.

Results

Although still in development, we believe EpiSentry™ will significantly advance contemporary response management. A system built around EpiSentry™ will be part of a bold strategy to meet information needs for the management of

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

outbreaks. It will provide an integrated, end-to-end, and collaborative solution that enables the use of advanced algorithms and simulation software to monitor and identify outbreaks from disparate data sources, plan intervention strategies, and monitor the effectiveness of the chosen course of action.

Conclusions

EpiSentry™ is a new type of surveillance system that combines data analysis and outbreak modeling, and response-planning software systems. Although promising, further work is needed to define the benefits of this integrated approach.

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.

Emerging Health Threats Journal 2011, **4**:s99. **doi:** 10.3134/ehtj.10.099
© 2011 AF McIntyre *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Quantifying the relationship between influenza-related emergency department visits and hospital admissions in BioSense

AF McIntyre[1], L Finelli[1], HS Burkom[2], SR Benoit[1], KL Kniss[1], and TA Kass-Hout[1]

[1]*Centers for Disease Control and Prevention, Atlanta, GA, USA; and* [2]*Johns Hopkins University Applied Physics Laboratory, Baltimore, MD, USA*
E-mail: AMcIntyre@cdc.gov

Objective

The purpose of this analysis is to gain understanding of the burden of influenza in recent years through analysis of clinically rich hospital data. Patterns of visits and severity measures such as the ratio of admissions related to influenza-like illness (ILI) by age group from 2007 to 2010 are described.

Introduction

Real-time emergency department (ED) data from the BioSense surveillance program for ILI visits and ILI admissions provide valuable insight into disease severity that bridges gaps in traditional influenza surveillance systems that monitor ILI in outpatient settings and laboratory-confirmed hospitalization, but do not quantify the relationship between ILI visits and hospital admissions.

Methods

All patients with ILI presenting to EDs participating in BioSense ($n = 650$ hospitals) during September 2007 through June 2010 were included in this analysis. ILI visits were defined as encounters with an influenza diagnosis code

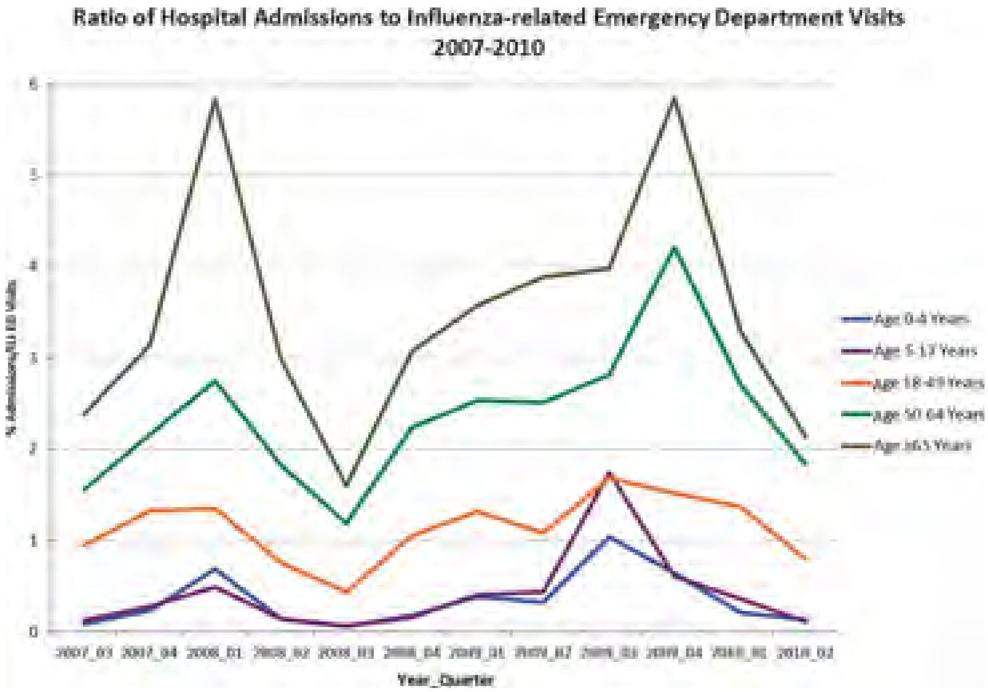


Figure 1 Ratio of hospital admissions to influenza-related emergency department visits.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(ICD-9-CM codes 487 and 488) or an ILI-related chief complaint ('influenza' or fever and cough/upper respiratory infection). Age groups were defined as 0–4 years, 5–17 years, 18–49 years, 50–64 years, and ⩾65 years. Those with unknown age or age >105 years were excluded. To account for age and seasonal differences, frequencies and mean were stratified by age group and by calendar year quarter to calculate the proportion of ILI-related ED visits resulting in hospitalization. Analysis of variance was performed to assess differences in mean hospitalization encounters by age group and calendar year quarter.

Results

There were 1,925,539 patients who sought care for ILI at EDs reporting to BioSense during September 2007 through June 2010. Among those with ILI, 59,294 (3.1%) were hospitalized. The admission to ILI ratio was similar ($P=0.58$) by quarter with 2.92% (range 2.65–3.14%) of visits admitted in Quarter 1, 3.00% (range 2.79–3.32%) in Quarter 2, 3.94% (range 2.86–5.33%) in Quarter 3, and 3.33% (range 2.18–4.84%) in Quarter 4. The mean admission to ILI ratio was highest in those aged ⩾65 years (M = 3.48%, s.d. = 1.31), followed by 50–64 years (M = 2.36%, s.d. = 0.77), 18–49 years (M = 1.34%, s.d. = 0.36), 5–17 years (M = 0.41, s.d. = 0.46), and 0–4 years (M = 0.34%, s.d. = 0.30), although there were no significant differences by age group. By quarter and age group, the highest admissions occurred during Quarter 1 among those aged ⩾65 years (M = 4.24%, s.d. = 1.40), 18–49 years (M = 1.35%, s.d. = 0.03), and 0–4 years (M = 0.43%, s.d. = 0.24), during Quarter 3 for 5–17 years (M = 0.64%, s.d. = 0.96), and during Quarter 4 for 50–64 years (M = 2.87%, s.d. = 1.16) (Figure 1).

Conclusions

Data on the relationship between hospitalizations and ILI ED visits provide an objective method to describe the burden of disease and to compare severity between influenza seasons, as seen in the leading peak for the 5- to 17-year group in the pandemic spring wave. Further analyses are underway to assess the degree to which (1) the 2009 H1N1 pandemic impacted these results and (2) comorbid conditions affected susceptibility to and exacerbation of influenza in different age groups.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s100. **doi:** 10.3134/ehtj.10.100
© 2011 J-B Meynard *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Evaluation of the NATO Disease Surveillance System in Kosovo in 2010

J-B Meynard[1], H Chaudet[2], L Pellegrin[2], Y Moreno[2], O Deboosere[2], P Mornand[2], K Pennin[3], S Stein[4], B Queyriaux[2], H-U Holterm[5], A Zuendorf[3], J Kocik[3], R Migliani[1], X Deparis[2], and G Texier[2]

[1]*Ecole du Val-de-Grâce, Paris, France;* [2]*Institut de Médecine Tropicale du Service de santé des armées, Marseille, France;* [3]*NATO Consultation, Command and Control Agency, The Hague, The Netherlands;* [4]*KFOR Headquarters, Pristina, Kosovo; and* [5]*Saniaetskommando I, Preventive Medicine Department, Kiel, Germany*
E-mail: jb.meynard@wanadoo.fr

Objective

The new NATO Disease Surveillance System (DSS) was deployed for the second time in Kosovo within the multinational armed forces in 2010 for a 3 days experiment. The objective of the survey was to continue the development of real-time disease surveillance capability for NATO forces, in parallel with the implementation of the NATO Deployment Health Surveillance Centre in Munich in 2010.

Introduction

The 2010 NATO DSS experiment was the second deployment of the French 'Alerte et Surveillance en Temps Réel' (ASTER) system within a multinational armed task force in real operational conditions. This experiment was scheduled within the ASTER evaluation program (Figure 1), as constructed by French and NATO Armed Forces after several previous works.[1–3]

Methods

The DSS was set up in May 2010 within the medical facilities of six NATO allied nations taking part in the multinational armed forces task force in Kosovo (KFOR): Austria, Czech Republic, France, Germany, Poland, and United States of America. Each nation received one laptop loaded with software that allowed medical data to be recorded. The recorded data was sent to a national data collection server at KFOR headquarters in Pristina. A permanent communication link was used to send the data simultaneously to the analysis centre level in Munich, where a multinational team was deployed. Real medical data were used, but also simulated data within the Polish medical facility, in order to assess the ability of the analysis level to detect and to manage a natural Norovirus outbreak. Three other types of surveys were realized: a knowledge–attitude–practice survey within the stakeholders, a study of social networks, and a Human Factor study (using Lewis, NASA, and SART questionnaires).

Results

The Norovirus outbreak was detected in real-time and adapted measures proposed by the analysis level (individual and collective hygiene measures, medical treatment, and adaptation of activities). In all, 42 people took part in the three studies (28 in Kosovo and 14 in Munich). The analysis of the results is currently in progress and will be presented during the congress.

Conclusions

This evaluation is an intermediary one, it should be strengthened by a final evaluation, promoting continuous improvement, before generalization for all the military NATO deployed units.

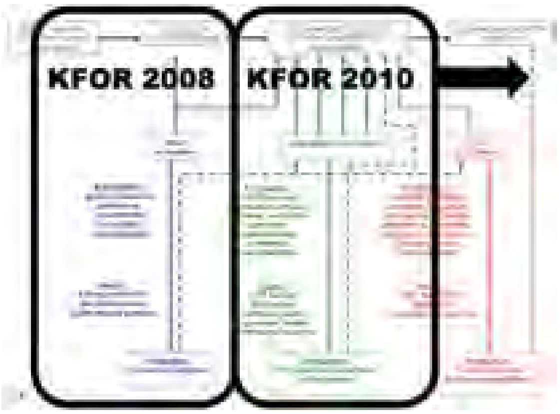


Figure 1 Intermediary evaluation for NATO 'Disease Surveillance System' deployed in Kosovo in November 2010 as a part of the whole NATO evaluation process.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 Meynard J-B, Chaudet H, Green Ad, Jefferson H, Texier G, Webber D, *et al*. Proposal of a framework for evaluating military surveillance systems for early detection of outbreaks on duty areas. *BMC Public Health* 2008;**8**:146.
2 Jefferson H, Dupuy B, Chaudet H, Texier G, Green Ad, De Lesquen M, *et al*. Evaluation of a syndromic surveillance for the early detection of outbreaks among military personnel in a tropical country. *J Public Health* 2008;**30**:375–83.
3 Meynard J-B, Chaudet H, Varcol C, Decker E, Andresen S, Holterm H-U, *et al*. Evaluation of the NATO Disease Surveillance System by its users in Kosovo in 2008. *Mil Med* 2010;**175**:466–8.

Emerging Health Threats Journal 2011, **4**:s104. **doi:** 10.3134/ehtj.10.104
© 2011 J Morin *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Creating a fast and flexible syndromic surveillance reporting system

J Morin[1,2], A Ali[3], D Ferguson[2], RG Gutierrez[4], A Riley[4], N Vinson[2], S McClinton[1], C McDermaid[3], K Ali[5], J Martin[2], and RF Davies[1]

[1]*University of Ottawa Heart Institute, Ottawa, Ontario, Canada;* [2]*National Research Council of Canada Institute for Information Technology, Ottawa, Ontario, Canada;* [3]*Ottawa Public Health (OPH), Ottawa, Ontario, Canada;* [4]*StataCorp LP, College Station, TX, USA; and* [5]*AMITA Corporation, Ottawa, Ontario, Canada*
E-mail: jason.morin@nrc.gc.ca

Objective

The objective of this study was to create and evaluate a system that uses customized scripts developed for commercial off-the-shelf (COTS) statistical and GIS software to (1) analyze syndromic data and produce regular reports to public health epidemiologists, containing the information they would need to detect and manage an ILI outbreak, and (2) facilitate the generation more detailed analyses relevant to specific situations using these data.

Introduction

Syndromic surveillance systems significantly enhance the ability of Public Health Units to identify, quantify, and respond to disease outbreaks. Existing systems provide excellent classification, identification, and alerting functions, but are limited in the range of statistical and mapping analyses that can be done. Currently available COTS statistical and GIS packages provide a much broader range of analytical and visualization tools, as well as the capacity for automation through user-friendly scripting languages. This study retrospectively evaluates the use of these packages for surveillance using syndromic data collected in Ottawa during the 2009 pH1NI outbreak.

Methods

Four Ottawa area hospitals have been reporting chief complaint and demographic data to Ottawa Public Health using a RODS-based syndromic surveillance system (ASSET) since January 2009. (ASSET is based on the University of Pittsburgh's Real-time Outbreak and Disease Surveillance system (http://www.rods.pitt.edu)). During the 2009 pH1N1 pandemic, a team comprised of epidemiologists, health care researchers, and NRC HCI specialists recognized the need for an automated syndromic reporting system that would free epidemiologists for other tasks. The team designed specific reports by developing customized scripts using Stata and ArcGIS. Data from ASSET were parsed using an ILI classifier (NRC) and processed using these scripts.

Results

The resulting ILI Watch report automatically produces epicurves, exploratory data analyses, aberration detection graphs, and color-coded maps that replace daily reports, which previously had to be generated manually. The system also makes the syndromic data set easily available to epidemiologists, and facilitates follow-up analyses. The system is currently in production in the Ottawa area (Figure 1).

Conclusions

An automated system developed using state-of-the-art COTS software can reduce the time and resource burdens on epidemiologists and IT staff during an outbreak by

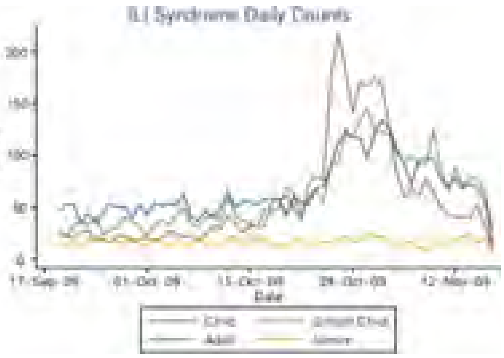


Figure 1 Sample data from the ASSET ILI Watch. Data is categorized by age group, and demonstrates both the high prevalence of ILI in children in general and school-aged children in particular, and the nominal infection rates seen in people aged ⩾60 years.

This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

automatically producing customized reports. It also provides the flexibility needed to rapidly meet changing information needs by providing a user-friendly scripting template for ad hoc analyses.

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.

Emerging Health Threats Journal 2011, **4**:s105. **doi:** 10.3134/ehtj.10.105
© 2011 E Westheimer *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Comparison of respiratory, febrile and influenza-like illness syndromes to detect laboratory-reported H1N1 and RSV, Influenza Season 2009–10, New York City

E Westheimer, TQ Nguyen, M Paladini, D Weiss, and S Balter

New York City Department of Health and Mental Hygiene, New York, NY, USA
E-mail: tnguyen@health.nyc.gov

Objective

To determine the correlation of the influenza-like illness (ILI) syndrome with laboratory-confirmed H1N1 and respiratory syncytial virus (RSV) during the October 2009 to March 2010 H1N1 season in New York City (NYC).

Introduction

The NYC Department of Health and Mental Hygiene (DOHMH) monitors visits daily from 49 of 54 NYC emergency departments (EDs), capturing 95% of all ED visits. ED visits for ILI have reflected influenza activity in NYC,[1] better than the more broadly defined fever/flu (FF) and respiratory (Resp) syndromes, but the correlation with H1N1 is unknown.

Laboratory-confirmed influenza and RSV were made reportable in NYC in February 2008. DOHMH receives electronic reports of positive tests.

As part of 2009–10 influenza surveillance, five hospitals were selected for 'sentinel' surveillance of hospitalized influenza cases, to test all patients with a respiratory condition for influenza. Sentinel hospitals ensured that patient medical record numbers (MRNs) were in the daily ED syndromic file and in the electronic laboratory reports.

Methods

Sentinel hospital ED visits were matched by MRN to corresponding cases of influenza and RSV reported electronically for 1 October 2009 through 31 March 2010. Trends were assessed citywide, across sentinel hospitals, and by hospital and age groups. Correlation coefficients were calculated.

Results

Between 1 October 2009 and 31 March 2010, the five sentinel hospitals represented 13.6% of 1,939,417 citywide ED visits. Of

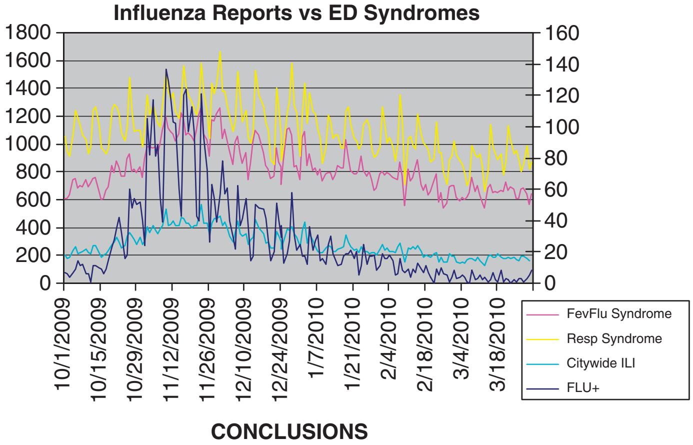


Figure 1 Weekly citywide ILI ED visits vs citywide laboratory reports of influenza and RSV, October 2009–March 2010, New York City.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the sentinel ED visits, 24,100 (9%) were categorized as FF, 29,859 (11%) were Resp, and 4996 (2%) were ILI; 21% of FF visits and 10% of Resp visits also met the ILI definition.

During the same period, the sentinel hospitals reported 550 (11%) of 5137 unique laboratory reports of influenza and 1348 (27%) of 5030 RSV reports. Citywide, ILI correlated best with influenza reports (0.82), followed by FF (0.78) and Resp (0.70) (see Figure 1).

With the sentinel data, 499 influenza reports and 1114 RSV reports were matched by MRN to an ED visit. The majority of influenza reports matched to the FF (54%) and Resp (23%) syndromes; 14% were ILI. RSV cases matched similarly to FF (41%), Resp (40%), and ILI (10%).

Conclusions

Although influenza cases were classified primarily as FF or Resp, the correlation of ILI with influenza was higher than the other syndromes. Therefore, ILI still best reflects influenza trends but not the magnitude of illness.

ILI trends enable citywide monitoring of influenza, but variations by age group and hospital catchment suggest that fine-tuning syndromes to better represent influenza and RSV activity is possible. Data matching was limited to only five hospitals, but demonstrates that matching ED visits to laboratory reports could enhance syndrome definitions and modeling.

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.

Reference

1 Olson DR, Heffernan RT, Paladini M, Konty K, Weiss D, Mostashari F. Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. *PLoS Medicine* 2007;**4**:1349–61.

Emerging Health Threats Journal 2011, **4**:s144. **doi:** 10.3134/ehtj.10.144
© 2011 G Rainisch and K Hayashi.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Early Aberration Reporting System (EARS) update: present and future

G Rainisch, and K Hayashi

Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA
E-mail: buu8@cdc.gov

Objective

The objective of this poster is to highlight recent upgrades to the Early Aberration Reporting System (EARS) software (US Centers for Disease Control and Prevention, EARS Program, MS C-18, Atlanta, GA, USA), and identify features planned for future releases.

Introduction

Early Aberration Reporting System (EARS) is a freeware surveillance tool that can be downloaded from the Center for Disease Control and Prevention's website (http://emergency.cdc.gov/surveillance/ears/). It was designed for quick set-up and customization for automated monitoring of emergency department and other syndromic data sources, including, but not limited to, 911 calls, school absenteeism, and over-the-counter medication sales. The United States' city, county, state health departments, and various international public health organizations, use EARS software to conduct daily, near-real time surveillance of conditions easily defined by patient-reported complaints, and physician diagnoses (for example, influenza-like illness, gastroenteritis, asthma, heat-related illness). It is also used to conduct suspect case finding during outbreaks, natural disaster responses, verify that potential threats are not manifested in communities, and for supporting *ad hoc* analyses and research.

Rate And Stratification

Do you want to calculate a rate (i.e. adjust counts by a denominator)?

○ No ● Yes

Specify how the denominator variable is summed

● One total per day

○ One total for each distinct value per day

Figure 1 Screen example of the EARS-SAS version5.0 graphical user interface.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 2 Screen example of the EARS-SAS version5.0 interactive graphs output.

Methods

EARS is undergoing significant changes focused on reducing processing times and expanding user-defined customization options. Installation and set-up have been simplified through the use of a 'wizard-like' graphical user interface that significantly restricts user-generated errors (Figure 1). Anomaly detection algorithms are being modified to increase both sensitivity and specificity (without increasing the alert rate) through use of a minimum standard deviation, a longer baseline/comparison period, an adjustment for total visits (or other denominator), and an option for stratifying expected value calculations by days of the week.[1] Output filtering options, based on both empirical and epidemiological criteria, are being added to allow the user to further manage the alert rate (Figure 2).

Results

Future enhancements include writing EARS in other languages to support an increasing international user-base. Rewrites will be on the basis of open-source tools that will not require Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) or SAS (Institute Inc., Cary, NC, USA) to run. Optional plug-ins will enable users to format and automate submission of aggregate-level data summaries to other systems (for example, DiSTRIBuTE, http://www.isdsdistribute.org/). The EARS program plans to also explore increasing EARS compatibility with varying data types, such as diagnoses and laboratory data through data-specific modules (based on the design of the current EARS Indicator Processing feature).

Conclusions

The potential utility and access to electronic health records, and syndromic data sources are expected to continue to expand. The EARS tool continues to improve to keep up with this demand. Collaboration on EARS design, with state and local public health departments, is a crucial component to its continued successful development.

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.

Reference

1 Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, *et al.* Enhancing time-series detection algorithms for automated biosurveillance. *Emerg Infect Dis* 2009;**15**:4.

Emerging Health Threats Journal 2011, **4**:s108. **doi:** 10.3134/ehtj.10.108
© 2011 D Rajeev *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Description of the quality of public health case reports received at a local health department and potential impact on workflow

D Rajeev[1], A Price[2], K Schnirel[2], M Hill[2], C Staes[1], and I Risk[2]

[1]*Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; and* [2]*Salt Lake Valley Health Department, Salt Lake City, UT, USA*
E-mail: deepthi.rajeev@hsc.utah.edu

Objective

This paper describes a comparison study conducted to identify quality of reportable disease case reports received at Salt Lake Valley health department (SLVHD) in 2009 and 2010.

Introduction

When a reportable condition is identified, clinicians and laboratories are required to report the case to public health authorities. These case reports help public health officials to make informed decisions and implement appropriate control measures to prevent the spread of disease. Incomplete or delayed case reports can result in new occurrences of disease that could have been prevented. To improve the disease reporting and surveillance processes, the Utah Department of Health is collaborating with Intermountain Healthcare and the University of Utah to electronically transmit case reports from healthcare facilities to public health entities using Health Level Seven v2.5, SNOMED CT, and LOINC.[1] As part of the Utah Center of Excellence in Public Health Informatics, we conducted an observation study in 2009 to identify metrics to evaluate the impact of electronic systems.[2] We collected baseline data in 2009 and in this paper we describe preliminary results from a follow-up study conducted in 2010.

Methods

We conducted two observation studies of the workflow associated with processing case reports at SLVHD, including from 6 July 2009 to 13 July 2009 and 7 July 2010 to 21 July 2010. The 2009 study occurred during the H1N1 outbreak. The studies involved direct observations of the workflow of the triage nurse at SLVHD. To ensure we were capturing the quality of the reports received initially at the health department, we used a data collection form to document whether certain core data elements such as ‘patient address’, ‘patient telephone number’, ‘hospitalization status’, ‘physician notes’ and so on, were missing. Currently, we are extracting data from the Utah statewide surveillance system (UT-NEDSS) to compute the ‘time to diagnosis of a case’, ‘reporting time delay’, ‘time to triage a report’, ‘time until the start of case investigation’ and so on,

Results

In 2009 ($n=380$ reports) and 2010 ($n=322$ reports), there were similar proportion of out-of-county reports (23 and 29%, respectively) and duplicate reports (19 and 20%, respectively). The quality of data in reports received in 2009 and in 2010 is described in Table 1. In 2009, Chlamydia, Influenza related cases, and Salmonella contributed to 68% of the reports and in 2010, Chlamydia, Giardia, and Salmonella represented 60% of the reports. The analysis of the timeliness of the reporting process is currently underway.

Discussion

The processing of out-of-county and duplicate reports continues to be a burden on the triage nurse. The hospitalization status and physician notes were more complete in 2009 than in 2010. Patient telephone number and patient address are generally required for all diseases but

Table 1 Comparison of quality of key data elements in reports received at SLVHD in 2009 and 2010

Data element	*% of (updated/new) reports with information included*	
	2009 (%)	**2010 (%)**
Patient telephone number	82	81
Patient address[a]	79	61
Hospitalization status[a]	86	20
Physician Notes[a]	82	8

[a]Significant difference at $\alpha=0.05$.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

hospitalization status and physician notes are not typically provided and/or required with the disease report. However, during the H1N1 outbreak these data were requested with the report to quickly identify routes of exposure to reduce spread. The results demonstrate the challenges and burden for public health to obtain additional data elements such as hospitalization status and physician notes.

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. Funding: (Grant# 1PO1HK000069–01) Rocky Mountain Center for Translational Research in Public Health Informatics, NLM Training grant # 5T15LM007124 (DR).

References

1 Rajeev D, Staes CJ, Evans RS, Mottice S, Rolfs R, Samore MH, *et al*. Development of an electronic public health case report using HL7 v2.5 to meet public health needs. *J Am Med Inform Assoc (JAMIA)* 2010;**17**:34–41.
2 Rajeev D, Zeller R, Price A, Reid J, Staes CJ, Risk I. *Evaluating the Impact of Electronic Disease Surveillance Systems on Local Health Department Work Processes*. Public Health Information Network (PHIN); Atlanta, GA, 2009.

Emerging Health Threats Journal 2011, **4**:s145. **doi:** 10.3134/ehtj.10.145

© 2011 M Rennick *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Asthma patterns in Boston emergency department visits for children age five and under

M Rennick[1], J Gunn[1], M Donovan[1], and J Salvia[2]

[1]*Boston EMS, Boston Public Health Commission, Boston, MA, USA; and* [2]*Communicable Disease Control Division, Raleigh, NC, USA*
E-mail: mrennick@bphc.org

Objective

The objective of this study is to report on the use of syndromic surveillance data to describe seasonal patterns of asthma and health inequities among Boston residents, age five and under.

Introduction

The burden of asthma on the youngest children in Boston is largely characterized through hospitalizations and self-report surveys.[1] Hospitalization rates are highest in Black and Hispanic populations under age five. A study of children living in Boston public housing showed significant risk factors, including obesity and pest infestation, with less than half of the study population being prescribed daily medication.[2]

Information on asthma visits for children ⩽5 years old was requested by the Boston Public Health Commission Community Initiatives Bureau. The information is being used to establish a baseline for an integrated Healthy Homes Program that includes pest management and lead abatement. There is limited experience in using syndromic surveillance data for chronic disease program planning.

Methods

The Boston syndromic surveillance system receives demographics, chief complaints, and disposition. To assess asthma visits in the emergency departments (ED) from 1 January 2008–19 June 2010, an asthma syndrome was built to catalog the following key words (and appropriate variations) from the chief complaint field: asthma, reactive airway disease, and wheezing. This was then restricted to all Boston residents, ages five and under, and compared with all ED visits of the same ages and residency. Syndrome counts were further stratified by race/ethnicity, geographic neighborhoods, gender, and disposition.

Boston Emergency Medical Service transports coded as likely asthma in the electronic trip sheet for Boston residents, ages five and under were also obtained.

Results

Asthma ED visits occurred in a seasonal pattern in Boston with highest levels of activity in September to early October in both 2008 and 2009 (max = 10.6% of all visits for week ending 20 September 2008), and was correlated with an increase in Emergency Medical Service transports ($R^2 = 0.23$). (Figure 1).

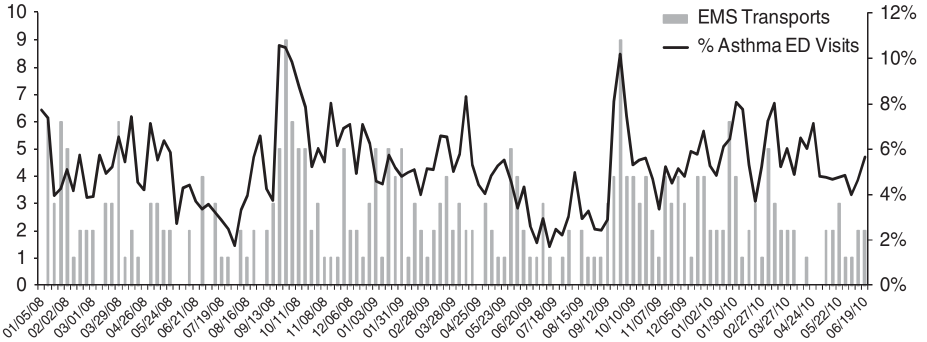


Figure 1 Weekly percent asthma of all Boston ED visits ⩽5 years old (line) overlayed with weekly count of EMS transports for asthma patients ⩽5 years old (bars), 1 January 2008–19 June 2010.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

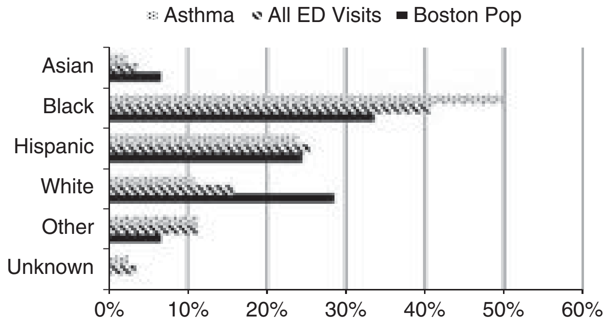


Figure 2 Race/ethnicity distribution of ED visits for asthma, and for all visits compared with Boston population ⩽5 years old, 1 January 2008–19 June 2010. Fifty percent of asthma patients are Black, 41% of all ED patients are Black, 34% of the age-specific Boston population are Black (2000 Census).

From 1 January 2008–19 June 2010, Black children were nearly twice as likely to present to an ED for an asthma syndrome visit compared with White children (odds ratio (OR) = 1.84 (1.65, 2.06)); Hispanic children were 40% more likely (OR = 1.40 (1.24, 1.57)) than White children to present at an ED for an asthma visit. Males accounted for 64% of all asthma syndrome visits compared with 54% for all ED visits. Between 29 November 2009 and 19 June 2010, 25% of the ED asthma syndrome visits required hospitalization. (Figure 2).

Conclusions

The Boston Public Health Commission syndromic surveillance system provided information on ED asthma syndrome visits for children ⩽5 years old. Seasonal patterns in asthma syndrome visits correlated with Emergency Medical Service data with peak activity in September. Environmental triggers including returning to school (either patients or siblings) may be associated with increased activity in September. A flexible chief complaint coding system with the availability of demographic information that included race/ethnicity was essential to providing timely baseline information for program planning.

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 Boston Public Health Commission Research and Evaluation Office. *Health of Boston 2009*. Boston Public Health Commission Research and Evaluation Office. Boston, MA, 2009.
2 Levy J, Welker-Hood LK, Clougherty JE, Dodson RE, Steinbach S, Hynes HP. Lung function, asthma symptoms, and quality of life in public housing in Boston: a case-series analysis. *Environ Health* 2004;**3**:13.

Emerging Health Threats Journal 2011, **4**:s146. **doi:** 10.3134/ehtj.10.146

© 2011 J Robertson and M Hill.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Evaluating the use of syndromic surveillance for the detection of influenza-like illness in Salt Lake County, Utah

J Robertson, and M Hill

Salt Lake Valley Health Department, Salt Lake City, UT, USA
E-mail: jrobertson@slco.org

Objective

The objective of this study is to compare the performance of syndromic surveillance with the United States Outpatient Influenza-like Illness Surveillance Network (ILINet), for the detection of influenza-like illness (ILI) during the fall 2009 wave of H1N1 influenza in Salt Lake County.

Introduction

Salt Lake Valley Health Department uses syndromic surveillance to monitor ILI activity as part of a comprehensive influenza surveillance program that includes pathogen-specific surveillance, sentinel surveillance, school absenteeism and pneumonia, and influenza mortality. During the 2009 spring and fall waves of novel H1N1 influenza, sentinel surveillance became increasingly burdensome for both community clinics and Salt Lake Valley Health Department, and an accurate, more efficient method for ILI surveillance was needed. One study found that syndromic surveillance performed, as well as a sentinel provider system in detecting an influenza outbreak[1] and syndromic surveillance is currently used to monitor regional ILI in the United States.[2]

Methods

Data were collected from 30 August 2009 to 26 December 2009. Weekly reported hospitalized H1N1 cases were summed from daily lab and provider reports. Daily sentinel provider ILI, defined as fever $\geqslant$100.4°F, cough and/or sore throat, and total patient visits were summed from 12 sentinel sites, and used to calculate weekly percentages of sentinel provider ILI. Daily text-based chief complaint data from 15 syndromic sites (emergency departments and urgent care centers) were obtained from EpiCenter[3] (funds provided by Utah Department of Health), mapped to an ILI disease category, defined as 'fever' and 'cough' or 'sore throat,' and summarized using the Early Aberration Reporting System (EARS).[4] Daily ILI and total patient visits were used to calculate weekly percentages of EARS/syndromic ILI. Spearman correlation analysis was performed to determine the relationships between EARS/syndromic ILI, sentinel provider ILI, and reported hospitalized H1N1 case counts for all ages and each ILI age group.

Results

Overall, weekly percentage of EARS/syndromic ILI strongly correlated with both weekly percentage of sentinel provider ILI ($r = 0.93$, $P < 0.0001$), and weekly reported hospitalized H1N1 case counts ($r = 0.97$, $P < 0.0001$) (Figure 1). Strong correlations of EARS/syndromic ILI, with sentinel provider ILI and H1N1 case counts were also evident in the 0–4, 5–24, 25–49, and 50–64 age groups.

Conclusions

These results suggest that syndromic surveillance is an accurate method for ILI surveillance. Salt Lake Valley Health

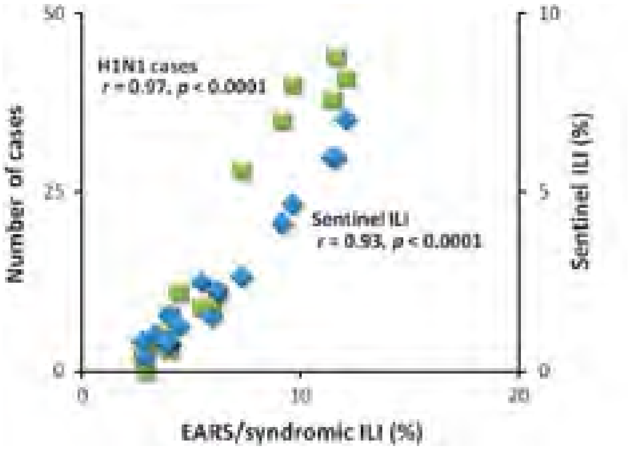


Figure 1 Correlation of EARS/syndromic % ILI with reported hospitalized H1N1 influenza cases (green), and sentinel % ILI (blue).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Department plans to collect ILI data from both syndromic and sentinel surveillance systems in future influenza seasons and establish an epidemic threshold percentage for syndromic ILI, with the eventual goal of replacing sentinel ILI surveillance with syndromic ILI surveillance.

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 Ritzwoller DP, Kleinman K, Palen T, Abrams A, Kaferly J, Yih W, *et al.* Comparison of syndromic surveillance and a sentinel provider system in detecting an influenza outbreak—Denver, Colorado, 2003. *MMWR Morb Mortal Wkly Rep* 2005;**54** (Suppl): 151–6.
2 International Society for Disease Surveillance (ISDS), DiSTRIBuTE project. http://isdsdistribute.org/index.php.
3 Health Monitoring Systems, Inc, EpiCenter, v 2.6.3. http://www.hmsinc.com/service/epicenter.html.
4 Centers for Disease Control and Prevention, Division of Preparedness and Emerging Infections, Early Aberration Reporting System v 45. http://www.bt.cdc.gov/surveillance/ears/.

Emerging Health Threats Journal 2011, **4**:s109. **doi:** 10.3134/ehtj.10.109

© 2011 C Rodriguez *et al*.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

School absenteeism and emergency department ILI rates in King County, WA 2003–2009

C Rodriguez[1,2], A Baer[1,2], J Baseman[2], I Painter[2], D Yanez[2], T Koepsell[2], and JS Duchin[1,2]

[1]*Public Health—Seattle & King County (PHSKC), Seattle, WA, USA; and* [2]*University of Washington, Seattle, WA, USA*
E-mail: carla.rodriguez@kingcounty.gov

Objective

To describe the relationship between emergency department (ED) visits for influenza-like-illness (ILI) and absenteeism among school-aged children.

Background

Absenteeism is regarded as an expedient and responsive marker of illness activity. It has been used as a health outcome measure for a wide spectrum of exposures and as an early indicator of influenza outbreaks.[1] A handful of studies have described its validity compared with traditional 'gold-standards' for influenza and ILI.[2,3] We sought to further quantify the relationship between ED ILI and school absenteeism because absenteeism, as it relates to illness, and subsequent loss in productivity and wages for parents, school staff and children, is an important public health outcome.

Methods

Data were drawn from Public Health—Seattle & King County's syndromic surveillance system and included the weekly number absent by school and the weekly number of ILI visits in the county. Eighteen of nineteen public school districts and 18 of 19 hospitals contributed data for school-aged children from 2003–2009. An ecological study design and generalized estimating equations for Poisson models were used to measure the marginal association between county-wide (log) ED ILI rates and school absenteeism over time, using school enrollment as an offset. Models included adjustment by school level, influenza season, (log) average baseline absenteeism count, pre- and post-holiday weeks, time and confounding by diarrheal illness. The 2003–04 and 2008–09 school years were analyzed separately because of distinct patterns in respiratory virology during those years. Sensitivity analyses for the number of ED visits per school, simulated using a normal distribution for a range of means, were conducted to test the reliability of our results. Signals, representing excess events, in both ED and absenteeism data streams were generated using the cumulative sum (CUSUM) method and compared with virology data from the University of Washington Clinical Virology Laboratory. Virology data were also used to define the start and peak of viral epidemics, including influenza.

Results

During seasonal influenza periods from 2003–08, a 3-fold increase in county-wide ED ILI rate among children aged 5–9 years was associated with 18% (95 CI: 16,20%) higher absenteeism in elementary schools. Associations were similar among elementary and middle school-aged children, and greater than those observed in high schools. During the spring 2009 H1N1 pandemic, a 3-fold increase in ED ILI rate was associated with 38% (95 CI: 33,43%) higher absenteeism in elementary schools. We generally observed only weak associations between absenteeism and ILI during non-influenza periods from 2003–08 and outside the pandemic period of 2009. Predicted absenteeism rates were less variable than ILI rates, but still sensitive to the rise and peak of ED ILI and influenza epidemics. ED ILI signals were better correlated with virology data for the start and peak of influenza season than predicted absenteeism.

Conclusions

To our knowledge, this is the first study to quantify the degree to which absenteeism increases when ED ILI activity increases over several viral epidemics. We also observed improved temporal correlation of ED ILI signals with influenza data over predicted absenteeism, supporting the use of ED ILI as a predictor for school absenteeism. Because data on ED visits were only available at the county-level, we were unable to control for potential confounding by ED visits at the school level. However, our results were robust to sensitivity analyses. This is a first step towards understanding the validity of school absenteeism to describe ILI activity. Future studies that better control for the effect of school-level ED visits on absenteeism and test the predictive ability

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of absenteeism for ILI are needed to verify and expand upon our results.

Acknowledgements

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

References

1 Weitzman M. School absence rates as outcome measures in studies of children with chronic illness. *J Chronic Dis* 1986;**39:** 799–808.
2 Glezen WP. Emerging infections: pandemic influenza. *Epidemiol Rev* 1996;**18:**64–76.
3 Quenel P, Dab W, Hannoun C, Cohen JM. Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics. *Int J Epidemiol* 1994;**23:** 849–55.

Emerging Health Threats Journal 2011, **4**:s110. **doi:** 10.3134/ehtj.10.110

© 2011 D Rumoro *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Impact of alternate diagnoses on the accuracy of influenza-like illness case definition used for H1N1 screening in the emergency department

D Rumoro[1], J Silva[1], S Shah[1], J Bayram[1], M Hallock[1], G Gibbs[1], M Waddell[2], and K Thomas[1]

[1]*Rush University Medical Center, Chicago, IL, USA; and* [2]*Pangaea Information Technologies, Chicago, IL, USA*
E-mail: dino_rumoro@rush.edu

Objective

This study investigates the impact of alternate diagnoses on the accuracy of the Centers for Disease Control and Prevention's (CDC) case definition for influenza-like illness (ILI) when used as a screening tool for influenza A (H1N1) virus during the 2009 pandemic, and the implications for public health surveillance.

Introduction

In June 2009, the CDC defined a confirmed case of H1N1 as a person with an ILI and laboratory confirmed novel influenza A H1N1 virus infection.[1] ILI is defined by the CDC as fever and cough and/or sore throat, in the absence of a known cause other than influenza.[1] ILI cases are usually reported without accounting for alternate diagnoses (that is, pneumonia). Therefore, evaluation is needed to determine the impact of alternate diagnoses on the accuracy of the ILI case definition.

Methods

This is a retrospective cross-sectional study design conducted from September 5, 2009 to May 5, 2010, at an emergency department of a large urban tertiary care academic medical center. During this period, 32,922 patients were seen, of which 1233 were tested for H1N1 using Polymerase Chain Reaction (PCR) testing for respiratory viruses. Geographic Utilization of Artificial Intelligence in Real-Time for Disease Identification and Notification (GUARDIAN), a syndrome surveillance program,[2] was utilized to assign ILI status to each patient.

Positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity were calculated with and without consideration of the following alternate diagnoses: pneumonia, respiratory syncytial virus, infectious mononucleosis, and streptococcal pharyngitis. Positive laboratory results, in addition to ICD-9 codes, were used to detect these alternate diagnoses. McNemar's and χ^2-test were then used to compare the results for statistical significance (Table 1).

Table 1 Positive and negative predictive value and sensitivity and specificity of the ILI case definition to detect H1N1 cases

ILI status		*H1N1 status*		*PPV (%)*	*NPV (%)*	*Sensitivity (%)*	*Specificity (%)*
		No	*Yes*				
Including alternate diagnoses	No	439	21	17.2	95.4	86.4	40.7
	Yes	640	133				
Excluding alternate diagnoses	No	638	35	21.3	94.8	77.3	59.1
	Yes	441	119				

Abbreviations: NPV, negative predictive value; PPV, positive predictive value. Based on the McNemar test for evaluation of classifiers, the improvement (especially in specificity) generated by ILI case definition excluding alternate diagnoses was statistically significant ($\chi^2=159$, $P<0.05$).

Results

Of the 1233 tested for H1N1, only 62.7% ($n=773$) had signs and symptoms consistent with ILI, before considering alternate diagnoses. This significantly decreased to 45.4% ($n=560$) ($\chi^2=74$, $P<0.001$) after extraction of cases with alternate diagnoses.

Conclusions

By excluding patients with alternate diagnoses, ILI case definition specificity was improved, without significant difference in other indices. Dual diagnoses of H1N1 and pneumonia were present in 15 patients, which complicated ILI status designation. In this study, these cases were considered ILI negative because of their alternate diagnoses. In addition, there was a 17.3% decrease in the overall ILI prevalence rate by excluding cases with an alternate diagnosis. This demonstrates how surveillance methodology may affect ILI rates reported by hospitals, and subsequently

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

may affect regional public health surveillance data, necessary for appropriate response.

Accurately identifying ILI patients with alternate diagnoses can be difficult because of lack of available data, specifically lab results, which can be delayed up to 48 h. By using GUARDIAN in this study, we were able to automatically and accurately identify ILI patients with other known causes, and increase our accuracy in identifying H1N1 cases. This study exemplifies the importance of an accurate and consistent clinical case definition for the diagnosis of H1N1, along with an automated real-time surveillance system.

Acknowledgements

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

References

1 Centers for Disease Control and Prevention. http://www.cdc.gov/.
2 Waddell MJ, Doseck SM, Silva JC, Rumoro DP. GUARDIAN: Geographic Utilization of Artificial intelligence in Real-Time for Disease Identification and Notification. *Adv in Dis Surveill* 2007;**4**:63.

Emerging Health Threats Journal 2011, **4**:s111. **doi:** 10.3134/ehtj.10.111
© 2011 E Samoff *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Access to and use of syndromic surveillance information at the local health department level

E Samoff[1], A Ising[1,2], A Waller[1,2], E Sickbert-Bennett[3], M Davis[1], M Park[1], SW Haas[4], L DiBiase[1], and PDM MacDonald[1,5]

[1]*North Carolina Preparedness and Emergency Response Research Center, Gillings School of Global Public Health, University of North Carolina at Chapel Hill (UNC-CH), Chapel Hill, NC, USA;* [2]*Carolina Center for Health Informatics, Department of Emergency Medicine, UNC-CH, Chapel Hill, NC, USA;* [3]*North Carolina Division of Public Health, Chapel Hill, NC, USA;* [4]*School for Information and Library Science, UNC-CH, Chapel Hill, NC, USA; and* [5]*Department of Epidemiology, UNC-CH, Chapel Hill, NC, USA*
E-mail: erika.samoff@unc.edu

Objective

To characterize use of syndromic surveillance information for key public health functions at the local health department level, and to make recommendations to facilitate use of syndromic surveillance data for these functions.

Introduction

Syndromic surveillance data have been widely shown to be useful to large health departments. Use at smaller local health departments (LHDs) has rarely been described, and the effectiveness of various methods of delivering syndromic surveillance data and information to smaller health departments is unknown. Syndromic surveillance data and information in North Carolina are available to all local public health staff by several routes. This report characterizes LHD access to syndromic surveillance data and information and their use for key public health purposes.

Methods

Structured interviews were conducted with local health directors and communicable disease nursing staff from a stratified random sample of LHDs during May through September 2009. The survey captured information on the use of the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) web application for direct data access and on the use of syndromic surveillance information for outbreak management, program management, and the creation of reports. Data on direct access to NC DETECT were available for all 85 LHDs in the state.

Results

Surveys were completed with 14 of 15 LHDs (93% response rate). Syndromic surveillance data were used by LHDs for outbreak management (two of 14 (15%)), program management (three of 14 (21%)), and creating reports for distribution to public health stakeholders and partners (four of 14 (26%)).

LHD staff obtained syndromic surveillance information from hospital-based public health epidemiologists (PHEs) and public health regional surveillance teams (PHRSTs), and syndromic surveillance data directly from NC DETECT. The information distributed by PHEs includes both reports and notification of specific events by telephone, and may include relevant information from patient–hospital records; 22 of 85 LHDs (26%) had access to a PHE. The information distributed by PHRSTs includes regularly distributed standardized reports. Direct access to NC DETECT allows LHDs to create and review aggregate data and to review single event details; direct access is voluntary. At the time of this survey, 29 of 85 LHDs (34%) had NC DETECT logins; staff at 10 of 85 LHDs (34%) accessed the system six or more times in the 6 months before the survey. Use of syndromic surveillance information for outbreak, program, and report uses was associated with receiving information from PHEs or PHRSTs.

Reasons given for not using syndromic surveillance information focused on the time or effort needed to obtain useful information from large NC DETECT data sets. Information from this survey and other North Carolina sources was gathered to support the design of an LHD web portal for NC DETECT. A draft version of this portal that attempts to automate some of the filtering role provided by PHEs and PHRSTs, will be presented.

Conclusions

Syndromic surveillance information is useful to the local level, as evidenced by use of syndromic surveillance information for outbreak, program, and report purposes. Syndromic surveillance information and data are available to

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

LHDs, and some LHD staff members directly access NC DETECT. However, functional use of this information for public health purposes occurs only where these data have been filtered and reviewed by public health personnel who are routine users of NC DETECT. These results show that distribution of syndromic surveillance information by state and regional staff is effective, and suggest that improvements that support more efficient filtering and interpretation of syndromic surveillance data, as well as provision of training, may increase use of this information at the local level.

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.

Emerging Health Threats Journal 2011, **4**:s147. **doi:** 10.3134/ehtj.10.147
© 2011 B Nichols *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Tracking H1N1 vaccine doses administered using CRA system

B Nichols[1], S Waite[2], H Hill[3], L Han-Lee[2], S Sapkota[2], and J Tropper[1]

[1]*Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA;* [2]*CDC Contractor from SRA International Inc., Fairfax, VA, USA; and* [3]*CDC Contractor from McKing Consulting, Fairfax, VA, USA*
E-mail: ssapkota@cdc.gov

Objective

The objective of this paper is to report the use of the Centers for Disease Control and Prevention's (CDC) Countermeasure and Response Administration System (CRA) to track and monitor H1N1 doses administered during the initial weeks of the 2009–2010 H1N1 Vaccine Program when supplies of the vaccines were limited, and before population-based surveys like Behavioral Risk Factor Surveillance Systems, and National H1N1 Flu Survey could effectively monitor vaccine coverage.[1]

Introduction

The novel strain of H1N1 Influenza A virus, which first caused localized outbreaks in parts of Mexico, was declared a pandemic in June 2009. CDC's CRA was used to track the H1N1 vaccine uptake across population age groups during the first eight weeks of the event (3 October to 21 November 2009). The CRA application was utilized to track vaccine doses administered in the initial period of H1N1 vaccine campaign, as there was no other method available to inform how well the vaccine was reaching target age groups.

Methods

Through several webinars and conference calls with vaccine-tracking coordinators, Project Areas were required to participate in reporting H1N1 vaccine doses-administered counts on a weekly basis. All 50 states, four major cities (Los Angeles, New York City, Chicago and Washington DC) and eight US islands and territories, also known as Project Areas, participated in the campaign. Reporting occurred along seven Advisory Committee on Immunization Practices (ACIP) recommended age groups: 6–23 months; 24–59 months; 5–18 years; 19–24 years; 25–49 years; 50–64 years; and 65 years and above. Each Project Area was asked to select an option from the following to transmit report/data to CDC: Option 1, data exchange; Option 2, web entry aggregate; Option 3, web entry detail.

Results

During the initial eight-week-period of the H1N1 vaccine response, from 3 October to 21 November 2009, a total of 14,788,795 vaccines were reported to CDC from 60 reporting Project Areas. Of the total doses reported, 35% of vaccines were administered to children aged 5–18 years, and over 83% were administered to persons less than 65 years. The average delay between administration of H1N1 vaccine and reporting of that data to CRA was from two to four weeks. Overall, by the final week of reporting (21 November 2009), Project Areas reported approximately 4.4% of H1N1 vaccine administered during the first eight weeks of the campaign using the CRA system.

Conclusions

CRA is a functioning informatics system for monitoring vaccine uptake in different age groups in the early stages of the H1N1 response when vaccine supply is inadequate. This vaccine reporting event demonstrated that CRA application is a useful system for national reporting of aggregate doses administered early on during a pandemic, providing an understanding of vaccine coverage before the traditional surveillance systems. Comparisons of data obtained from NHFS validated data from CRA application.[2] Improvement on individual Project Area automation would contribute significantly towards enhancing doses-administered reporting systems.

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 Tropper J, Shimabukuro T, Sapkota S, Williams W, Williams CE, Stanley T. CDC's Countermeasure and Response Administration system for monitoring H1N1 vaccine doses administered. *Disaster Med Public Health Prep* 2009;**2** (Suppl): S107–8.
2 CDC. Interim Results: Influenza A (H1N1) 2009 Monovalent Vaccination Coverage—United States, October—December 2009. *MMWR* 2010;**59(02)**:44–8.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s149. **doi:** 10.3134/ehtj.10.149
© 2011 REN Sembajwe *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

User requirements for a user-centered design (UCD) redesign of a public health surveillance system: BioSense

REN Sembajwe, TR Farris, OL Rojas-Smith, BL Massoudi, AK Banger, and A Casoglos

RTI International, Research Triangle Park, NC, USA
E-mail: rsembajwe@rti.org

Objective

The objective of this poster presentation is to provide information on the efforts and results of the user requirements gathering process, pursued in the redesign of the BioSense system employing a User-Centered Design (UCD) approach. A UCD system development approach studies the people that the system will serve, and involves them through the planning, design, and development processes.

Introduction

The Centers for Disease Control and Prevention (CDC) launched the BioSense program in 2003 to establish an integrated system of nationwide public health surveillance for the early detection and assessment of potential bioterrorism-related illness. Over the ensuing years, the original aims of BioSense were broadened to meet evolving public health surveillance needs and priorities.

In 2009, CDC embarked on a redesign of the BioSense system that would retain the original purpose of early event (or threat) detection and characterization, but enhance the capacity for situation awareness, event response, and routine public health practice. The BioSense redesign will be accomplished through continuous involvement of state and local users, and technical experts in the planning, design, development, and testing phases of the new BioSense system.

Methods

Information regarding state and local biosurveillance practices was gathered from end users and other stakeholders through user sessions convened at conferences, workshops, and *ad-hoc* meetings throughout 2009 and 2010. Participants in user sessions were primarily state and local health department stakeholders. Information on biosurveillance practices was also gathered from General Accounting Office (GAO) reports, 2009 BioSense evaluations, and Technical Expert Panel feedback. The information was grouped, analyzed, and weighted/ranked on the basis of.frequency, and then prioritized and translated into meaningful user requirements for the system redesign.

Results

The user requirements analysis identified challenges and suggestions in five key areas:[1] capacity (workforce presence, skill set, and funding resources),[2] governance (stakeholder engagement, design, and purpose of system),[3] policy (regulations and legislation),[4] standards and definitions (taxonomies, business rules, and case definitions),[5] data (sources, quality, and access), technological infrastructure, and application design (functionality and usability).

Conclusions

The user requirements gathering process identified gaps in biosurveillance practices and systems that BioSense can directly address in the redesign, which will result in more effective and timely public health surveillance at the local, state, and national levels.

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 Buehler JW, Whitney EA, Smith D, Prietula MJ, Stanton SH, Isakov AP. Situational uses of syndromic surveillance. *Biosecur Bioterror* 2009;**2**:166–77.
2 Lesesne DB, Leese PJ, Rojas-Smith OL, Rein DB. Characteristics and geographic dispersion of syndromic surveillance in the United States in 2008, 2009. Unpublished manuscript.
3 Uscher-Pines l, Farrell CL, Cattani J, Hsieh YH, Moskal MD, Babin SM. A survey of usage protocols of syndromic surveillance

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

systems by state public health departments in the United States. *J Public Health Manag Pract* 2009;**5**:432–8.

4 Usability Professionals' Association, Promoting usability concepts and techniques worldwide: What is user-centered design? Available at: http://www.upassoc.org/usability_resources/about_usability/what_is_ucd.html. Accessed September 1, 2010.

5 U.S. General Accounting Office. Health information technology: More detailed plans needed for the Centers for Disease Control and Prevention's redesigned BioSense program. GAO-09-100. Washington, DC: U.S. General Accounting Office, 2008. Available at: http://www.gao.gov. Accessed August 16, 2010.

6 U.S. General Accounting Office. Information technology: Federal agencies face challenges in implementing initiatives to improve public health infrastructure. GAO-05-308. Washington, DC: U.S. General Accounting Office, 2005. Available at: http://www.gao.gov. Accessed August 16, 2010.

Emerging Health Threats Journal 2011, **4**:s113. **doi:** 10.3134/ehtj.10.113
© 2011 A Signorini *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Using Twitter to estimate H1N1 influenza activity

A Signorini[1], PM Polgreen[2], and AM Segre[1]

[1]*Departments of Computer Science, The University of Iowa, Iowa City, IA, USA; and* [2]*Internal Medicine, The University of Iowa, Iowa City, IA, USA*
E-mail: alessio-signorini@uiowa.edu

Objective

This paper describes a system that uses Twitter to estimate influenza-like illness levels by geographic region.

Introduction

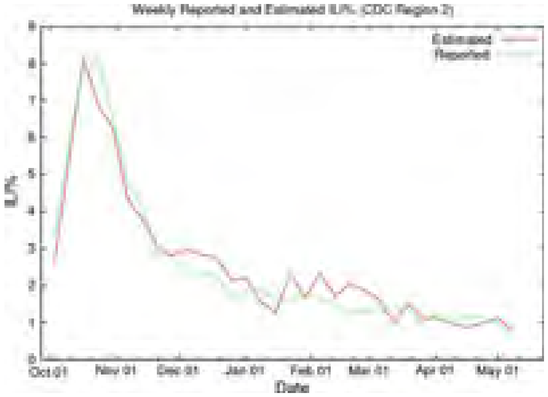
Twitter is a free social networking and micro-blogging service that enables its millions of users to send and read each other's 'tweets', or short messages limited to 140 characters. The service has more than 190 million registered users and processes about 55 million tweets per day.[1] Despite a high level of chatter, the Twitter stream does contain useful information, and, because tweets are often sent from handheld platforms on location, they convey more immediacy than other social networking systems.

Methods

We collected and stored all public tweets beginning October 1, 2009 that matched a set of pre-specified search terms (for example, flu, swine, influenza, tamiflu, oseltamivir, H1N1, and so on). After culling, for example, non-US and non-English tweets, and applying appropriate stemming algorithms, the remaining tweets were used to produce a dictionary of English word equivalents. We compiled daily and weekly usage statistics for each dictionary term, both nationally and at the CDC's influenza reporting-region level.[2] These weekly term-usage statistics were then used to estimate weekly levels of influenza-like illness (ILI). Using Support Vector Regression,[3] a supervised machine-learning method generally applied to solve classification problems,[4] we trained our system using weekly term-frequency statistics from tweets issued outside of CDC Region two (New York and New Jersey) and ILI values reported by the CDC for the weeks October 4–10, 2009 through May 16–22, 2010. We then used the resulting system to estimate ILI in CDC Region two, thus performing an out-of-sample validation.

Results

The Figure 1 shows estimated weighted ILI values for CDC Region two (New Jersey and New York) produced by our system when trained on Twitter traffic exclusive of CDC Region two. The predicted weekly ILI values are shown in red, with ILI values reported later by the CDC in green. Our regional model approximates the epidemic curve reported by ILI data with an average error of 0.37% (min = 0.01%, max = 1.25%), and a standard deviation of 0.26%. Similar results were obtained when estimating ILI at a national level.



Conclusions

Our results demonstrate that Twitter traffic can be used to provide real-time estimates of disease activity. Our ability to quickly detect trends, which are then confirmed by observations from traditional surveillance approaches, make this new form of surveillance a promising area of research at the interface between computer science, epidemiology, and medicine.

Acknowledgements

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

References

1 Lorica B. Twitter by the numbers. O'Reilly Radar, http://radar.oreilly.com/2010/04/twitter-by-the-numbers.html, 14 April 2010 (accessed 15 August 2010).
2 Centers for Disease Control and Prevention. Overview of Influenza Surveillance in the United States, http://www.cdc.gov/flu/weekly/fluactivity.htm (accessed 15 August 2010).
3 Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V (1997) 'Support Vector Regression Machines'. Advances in Neural Information Processing Systems 9, NIPS. MIT Press: Cambridge, MA, USA, 1996, pp 155–61.
4 Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press: Cambridge, England, 2000.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s150. **doi:** 10.3134/ehtj.10.150
© 2011 O Simwale.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

How did specific specialties and new sentinel providers affect ILINet data post 2009 A/H1N1 recognition

O Simwale

Pennsylvania Department of Health, Bureau of Epidemiology, Division of Infectious Disease Epidemiology, Harrisburg, PA, USA
E-mail: owsimwale@state.pa.us

Objective

The objective of this study is to describe changes in influenza-like illness (ILI) surveillance, eight weeks before and after the 2009 A/H1N1 pandemic influenza outbreak. We examined changes in provider recruitment, composition, reporting of ILI, and we characterize ILI data in terms of timeliness, and ILI baselines by type of sentinel provider.

Introduction

The United States outpatient Influenza-like Illness Surveillance Network (ILINet) is one of the five systems used for influenza surveillance in the United States. In Pennsylvania, ILINet providers are asked to report, every Monday, the total number of patients seen for any cause, and the number of patients with influenza-like illness (ILI) by age group. In order to encourage timely reporting, weekly reminders along with a data summary were sent to all sentinel providers post-outbreak recognition. Through the study period, recruitment of new sentinel sites was done through local health departments, health alerts, and training sessions. Sentinel providers were not restricted from submitting specimens to the state lab before and after the outbreak, whereas non sentinel providers had strict restrictions.

Methods

We examined ILINet data pre-outbreak recognition (January to March 2009) and post-outbreak recognition (April to June 2009) changes in provider recruitment, composition, reporting of ILI, and we characterized ILI data in terms of timeliness, and computed ILI baselines by type of sentinel provider. In this study, we defined timeliness as the number of providers reporting by Close of Business every Monday.

Results

In multivariate analyses of preliminary data, and after controlling for observed covariates, we noted a two-fold increase in number of sentinel providers (60–136, $P > 0.05$). Whereas before the outbreak the composition of sentinel providers was limited to pediatricians and family practices physicians, a variety of practices including Obstetricians and Gynecologists, colleges, and emergency rooms joined the sentinel providers post-outbreak recognition. Sentinel colleges reported significantly higher levels of ILI, pediatricians, emergency rooms, and family practice offices (8 versus 3 versus 2%). There was a significant increase in the proportion of sentinel providers reporting ILI every week post-outbreak recognition (50 versus 20%, $P > 0.5$). There was a week correlation between the numbers of sentinel providers reporting every week and the proportion of ILI visits reported ($r = 0.05$). We also noted a significant difference between the levels of ILI reported every Monday compared with what is reported by Thursday of each week ($P > 0.5$).

Conclusion

Our study highlights several changes in ILI surveillance post A/H1N1 recognition. The number and composition of sentinel providers increased significantly post-outbreak recognition. Timely reporting of ILI should be encouraged in order to allow for near-real time use of ILINet data. An ILINet data should be analyzed at provider level in order to establish provider-specific and region-specific ILI baselines.

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.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s114. **doi:** 10.3134/ehtj.10.114
© 2011 AJ Siniscalchi.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Use of a public health working group to coordinate multi-jurisdiction response to bioterrorism surveillance signals and influenza outbreaks

AJ Siniscalchi

State of Connecticut, Department of Public Health, Hartford, CT, USA
E-mail: alan.siniscalchi@ct.gov

Objective

The rapid and effective coordination of the multi-jurisdictional communications and response to a surveillance system signal are an important goal of public health preparedness planning. This goal is particularly challenging if the signal indicates a possible risk that could adversely affect populations in multiple states and municipalities. This paper examines the value of a regional workgroup in the activation, integration, and coordination of multiple surveillance systems along with efforts to coordinate risk communication messaging. Recommendations for the development of similar groups in other regions are discussed.

Introduction

The use of syndromic surveillance systems to detect illness and outbreaks in the mid 1990s in New York City resulted in recommendations for increased use of these systems for detection of bioterrorist agents, and tracking influenza throughout the region.[1,2] Discussions on approaches to best respond to surveillance system signals led to initial efforts to organize a coordinating group of various public health agencies throughout the New York City region. These efforts were strengthened after the events of September 11, 2001, and resulted in the development of a regional workgroup consisting of epidemiologists and other staff from all state, county, and municipal health departments who operate, respond to, or oversee public health preparedness surveillance systems throughout the greater New York City metropolitan area.

Methods

Syndromic surveillance system data from hospital-based networks in state and large municipal public health jurisdictions are reviewed on a daily basis along with information from laboratory-based and remote-sensing systems. Information on surveillance system signals and other indicators of bioterrorism events or emerging infectious disease outbreaks are shared among members of the workgroup, as appropriate. This information can be used for an enhanced review of syndromic surveillance system data in other jurisdictions, notification of emergency department staff to look for patterns of illness, and recommendations for increased laboratory testing. Findings of laboratory-confirmed bioterrorist agents and evidence of outbreaks of emerging illness would result in development of coordinated messaging among member jurisdictions of the work group. Periodic meetings and exercises are scheduled among members that include representatives of federal, state, and local law enforcement agencies. Joint public health and law enforcement emergency response protocols have also been developed and tested.

Results

Members of the work group have coordinated a number of joint multi-jurisdictional surveillance system signal response investigations. These include the regional response to the 2001 anthrax mail attacks (Amerithrax),[3] the identification of inhalation[4] and cutaneous[5] anthrax cases in individuals exposed to contaminated goatskins used in African drums, targeted regional surveillance for detection of influenza and other disease events following the 2009 US Presidential Inauguration[6] and tracking of the novel 2009 influenza A (H1N1) pandemic.[6]

Conclusions

The use of a multi-jurisdictional public health working group has resulted in the enhanced coordination of a regional response to various surveillance system signals and emerging disease outbreaks. Development of similar working groups is recommended for other regions.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 Heffernan R, Mostashari F, Das D, Besculides M, Rodriguez C, Weiss D, *et al*. System descriptions New York City syndromic surveillance systems. *MMWR* 2004;**53** (Suppl): 23–7.

2 Hadler JL, Siniscalchi A, Dembek Z. Hospital admissions syndromic surveillance: experience monitoring serious rash illness and influenza. *MMWR* 2005;**54** (Suppl): 169–73.

3 CDC. Update: investigation of bioterrorism-related anthrax — Connecticut, 2001. *MMWR* 2001;**50**:1077–9.

4 CDC. Inhalation anthrax associated with dried animal hides — Pennsylvania and New York City, April 2006. *MMWR* 2006;**55**:280–2.

5 CDC. Cutaneous anthrax associated with drum making using goat hides from West Africa — Connecticut, 2007. *MMWR* 2008;**57**:628–31.

6 Siniscalchi A, Andrews M, Purviance K, Esponda B, Brennan T, Nelson R, *et al*. Yes we can! use of multiple surveillance systems for detection of post inaguration and late season influenza events, 2009. ISDS Annual Conference.

Emerging Health Threats Journal 2011, **4**:s151. **doi:** 10.3134/ehtj.10.151

© 2011 D Peace *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

A comparison of syndromic surveillance chief complaint data and discharge data in a pediatric hospital system during 2009 H1N1

D Peace[1], W Smith[2], A Reeves[2], D Little[2], K Soetebier[2], and C Drenzek[2]

[1]*Children's Hospital of Atlanta, Atlanta, GA, USA; and* [2]*Georgia Division of Public Health, Atlanta, GA, USA*
E-mail: delittle@dhr.state.ga.us

Objective

The objective of this study is to describe the difference between patient chief complaint (CC), influenza- like- illness (ILI) data provided daily to the Georgia Syndromic Surveillance Program (SSP) during the 2009 H1N1 pandemic, and patient discharge data (DD) subsequently provided for comparison with the SSP from its participating pediatric hospital system, and its two affiliated emergency rooms.

Introduction

The Syndromic Surveillance Program (SSP) of the Acute Disease Epidemiology Section of the Georgia Division of Public Health, provides electronic ILI data to the Center for Disease Control and Prevention's Influenza-like Illness Surveillance Network (ILINet) Program that characterizes the burden of influenza in states on a weekly basis.

ILI is defined as a fever of 100°, plus a cough or sore throat. This definition is used to classify ILI by the SSP, as well as in diagnosis at the pediatric hospital system. During the 2009 H1N1 pandemic, the SSP was provided a daily data transfer to the Center for Disease Control and Prevention to heighten situational awareness of the burden of ILI in Georgia. Throughout the peak of the pandemic, data from the pediatric hospital system identified when the percentage of daily visits for ILI had substantively increased. The data includes patient CC data from emergency department visits for two facilities at Facilities A and B. The data received by SSP does not include diagnosis data.

Patient emergency department DD for 'FLU' was provided to SSP retrospectively to compare with the CC data routinely collected and analyzed. The data was derived from the pediatric health system's month end, internal, syndromic surveillance report based upon emergency department visits, and including physician's diagnosis at the time of patient's discharge. The case definition of 'FLU' from the pediatric health system facilities is acute onset of fever, with cough and/or sore throat in the absence of a known cause other than influenza.

Methods

The data were evaluated by analyzing the percentage of 'FLU'–DD visits during 2009, in contrast to the percentage of ILI–CC visits provided to SSP daily from Facilities A and B. The total percentage of ILI visits to both facilities for CC and DD were then compared and correlated by Facilities A and B, observing 'FLU'–DD to the SSP–ILI–CC data. The CC data were then assessed for its ability to accurately identify changes in actual influenza activity in the two facilities during corresponding time period using Pearson's correlation coefficient. Finally, CC–ILI data were compared with CDC's National Respiratory and Enteric Virus Surveillance System (NREVSS), and ILINet data for Georgia.

Results

The differences between ILI–DD and CC, from facilities A and B were substantial. When comparing Facility A and B, ILI data separately or combined, the burden of ILI based on CC was substantively higher than the observed DD for the same time period. Interestingly, patients from Facility A were more likely to receive an influenza diagnosis than patients from Facility B. The case definition between the facilities is the same; the reason for the difference is not clear (Figure 1).

The SSP–ILI–CC data that were compared with Georgia's NREVSS, and ILINet data overestimated the burden of influenza, as expected. However, the SSP–CC–ILI data were effective in reflecting both increases and decreases in influenza activity that were shown in NREVSS and ILINet data.

Conclusions

The advantages of using electronic ILI–CC data during an evolving event, such as pandemic, are that it is readily available, and allows public health practitioners to characterize the health seeking behaviors of the population. Although it

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

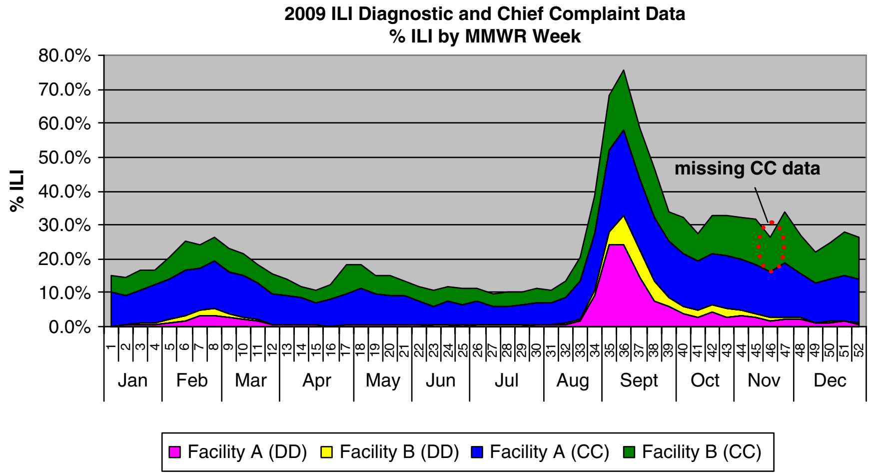


Figure 1 Weekly graph of % ILI for chief complaint (blue/green), and discharge diagnosis (pink/yellow) visits for facilities A and B in 2009.

still remains difficult to accurately quantify the amount of influenza activity because of the overestimation of ILI disease burden that CC data produces, its ability to reflect trends in the burden of diagnosed influenza in the population is valuable.

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.

ISDS

Emerging Health Threats Journal 2011, **4**:s116. **doi:** 10.3134/ehtj.10.116
© 2011 C Staes *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Design and development of a standards-based model to publish public health reporting criteria

C Staes[1], D Rajeev[1], R Bradshaw[1], S Mottice[2], Z Lui[2], N Hulse[1], and S Narus[1]

[1] *Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; and* [2] *Utah Department of Health, Salt Lake City, UT, USA*
E-mail: catherine.staes@hsc.utah.edu

Objective

In this paper, we describe the content and functional requirements for a knowledge management system that can be authored by public health authorities to inform reporting facilities 'what's reportable where'.

Introduction

State laws mandate clinicians and laboratories to report occurrences of reportable diseases to public health entities. For this purpose, a set of case-reporting specifications are published and maintained by public health departments. There are several problems with the existing case-reporting specifications: (1) they are described on individual state websites and posters and not structured or executable; (2) the specifications are often misleading representing case classification rather than case reporting criteria; (3) they vary across jurisdictions and change over time; and (4) reporting facilities are required to interpret the criteria and maintain logic in their own systems.[1] To overcome these problems, we are designing and developing a prototype system to represent case-reporting specifications that can be authored and maintained by public health and published openly.

Methods

To determine the content and functional requirements, we reviewed existing reportable disease lists and rules, fact sheets, and CSTE Position Statements. We used ethnographic methods to obtain feedback from public health authorities. To model the knowledge using standards, we reviewed existing HL7-structured documents including the Clinical Document Architecture. The knowledge management system allows public health authorities to author, store, and publish knowledge-concerning reporting logic and specifications (Figure 1). For this demonstration project, the reporting entities will be able to view human-readable specifications, download structured content using web services for execution within their own systems, and subscribe or query for updates. Knowledge authoring, publication, and access are provided through service-oriented architecture.

Results

After review of the websites from Utah, LA County, Colorado and Washington State for communicable disease and environmental/occupational diseases, we identified 103 reportable events, of which only 62 were included among the nationally notifiable list. The reportable events included diagnoses, laboratory results, clinical observations, interventions, outbreaks, and intent. We identified the following content requirements to represent reporting specifications: (a) detection criteria that includes clinical and laboratory findings; (b) constraints including characteristics of the patient, encounter, laboratory or clinical setting; (c) reference resources; and (d) reporting action details including the reporting time frame and so on. We have designed a model workflow for public health authorities to author reporting specifications, allowing for default content based on the CSTE disease specific 'Position Statements'. We are developing the use case for accessing the information in both human-readable and machine-processable format. We are storing the knowledge assets in XML and exploring the HL7-structured document format to structure reporting specifica-

Figure 1 Future process for laboratories to determine 'what is reportable where' and maintain logic.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

tions. We are exploring models to represent laboratory criteria for improved linkage to the relevant pairs of LOINC/SNOMED mappings.

Discussion

We have demonstrated the modeling of a knowledge management system using existing standards. The development of a prototype knowledge management system that allows public health authorities to author, store, and publish knowledge-concerning reporting specifications will help inform the development and implementation of a national system.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010. Funding: (Grant# 1PO1HK000069-01) Rocky Mountain Center for Translational Research in Public Health Informatics, NLM Training Grant # 5T15LM007124 (DR).

Reference

1 M'ikanatha NM, Welliver DP, Rohn DD, Julian KG, Lautenbach E. Use of the web by state and territorial health departments to promote reporting of infectious disease. *JAMA* 2004;**291**:1069–71.

Emerging Health Threats Journal 2011, **4**:s118. **doi:** 10.3134/ehtj.10.118
© 2011 D Swenson *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Use of emergency department data for case finding following a community anthrax exposure

D Swenson[1], LK Mayo[2], K Dufault[1], ER Daly[1], S Bascom[1], J Dionne-Odom[1,2], and EA Talbot[1,2]

[1]*New Hampshire Department of Health and Humans Services, Concord, NH, USA; and* [2]*Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA*
E-mail: dswenson@dhhs.state.nh.us

Objective

This paper describes the use of customizable tools to query electronic emergency department data, as part of case finding, during the response to a community anthrax exposure in New Hampshire.

Introduction

On 24 December 2009, a female New Hampshire resident was confirmed to have gastrointestinal anthrax on the basis of clinical findings and laboratory testing.[1] Her source of anthrax was not immediately known, so the New Hampshire Department of Health and Human Services (DHHS), in conjunction with several other state and federal agencies, conducted a comprehensive epidemiologic investigation, which included active surveillance to identify any additional anthrax cases from a similar exposure. It was determined that the index patient participated in a drumming event with animal-hide drums on 4 December, one day before the onset of symptoms. Two drums used at the event were later found to be contaminated with *Bacillus anthracis*.

Methods

A retrospective review of emergency department encounters between 1 October and 26 December was conducted immediately after confirmation of the index anthrax case to identify possible anthrax cases that had already presented but were not confirmed because of the rarity of clinical anthrax. In addition, active, prospective surveillance for possible cases continued between 26 December and 31 January, 2010. DHHS's statewide emergency department data system's customizable query tools were used to mine data daily for clinical syndrome terms in chief complaint text fields and ICD-9-CM codes associated with the three manifestations of anthrax (inhalation, cutaneous, and gastrointestinal). Chief complaint text and ICD-9-CM query results were reviewed by surveillance staff and a physician using a clinical criteria protocol to decide whether the case warranted follow-up with hospitals to collect additional clinical and laboratory data.

Results

Of 1,83,973 emergency department encounters between 1 October 2009 and 31 January 2010, 7111 (3.9%) met either the chief complaint or ICD-9-CM search criteria. Of these 7111, 62 (0.9%) warranted collection of additional information based on the clinical criteria protocol. No additional anthrax cases were identified. Three persons were found seeking anthrax post-exposure prophylaxis at an emergency department. Two of these persons were workers involved in the public health response and one person had attended the drumming event and was previously identified during the contact investigation (Figure 1).

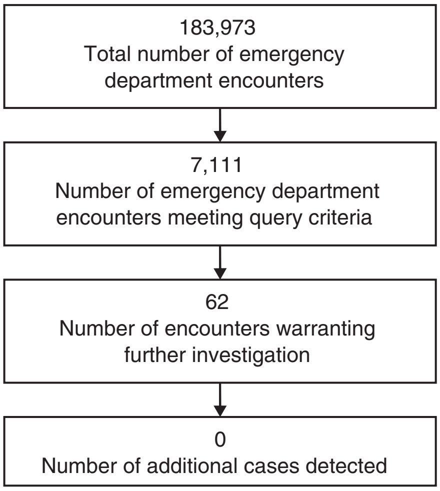


Figure 1 Investigation outcome of emergency department encounters meeting anthrax query and clinical criteria.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

New Hampshire's emergency department surveillance system was quickly adapted, through the development of custom queries, to investigate an emergent public health threat. The query tool and clinical criteria protocol allowed for efficient follow-up of only those encounters that were most suspicious for anthrax. These tools narrowed down a pool of nearly 2,00,000 emergency department visits to 62 possible cases requiring further investigation, allowing for more efficient use of health department resources.

Acknowledgements

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

Reference

1 CDC. Gastrointestinal anthrax after an animal-hide drumming event—New Hampshire and Massachusetts, 2009. *MMWR* 2010; **59**:872–7.

Emerging Health Threats Journal 2011, **4**:s153. **doi:** 10.3134/ehtj.10.153
© 2011 M Taddie and I Yamada.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Prospective space-time analysis of the 2007 cryptosporidiosis outbreak, Salt Lake County, Utah

M Taddie[1,2], and I Yamada[1,2]

[1]*Geography Department, University of Utah, Salt Lake City, UT, USA; and* [2]*Center for Spatial Information Science, University of Tokyo, Tokyo, Japan*
E-mail: mtaddie@earthlink.net

Objective

The objective of this study was to investigate if prospectively applied space-time surveillance could have detected significant, emerging clusters as cryptosporidiosis, cases were reported to the Salt Lake Valley Health Department during the 2007 outbreak.

Introduction

Cryptosporidiosis is a gastrointestinal illness due to a protozoan parasite that is highly contagious, and resistant to multiple disinfectants.[1] Utah experienced a large, community-wide outbreak of cryptosporidiosis between June and December of 2007.[2,3] During this time period, the Utah Department of Health received reports of 1,902 laboratory-confirmed cryptosporidiosis cases across the state.[2] Nearly 40% of these cases occurred in Salt Lake County (SL County), Utah. In past years, SL County averaged fewer than five cases annually;[2] however, the incidence rate in the county for this entire outbreak was 125.9 per 100,000 person–years.[3]

Methods

This study utilized a space-time scan statistic implemented in SaTScan[4] to test for the occurrence and location(s) of cryptosporidiosis clusters using time-periodic prospective surveillance, and a Poisson probability model. The study area included the jurisdictional purview of the Salt Lake Valley Health Department, which is SL County. Although true disease outbreaks should be characterized by the onset date of each case, onset dates are not collected until after an investigation is started, and the date of the case reported is generally the first piece of case-specific information available to many health departments. Therefore, report dates were used in the prospective space-time analysis to replicate the realistic surveillance processes that occur in health departments, and to mimic a near real-time surveillance system.

Results

The first cluster signaled approximately 20 days after the first reported case in SL County. This cluster occurred two days before a statewide press release was issued, and 21 days before the implementation of major intervention measures. From August through mid-September, many significant clusters were detected throughout SL County (Figure 1).

Conclusions

The results of this study suggest that there were distinct spatial-temporal patterns throughout the outbreak period. Therefore, space-time analysis would have been a valuable and complementary tool to temporal surveillance because it could have detected spatial clusters and high-risk areas of disease as they were reported, or emerged. In addition, it may have been useful for targeting intervention strategies and prioritizing investigations during this large community-wide outbreak.

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 American Public Health Association. [Cryptosporidiosis]. In Heymann DL (ed). *Control of Communicable Diseases Manual*, 18th edn. APHA: Washington, DC, 2004, pp 138–41.
2 Utah Department of Health, Bureau of Epidemiology. *Report of Investigation: Cryptosporidium Outbreak Linked to Multiple Recreational Water Venues: Utah, 2007*. UDOH: Salt Lake County, UT, 2008.
3 Salt Lake Valley Health Department, Bureau of Epidemiology. *2007 Cryptosporidiosis Outbreak Report.* SLVHD: Salt Lake County, UT, 2008.
4 Kulldorff M. *Information Management Services, Inc: SaTScan (TM) v7.0.3: Software for the spatial and space-time scan statistics*. National Cancer Institute: Bethesda, MD, 2007, www.satscan.org.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

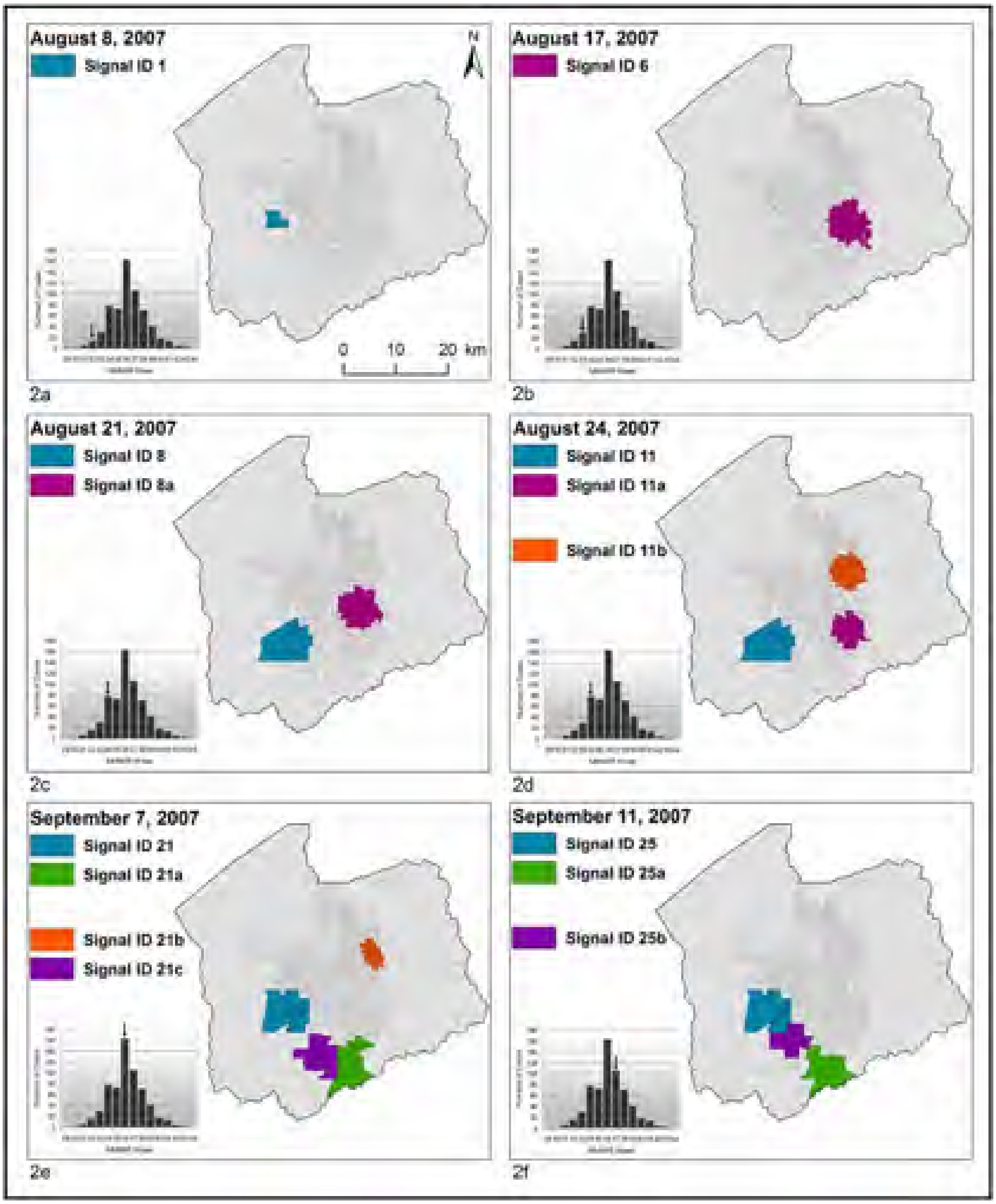


Figure 1 Significant space-time clusters detected during six 28-day time steps from 8 August 2007 to 11 September 2007.

Emerging Health Threats Journal 2011, **4**:s117. **doi:** 10.3134/ehtj.10.117
© 2011 F-C Tsui *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

An automated influenza-like-illness reporting system using freetext emergency department reports

F-C Tsui[1], H Su[1], J Dowling[1], RE Voorhees[2], JU Espino[1], and M Wagner[1]

[1]*RODS Laboratory, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA; and* [2]*Office of Epidemiology and Biostatistics, Allegheny County Health Department, Pittsburgh, PA, USA*
E-mail: tsui2@pitt.edu

Objective

This paper describes an automated Influenza-Like-Illness (ILI) reporting system based on natural language processing of transcribed ED notes and its impact on public health practice at the Allegheny County Health Department (ACHD).

Introduction

Current methods for influenza surveillance include laboratory confirmed case reporting, sentinel physician reporting of ILI and chief-complaint monitoring from emergency departments (EDs).

The current methods for monitoring influenza have drawbacks. Testing for the presence of the influenza virus is costly and delayed. Specific, sentinel physician reporting is subject to incomplete, delayed reporting. Chief complaint (CC) based surveillance is limited in that a patient's chief complaint will not contain all signs and symptoms of a patient.

A possible solution to the cost, delays, incompleteness and low specificity (for CC) in current methods of influenza surveillance is automated surveillance of ILI using clinician-provided free-text ED reports.

Methods

Our ILI reporting system has five parts: an HL7 message parser, the MedLEE natural language processing algorithm, a database, rule-based logic, chart engine and email server. The HL7 message parser extracts ED reports from HL7 messages. Then, MedLEE[1] finds medical terms contained in each report, including significant negative findings. We store the NLP results in a database using an entity attribute value schema. The rule-based case definitions represent CDC's ILI case definition, (*Fever* or *Chills*) and (*Cough* or *Sore throat*),[2] and the symptoms found by the NLP for each patient are tested against this logic. The system plots a 6-month time series graph of the percentage of total ED visits per day with ILI detected by the rule-based logic. The graph also includes a 5-day moving average (centered on the 3rd day) of the daily ILI time series to follow existing practices at ACHD. We employ Jfreechart to create the chart with the two time series. Every day the system sends the report to ACHD using a mail server.

We performed a preliminary evaluation of the system comparing ILI rule-based logic output with manual review of the transcribed ED reports by a physician board, certified in infectious diseases.

Results

We deployed the ILI reporting system in early May 2009 in Allegheny County. It receives transcribed ED reports by HL7 messages from the seven EDs of the UPMC Health System. The average number of daily ED visits with ED reports between July and Dec. 2009 was 569. Figure 1 shows the ILI chart for 31 December 2009 report that comprises daily ILI percentage and its 5-day moving average between 15 June 2009 and 30 December 2009.

Our evaluation using 140 randomly selected reports (78 positives) found a sensitivity of 94.9% (95%CI: 87.54–97.99%), a specificity of 100% (95% CI: 94.17–100%), and a positive predictive value of 100% (95% CI: 94.17–100%).

These charts had three impacts on practice at ACHD: (1) ACHD had daily updates instead of weekly reports (from sentinel physicians). (2) ACHD provided the charts to local media on a regular basis.[3] (3) ACHD reduced staff time as they no longer had to manually compile ILI reports from sentinel ILI reports (2 days of work for each weekly report).

Figure 1 Daily reporting chart showing daily percentage of ED visits with ILI (green) and its 5-day moving average (black) between 15 June 2009 and 30 December 2009.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

We demonstrate utility and method of using free text ED reports for ILI reporting and flu surveillance.

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. The project is funded by CDC Grants P01 HK000086 and 1U38 HK000063-01.

References

1 Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004;**11**:392–402.
2 Bodnar UR, Fielding KL, Navin AW, Maloney SA, Centron MS, Bridges CB, *et al*. *Preliminary Guidelines for the Prevention and Control of Influenza-like Illness among Passengers and Crew Members on Cruise Ships*. Center for Disease Control and Prevention, Atlanta, GA, 1999.
3 Two more people die from H1N1 in Pittsburgh area [online] 2009 cited. Available from: http://kdka.com/health/H1N1.flu.deaths.2.1321133.html.

Emerging Health Threats Journal 2011, **4**:s155. **doi:** 10.3134/ehtj.10.155
© 2011 KS Wallace *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

The United States Department of Veterans Affairs Integrated Operations Center (VA IOC): collaborations for surveillance, analysis, and prediction for infectious disease threat preparedness—pilot review of dengue occurrence

KS Wallace[1], GA Roselle[2,3,4], S Gamage[2], D Green[5], LA Simbartl[2], M Villeneuve[1], J Lacey[1], SM Kralovic[2,3,4], and JD Riojas[1]

[1]*Office of Operations, Security and Preparedness (OSP), VA Central Office, Washington, DC, USA;* [2]*National Infectious Diseases Program Office (IDPO), Veterans Health Administration (VHA), Washington, DC, USA;* [3]*Cincinnati VA Medical Center, Cincinnati, OH, USA;* [4]*University of Cincinnati College of Medicine, Cincinnati, OH, USA; and* [5]*Office of Climate, Water, and Weather Services, National Weather Service (NWS), National Oceanic and Atmospheric Administration (NOAA), Silver Spring, MD, USA*
E-mail: katherine.wallace@va.gov

Objective

The objective of this study is to describe Veterans Affairs Integrated Operations Center-enabled collaborations to enhance the synergy of relevant data/information from Veterans Affairs (VA) and non-VA partners for improved early warning, and situational awareness of infectious disease threats.

Introduction

Under leadership of the Secretary of Veterans Affairs (VA), Office of Operations, Security and Preparedness has established the Veterans Affairs Integrated Operations Center, with the goal of enhancing integration and analysis of data, and information from VA's preparedness partners, both internal and external, for timely decision support.

The Office of Operations, Security and Preparedness oversee emergency preparedness for the VA, which includes responsibility for preparedness activities at Veterans Health Administration (VHA). The VHA provides medical care to over 5 million patients a year at 153 medical centers, and over 900 outpatient clinics in the United States, and the United States territories. The Office of Operations, Security and Preparedness is developing a VA–Subject Matter Expertise Center for Biological Events in collaboration with the VHA–National Infectious Diseases Program Office. The Subject Matter Expertise Center for Biological Events is initiating pilot projects to examine data sources, integration, and predictive analysis. The recent increase in dengue cases internationally prompted the Office of Operations, Security and Preparedness, and the Subject Matter Expertise Center for Biological Events to establish collaborations, and investigate factors influencing dengue disease patterns in VHA facilities.

The National Weather Service has the mission to provide weather, water and climate data, forecasts and warnings for the protection of life and property, and enhancement of the national economy. The Veterans Affairs Integrated Operations Center enabled collaboration with the National Weather Service for integration of weather, water and climate data, and retrospective analysis into preparedness activities.

Methods

The VHA patient treatment file administrative dataset was queried for occurrence of dengue, and dengue-like diagnosis codes in VHA inpatients across the nation for Federal Fiscal Years (FYs) 2008, 2009, and 2010 (up to July).

The National Weather Service used Advanced Hydrologic Prediction Service precipitation data. This quality-controlled data is projected into gridded images to provide spatial information. Analysis fields included departures from average, and percent of average (daily, monthly, and for a water year). Spatial variation data were contrasted with the rain gauge point data available from National Oceanic and Atmospheric Administration's National Climatic Data Center archive.

The environmental data for FYs 2008–2010 were added to the VHA dengue case data to evaluate if weather and water influences can be associated with disease occurrence in a

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

patient population, with the intent of laying the building blocks for a predictive model of disease.

Results

As expected, dengue or dengue-like cases in VHA inpatients primarily occurred in Puerto Rico, where the disease is endemic. Thus, far in FY 2010, 35 cases have been recorded; FYs 2008 and 2009 had 31 and 14 cases, respectively. Interestingly, the case occurrence pattern was different for FY 2010 compared with FYs 2008 and 2009: cases in FY 2010 occurred in peaks in the winter, spring, and summer, whereas occurrence in the other years peaked only in the fall with low case numbers the rest of the year. The Advanced Hydrologic Prediction Service precipitation web portal provided access to images of observed precipitation, period normal, and departure from normal and percent of normal precipitation pre-compiled for time periods from one day to the water-year to date, to guide analysis of dengue fever outbreaks.

Conclusions

This work demonstrates the value and potential of Integrated Operations Center initiatives within VA (for example, Subject Matter Expertise Center for Biological Events), and external collaborations (for example, National Oceanic and Atmospheric Administration) toward integrated disease threat prediction, prevention, and mitigation capabilities. The outcomes will drive policy to enhance care of VA's Veteran patients, and support preparedness activities that will improve national decision-making in support of the National Response Framework.

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.

Emerging Health Threats Journal 2011, **4**:s157. **doi:** 10.3134/ehtj.10.157

© 2011 WG Weppner *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Syndromic surveillance of influenza-like illness using automated VA data preliminary results from the Idaho infectious disease reporting network

WG Weppner[1,2], AP Wilper[1,2], J Mitten[1], J Tiumalu[1], R Tivis[3], and D Stevens[1,2]

[1]*Boise VA Medical Center, Boise, ID, USA;* [2]*University of Washington School of Medicine, Seattle, WA, USA; and* [3]*Idaho State University, Pocatello, ID, USA*
E-mail: William.Weppner@va.gov www.IIDRN.org

Objective

The objective of this paper is to study whether syndromic surveillance using data from the Veterans Administration electronic medical record computerized patient record system (CPRS) correlates to officially reported influenza activity levels in the State of Idaho.

Introduction

Current influenza-like illness monitoring in Idaho is on the basis of syndromic surveillance using laboratory data, combined with periodic person-to-person reports collected by Idaho state workers. This system relies on voluntary reporting.

Electronic medical records with relational databases offer a method of obtaining data in an automated fashion. Clinical data entered in CPRS includes real-time visit information, vital signs, ICD-9, pharmacy, and labs. ICD-9 and vital signs have been used to predict influenza-like illness in automated systems.[1,2] We sought to combine these with lab and pharmacy data as part of an automated syndromic surveillance system.

Methods

The Boise Veterans Affairs Medical Center provides care to over 20,000 veterans living in Idaho, with clinics in Boise and surrounding sites. Using data from the Veterans Integrated Service Network (VISN 20) data warehouse for influenza from 2009, we identified influenza-like illness cases from these clinics using ICD-9 codes collected as weekly counts. Additional counts of fever ($>100.5°$F), hypoxia (O_2 $<92\%$), lab tests for influenza (A/B antigen, culture, novel flu), and prescriptions for antivirals (oseltamivir) were summed individually, and in an unweighted fashion, as total weekly counts. Spearman correlation, and multivariate logistic regression were used with predictors from the same week, and preceding week. This was correlated with weekly flu activity as reported by the Center for Disease Control and Prevention; this reports geographic spread of influenza, as reported by state epidemiologists.[3] The Veterans Administration Puget Sound Institutional Review Board (IRB) approved this study.

Results

Using comparisons with epidemiologist-reported flu activity level, all clinical data elements had statistically significant associations using Spearman correlation: sum of total counts of predictors $r=0.57$ ($P<0.0001$); lab tests $r=0.51$ (0.0001); ICD-9 codes $r=0.47$ ($P=0.0003$); prescription $r=0.38$ ($P=0.005$); fever $r=0.31$ ($P=0.02$); hypoxia $r=0.28$ ($P=0.04$). Total counts accounted for one-third ($r^2=0.3$) of variance. Similar results were found for preceding week counts. In logistic regression, both ICD-9, and lab counts were significant predictors (Wald $\chi^2=6.80$, $P=0.009$ and Wald $\chi^2=7.15$, $P=0.007$, respectively) (Figure 1).

Limitations

We sampled a small percent of the overall state population using Veterans Administration data; young people, women, and children are underrepresented.

Conclusions

Data obtained from electronic health records may be useful in predicting influenza-like illness on a regional basis. The combination of ICD-9 codes, vital signs, lab, and pharmacy data provided the best correlation with influenza. ICD-9 and lab counts both contribute independently to prediction, and should be considered to build a stronger model of prediction in our data.

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.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

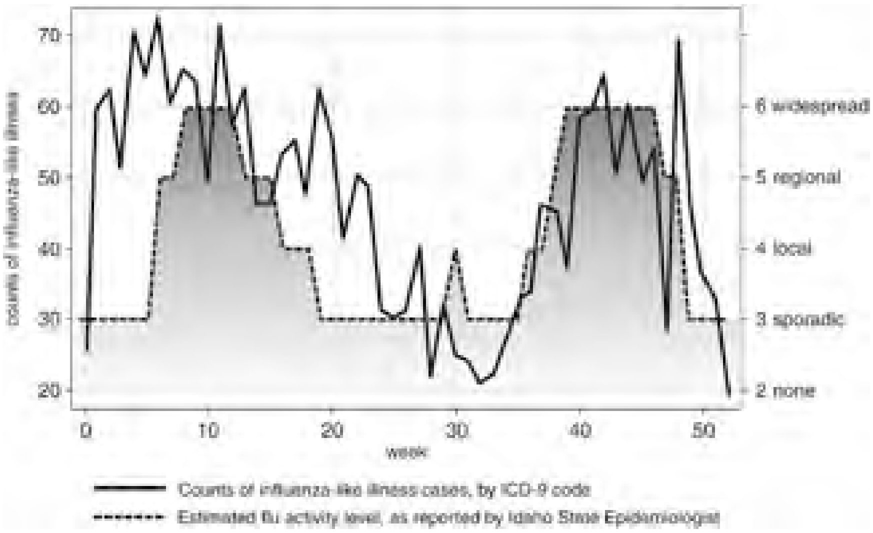


Figure 1 Veterans Administration influenza-like illness cases versus reported weekly influenza activity in the state of Idaho, 2009.

References

1 Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenza-like illness by International Classification of Diseases, ninth revision. *Emerg Infect Dis* 2007;**13**:207–16.

2 Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, Kruskal BA, *et al.* Evaluating real-time syndromic surveillance signals from Ambulatory Care Data in four states. *Public Health Rep* 2010;**125**:111–25.

3 CDC. Fluview. Weekly US Map: Influenza Summary Update. http://www.cdc.gov/flu/weekly/usmap.htm.

Emerging Health Threats Journal 2011, **4**:s123. **doi:** 10.3134/ehtj.10.123
© 2011 Y Zhang *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Evaluating University syndromic surveillance systems during the 2009 H1N1 influenza pandemic

Y Zhang[1,2], L May[1,2], and MA Stoto[1,2]

[1]*Georgetown University, Washington, DC, USA; and* [2]*George Washington University, Washington, DC, USA*
E-mail: yz62@georgetown.edu

Objective

To describe the 2009 H1N1 outbreak at Georgetown University (GU) and George Washington University (GWU) in Fall 2009. Identify the datasets that most accurately depict 2009 H1N1 disease in real time.

Introduction

Syndromic surveillance has been widely adopted as a real-time monitoring tool in early response to disease outbreaks. In order to provide real-time information on the impact of 2009 H1N1 during the Fall 2009 semester, GU and GWU employed syndromic surveillance systems incorporating a variety of data sources.

Methods

The data series include:

- Student influenza-like illness (ILI) cases reported to medical staff: ED visits (GU and GWU Hospital emergency room visits records for patients aged 17–24 years with ILI), SHC (total of visits to the GU and GWU student health centers), off-hour calls to the GU student health center, and calls to the nurse-operated GU H1N1 advice line.
- GU student absenteeism: ILI cases reported to deans at four undergraduate colleges, athletic trainers, and resident assistants.
- GU employee absenteeism data: Real-time employee absenteeism based on call-in sick log at the Facilities Office and Dining Services, and retrospective employee absenteeism in 2008 and 2009 based on payroll data.
- External surveillance data: American College Health Association Pandemic Influenza Surveillance Network, and the CDC ILINet.

Results

- ILI cases reported to medical staff, especially ED visits, peaked first in early September and dropped sharply afterwards. This is similar at GU and GWU, and corresponds to regional and national ACHA data, but

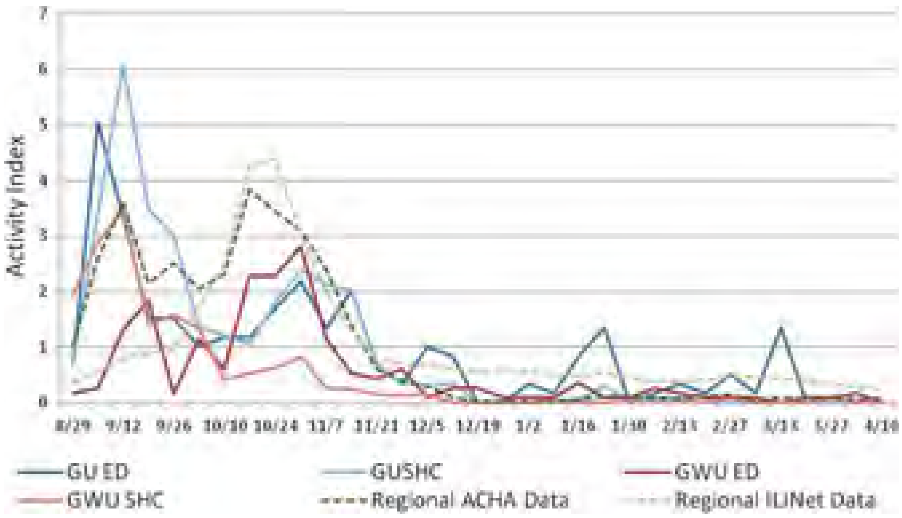


Figure 1 GU and GWU ILI cases.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

not Region 3 (DE, DC, MD, PA, VA, WV) ILINet data (Figure 1).

- Student self-reports collected by deans exhibited a second peak in weeks around midterms.
- Retrospective employee absenteeism patterns are similar to those of the previous year (Figure 2).

Discussion

- 2009 H1N1 activity at both GU and GWU seems to peak right after school begins, which can be observed in a variety of data sources. The second peak in late October corresponds to regional ILINet data but does not show the same scope of impact on campus. This might reflect the epidemic in the community and the increasing need for student medical notes during midterm weeks (Figure 3).
- Syndromic surveillance data can be strongly influenced by factors that influence people's reporting behavior such as communication and intervention policies, and incentives and barriers associated with presenting oneself to the reporting system.

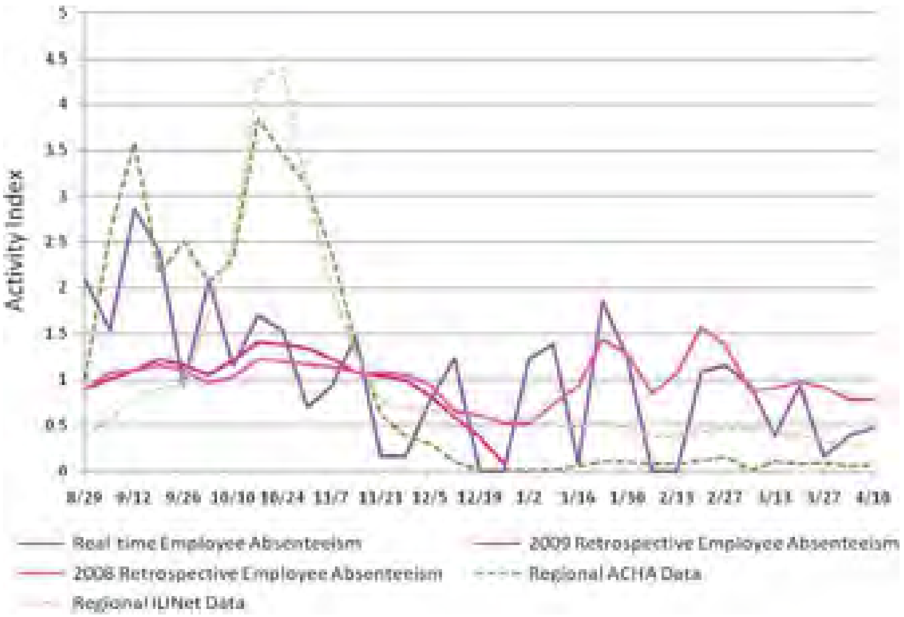


Figure 2 GU employee absenteeism data.

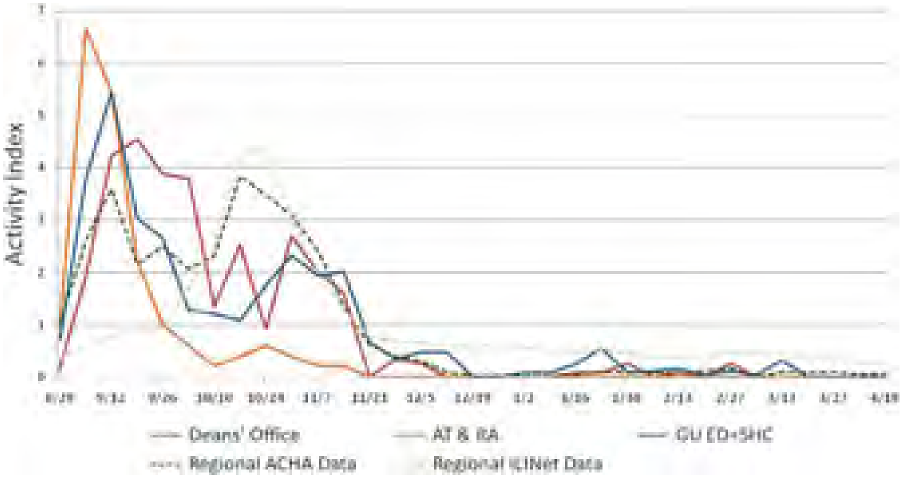


Figure 3 GU student absenteeism data.

- The SHC data may be the most reliable data source in the early stages of an outbreak, but completeness of reporting seems to fall off with policy changes and as students may subsequently see no value to seeking medical attention.

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.

International Society for Disease Surveillance Conference 2010

Track 3: Applications of Methodologies to New Domains

Emerging Health Threats Journal 2011, **4**:s120. **doi:** 10.3134/ehtj.10.120
© 2011 R Blouin *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Using administrative databases to identify cases of chronic kidney disease: a systematic review

R Blouin[1], M Hall[1], and S Saydah[2]

[1]*SciMetrika, Research Triangle Park, NC, USA; and* [2]*Centers for Disease Control and Prevention, Atlanta, GA, USA*
E-mail: rblouin@scimetrika.com

Objective

This poster summarizes a systematic literature conducted to (1) describe published methods for researching chronic kidney disease (CKD) in administrative databases and (2) summarize the reported validity of methods of searching for CKD in administrative databases.

Introduction

CKD is currently the ninth leading cause of death in the United States. The prevalence of end-stage renal disease, the most severe form of CKD, has doubled in the last decade.[1] Early detection and treatment of CKD is critical to slowdown the progression of the disease and to decrease the risk of other chronic conditions, such as cardiovascular disease.[2] One accessible and cost-effective method for health research activities involves use of medical administrative databases, such as insurance claims databases and institutional medical record systems. Individuals with diabetes, for example, have been accurately identified in Medicare and Veterans' Health Administration databases using clearly defined and highly valid search algorithms.[3] However, little is known about the validity of administrative databases for identifying CKD. A systematic review of the literature was conducted to identify the validity of published methods for searching administrative databases for cases of CKD.

Methods

Publications were identified through use of Medline, Cochrane Library, EMBASE, EconLit, Cumulative Index to Nursing and Allied Health, and Web of Knowledge. Articles were included if they were published in peer reviewed journals, constituted original research, focused on adults aged 18 years or older, contained some description of the search algorithm used to identify CKD within an administrative database, and were written in or translated into English. Case studies, news or magazine articles, or clinical articles were excluded. A multiphase process was implemented to review and determine final relevancy of each identified article. The content of all relevant articles was recorded in an Excel-based data abstraction form.

Results

The literature review yielded 38 articles that described methods for identifying CKD in a wide variety of administrative databases (that is, electronic medical record systems, Veteran's Administration files, Medicare claims databases, health management organization claims databases, and disease registries). Algorithms were identified for general CKD, stages 3–5 CKD, renal failure, and diabetic nephropathy. Most of the identified algorithms used International Classification of Diseases, 9th Revision (ICD-9) diagnostic codes as the algorithms' foundation. Algorithms for Stages 3–5 CKD, however, relied on laboratory serum creatinine values converted to estimated glomular filtration rates rather than ICD-9 codes. There was wide variation in the performance of published algorithms. Overall, the algorithms found for general CKD, renal failure, and diabetic nephropathy tended to demonstrate high specificity but lower sensitivity.

Conclusions

Historically, ICD-9 codes have not differentiated between the five stages of CKD, but new stage-specific ICD-9 codes were introduced in 2005. Our literature review did not identify any validated ICD-9 algorithms for stage-specific CKD; both the frequency of use and validity of these new codes are unclear. Researchers should focus on search methods and algorithms that match their own working definition of CKD, have been validated in similar databases, and have been used with similar patient populations. The high specificity and lower sensitivity of the algorithms suggest

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

that the algorithms might, for research studies and to assess the quality of care, identify cohorts of individuals with CKD, but would not perform well for surveillance purposes.

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 Powe N, Plantinga L, Saran R. Public health surveillance of CKD: principles, steps, and challenges. *Am J Kidney Dis* 2009;**53:**S37–45.

2 Saran R, Hedgeman E, Plantinga L, Burrows NR, Gillespie BW, Young EW, *et al*. Establishing a national chronic kidney disease surveillance system for the United States. *Clin J Am Soc Nephrol* 2010;**5:**152.

3 Sambamoorthi U, Tseng CL, Rajan M, Anjali T, Findley PA, Pogach L. Initial nontraumatic lower-extremity amputations among veterans with diabetes. *Med Care* 2006;**44:**779–87.

Emerging Health Threats Journal 2011, **4**:s12. **doi:** 10.3134/ehtj.10.012
www.eht-journal.org

© 2011 R Chunara *et al.*; licensee Emerging Health Threats Journal.

ABSTRACT

OutbreakMD: tracking and identifying disease outbreaks in post-earthquake Haiti

R Chunara, C Freifeld, and JS Brownstein

Children's Hospital Informatics Program, Children's Hospital Boston, Boston, MA, USA
E-mail: chunara@gmail.com

Objective

OutbreakMD is a mobile Web application that was piloted in post-earthquake Port-au-Prince, Haiti. The application is designed for collecting, organizing and visualizing clinical information from individual patients to better monitor emerging infectious disease in disaster situations, in situations with limited public health infrastructure and unreliable Internet connectivity.

Introduction

HealthMap is a real-time disease epidemic intelligence tracking and visualization system that collects information from general news media, individual first-hand reports and public health sources around the world. Gaps in this effort clearly occur during times of crisis where traditional mechanisms may be dismantled. Clinical information gathered by deployed physicians can play a key role in providing early insight on emerging public health threats. We developed OutbreakMD to gather such information in real-time and combine with existing HealthMap informal and formal surveillance techniques.

With Haiti already carrying a significant endemic burden of infectious disease, including tuberculosis, malaria, and HIV, the aftermath of the earthquake of January 12, 2010 has exacerbated the spread of disease while treatment capacity has diminished. This application allows a clinician to rapidly enter patient demographics and case information, and then identify high-risk areas and times by examining the aggregated data spatially and temporally. These insights can be coupled with other public health information sources and can be used to inform responders on how to best target prevention such as vector control and water purification, as well as to identify internally displaced person camps and other specific communities where additional resources are needed.

Methods

Implemented in HTML5, a new Web standard that supports offline data storage, OutbreakMD can be used to collect data

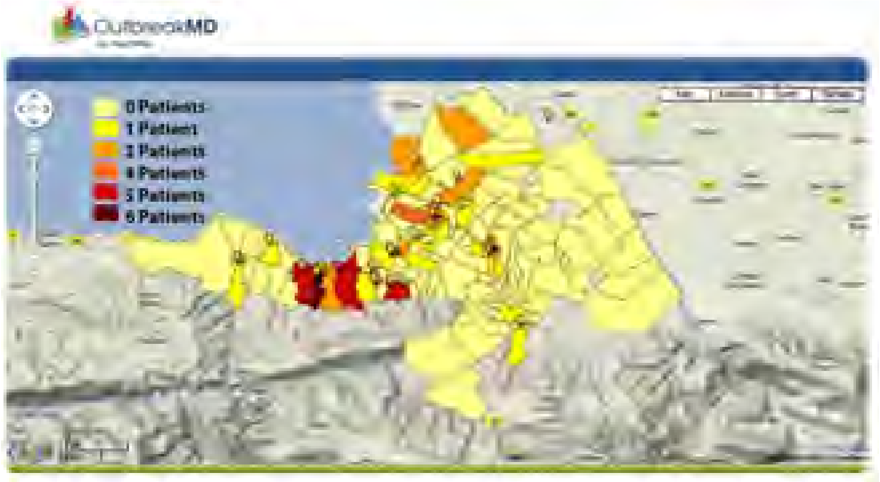


Figure 1 Distribution of all patients documented. Patient locations of origin are denoted by a marker (yellow for patients presenting with tuberculosis symptoms, red for malaria) for cases where exact geolocation was possible; else more generically, by the Quartier of origin (polygon).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

even when no network connection is available. Once a connection becomes available, the system automatically uploads stored reports to a secure online database, capturing patient demographics, address, symptom and clinical test results. Seamless online and offline use makes the application particularly suited to resource poor settings where both Internet and cell phone connectivity are available but often unreliable. Although the current version is designed for the Apple iPad, OutbreakMD is compatible with a range of mobile devices such as iPhone, Android and desktop browsers, so the application can be easily carried and used in the clinic. Once the data are captured and synchronized online, a time-series view of the number of unique patients presenting with symptoms in each of the disease categories, as well as spatial visualization of patient locations (Figure 1) are generated in real-time from the electronic records and made available online.

Results

Over the course of three months subsequent to the Haiti earthquake, data was collected from 117 unique patients. We found that the predominant disease burden was of Tuberculosis, with distribution highest in South Central Port-au-Prince; however a greater number of cases is required for statistical significance. We are currently examining how resolution (exact latitude and longitude versus suburb-level location) affects spatial visualization of the information. This will be relevant in order to automate data processing for locations that are not readily mapped (such as informal camps) while maintaining the quality of information. With enough volume of patient information we will be able to evaluate which levels of spatial resolution will be useful for population public health surveillance.

Conclusion

OutbreakMD is a platform for real-time collection of clinical data in the field with the goal of rapid public health surveillance in disasters and resource-poor settings, and the potential to play a key role in closing the loop between clinical providers and responders.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s14. **doi:** 10.3134/ehtj.10.014
© 2011 C Dangel *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

911 and EMS data for detection of drinking water contamination

C Dangel[1], D Gibbons[2], SC Allgeier[1], A Haas[2], and T Ellis[3]

[1]*US EPA Water Security Division/Office of Water, Cincinnati, OH, USA;* [2]*CSC Science & Engineering, Mission Support Group, Alexandria, VA, USA; and* [3]*Emergint Technologies, Inc, Louisville, KY, USA*
E-mail: dangel.chrissy@epa.gov

Objective

This paper describes the design, application and use of 911 and Emergency Medical Services (EMS) data in a drinking water contamination warning system.[1]

Introduction

The Public Health Surveillance (PHS) component (one of five monitoring and surveillance components deployed in the Cincinnati drinking water contamination warning system) functions to detect public health incidents resulting from exposure to toxic chemicals that produce a rapid onset of symptoms. Within the PHS component, four data streams were monitored: 911 calls, Emergency Medical Services (EMS) logs, Local Poison Control Center call data, as well as Emergency Department data (via EpiCenter). The focus of this paper centers on the 911 and EMS surveillance tools. The 911 data is dependent on information provided by the caller and the information entered by the dispatcher. EMS data, on the other hand, is recorded by a medical professional, and although not provided as rapidly as 911 data, provides more detailed information. The data included in 911 and EMS alerts, when utilized together, can provide timely and beneficial information during investigation of a possible drinking water contamination incident.

Methods

911 calls are captured electronically via a computer aided dispatch system. Calls are classified into standard incident code categories based on information a caller reports to the 911 dispatcher. Call data that meets established filtering criteria is transferred from the Cincinnati Fire Department's (CFD) source database to a dedicated application database at the local water utility. Automated surveillance once every hour using the SaTScan (SaTScan Software, Boston, MA, USA) space-time scan statistics searches for unusual clusters of 911 calls assigned to selected symptom categories. If an unusual cluster is detected, an alarm is generated.

CFD emergency medical technicians (EMT) collect EMS run information using a wireless tablet, and capture patient data such as age, gender, vital signs, chief medical complaint, EMT medical observations, and incident zip code. EMS run data are transferred from CFD's source database to a dedicated application database at the local water utility. The EMT's medical observations are categorized into syndromes and analyzed by the CDC's Early Aberration Reporting System (EARS) software (Early Aberration Reporting System, Atlanta, GA, USA). Surveillance of EMS data occurs on an hourly basis. If unusual conditions are detected in the EMS runs, an alarm is generated.

For both 911 and EMS, once an alarm is generated, an email notification is sent to the local public health partners who have primary responsibility to lead the ensuing investigation.

Results

Alarms from the 911 and EMS surveillance tools were investigated by the public health partners. Although no true water contamination event occurred during the study, a handful of public health events were detected. Various drills and exercises allowed the public health partners and local water utility to investigate alerts, which suggested possible water contamination, and to gain a better understanding of the surveillance tools' capabilities. The 911 and EMS alarm occurrence was tracked during the study, and it was determined that there was a need to modify the existing alerting criteria for each tool given the high rate of false alarm occurrence. Upon implementation of additional alerting restrictions, alarm frequency decreased significantly and was accepted as more beneficial to detecting a waterborne public health event.

Conclusions

The results of the pilot in Cincinnati demonstrate that 911 and EMS surveillance tools can produce timely and beneficial data in detecting a drinking water contamination incident, and provide dual use benefits for detection of other non-water related public health outbreaks.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

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

Reference

1 US EPA. Water security initiative: Cincinnati pilot post-implementation system status report, 2008 EPA 817-R-08-004.

Emerging Health Threats Journal 2011, **4**:s19. **doi:** 10.3134/ehtj.10.019

© 2011 G Divita and QZ Trietler.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Finding medically unexplained symptoms within VA clinical documents using v3NLP

G Divita[1,2], and QZ Trietler[1,2]

[1]*Salt Lake City VA Health Care System, Salt Lake City, UT, USA; and* [2]*University of Utah, Salt Lake City, UT, USA*
E-mail: guy.divita@hsc.utah.edu

Objective

Pro-WATCH (protecting war fighters using algorithms for text processing to capture health events), a syndromic surveillance project, includes a task to identify medically unexplained symptoms. The v3NLP entity extraction tool is being customized to identify symptoms, then to assign duration assertions to address part of this project. The v3NLP tool was recently enhanced to find problems, treatments, and tests for the i2b2/VA challenge. The problem capability is being further refined to find symptoms. Machine learning models will be developed using an annotated corpus currently in development to find duration assertions.

Introduction

Pro-WATCH (protecting war fighters using algorithms for text processing to capture health events), a syndromic surveillance project for veterans of operation enduring freedom (OEF)/operation Iraqi freedom (OIF), includes a task to identify medically unexplained symptoms (MUS). The v3NLP entity extraction tool is being customized to identify symptoms within VA clinical documents, and then refined to assign duration. The identification of medically unexplained symptoms and the aggregation of this information across documents by patient's is not addressed here.

Methods

The v3NLP tool, previously known as HITEXt,[1] includes the capability to identify medical statements from the notes sections within VA clinical documents. The v3NLP is built using GATE[2] pipelines. It includes cTAKES[3] POS tagger and noun phrase parser. The v3NLP recently adopted NLM's MetaMap[4] phrase to concept mapping tool to map to Unified Medical Language System (UMLS) concepts (http://www.nlm.nih.gov/research/umls/). The v3NLP includes a section identification component.

The tool has been further developed to address the 2010 i2b2/VA NLP challenge. This challenge called for the identification of problems, treatments and tests within clinical documents. The challenge also called for the assignment of given assertions associated with medical problems and the identification of relationships between the problem, treatment and tests. The tool was augmented with a statistical machine learning component to improve its performance at identifying problems, treatments and tests, trained on annotated text. The features used included the presence of frequently occurring salient words, concepts and semantic types returned by MetaMap, generalizations of the semantic types, the parts of speech and words around medical concepts, document type and section headings.

For this Pro-WATCH task, the problems found using v3NLP will be further refined down to symptoms, partially by semantic type assignment, but also possibly by the use of an additional machine learning model, trained using annotations from a training corpus currently under development.

A statistical machine learning model will be developed to address the duration component of the symptom identification. The training corpus is expected to have duration assertions. The presence of salient duration words and concepts will be considered some of the features to learn from.

Conclusion

The VINCI team is leveraging the knowledge gained, the methodology, and the software components from the i2b2/VA challenge to address the ProWATCH MUS task. It will further extend v3NLP's capabilities based on machine learning from human annotations.

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 Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R, *et al.* Extracting principal diagnosis, co-morbidity and smoking

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak* 26 July 2006;**6**:30.
2 Cunningham H, Maynard D, Bontcheva K, Tablan V. *GATE: a framework and graphical development environment for robust NLP tools and applications*. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics 2002.
3 Savova GK, Kipper-Schuler K, Buntrock JD, Chute CG. UIMA-based clinical information extraction system. LREC 2008: towards enhanced interoperability for large HLT systems: UIMA for NLP; 2008; Marrakech, Morocco; 2008.
4 Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001, pp. 17–21.

Emerging Health Threats Journal 2011, **4**:s21. **doi:** 10.3134/ehtj.10.021
© 2011 Z Faigen *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Identification and tracking of heat-related illnesses using syndromic surveillance

Z Faigen, I Ajit, S Aslam, and S Adams

Office of Preparedness and Response, Maryland Department of Health and Mental Hygiene, Baltimore, MD, USA
E-mail: zfaigen@dhmh.state.md.us

Objective

This paper describes the use of the electronic surveillance system for the early notification of community-based epidemics, a syndromic surveillance system, to monitor heat-related illnesses throughout the state of Maryland during the summer of 2010.

Introduction

The summer of 2010 in Maryland was characterized by unusually high temperatures. This type of increased and prolonged heat can potentially make residents sick, and extreme exposure can even kill people at highest risk. Numerous deaths throughout the state were attributed to this heat wave. The Maryland Department of Health and Mental Hygiene (DHMH) addressed this public health issue by using public messaging and maintaining constant situational awareness through the electronic syndromic surveillance. Thus, the electronic surveillance system for the early notification of community-based epidemics (ESSENCE) was used to monitor heat-related illnesses throughout the state.

Methods

All acute care hospitals in the state of Maryland ($n = 45$) report data on emergency department (ED) visits to ESSENCE in near real-time manner. Therefore, ED visits for heat-related illnesses were reviewed closely on a daily basis throughout the summer months (1 May to 8 August 2010). DHMH used its syndromic surveillance system as a tool to enhance case finding efforts as well as to monitor levels of heat-related illnesses across the state. Daily reports were provided to public health leaders to enhance situational awareness and inform decision making. The counts of ED visits for heat-related illnesses were also compared with the daily temperature data collected at the Baltimore–Washington International Thurgood Marshall Airport, located in Anne Arundel County, to assess the relationship between air temperature and the number of ED visits for heat-related illnesses.

Results

The age group with the highest number of heat-related illnesses was 65 + years old, with 18–44 years old and 45–64 years old having the next highest number of cases, respectively. The region of Maryland with the highest proportion of cases was Baltimore City. There were more female cases than male cases. The peak of emergency department visits for heat-related illnesses occurred on 7 July 2010, corresponding with a 3 day stretch of extremely high temperatures from 5 July to 7 July 2010, with maximum temperatures ranging from 100°F to 105°F. Overall, there was a positive correlation between air temperature and the number of ED visits for heat-related illnesses (Figure 1).

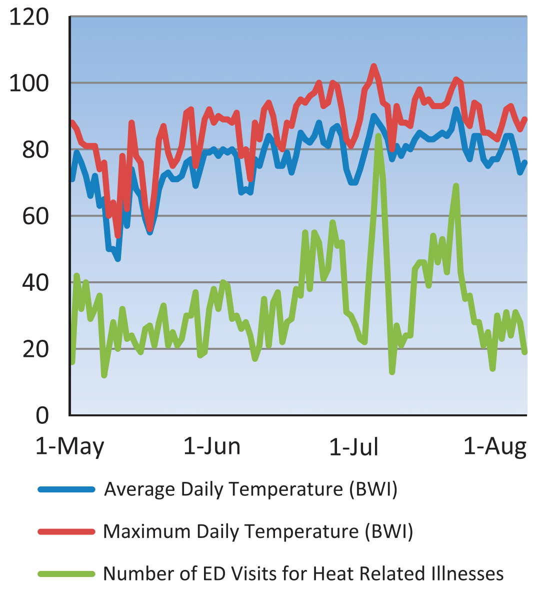


Figure 1 Daily Temperature and Emergency Department (ED) visits for Heart Related illness in Maryland During the Summer of 2010.

Conclusions

Syndromic surveillance proved to be a vital tool in supporting decision makers by providing timely and ongoing

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

situational awareness of heat-related illnesses. The data indicated that there was a positive correlation between air temperature and the number of ED visits for heat-related illnesses. These results demonstrate yet another useful way in which electronic surveillance, and specifically ESSENCE, can be used to help protect the public's health. DHMH is continuing to monitor this data closely to maintain situational awareness of heat-related illness activity in Maryland.

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.

Emerging Health Threats Journal 2011, **4**:s24. **doi:** 10.3134/ehtj.10.024

© 2011 JA Fries *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

The use of Craigslist posts for risk behavior and STI surveillance

JA Fries, A Segre, L Polgreen, and P Polgreen

Computational Epidemiology Research Group, University of Iowa, Iowa, IA, USA
E-mail: jason-fries@uiowa.edu

Objective

This paper describes a novel method of obtaining large scale, geographically diverse behavioral data about Men who use the Internet to seek Sex with Men (MSM) by examining anonymous Craigslist message posts to predict HIV/AIDS.

Introduction

The rise and associated risks of using the internet to find sexual partners among men who have sex with men (MSM) has been noted by many researchers.[1,2] The anonymity and relative ease of finding partners on the internet has facilitated casual sexual encounters that can encompass a variety of unsafe sexual practices, from anonymous partners to 'Party and Play' activities (PNP), slang for illegal drug use, unprotected sex, group sex and so on. These anonymous sexual encounters make it more difficult for public health officials to notify exposed partners.

In addition, detailed data regarding risk behaviors are generally obtained via conventional survey techniques, which are expensive to conduct. Thus, a general method of empirically deriving large scale, location-specific behavioral data could be immensely useful in understanding or anticipating STI outbreaks.

Craigslist is a website specializing in online classified advertisements around the world. Our hypothesis is that Craigslist contains rich behavioral data regarding MSM communities and that such information can function as proxy for external prevalence rates for diseases (that is, HIV/AIDS).

Methods

Beginning 1 July, 2009, daily Craigslist RSS feeds were collected for eight personals categories in 416 local Craigslist sites around the United States. As of 1 September, 2010, 54,450,547 individual posts have been collected. Of the 6,951,603 posts in California, $\sim 57\%$ can be classified as MSM specific. Approximately 60% of these MSM posts can be naively geocoded to specific counties using the Google Maps API. All geocoded messages are then searched to identify two message categories: (1) self-disclosed, positive HIV status, (2) PNP.

California's department of public health provides quarterly statistics on HIV/AIDS cumulative case counts for all counties in the state. March 2010 summary data were used to calculate the prevalence rate used in this analysis.[3]

To find proxies for HIV prevalence in each county in California, we used ordinary least squares (OLSs). The dependent variable in all models was the actual HIV rate in each county (number of HIV cases divided by the county population). Independent variables included broadband penetration by housing unit as well as variables generated from Craigslist. We considered the number of MSM posts, and within the MSM categories, we also considered the number of posts with self-disclosed HIV status and PNP.

Results

MSM posts with HIV status disclosed as a fraction of MSM personal posts is a positive predictor of HIV rates (coeffecient $=5.4$; $P<0.001$; $R^2=0.85$). MSM posts as a fraction of all personal posts is a positive predictor of HIV rates but the R^2 was relatively low (coeffecient $=0.0068$; $P=0.008$; $R^2=0.24$). In contrast, PNP posts as a fraction of all personal posts is a positive predictor of HIV rates with a higher R^2 (coeffecient $=0.56$; $P<0.001$; $R^2=0.75$).

Conclusions

Our Craigslist HIV-positive self-disclosure rate is a proxy for HIV among Californian MSM communities at the county level. Future models will explore this relationship in more depth. The second two models show that there are meaningful behavioral data embedded within messages. When considering just counts of MSM posts, the amount of variation that can be explained is rather modest but including posts denoting high-risk behavior, the R^2 increases considerably—much more of the variability in HIV rates is explained. These results suggest that more sophisticated data mining techniques could yield an important source of behavioral data to help understand and perhaps anticipate STI activity.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

1 McFarlane M, Bull SS, Rietmeijer CA. The internet as a newly emerging risk environment for sexually transmitted diseases. *J Am Med Assoc* 2000;**284:**443–6.

2 Rosser BR, Oakes JM, Horvath KJ, Konstan JA, Danilenko GP, Peterson JL. HIV sexual risk behavior by men Who use the internet to seek Sex with men: results of the MEN'S INTernet Sex Study II (MINTS-II). *AIDS Behav* **13:**488–98.

3 California Department of Public Health, Office of AIDS, HIV/AIDS Surveillance Section. http://www.cdph.ca.gov/data/statistics/pages/oahivaidsstatistics.aspx.

Emerging Health Threats Journal 2011, **4**:s26. **doi:** 10.3134/ehtj.10.026

© 2011 A Haas *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Identifying water contamination from syndromic surveillance signals

A Haas[1], C Dangel[2], SC Allgeier[2], and D Gibbons[1]

[1]*CSC Science & Engineering, Mission Support Group, Alexandria, VA, USA; and* [2]*US EPA Water Security Division/Office of Water, Cincinnati, OH, USA*
E-mail: ahaas2@csc.com

Objective

To develop standard operating procedures (SOPs) to identify or rule out possible water contamination as a cause for a syndromic surveillance alarm.

Introduction

The EPA Water Security initiative contamination warning system (CWS)[1] detection strategy involves the use of multiple monitoring and surveillance components for timely detection of drinking water contamination in the distribution system. The public health surveillance (PHS) component of the contamination warning system involves the analysis of health-related data to identify disease events that may stem from drinking water contamination. Public health data include hospital admission reports, infectious disease surveillance, emergency medical service reports, 911 calls and poison control center (PCC) calls. Automated analysis of these data streams results in alerts, which are investigated by health department epidemiologists. A comprehensive operational strategy was developed to describe the processes and procedures involved in the the initial investigation and validation of a PHS alert. The operational strategy established specific roles and responsibilities, and detailed procedural flow descriptions. The procedural flow concluded with the determination of whether or not an alert generated from surveillance of public health data streams is indicative of a possible water contamination incident.

Methods

Cincinnati was chosen to be the first pilot city for implementation of a drinking water CWS. Over the course of development and deployment of the PHS component, local partners from the Greater Cincinnati Water Works (GCWW), Cincinnati Fire Department (CFD), city and county health departments, PCC and local and federal law enforcement met quarterly to develop and test the PHS alert response strategy. Initially, alerts were received by the health department or PCC. Owing to the branching and looping flow patterns within water distribution systems, water contamination, particularly with extremely toxic chemicals, will likely feature spatial clustering of cases with similar medical complaints. If the underlying data for the alert demonstrates clustering of cases with similar symptoms, the epidemiologist will investigate other public health data streams for corresponding trends. If the investigator believes that water contamination could be the cause of the alert, regardless of whether it is the primary suspected cause, all local partners convene to discuss the alert. GCWW provides information on water quality data and customer complaints from the impacted area. CFD reports or confirms unusual activity from emergency responders. PCC specialists provide information on PCC calls as well as expert toxicological interpretation of the data. On the basis of all information gathered, the utility and local partners make a joint decision regarding whether water contamination is possible.

Results

In addition to routine false alarm investigations, the operational procedures were evaluated during exercises involving simulated water contamination. The exercises illustrated the value of leveraging knowledge from multiple disciplines. Possible water contamination determinations were made in ~90 min based on the correlation of public health alerts with anomalous water quality data. Additionally, PCC specialists assisted GCWW to prioritize its sampling strategy on the basis of suspected contaminants. During exercise debriefs, all partners agreed that the communication resulted in increased confidence in the joint decision that water contamination was possible. Although water contamination is often ruled out for PHS alerts on the basis of initial criteria, in one instance the partners were convened to discuss an alert generated by multiple EMS runs involving respiratory distress in the central business district of Cincinnati. The investigator suspected that the alert was the result of a weekend festival coinciding with a heat wave. However, the alert met the clustering of similar cases criteria.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

After convening a 15 min conference call in which GCWW reported no anomalous water quality or utility conditions in and around the area, the alert was attributed to the heat wave.

Conclusions

The Cincinnati pilot applied an effective procedure for identifying possible water contamination from syndromic surveillance alerts.

Acknowledgements

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

Reference

1 US EPA. Water Security Initiative: Interim Guidance on Developing an Operational Strategy for Contamination Warning Systems. EPA 817-R-08-002 2008.

Emerging Health Threats Journal 2011, **4**:s30. **doi:** 10.3134/ehtj.10.030
© 2011 A Ising *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

The use of NC DETECT ED data to examine heat-related illness

A Ising[1], S Rhea[1], L Deyneka[2], and H Vaughan-Batten[2]

[1]*Department of Emergency Medicine, University of North Carolina, Chapel Hill, NC, USA; and* [2]*Communicable Disease Branch, NC Division of Public Health, Raleigh, NC, USA*
E-mail: isina@med.unc.edu

Objective

To examine the utilization of NC emergency departments for heat-related illness by age, disposition and cause based on chief complaint and triage note categorization.

Introduction

NC DETECT provides near-real-time statewide surveillance capacity to local, regional and state level users across NC with twice daily data feeds from 119 (99%) emergency departments (EDs), hourly updates from the statewide poison center, and daily feeds from statewide EMS runs, select urgent care centers and veterinary lab data. The NC DETECT Web Application provides access to aggregate and line listing analyses customized to users' respective jurisdictions. Several reports are currently available to monitor the health effects of heat waves. Heat wave surveillance is essential as temperature extremes are expected to increase with climate change.[1]

Methods

NC DETECT has three reports related to heat waves: heat-related illness (HRI) by keyword, by ICD-9-CM codes and dehydration (keyword). The ICD-9-CM report captures the most visits for HRI compared with the HRI keyword and dehydration reports. Statewide data were extracted from the ICD-9-CM report for the months of May to September for 2008 ($n=2314$) and 2009 ($n=1671$) and May to July for 2010 ($n=2168$). The burden of HRI ED visits compared with total ED visits was examined by age group (0–9; 10–14; 15–18; 19–24; 25–44; 45–64; 65+) for each month under study. Patient disposition for HRI visits was also examined by age group. Finally, line-listing data were used to assign each HRI ED visit to one of the following categories: exercise/recreation, work, home (maintenance), negligence, no A/C, substance abuse or undefined. Visit categories were examined by age group over the study period.

Results

From May to September 2008 and 2009 and May to July 2010, visits by patients 25–44 years of age reflected the highest proportion of ED visits compared with the other age groups (range of 33.8–45.3%). Children 15–18 years of age had the highest percentage of HRI visits compared with all ED visits within their age group (0.23% average from 2008–2010). Those patients of 65 years and older were more commonly admitted to the hospital after a HRI ED visit than patients from other age groups (Figure 1). Approximately 79% of HRI visits had an undefined cause for the visit (Figure 2). The most common identified causes for all ED HRI visits were exercise/recreation (~10%) followed by work (~8%) and home maintenance (~3%).

Discussion

By conference time, ICD-9-CM comorbidities for heat-related ED visits will be examined and an analysis of NC regional temperatures and heat-related ED visits will be complete.

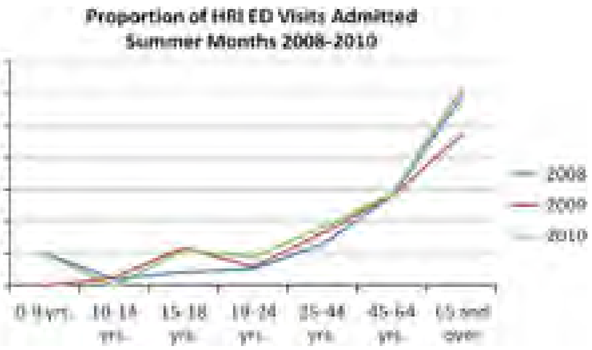


Figure 1 Proportion of HRI ED visits admitted.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

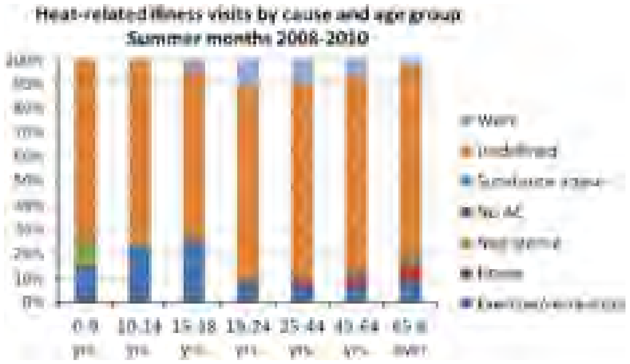


Figure 2 HRI by cause and age group.

Conclusions

NC DETECT receives triage notes for roughly 30% of all ED visits and this data element provide more detail than the chief complaint; however, detailed information about HRI cause was often absent. Of those that are able to be categorized, exercise/recreation is the most common cause of HRI ED visits. Prevention messages targeted to those supervising organized sports as well as those exercising individually may help to reduce these types of visits.

Acknowledgements

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

Reference

1 Luber G, McGeehin M. Climate change and extreme heat events. *Am J Prev Med* 2008;**35**:429–35.

Emerging Health Threats Journal 2011, **4**:s32. **doi:** 10.3134/ehtj.10.032
© 2011 RS King *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Use of syndromic surveillance systems for oral health surveillance

RS King[1], L Deyneka[1], H Vaughan-Batten[1], L Barker[2], V Robison[2], S Benoit[2], and HS Burkom[2]

[1]*North Carolina Department of Health and Human Services, Raleigh, NC, USA; and* [2]*Centers for Disease Control and Prevention, Atlanta, GA, USA*
E-mail: rebecca.king@dhhs.nc.gov

Objective

This paper describes use of national and state syndromic surveillance systems for monitoring and evaluating usage of hospital emergency departments for ambulatory care sensitive dental problems.

Introduction

National and state surveillance systems for oral health have relied on sample-based screenings and self-reported surveys.[1] Recent publications suggest the need and potential for use of data from syndromic surveillance systems and insurers to monitor indicators of oral health status, utilization of care, and costs of treatment.[2,3] Few consensus indicators for oral health derived from these data sources exist, with the exception of a set of five ICD-9 codes comprising ambulatory care sensitive dental problems (ACS-DP).[4] This paper describes North Carolina's Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) data analyzed within CDC's BioSense System to report state and county population-based rates of hospital emergency department (ED) utilization for ACS dental conditions.

Methods

The total number of ED visits to facilities reporting to NC DETECT and the number of visits containing ICD-9 codes 521, 522, 523, 525, and 528 in any of 11 final diagnosis fields were tabulated for 2009. These ICD-9 codes are related to tooth decay, gum diseases, and oral abscesses. Visit records were deduplicated by a system generated unique identifier. The number of ED visits for ACS-DP was tabulated by unique visitor and percentage of repeat visits was calculated. NC DETECT is a statewide early event detection system with 112 out of 112 24/7 EDs reporting in 2009, allowing population-based rates of ED utilization for ACS-DP. Rates per 10,000 population were calculated for the state of North Carolina and each of the 81 counties where facilities are located. Statewide population rates were age adjusted to the NCHS 2000 standard population. County rates may include visits by patients residing in other counties.

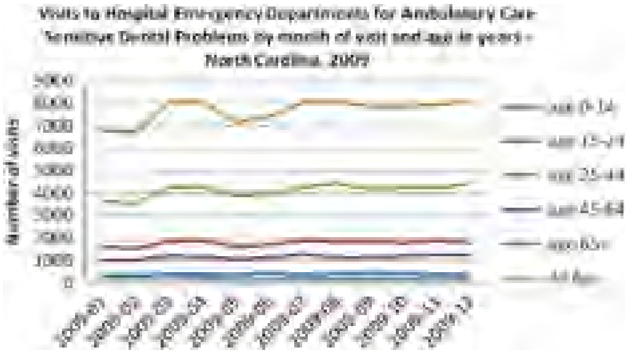


Figure 1 Monthly plot of visits to hospital emergency departments for ambulatory care sensitive dental problems by age – North Carolina, 2009.

Results

County rates per 10,000 population ranged from 2.2 to 3627.7 (median 105.3), with statewide rate of 114.5 (112.2 age adjusted). The greatest utilization was among 25–44 year olds (Figure 1). More detailed results will address the burden of repeat visits and regional and coverage-related effects.

Conclusions

The NC DETECT and CDC BioSense syndromic surveillance systems could be considered for routine surveillance of emergency department use for ambulatory care sensitive dental conditions. Population-based rates can be calculated for areas with high-population coverage among reporting hospital systems; visit-based rates can be calculated for all participating hospital systems.

Acknowledgements

This paper was presented as an oral presentation at the 2010, International Society for Disease Surveillance

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conference, held in Park City, Utah, USA on 1–2 December 2010.

References

1 Malvitz DM, Barker LK, Phipps KR. Development and status of the National Oral Health Surveillance System. *Prev Chronic Dis* 2009;**6**. http://www.cdc.gov/pcd/issues/2009/apr/08_0108.htm.

2 Quiñonez C, Gibson D, Jokovic A, Locker D. Emergency department visits for dental care of nontraumatic origin. *Community Dent Oral Epidemiol* 2009;**37**:366–71.

3 Cohen LA, Bonito AJ, Eicheldinger C, Manski RJ, Macek MD, Edwards RR, *et al.* Comparison of patient visits to emergency departments, physician offices, and dental offices for dental problems and injuries. *J Public Health Dent* 2010 [e-pub ahead of print].

4 Michael Millman (ed). Committee on Monitoring Access to Personal Health Care Services, Institute of Medicine. Access to Health Care in America 1993. National Academy Press: Washington, DC, p 221.

Emerging Health Threats Journal 2011, **4**:s33. **doi:** 10.3134/ehtj.10.033

© 2011 A Kite-Powell *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Public health surveillance after the Haiti earthquake: the Florida experience

A Kite-Powell, J Hamilton, L Eisenstein, and R Hopkins

Florida Department of Health, Bureau of Epidemiology, Tallahassee, FL, USA
E-mail: Aaron_Kite-Powell@doh.state.fl.us

Objective

To describe the public health surveillance efforts designed to measure syndromic and disease-specific conditions in patients who were in Haiti during or after the earthquake, and evacuated to Florida.

Introduction

On 12 January 2010 at 1652 hours local time, an earthquake measuring 7.0 on the Richter scale struck near the Haitian capital of Port-au-Prince and created enormous devastation. Florida, as the state closest to Haiti, became an initial focal point for assisting the federal repatriation and humanitarian parolee efforts. Florida supported shipments of personnel and relief supplies into Haiti, and served as the point of entry for repatriated U.S. citizens and those evacuated from Haiti for medical care. As of 9 February 2010, there were ~22 500 arrivals in Florida from Haiti related to repatriation and medical humanitarian missions. These patients primarily arrived in Miami, Tampa and Orlando areas. Approximately 650 people arrived in Florida during this period as medical evacuees, and were transported to and often directly admitted to hospitals for treatment of severe or complicated injuries. Some of these patients also generated emergency department (ED) visits on arrival.

Methods

Two Florida disease surveillance systems were used to monitor the impact of the Haitian earthquake. (1) ED chief complaint data captured in the statewide Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). On 15 January 15 2010, the Florida Department of Health, Bureau of Epidemiology, requested that hospitals reporting to ESSENCE add the word 'Haiti' to chief complaints of visits to EDs in persons who were in Haiti on or after the earthquake. Free text queries were then written to categorize chief complaints with either 'Haiti' or 'earthquake', along with the reason for visit, into appropriate syndrome categories. The intent of this surveillance was to capture data on all persons presenting to the ED for care, including care provided for U.S. citizens returning from Haiti as well as any foreign nationals who present to Florida EDs for care. (2) A request was made to county health department (CHD) staff to document cases of reportable diseases in the state web-based reportable disease surveillance system (Merlin) for all persons (regardless of whether they were Florida residents). For this event, a case of disease in a person returning from Haiti after the earthquake was defined as: a person (including relief workers) with a reportable disease per Rule 64D-3 Florida Administrative Code, who was in Haiti on 12 January 2010 or had been in Haiti since that date.

Results

From 15 January 2010 to 1 March 2010, 30 hospitals in 11 counties reported 239 patients where Haiti or earthquake was mentioned in the chief complaint. ED visits that were injury-associated accounted for 47.7%, and another 23.8% were classified as possibly infectious disease related. Of these, 57.9% were categorized as gastrointestinal illness, 19.3% included fever, 17.5% as respiratory illness and the remaining 5% fell outside of those categories. In all, 28% of the ED visits with Haiti or earthquake mentioned were not categorized.

In all, 42 reportable diseases were reported to the Florida Department of Health among people returning from Haiti after the earthquake. More than a third of the illnesses were malaria (16). Other diseases and conditions reported include lead poisoning (8), giardiasis (5), dengue fever (2), hepatitis A (2), possible exposure to rabies (2), shigellosis (2), tetanus (2), cryptosporidiosis (1), *Haemophilus influenzae* invasive disease (1) and salmonellosis (1). In several instances these illnesses were identified in individuals who originally presented for care for trauma or injuries. When compared with reportable disease reports associated with travel to Haiti in 2009, these data show a small increase in reports of malaria and lead poisoning.

Conclusions

The findings from these two surveillance systems underscore previous knowledge that earthquakes cause injuries because of the impact and because of the rescue and clean-up

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

activities. In addition, the findings also indicate that due to the pre-earthquake water and sanitation and other environmental issues, persons with travel to Haiti or who are from Haiti are at risk of vectorborne and waterborne diseases such as malaria and giardia. Relatively minor modifications to surveillance systems enabled FDOH to monitor the impact of this event on their communities. This surveillance approach did not cover those persons who were medically evacuated on life-flights and directly admitted to facilities for life-threatening care.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s34. **doi:** 10.3134/ehtj.10.034

© 2011 M Klompas *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Automated surveillance and public health reporting for gestational diabetes incidence and care using electronic health record data

M Klompas[1], J McVetta[2], E Eggleston[1], T Mendoza[3], PR Daly[3], P Oppedisano[3], B Beagan[3], K Golden[3], R Lazarus[2], and R Platt[1]

[1]*Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, USA;* [2]*Channing Laboratory, Brigham and Women's Hospital, Boston, MA, USA; and* [3]*Diabetes Prevention and Control Program, Massachusetts Department of Public Health, Boston, MA, USA*
E-mail: mklompas@partners.org

Objective

To develop an electronic, prospective surveillance system to describe the incidence, care, and complications of gestational diabetes using live electronic health record data from a large defined population.

Introduction

Public health departments have a strong interest in monitoring the incidence, care, and complications of gestational diabetes, as it is associated with poor outcomes for infants and increased risk of diabetes type II for mothers. Gestational diabetes rates are also a possible early marker for changes in the incidence of diabetes type II in the general population. However, diabetes is not generally a reportable condition and therefore, public health surveillance is limited to periodic telephone surveys (subject to self-report inaccuracies), sponsored clinical examinations (expensive, small sample size, no information about processes of care), and occasional research studies. Automated analysis of electronic health record data is a promising method to complement existing surveillance tools with longitudinal, continually updated, clinically rich data derived from large populations. We describe a pilot project to automatically survey electronic health record data in order to identify cases of gestational diabetes, describe their patterns of care and complications, and report summary data to the state health department.

Methods

The Electronic medical record Support for Public Health (ESP) system is a generalizable public health surveillance platform that can continually extract structured data from any electronic health record system, analyze the data for conditions of public health interest, and then transmit individual case reports or summary data to designated public health agencies. ESP was originally designed for notifiable disease detection and reporting, but modules have since been added for syndromic surveillance and vaccine adverse event detection. We created a new ESP module to detect gestational diabetes. The ESP diabetes module was developed within the ESP installation at Atrius Health, a multisite ambulatory practice based in Eastern Massachusetts with over 700,000 patients. Data spanning from June 2006 to March 2010 were queried. A case of gestational diabetes was defined as a pregnant patient with a positive oral glucose-tolerance test per American Diabetes Association standards. We also assessed the added value of ICD9 codes and suggestive prescriptions to augment case detection. The sensitivity and positive predictive value of each case detection method was assessed by full text chart review of a sample of patients. When patients met case criteria for gestational diabetes, ESP summarizes patients' age, race/ethnicity, postal code, referral for medical nutrition therapy, prescriptions for insulin, and postpartum testing for frank diabetes. Results are summarized and stratified by race, age or zip code. Summary data will be presented visually to the state health department via secured website.

Results

Screening for positive oral glucose-tolerance tests in pregnant women detected 706 cases of gestational diabetes between June 2006 and March 2010. Over 2300 patients had ICD9 codes for gestational diabetes during this same period, but on chart review the positive predictive value for this ICD9 code was only 38%. An ICD9 code for gestational diabetes and concurrent prescription for insulin test strips or lancets, however, had a positive predictive value of 91%. Adding this criterion to oral glucose-tolerance test criteria increased case yield by 38%. Analysis of care showed that 70% of patients were referred to

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

medical nutrition therapy, 28% were prescribed insulin, but only 25% received recommended postpartum testing for frank diabetes. Of these, 5% were found to have frank diabetes.

Conclusions

Electronic health record data can provide a comprehensive picture of the epidemiology of a chronic disease including incidence, care, and complications. This has the potential to provide more accurate, timely, complete, and detailed information to public health practitioners compared with existing surveillance methods.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s36. **doi:** 10.3134/ehtj.10.036

© 2011 JN Lauritzen *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Exploring illness prediction in type 1 diabetes mellitus pre-symptom onset

JN Lauritzen[1,2], E Årsand[3], JG Bellika[2,3], K Van Vuurden[1,2], OK Hejlesen[4], and G Hartvigsen[1,2,3]

[1] *Tromsø Telemedicine Laboratory (TTL), University of Tromsø, Tromsϕ, Norway;* [2] *Department of Computer Science, University of Tromsø, Tromsϕ, Norway;* [3] *Norwegian Centre for Integrated Care and Telemedicine (NST), University Hospital of North Norway, Tromsϕ, Norway; and* [4] *Department of Health Science and Technology, Aalborg University, Aalborg, Denmark*
E-mail: jonas@fauerbylauritzen.dk

Objective

To develop an illness prediction model that can predict illness in type 1 diabetes mellitus (T1DM) patients before onset of symptoms, using the patient's observable parameters.

Introduction

Illnesses like infections, cold, influenza and so on in type 1 diabetes mellitus (T1DM) patients, can compromise the daily patient administered diabetes treatment. This in turn may result in fluctuating blood glucose concentrations, especially hyperglycemia for prolonged periods, which over time can cause serious late complications.[1,2] The illness prediction project at Tromsø Telemedicine Laboratory aims to construct a prediction model that, through use of patient observable parameters, for example, blood glucose, insulin injections and body temperature, can significantly identify risk of developing illnesses, before onset of symptoms and before illness onset.

Such a model could potentially enable T1DM patients to fight the illnesses, and prepare for an adequate change in the T1DM-management earlier on.

Methods

By collecting the patient's monitored data concerning the health and illness of T1DM patients, the project group is working towards a classification of a patient's health into three stages: 'healthy', 'in risk of developing illness' and 'ill, with presence of symptoms'. To obtain this data, a Smartphone application was developed, intended for monitoring diabetes treatment (blood glucose concentrations, insulin injections, food intake and physical activity) and general health status (symptoms of illness) for T1DM patients. The application was based on a newer iteration of the Few Touch Application by Årsand *et al.*,[3] but modified to address T1DM patients and additionally collect symptom data. The application was tested initially by two T1DM patients with connection to the project and subsequently refined, in order to prepare the application for use in a large field study.

Results

The results of the initial user study of two T1DM patients showed that the developed application was usable for collecting diabetes and health related parameters and that the users did register symptoms. Although they did not consider the application difficult to use, the users expressed a wish for more predefined parameters to choose from, instead of manually symptom input via text. These findings, along with minor issues have been fixed and the application symptom input features is being expanded to include the following symptoms: headache, abdominal pain, sore throat, joint/muscle pain, presence of fever, degree of fever, presence of cold, feeling weak, illness, influenza and free text input for other observations.

The project group will conduct a study in Seattle in autumn 2010, two studies in Norway 2010–2011 and a study in Denmark 2011. Utilizing the refined application, the project group hopes to achieve the following from these studies:

(1) Provide T1DM patients with a system that offers assistance in the daily patient administered diabetes management, including an overview of current treatment and health situation. (2) Produce data sets for analysis, as input to the development of the illness prediction model. (3) Produce a model that can significantly predict illness in the individual T1DM patients, using available information and patient measurable parameters. (4) Implementing the prediction model in the application and conduct a field study, to assess the patient usefulness of the model.

Conclusions

Developing a model for predicting illness in T1DM patients, before onset of symptoms, may help patients avoid illness and hyperglycemias. To develop a model, user studies centered on data gathering must be conducted to determine if it is possible to uniquely classify a patient as being 'healthy', 'in risk of developing illness' or 'ill, with presence

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of symptoms' at a given time. Such studies are being prepared and will be conducted in near future.

Acknowledgements

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

References

1 Vinik AI, Vinik E. Preventions of the complications of Diabetes. *Am J Manag Care* 2003;**9** (3 Suppl): S63–80.
2 Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. *Phys Ther* 2008;**88:**1254–64.
3 Årsand E, Tatara N, Østengen G, Hartvigsen G. Mobile-phone-basen self-management tools for type 2 diabetes—the few touch application. *J Diabetes Sci Technol* 2010;**4:**328–36.

Emerging Health Threats Journal 2011, **4**:s37. **doi:** 10.3134/ehtj.10.037
© 2011 R Law *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Use of the National Poison Data System for surveillance of human health effects from the Deepwater Horizon Oil Spill

R Law[1], C Martin[2], A Wolkin[1], and A Bronstein[3]

[1]*Centers for Disease Control and Prevention, Chamblee, GA, USA;* [2]*TKC Integration Services, Herndon, VA, USA; and* [3]*American Association of Poison Control Centers, Alexandria, VA, USA*
E-mail: hua1@cdc.gov

Objective

The objective of this study was to describe how National Poison Data System (NPDS) was used for surveillance of human health effects associated with crude oil and dispersant exposures during the Deepwater Horizon Oil Spill.

Introduction

NPDS is the only source for national surveillance data regarding human exposures to hazardous substances and their health effects. It is a near real-time surveillance system operated by the American Association of Poison Control Centers (AAPCC) in cooperation with CDC's National Center for Environmental Health. The system receives, analyzes, and displays data from 60 regional Poison Centers (PCs). On 20 April 2010, an explosion occurred on the Deepwater Horizon oil rig, causing oil to be continuously spilled into the Gulf of Mexico. In response, AAPCC created a code that was sent to all 60 PCs, allowing the centers to identify and properly code all calls associated with the oil spill at the local level. This enabled CDC to track all spill-related exposure and information calls.

Methods

Beginning 30 April, CDC used NPDS to monitor daily oil spill-associated calls (identified by querying the system using the oil spill code) for trend analysis. CDC also used NPDS to identify anomalies through automated analyses of callers' self-reported clinical data, which include 131 pre-coded signs, symptoms, and laboratory abnormalities referred to as clinical effects (CEs). On 4 May, a CE definition was created for the Deepwater Horizon oil spill to identify statistically significant increases in the volume of CEs for the four affected Gulf States (Alabama, Florida, Louisiana, and Mississippi). This CE definition automatically monitored all 131 CEs for the Gulf States over a 24-h window and sent notifications when the observed number of calls for a CE exceeded the historical baseline plus two standard deviations. Anomalies triggered by the CE definition were then reviewed to determine whether they were related to the oil spill, meaning that at least one call was related to the oil spill. We use descriptive statistics to summarize the findings.

Results

From 30 April to 31 July, NPDS identified 1675 calls in total with the oil spill code, 1028 of which were identified as calls reporting a potential exposure. The top 10 reported health effects included headache (247), nausea (165), coughing and choking (114), shortness of breath (109), throat irritation (108), vomiting (86), dermal irritation and pain (74), dizziness (60), diarrhea (52), and rash (40). Oil spill calls peaked for all four Gulf States during the week of 20 June to 26 June, after which there was a steady decline in calls. The percent of calls by state were Alabama (24.5%), Louisiana (24.0%), Mississippi (22.2%), and Florida (21.5%).

From 4 May to 31 July, NPDS identified 88 CE anomalies within the Gulf States. Of these 88, 44 (50%) were determined to be related to the oil spill. The most common of these was headache, which triggered the CE definition 16 times (36%).

Aggregate NPDS data were reported to CDC's Emergency Operations Center and state-specific NPDS case line lists were reported to state PH officials on a daily basis. Monitoring and trend analysis associated with the oil spill is still ongoing.

Conclusions

Summary analyses of oil spill data from NPDS were shared with federal, state, and/or local PH officials to improve situational awareness and inform decision making for interventions as necessary. Collaboration between PH and PC staff was crucial for surveillance success. During the response to the Deepwater Horizon oil spill, NPDS demonstrated utility for conducting near real-time human health effects surveillance associated with a known public health emergency affecting four states.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s138. **doi:** 10.3134/ehtj.10.138
© 2011 MC Rangel *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Adapting a syndromic biosurveillance system to monitor veterans' health impact associated with the gulf coast oil spill

MC Rangel[1,2], RA Martinello[2,3], C Lucero[1,2], G Oda[1,2], M Holodniy[1,2], M Peterson[2], and VJ Davey[2], the VA Gulf Coast Oil Spill Task Force

[1]*Department of Veterans Affairs, Office of Public Health Surveillance and Research, Palo Alto, CA, USA;* [2]*Department of Veterans Affairs, Office of Public Health and Environmental Hazards, Washington, DC, USA; and* [3]*Yale School of Medicine, New Haven, CT, USA*
E-mail: maria.rangel@va.gov

Objective

To describe a surveillance system created to identify acute health issues potentially associated with the Deepwater Horizon oil spill among Veterans in the Gulf of Mexico coastal region.

Introduction

On 20 April 2010, an explosion on an offshore drilling rig in the Gulf of Mexico led to a prolonged uncontrolled release of crude oil. Both clean-up workers and coastal residents were potentially at high risk for respiratory and other acute health effects from exposure to crude oil and its derivatives, yet there was no surveillance system available to monitor these health effects. The Department of Veterans Affairs (VA) conducts routine surveillance for biological threats using the *Electronic Surveillance System for Early Notification of Community Based Epidemics* (ESSENCE).[1] ESSENCE captures specific patient care visit ICD-nine codes belonging to selected conditions that could represent a biological threat. VA operates 153 medical centers and over 1000 free standing patient care facilities across the United States. We describe the adaptation of ESSENCE to allow surveillance of health conditions potentially related to the oil spill.

Methods

VA facilities in the immediate gulf coast region were identified. The ICD-nine codes resulting from patient visits to 34 VA coastal emergency departments and outpatient clinics along the Florida gulf coast and panhandle and coastal areas of Alabama, Mississippi, Louisiana and Texas were downloaded weekly from ESSENCE starting on *MMWR* week 32 (27 June–03 July). ICD-nine codes were grouped into five syndrome categories; respiratory (excluding asthma), asthma, gastrointestinal, heat effects and neurological. Weekly visits counts for each of these syndromes were summarized in tables by state. The current frequency for each syndrome was compared with the frequency during the same week in 2009. Current measures that were significantly above modeling predictions for at least 1 day during the week were flagged as yellow (above 95% CI) or red (above 99% CI) alerts, respectively, for further investigation.

Table 1 Summary of yellow and red alerts detected by ESSENCE at 34 VA gulf coast facilities, 26 June–28 August, 2010

State	*Syndrome*	*Dates*	*Sustained?*
Florida—Gulf Coast	Respiratory	July 4–10	No
Florida—Gulf Coast	Respiratory	July 11–17	No
Louisiana	Asthma	July 11–17	No
Mississippi	Heat Effects	July 25–31	No
Florida—panhandle	Respiratory	August 1–7	No
Louisiana	Gastrointestinal	August 15–21	No
Texas	Respiratory	August 15–21	No
Texas	Asthma	August 15–21	YES
Texas	Asthma	August 21–28	Pending
Mississippi	Respiratory	August 21–28	Pending
Louisiana	Asthma	August 21–28	Pending

Results

To date, there has been one sustained significant increase in the frequency of asthma noted in coastal Texas. Sporadic, non-sustained, increases in respiratory, digestive and asthma syndromes have been identified in other geographic regions and are being monitored (Table 1).

Conclusions

Monitoring acute health effects potentially associated with the Gulf oil spill is critical to provide an understanding of the

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

impact of the spill and improving the ability of the VA to care for those residing and/or working in the gulf coast area. VA has adapted ESSENCE to perform surveillance for medical syndromes to provide an early warning for acute health effects potentially associated with the spill. The ESSENCE modifications were rapidly developed and implemented using existing resources. The modified system is able to provide real-time indication of the health impact because of either natural or man-made environmental crises.

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.

Reference

1 Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems. *MMWR* 2004;**53** (Suppl): 159–65.

Emerging Health Threats Journal 2011, **4**:s39. **doi:** 10.3134/ehtj.10.039

© 2011 SR Mekaru *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Anticipatory surveillance for mass gatherings: a novel application of mass media surveillance

SR Mekaru[1], K Khan[2], N Nelson[3], D Hartley[3], and J Brownstein[1]

[1]*Children's Hospital Boston, Boston, MA, USA;* [2]*St Michael's Hospital, Toronto, ON, Canada; and* [3]*Georgetown University Medical Center, Washington, DC, USA*
E-mail: Sumiko.Mekaru@childrens.harvard.edu

Objective

To present the value of early media-based surveillance for infectious disease outbreaks during mass gatherings, and enable participants and organizers to anticipate public health threats.

Introduction

Public health and medical research on mass gatherings (MGs) are emerging disciplines. MGs present surveillance challenges quite different from routine outbreak monitoring, including prompt detection of outbreaks of an unusual disease. Lack of familiarity with a disease can result in a diagnostic delay; that delay can be reduced or eliminated if potential threats are identified in advance and staff is then trained in those areas. Anticipatory surveillance focuses on disease threats in the countries of origin of MG participants. Surveillance of infectious disease (ID) reports in mass media for those locations allows for adequate preparation of local staff in advance of the MG. In this study, we present a novel approach to ID surveillance for MGs: anticipatory surveillance of mass media to provide early reconnaissance information.

Methods

First, we identify the geographic focus and scope of symptomology for the surveillance system. The ideal anticipatory surveillance system monitors for both diseases and symptoms in all countries of origin of visitors to the MG, preferably well in advance. These locations may be identified through study of historic air travel patterns or travel visa distribution. Local media sources are likely to report early signs of an outbreak and therefore are a critical component of the system. Disease inclusion criteria may be on the basis of virulence, R_0 value, herd immunity, or diversity of serotype distribution. For a novel disease, mass media keywords searches based solely on disease names will miss a potential threat; broad search terms must also be included (for example, 'sick' or 'ill').

Figure 1 Image of alert display for FIFA 2010.

Traditional mass media surveillance systems like HealthMap may examine tens of thousands of internet addresses hourly. Those systems must maintain a high level of specificity to avoid a flood of irrelevant material. For an anticipatory surveillance system, geographic restriction reduces the number of sources examined. After constructing the catalog of media, a multi-step automated search process with lower specificity ultimately achieves greater sensitivity than large-scale mass media surveillance.

Finally, human review of the data set flags potentially important articles.

Results

We have created anticipatory surveillance systems for two MGs: the 2010 Winter Olympics in Vancouver[1] and the 2010 FIFA World Cup in South Africa. We created internet accessible maps displaying potentially important public health alerts from the selected geographic focus (Figure 1). Through inclusion of context important information on the maps (for example stadium locations or high volume countries), these maps provide public health and medical specialists timely, context-rich surveillance intelligence that is normally inaccessible to them.

These anticipatory surveillance systems identified the danger of measles in Oceania before the Winter Olympics;

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

British Colombia experienced a subsequent measles outbreak of the same variant.

Conclusions

Anticipatory surveillance is a logical extension of research in MGs. High volume, automated internet search technology supports a level of surveillance inconceivable at the start of MG studies. Geographic restriction of sources allows for an expanded dictionary of disease related keywords. Multi-step review, both automated and human, improves both sensitivity and specificity of the system.

Acknowledgements

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

Reference

1 Khan K, Freifeld CC, Wang J, Mekaru SR, Kossowsky D, Sonricker AL, *et al.* Preparing for infectious disease threats at mass gatherings: the case of the Vancouver 2010 Olympic Winter Games. *CMAJ* 2010;**182**(6):579–83.

Emerging Health Threats Journal 2011, **4**:s40. **doi:** 10.3134/ehtj.10.040
© 2011 P Mulay *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Enhancement of ESSENCE-based chemical surveillance by incorporation of real-time poison information center data

JL Schauben[1], Sharon M Watkins[2], Carina Blackmore[3], and PR Mulay[4]

[1]*Department of Emergency Medicine, Florida/USVI Poison Information Center, Jacksonville Shands Jacksonville Medical Center, College of Medicine, University of Florida Health Science Center, Jacksonville, FL, USA;* [2]*Florida Department of Health, Analytical Environmental Epidemiology Program, Bureau of Environmental Public Health Medicine, Tallahassee, FL, USA;* [3]*Bureau of Environmental Public Health Medicine, Florida Department of Health, Tallahassee, FL, USA; and* [4]*Florida Department of Health, Bureau of Environmental Public Health Medicine, Tallahassee, FL, USA*
E-mail: Prakash_Mulay@doh.state.fl.us

Introduction

Florida Department of Health (FDOH) has developed a statewide syndromic surveillance system based on the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE). Authorized users can currently access data from the Florida Poison Information Center Network (FPICN), Emergency Room chief complaints (ED), Florida reportable disease system (Merlin) and the Florida death records through ESSENCE under one portal. The purpose of this paper is to summarize efforts to enhance statewide real-time chemical surveillance by incorporating FPICN data into ESSENCE.

Methods

FDOH began by selecting elements from FPICN data fields needed for chemical surveillance. Chemical substances were reviewed and prioritized for real-time monitoring. Case definitions for priority substances/syndromes were created using combinations of clinical effects and laboratory diagnosis. The interim system was tested and released to Florida county health department users on March 2009. Surveillance standards were developed for reportable and non-reportable chemical exposure-related illnesses. The ESSENCE system was upgraded with the implementation of new syndrome categories, as well as advanced query and analytical functions on August 2010. All users are required to complete 3 h of online training before gaining access to FPICN data through ESSENCE.

Results

Priority chemicals are currently monitored through FPICN on a daily basis and results are compared with other data sets in ESSENCE. FDOH monitors about 50 chemical agents in near-real time including carbon monoxide, pesticides, arsenic, mercury, aquatic toxins and food-borne illness. Exposure calls with acute health effects are also monitored in addition to chemical substances. Poison control data have been actively monitored during the Deep Water Horizon event as a key surveillance component. Currently, there are more than 85 users in 31 Florida counties with training and access to FPICN data through ESSENCE.

Challenges addressed during system development included data field selection and mapping between systems, case definition and syndrome alignment, initiation and maintenance of consistent data uploads to ESSENCE and assurance of accurate data transcription between the two systems. Incorporation of the poison center data component into ESSENCE has resulted in a more integrated working relationship between FDOH and FPICN. Both parties worked to address data and quality control issues.

Conclusions

The use of poison center data offers unique opportunities over conventional data sets used for surveillance providing data in real time from a non-hospital source. FPICN data along with ED data was extensively used to identify clusters of carbon monoxide and pesticide poisoning during year 2009–10. Additionally, FPICN data have been used to monitor national emergencies like the H1N1 flu outbreak and the Deep Water Horizon event.

Limitations include non-alignment of substance categories in the disparate systems and differences in definitions of exposure and illness, which limits the ability to determine true exposures from non-exposures and true illness from non-illness, based solely on FPICN call data. The collaboration between FDOH and FPICN provided an opportunity to successfully bridge the differing terminologies and ideologies embedded in each system. ESSENCE-PIC provides an epide-

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

miological interface to the FPICN data with flexibility to compare results among different data sources. Analysis of near-real-time FPICN data provides situational awareness of chemical-related illnesses by rapidly identifying potential clusters of poisoning or injury and allows public health officials to prevent additional morbidity or mortality from occurring in Florida. Florida's experience with real-time electronic surveillance demonstrates that ESSENCE-PIC will help identify seasonal variation and occurrence of unusual chemical events.

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.

Emerging Health Threats Journal 2011, **4**:s41. **doi:** 10.3134/ehtj.10.041
© 2011 P Mulay *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Utilization of Florida poison control data and Emergency Department chief complaint data to identify clusters of carbon monoxide poisoning

P Mulay, S Watkins, and C Blackmore

Bureau of Environmental Public Health Medicine, Florida Department of Health, Tallahassee, FL, USA
E-mail: Prakash_Mulay@doh.state.fl.us

Introduction

The Florida Department of Health (FDOH) previously monitored Florida Poison Information Center (FPICN) data for timely detection of increases in carbon monoxide (CO) exposures before, during, and after hurricanes. Recent analyses have noted that CO poisonings have also increased with generator use and improper heating of homes during cold winter months in Florida. Similarly, increases in CO poisoning cases related to motor vehicles have been observed during summer months. CO is an odorless, colorless, poisonous gas causing sudden illness and death, if present in sufficient concentration in ambient air. The most common signs and symptoms include headache, nausea, lethargy/fatigue, weakness, abdominal discomfort/pain, confusion, and dizziness. This presentation summarizes Florida's experience in identifying CO poisoning clusters using ESSENCE-based syndromic surveillance.

Methods

The ESSENCE-based syndromic surveillance system provides access to data from the FPICN, Emergency Department (ED) chief complaints and the FDOH reportable disease surveillance system (Merlin). ESSENCE was monitored for CO exposures during September 2009 through to August 2010, and all identified CO clusters were followed up and investigated. A cluster was defined as any two or more related cases of CO exposure. All CO cases matching the current case definition for CO poisoning were identified and entered into the FDOH disease reportable system, Merlin. Case ascertainment from these various sources was compared.

Results

The total number of CO clusters identified through ESSENCE-ED and FPICN data before investigation were 34 (88 reports of CO poisoning) and 19 (52 reports of CO poisoning), respectively. A total of 13 clusters were identified by both the sources (total reports of CO poisoning: ED-35 and FPICN-46). Most of the clusters (ESSENCE-ED-15, FPICN-12, and both-2) identified through ED and FPICN data did not contain cases that matched the Florida case definition for CO poisoning and were therefore not counted as cases or confirmed clusters. About 11 clusters reported by ED and FPICN data (ED-4, FPICN-5, and both-2) did not have sufficient patient information for follow-up and were not confirmed.

Among all the reports that contained confirmed cases after investigations, 156 cases (confirmed-118, probable-16, and suspected-22) were reported from 1 September 2009 to 31 October 2010 in Florida. A total of 114 cases (confirmed-86, probable-13, and suspected-15) were associated with 34 different clusters. In all, 42 reported cases were not associated with cluster. Among all the confirmed clusters, ESSENCE-ED data have identified 15 clusters (61 cases), FPICN identified 2 clusters (6 cases), and 9 clusters (29 cases) were identified through both the sources. The remaining 8 clusters (18 cases) were identified by other sources such as hospitals (3), medical examiners' office (2), and a call from citizen/Media (3).

Conclusions

Additional data sources such as FPICN and ED chief complaint data are very efficient in detecting reportable conditions like CO poisoning. Three out of four recent clusters were detected using ESSENCE. Timely investigation and reporting of CO poisoning can be ensured by early notification to the local CHDs.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s50. **doi:** 10.3134/ehtj.10.050
© 2011 AR Reilly *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Application of biosurveillance methodology to the 2010 Asian foot-and-mouth disease outbreaks

AR Reilly[1], EA Iarocci[1], A Endo, DM Hartley, and NP Nelson

Imaging Science and Information Systems Center, Division of Integrated Biodefense, Georgetown University, Washington DC, USA
E-mail: reilly@isis.georgetown.edu

Objective

Our objective is to demonstrate how biosurveillance, using direct and indirect indications and warning (I&W) of disease within vernacular internet news media, provides early warning and situational awareness for infectious animal diseases that have the potential for trade and economic implications in addition to detecting social disruption.[1,2] Tracking of I&W during the 2010 Japan foot-and-mouth disease (FMD) epidemic and outbreaks in other Asian countries was selected to illustrate this methodology.

Introduction

Argus is an event-based, multi-lingual, biosurveillance system, which captures and analyzes information from publicly available internet media. Argus produces reports that summarize and contextualize direct, indirect, and enviroclimatic I&W of human, animal, and plant disease events, and makes these reports available to the system's users. Early warning of highly infectious animal diseases, like FMD, is critical for the enactment of containment and/or prevention measures aiming to curb disease spread and reduce the potential for devastating trade and economic implications.[1–3]

Methods

Argus reports meeting the inclusion criteria identified below were retrieved from the Argus archive: (1) Disease Entities: FMD and undiagnosed animal disease (2) Location: Japan (3) Time Period: January–August 2010. The reports were reviewed for relevant I&W of FMD, with the goal of identifying factors that contributed to disease spread. For geospatial visualization, the location was expanded, and Argus reports were used to create a map of FMD type O outbreaks in East Asian countries.

Results

After 10 years without incidence of FMD, suspected cases were reported in April 2010 among beef cattle in Miyazaki prefecture, Japan.[4] Following the initial identification of cases, unusual events that were socially disruptive occurred, including the cancellation of local festivals, sporting events, and restrictions in public places. These events demonstrate an uncommon event where an animal disease affected the normal functioning of human society.

Despite the implementation of quarantine and prevention measures, disease spread rapidly within the prefecture and Japanese officials theorized that among other factors, delays in early detection greatly contributed to this outcome.[4] In May, Japan implemented its first ever FMD vaccination campaign, suspending its ability to engage in agricultural trade with the U.S. and other countries in accordance with the World Animal Health Organization (OIE) *Terrestrial Animal Health Code*; on 28 May 2010, the United States Department of Agriculture placed import restrictions on beef from Japan.[4,5] With the outbreak reportedly coming to a close, Miyazaki prefecture is estimated to have suffered nearly USD$583–699 million in damages and must endure the disease's stigma.[3–5]

By July, 2010, suspected or confirmed FMD type O outbreaks were reported in at least six East Asian countries: Japan, China, South Korea, North Korea, Taiwan, and Hong Kong, and affected additional animal species, especially swine.[4,5] Causal links among these outbreaks is not confirmed.

Conclusions

Socially disruptive events found in publicly available media suggest that there were I&W of the 2010 Japan FMD type O outbreak. Attention to such I&W in the future may enable more attuned surveillance for early warning of outbreaks. Further, identifying patterns in I&W between the East Asian outbreaks will aid in regional situational awareness to FMD events.

Acknowledgements

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

References

1 Wilson K, Brownstein JS. Early Detection of disease outbreaks using the Internet. *CMAJ* 2009;**180**:829–31.

---

[1] These authors contributed equally to this work.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2 Nelson NP, Brownstein JS, Hartley DM. Event-based biosurveillance of respiratory disease in Mexico 2007-2009 Connection to the 2009 Influenza A(H1N1) Pandemic? *Eurosurveillance* 2010;**15**.
3 The US Department of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services. Foot-and-Mouth Disease Factsheet. February 2007. Available at http://www.aphis.usda.gov/publications/animal_health/content/printable_version/fs_foot_mouth_disease07.pdf.
4 World Organisation for Animal Health. Terrestrial Animal Health Code 2010. Vol. 2. Available at http://www.oie.int/eng/normes/mcode/en_sommaire.htm.
5 World Animal Health Organisation (OIE).

Emerging Health Threats Journal 2011, **4**:s51. **doi:** 10.3134/ehtj.10.051
© 2011 DK Richardson *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Increased emergency department presentations for head trauma following media coverage of a fatal epidural hematoma

DK Richardson[1], M Paladini[2], T Azarian[3], A Baer[4], M Rennick[5], and D Weiss[2]

[1]*New York Presbyterian Hospital, Department of Emergency Medicine, New York, NY, USA;* [2]*New York City Department of Health and Mental Hygiene, New York, NY, USA;* [3]*Duval County Health Department, Duval county, FL, USA;* [4]*Public Health, Seattle and King County, Seattle, WA, USA; and* [5]*Boston Public Health Commission, Boston, MA, USA*
E-mail: dkr9002@nyp.org

Objective

This study describes an increase in head trauma-related visits to emergency departments (ED) in New York City, New York; Boston, Massachusetts; Duval County, Florida; and Seattle, Washington following the widespread media coverage of actress Natasha Richardson's head injury and subsequent fatal epidural hematoma.

Introduction

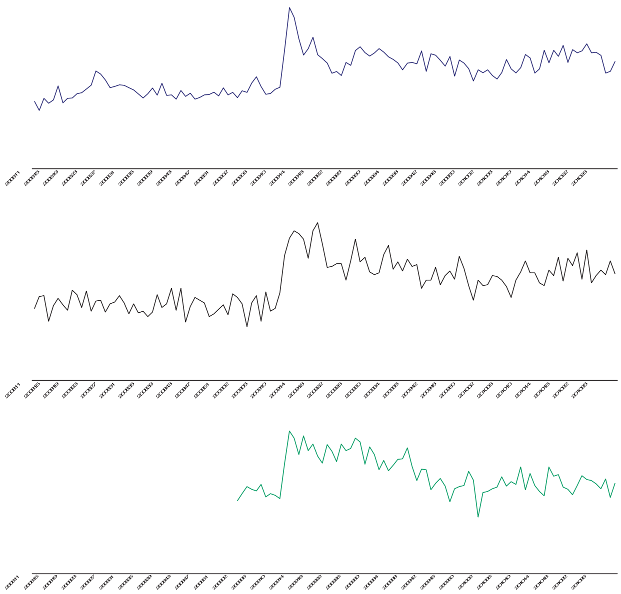
Previous reports have demonstrated the media's influence on ED visits in situations such as dramatized acetaminophen overdose, media report of celebrity suicides, television public announcements for early stroke care and cardiac visits following President Clinton's heart surgery. No previous study has demonstrated the influence of media-publicized trauma on ED visits. On 16 March 2009, the actress Natasha Richardson suffered a traumatic brain injury leading to her death on 18 March; these events were widely publicized by national news sources. The health departments of New York City, Boston, Duval County and Seattle monitor ED visits daily, and capture 95, 100, 100 and 95% of all ED visits, respectively. The data collected include basic demographic information, chief complaint and in some cases ICD-9 diagnosis codes.

Methods

Each health department independently reviewed their ED visit data from January 2008 (when possible) through August 2010 for head injury-related visits. The head injury syndrome consisted of any visit with a chief complaint of 'head trauma,' 'head injury,' 'struck head,' or 'hit head,' or an ICD-9 code of 959.01 (injury to head). Visits meeting the syndrome criteria were compared by week and age. The data were analyzed across all sites by week of visit and age group.

Results

The figure below shows the weekly plot of total ED visits for head trauma 2008–2010; NYC, Duval County, Seattle.



Conclusions

Following this event, a significant increase in ED visits for head trauma was noted in the four sites participating in this study. This surge of presentations for head trauma was most pronounced in pediatric patients, despite the event being the death of an adult, possibly reflecting heightened concern over head-related injuries by parents. EDs should prepare for a surge in visits following news media reports of head trauma, particularly among pediatric populations. While increased awareness of trauma is important to public health,

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

this may need to be balanced with community education about need for emergent hospital-based evaluation. Further studies are needed to see if any increased pathology is found following similar events.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s162. **doi:** 10.3134/ehtj.10.162

© 2011 IN Soyiri and DD Reidpath.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Predicting extreme asthma events in London using quantile regression models

IN Soyiri[1,2], and DD Reidpath[2]

[1]*University of Ghana, School of Public Health, Legon Accra, Ghana; and* [2]*Monash University, School of Medicine and Health Sciences, Kuala Lumpur, Malaysia*
E-mail: soyiriin@yahoo.com

Objective

This paper describes a framework for creating a time series data set with daily asthma admissions, weather and air quality factors; and then generating suitable lags for predictive multivariate quantile regression models (QRMs). It also demonstrates the use of root mean square error (RMSE) and receiver operating characteristic (ROC) error measures in selecting suitable predictive models.

Introduction

The burden of asthma is a major public health issue, and of a wider interest particularly to public health practitioners, health care providers and policy makers, as well as researchers. The literature on forecasting of adverse respiratory health events like asthma attacks is limited. It is an unclear field; and there is a need for more research on the forecasting of the demand for hospital respiratory services.

Methods

Asthma hospital admissions from the hospital episode statistics (HES) database in the UK, weather and air quality measures obtained via the UK Met Office databases were used in the analysis (2005–2006). The key variables in the data were examined for their distribution and properties. Suitable time series lags were generated and converted into non-time series for bivariate quantile regression analysis. Multivariate QRMs were developed to predict extreme asthma events defined as $>$the 90th percentile. RMSE and ROC curves were used to compare error measures of each predictive model.

Results

All the potential predictors were independently significant with asthma daily admissions ($P<0.01$), however most proved to be insignificant in multivariate analysis when controlling for the other factors. Three significant predictive models were constructed: Model-I involved an 18-day lag barometric vapour pressure and a 3-day lag Nitrogen dioxide. Model-II included barometric pressure (18-day lag), Nitrogen dioxide (3-day lag) and air temperature (21-day lag), and then Model-III had barometric pressure (18-day lag), Nitrogen dioxide (3-day lag), air temperature (21-day lag), humidity (4-day lag) and formaldehyde (3-day lag). But for humidity and formaldehyde concentrations, all the variables were at least statistically significant at $P<0.01$ (see Table 1). Model-II had better predictive values for both normal and extreme asthma events compared with the other two

Table 1 QRM predictions of extreme ($>$90th percentile) asthma events in London

Asthma predictive models	*I*		*II*		*III*	
	s.e.	*Coef*$^{\Psi}$	*s.e.*	*Coef*$^{\Psi}$	*s.e.*	*Coef*$^{\Psi}$
Barometric pressure HPa (18-day lag)	−0.18	2.00***	−0.3	1.44***	−0.24	0.85***
Nitrogen dioxide§ (3-day lag)	−7.20e+07	1.8e+08*	−7.50e+07	4.4e+08***	−8.90e+07	2.9e+08**
Air temperature °C (21-day lag)			−0.18	−0.84***	−0.16	−1.11***
Humidity % (4-day lag)					−0.05	−0.04
Formal dehyde§ (3-day lag)					−2.30e+08	3.70e+08
Specification test: (Linktest) *hatsq* *P*-value	0.155		0.922		0.136	

$^{\Psi}$Expected change in log count for a one-unit increase in variable and degree of significance: *$P<0.1$, **$P<0.01$, ***$P<0.001$; §kgm^{-3}; standard error (s.e.).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

models, and again had lower error measures compared with Model-I and Model-III.

Conclusions

Asthma daily admissions can be predicted from a combination of weather and air quality indicators including average daily measures of barometric pressure, Nitrogen dioxide, air temperature, humidity, and formaldehyde using QRMs. Barometric pressure, nitrogen dioxide, and air temperature were the best predictors of asthma daily admissions.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s54. **doi:** 10.3134/ehtj.10.054

© 2011 A Sprecher *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Evaluation of a post earthquake internally displaced persons surveillance system (IDPSS)—Haiti, 2010

A Sprecher[1], C El Bcheraoui[1], M Willis[1], E Jentes[1], R Tohme[1], J Walldorf[1], Beckett G[1], C Dubray[1], S Wiktor[1], S Blank[1], and R Magloire[2]

[1]*Centers for Disease Control and Prevention, Atlanta GA, USA; and* [2]*Ministère de la Santé Publique et de la Population, Port-au-Prince , Haiti*
E-mail: asprecher@cdc.gov

Objective

We evaluated internally displaced persons surveillance system (IDPSS) to determine its suitability for use during a complex humanitarian emergency.

Introduction

On January 12, 2010, a magnitude 7.0 earthquake struck Haiti, killing $>230\,000$ persons and placing an estimated 1.5 million into internally displaced persons (IDP) camps. IDPs are at increased risk for communicable diseases resulting from unhealthy living conditions. The Haitian Ministry of Public Health and Population (MSPP) established the IDPSS to detect outbreaks and characterize disease trends within these camps.

IDPSS gathers data on IDPs seen in clinics run by non-governmental agencies (NGOs). Physician tally sheets are totaled and sent to the MSPP by E-mail or on paper for those without internet connection. Each Monday, analyses of surveillance data through the preceding week are distributed. Reports, guidelines, and forms are sent to MSPP partners and NGOs through the system's Google group (Mountain View, CA, USA), an internet-based discussion forum.

Methods

We assessed the suitability of IDPSS by examining its simplicity, flexibility, acceptability, and timeliness because these criteria were considered the most relevant in an emergency context. We interviewed NGO surveillance managers and Centers for Disease Control and Prevention and MSPP staff, and analyzed IDPSS clinic report submission data for February 18 (date of inception)–April 22. We also reviewed materials available on the system's Google group internet site (reporting forms, case definitions, guidelines, and weekly reports).

Results

IDPSS' use of aggregate data, stratification into only two age groups, ages <5 and $\geqslant 5$ years, and use of case definitions that allowed NGOs to make a reporting determination at the point of care, make the system simple. The flexibility of the system was demonstrated by a transition to a new reporting form, and changes in case definitions and reporting frequency (from daily to weekly) that were all implemented without an interruption of the system. Although IDPSS added work and a redundancy existed between it and the surveillance conducted by certain NGOs, NGO representatives reported that IDPSS was acceptable to them, even modifying their own surveillance systems to integrate better with IDPSS. However, completeness of report submission from NGOs was low and varied by site. Throughout February 18–April 9, a total of 78 clinics submitted at least one daily report. Among these, the median number of reports submitted per site was only three (range, 1–33). The most commonly stated reason for low report submission was the added work required. System managers reported an improvement with the switch to weekly reporting. This also had an effect on timeliness. During the daily reporting period, managers received reports within 1 day and the majority within 1 week. With the switch to weekly reporting, the majority of reports were received each week. However, after 2 months, reports were frequently submitted 2–3 weeks late.

Conclusions

IDPSS provided communicable disease surveillance in the IDP camps well suited for humanitarian emergencies, meeting an international standard of practice. It is timely enough to allow for rapid intervention and flexible enough to respond to the demands of a dynamic situation. Although it is a simple system, its workload diminishes its acceptability to NGOs. Using the Google group improved simplicity, acceptability, and timeliness by providing a common conduit for surveillance communication. Future surveillance systems for humanitarian emergencies should implement similar internet-based forums (for example, the IDPSS Google group) to facilitate communication. Surveillance

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

systems should take advantage of the NGOs' internal surveillance to avoid duplication of effort; however, this requires standardization of surveillance procedures by the humanitarian community.

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.

Emerging Health Threats Journal 2011, **4**:s152. **doi:** 10.3134/ehtj.10.152
© 2011 WE Storm *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Using chief complaint data to evaluate the effectiveness of a statewide smoking ban

WE Storm, BL Bennett, and BE Fowler

Bureau of Infectious Disease Epidemiology and Surveillance, Center for Public Health Statistics and Informatics (CPHSI), Ohio Department of Health, Columbus, OH, USA
E-mail: William.Storm@odh.ohio.gov

Objective

The objective of this study, after completion of the preliminary analysis, was to evaluate whether or not the smoke-free law in Ohio has made a positive change in reducing the effects of secondhand smoke exposure by comparing syndromic surveillance data (trends for emergency department, and urgent care chief complaint visits), related to heart attack and/or acute myocardial infarction (AMI) before and after the smoking ban.

Introduction

In November 2006, Ohioans supported a statute that set into law a requirement that all public places, and places of employment in Ohio prohibit smoking.[1] The law took effect in December 2006; however, the rules for implementation were not finalized until June 2007. The primary purpose of the law was to protect employees in all workplaces from exposure to environmental tobacco smoke. When determining how best to evaluate the health impact of a smoke-free law as it relates to secondhand smoke exposure, most studies have reviewed the incidence of heart attacks or AMIs. In the 2006 Surgeon General's Report, 'The Health Consequences of Involuntary Exposure to Tobacco Smoke,'[2] secondhand smoke exposure is causally associated with cardiovascular events, including AMI. The Institute of Medicine also released a report in 2009 from a meta-analysis, 'Secondhand Smoke Exposure and Cardiovascular Effects: Making Sense of the Evidence,'[3] of 11 epidemiologic studies, reviewing the incidence of acute coronary events following the passing of a smoke-free law. Each of the 11 studies showed a decrease in heart attack rates after implementation of smoke-free laws. The purpose of this study was to evaluate this relationship in Ohio.

Methods

Syndromic surveillance data from hospital emergency department and urgent care chief complaints were collected and analyzed from the state of Ohio's EpiCenter system, for 2005–2010. Although these data types are pre-diagnostic in nature, they are more readily accessible than discharge data. Heart attack and AMI were defined rather specifically in the analysis (chief complaints must have included a reference to heart attack/pain/problems or AMI, and excluded common visits solely for cardiac conduction, or volume concerns, or general respiratory problems). These data were combined and analyzed as a total percentage of visits by month, using SAS v 9.2 (SAS Institute Inc., Cary, NC, USA). Data analyses were performed in 87 of Ohio's 88 counties. Franklin County was excluded from analyses as Columbus, Ohio (located within this county) passed its own smoke-free ban before the state ban.

Results

Figure 1 below shows the trends of total percentage of emergency department and urgent care visits related to heart

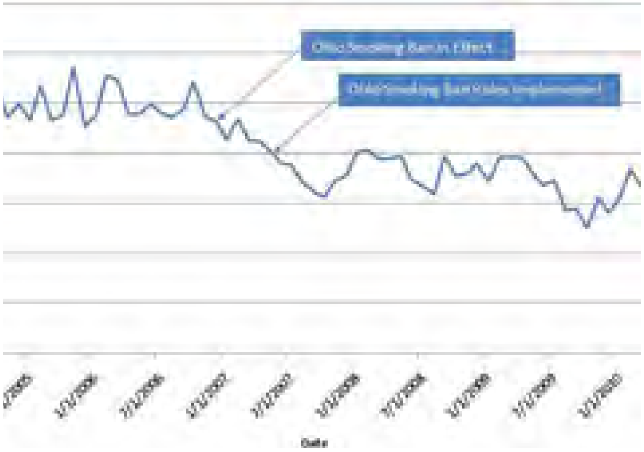


Figure 1 Total Percentage of ED Visits Related to Heart Attack/AMI in All Ohio Counties (excluding Franklin), 2005–2010, by month.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

attack/AMI from 2005–2010 for all Ohio counties, excluding Franklin County. When comparing the means pre- and post-smoking ban, the data showed almost a 30% reduction in mean total percentage of visits for heart attack/AMI post-smoking ban.

Conclusions

On the basis of these results, the data suggest since the smoke-free law in Ohio has been passed, a reduction in the harmful effects of secondhand smoke exposure has also been observed by reducing heart attack and AMI, as defined by pre-diagnostic chief complaint data; however, no causal assumptions can be made. Additional analyses should be completed to further evaluate this relationship, and to control for age and gender of the patients. Further, collection of patient diagnosis from the healthcare facilities would provide strength in validating the observed results.

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 Ohio Administrative Code. Chapter 3701-52 *Clean Indoor Air (Smoking Regulations).* http://www.odh.ohio.gov/rules/final/f3701-52.aspx.
2 A Report of the Surgeon General. 2006 *'The Health Consequences of Involuntary Exposure to Tobacco Smoke.'* http://www.surgeongeneral.gov/library/secondhandsmoke/report.
3 IOM (Institute of Medicine). 2010 *Secondhand Smoke Exposure and Cardiovascular Effects: Making Sense of the Evidence*. The National Academies Press: Washington, DC.

Emerging Health Threats Journal 2011, **4**:s154. **doi:** 10.3134/ehtj.10.154

© 2011 MH Torres-Urquidy *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

FluView interactive: using user-centered design and usability heuristics to improve visualization of influenza information

MH Torres-Urquidy, K Kniss, A McIntyre, L Brammer, and L Finelli

Centers for Disease Control and Prevention, Influenza Division, Atlanta, GA, USA
E-mail: jvu5@cdc.gov

Objective

The objective of this study is to report on the use of User-Centered Design and Usability Heuristics[1] to improve visualization of influenza-related information at the national level. The intention of the prototype is to make data more accessible to different stakeholders including the general public, public health officials at the local and state level, and other experts.

Introduction

Given the periodic nature of influenza activity, it is important to develop visualization tools that enable enhanced decision-making. User-Centered Design is a set of software development methodologies that primarily employ user needs to develop applications.[1] Similarly, Usability Heuristics provide a set of rules that increase the performance of user interfaces, and ease of use.[1] We combined some of these techniques to develop FluView Interactive, a prototype that will enable users to better understand influenza information.

Methods

Our goal was to improve the visualization of nationally reported data about influenza. In particular, we intended to enhance the accessibility of information presented in FluView.[2] FluView is one of the websites maintained by CDC, and it presents information collected by different reporting entities across the United States. First, we identified the need to present data interactively, given the particular needs of different stakeholders. Given the information already available, we conducted interviews with the staff members supporting influenza-like illness surveillance network. From these interviews, we elaborated a Paper prototype for testing (Figures 1 and 2). This prototype was tested with a user who had experience working at local public health department. The development team analyzed the subject feedback, and made modifications to the prototype that addressed the user needs. After these changes, one of the developers evaluated the application using Nielsen's Heuristics,[3] and suggested further modifications that improve usability.

Results

During paper testing, the user was able to identify the level of influenza activity. However, certain quality measures were not clear. In addition, comparing information between seasons was not easily achieved. After modifications, we conducted another evaluation using Usability Heuristics. This led to the inclusion of seasonal information for enabling immediate assessment and evaluation of trends. We included buttons to allow seasonal comparison.

Conclusions

Our prototype experienced various changes, given the different methodologies used. These methodologies allowed us to rapidly create an application while improving information access and reducing development costs.

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. We thank the Center for Disease Control and Prevention's GRASP team, James Jellison, Kang Sang and Venkata Vundi for their technical assistance.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 1 Paper prototype.

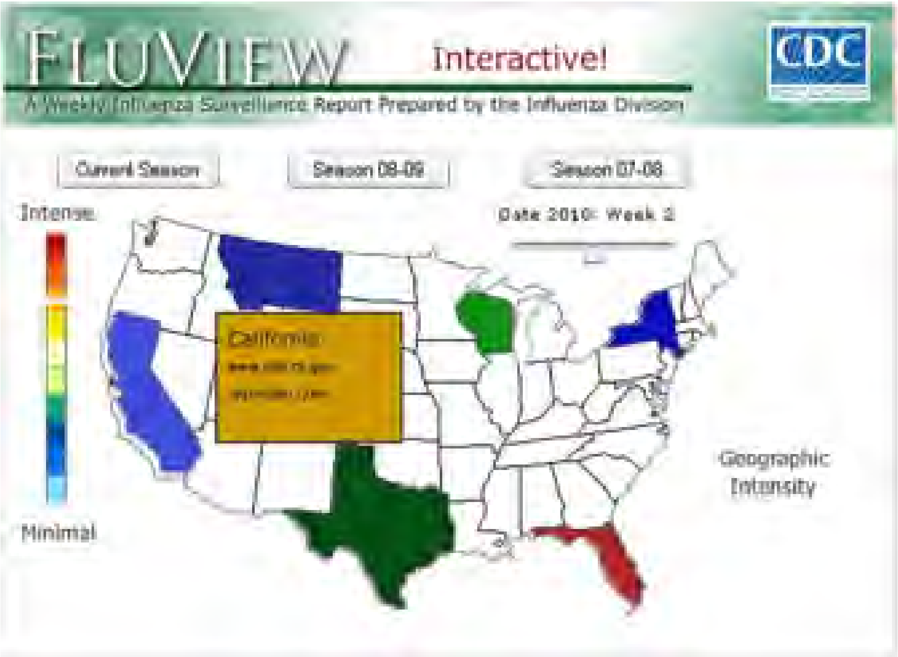


Figure 2 Electronic prototype.

References

1 Kinzie MB, Cohn WF, Julian MF, Knaus WA. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. *J Am Med Inform Assoc* 2002;**9**:320–30.

2 FluView [Internet]. Centers for Disease Control and Prevention: Atlanta (GA). [cited 2010 August [19]. Available from: http://www.cdc.gov/flu/weekly/.

3 Nielsen J. *Usability Engineering*, 1st edn. Academic Press: San Diego, 1993.

Emerging Health Threats Journal 2011, **4**:s58. **doi:** 10.3134/ehtj.10.058
© 2011 C Weerasinghe *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Challenges of introducing disease surveillance technology in developing countries: experiences from India and Sri Lanka

C Weerasinghe[1], N Waidyanatha[2], A Dubrawski[2], and M Baysek[2]

[1]*LIRNEAsia, Colombo, Sri Lanka; and* [2]*Auton Lab, Carnegie Mellon University, Pittsburgh, PA, USA*
E-mail: chamindusampath@gmail.com

Objective

We review challenges faced during the initial period of implementation of a Real-Time Biosurveillance Program (RTBP) in developing countries.

Introduction

Modern information and communication technologies have increasingly prominent roles in health care systems. To capitalize on attainable benefits, it is essential to thoroughly and purposefully weave them into the existing business processes. The challenges of doing so can be exacerbated by specific local circumstances of developing countries. We share our experiences from fielding a system designed to support real-time collection and analysis of public health data in rural areas of Sri Lanka and India. Its strong transformational potential has been proven, however, success of the ultimate field use requires overcoming multiple organizational and utility challenges.

Methods

RTBP comprises of three functional modules: data collection, event detection and analysis, and alerting.[1] Data collection relies on a mobile phone application mHealthsurvey to transfer paper records of outpatient visits to a central database. Data are then analyzed using T-Cube Web Interface (TCWI)—a web browser-based tool for event detection and interactive statistical analysis and visualization. Alerts raised using TCWI are disseminated through the Sahana Alerting Broker to designated responders and health officials via SMS, Email, and Web postings. RTBP business process involves data collection personnel, analysts, and responders. Technology must be tailored to match the capabilities of human resources and fit within limitations of available infrastructure.

Results and conclusion

Health officials were initially exposed to a generic version of TCWI. It provided a wide range of richly parameterized analytic algorithms to interactively process daily loads of data. The RTBP team was quick to realize that health officials in India and Sri Lanka had limited hands-on experience with advanced statistical analyses, despite having learned advanced statistics as part of their curricula. On the other hand, the primary duty of an epidemiologist is to monitor a set of routine hypotheses involving escalating fever diseases, communicable diseases, and unusual clusters of patients reporting with common symptoms. The subsequent revision of TCWI featured a vastly simplified and hence less intimidating user interface. It included one-button invocation of the results of routine surveillance using massive screening and pivot tables functions. This automated process presents the epidemiologist with a daily ranked list of possible disease outbreaks and simple means of interactive evaluation of findings to decide whether any required response or mitigation actions.

Health officials have been comfortable with the pre-existing paper-based disease notification process. It is limited to monitoring, reporting, and responding to some 25 notifiable disease cases, so the comprehensiveness of RTBP imposed a change of the mindset of its users. Initially, it was not easy to sell them on the importance of syndromic surveillance. But, doctors in developing countries examine about 100 patients a day and even experienced physicians cannot fully diagnose more than 20% of their patients. The rest receives uncertain preliminary diagnoses. As unlike the original system RTBP can record syndromes, it can identify escalating clusters of disease despite the limited accuracy of diagnostic data.

Another issue affecting reliability of event detection was the ability of health workers to submit accurate data. Initially, we saw multiple instances of systematic data entry errors such as entering a lexically close but incorrect disease name (whooping cough for worm infestation), misrepresenting health events (reporting toxoid vaccinations as cases of

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

typhoid), and the actual patient visitation date, or delays in data submission combined with errors in case date stamps.

Successful deployment of a sophisticated system like RTBP in a developing country requires a top-down-bottom-up, iterative, use-case-oriented approach. Close collaboration between developers, implementers, and users, and extensive field testing allow identifying challenges, overcoming them, and achieving practical solutions. RTBP is currently used in the field benefiting vulnerable populations in pilot regions of India and Sri Lanka.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010. Work supported in part by the International Development Research Centre of Canada (Award 105130) and by National Science Foundation (Grant 0911032).

Reference

1 Waidyanatha N, Sampath C, Dubrawski A, Prashant S, Ganesan M, Gow G. Affordable system for rapid detection and mitigation of emerging diseases. *Int J E-Health Med Commun* 2010;**2**:73–90.

International Society for Disease Surveillance Conference 2010

Track 4: Informatics

Emerging Health Threats Journal 2011, **4**:s64. **doi:** 10.3134/ehtj.10.064

© 2011 RS Bhatia *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Automatically tracking diabetes using information in physicians' notes

RS Bhatia, S McClinton, and RF Davies

Cardiology Research, University of Ottawa Heart Institute, Ottawa, Ontario, Canada
E-mail: rbhatia@ottawaheart.ca

Objective

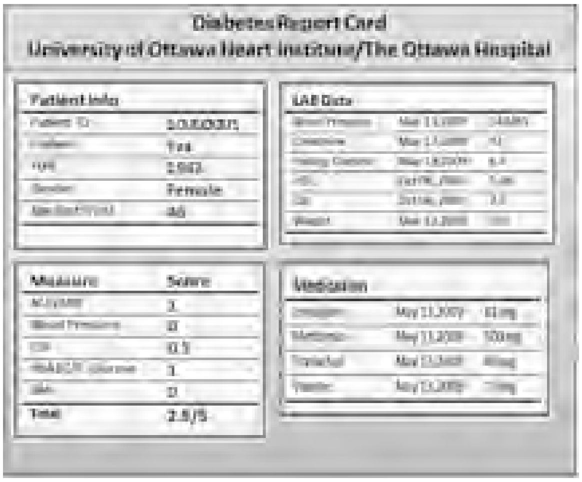
This paper describes DMReporter, a medical language processing system that automatically extracts information pertaining to diabetes (demography, numerical measurement values, medication list, and diagnoses) from the free text in physicians' notes and stores it in a structured format in a MYSQL database.

Introduction

Patient consultations recorded as voice dictations are frequently stored electronically as transcriptions in free text format. The information stored in free text is not computer tractable. Advances in artificial intelligence permit the conversion of free text into structured information that allows statistical analysis.

Methods

DMReporter uses machine learning and natural language processing tools to do information extraction. Numerical measurement values extracted include blood pressure, weight, LDL, HDL, total cholesterol, HbA1C, fasting glucose, glucose (unspecified), and creatinine. These values are present as multiple readings, potential target values, values over a period of time, varied location, values reflecting family history, changes in value versus absolute value, and so on. The methods developed in Bhatia *et al.*[1] extract and disambiguate these values while producing informative label–value pairs. The system extracts 290 medications in six categories using regular expressions and edit distance algorithm. The diagnosis detection uses negation detection in the sentences that mention diabetes using the NegEx algorithm.[2] The dataset used by the program consists of



Diabetes Report Card
University of Ottawa Heart Institute/The Ottawa Hospital

Patient Info

Female

LAB Data

Measure Score

Total 2.5/5

Medication

Figure 1 Diabetes report card.

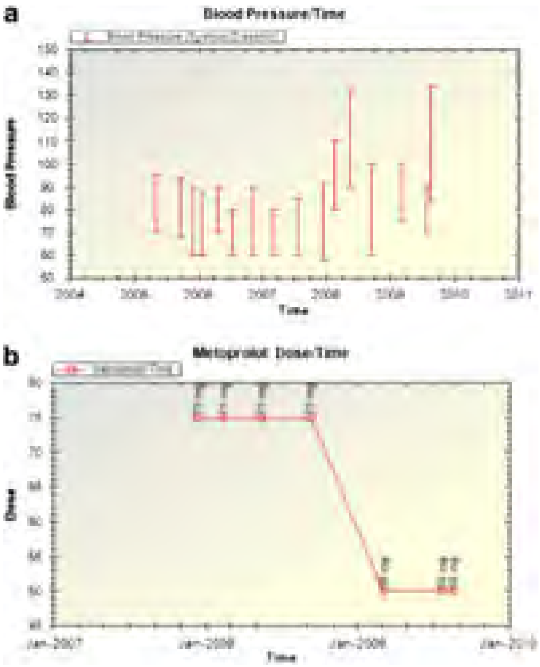


Figure 2 (**a**) Blood pressure recorded over time for a patient. (**b**) Metroprolol doses recorded over time for a patient.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

81 932 transcribed outpatient notes collected from 30 459 patients over 5 years.

Results

Currently, the system automatically extracts diabetes diagnoses, medications along with dosage and frequency information and nine numerical lab values. It produces a report card for every patient (Figure 1), assigns a score out of five representing treatment and health, and tracks each variable extracted over the entire recorded period of patient history (Figures 2a and b) in addition to population reports.

Conclusion

DMReporter allows monitoring of diabetic patients' laboratory values and medications over the entire documented history of the patient. The solutions developed for this project can be applied to documents from other institutions and to other diseases. It demonstrates the ability to develop patient-chart abstractors within a specified restricted domain.

Acknowledgements

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

References

1 Bhatia R, Graystone A, Davies RA, McClinton S, Morin J, Davies RF, *et al*. Extracting information for generating diabetes report card from free text in physicians notes. NAACL HLT 2010 Second Louhi Workshop, 2010, pp 8–14. Available at http://aclweb.org/anthology-new/W/W10/W10-11.pdf.

2 Chapman W, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. Evaluation of negation phrases in narrative clinical reports. *Proc AMIA Symp* 2001, pp 105–14. Available at http://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC2243578/.

Emerging Health Threats Journal 2011, **4**:s67. **doi:** 10.3134/ehtj.10.067
© 2011 H-G Chang *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

The evaluation of influenza surveillance data elements for the health information exchange minimum data set

H-G Chang, J-H Chen, S Jin, and PF Smith

New York State Department of Health, Albany, NY, USA
E-mail: hgc04@health.state.ny.us

Objectives

Evaluate the availability, timeliness, and accuracy of MDS data elements received from one RHIO for emergency department (ED), in-patient, and outpatient visits. Compare the characteristics of patients meeting the HIE influenza-like illness definition who were admitted to the hospital or expired versus those discharged home.

Introduction

The American Health Information Community Harmonized Use Case for the Biosurveillance minimum data set (MDS) was implemented to establish data exchange between regional health information organizations (RHIOs) and the New York State Department of Health (NYSDOH) for accelerating situational awareness through the Health Information Exchange (HIE) Project. However, the completeness, timeliness of the reporting and quality of data elements in the MDS through RHIOs are still unknown and need further validation before we can utilize them for NYSDOH public health surveillance.

Methods

In-patient, outpatient, and ED visit data from one hospital in RHIO A between June and December 2009 were extracted. MDS data elements were evaluated for their availability and timeliness. Patients who had discharge disposition coded as admitted to the hospital or expired were selected as cases, and a random sample of age and patient-class matched patients with disposition coded as discharge home were selected as controls. Using chart review as gold standard, MDS data elements were evaluated for accuracy and completeness. Characteristics of patients with influenza, influenza-like illness, and pneumonia were compared for those admitted or expired versus non-admitted.

Results

There were 10,808, 15,429, and 17,901 patients seen from in-patient, outpatient, and ED, respectively, during the study period. In all, 74 patients from ED visits were admitted to hospitals, 38 died while in the ED, and 15,547 patients were discharged home. There were 11,623 outpatients discharged home, 58 were admitted to hospitals, and one patient died. In all, 301 of the 10,718 in-patients died while hospitalized. The availability of MDS data elements is listed in Table 1. The median days from patient visit date to NYSDOH received date is 1 day for ED, and 2 days for in-patients and outpatients; the mean number of days were 5, 7, and 6

Table 1 Percentage of availability and accuracy of MDS data elements

MDS data elements availability	*In-patient (%)*	*Outpatient (%)*	*ED (%)*
Age	100	100	100
Gender	100	100	100
Zip code	100	100	100
Symptoms onset date	0	0	0
Chief complaint: free text	100	94	100
Diagnosis: free text	100	90	95
Diagnosis: ICD9	97	76	95
Discharge disposition	94	77	96
Temperature	0	0	0
Pulse oximetry	0	0	0
MDS accuracy			
Age	100	100	100
Gender	100	100	100
Zipcode	96	96	87
Diagnosis:ICD9	34	91	93
Discharge disposition	76	94	94

Abbreviations: ED, emergency department; MDS, minimum data set.

Table 2 Influenza patient characteristics by admitted or expired versus discharge home

Characteristics	*Admitted or expired* N *(%)*	*Discharge home* N *(%)*	P-*value*
Mean age	64	54	0.16
Temperature ≥100° F	7 (35)	13 (41)	0.69
Sore throat	2 (11)	4 (12)	0.86
Cough	8 (42)	15 (45)	0.81
Shortness of breath	12 (60)	11 (33)	0.05
Myalgia	0 (0)	6 (18)	0.05
Sepsis	3 (17)	0 (0)	0.04
Pneumonia/ARD	11 (58)	9 (27)	0.03

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

days for ED, outpatient, and in-patient, respectively. After excluding duplicate records, a total of 189 patients admitted to hospitals or expired were selected as cases; 195 patients matched by 5-year age groups and patient class were selected as controls. The accuracy of MDS data elements is listed in Table 1. There were 24 cases and 33 controls had influenza-like illness, 13 cases were admitted to ICU. Characteristics of cases and controls were shown in Table 2.

Conclusion

Most of the MDS data elements were available and complete except onset date, temperature, and pulse oximetry. HIE data (50%) were received within 2 days of visit date. The discrepancy of zip code, diagnosis code, and discharge disposition were due to multiple messages in the HIE data. Influenza patients died or admitted were more likely to be admitted to ICU and to have complications.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s68. **doi:** 10.3134/ehtj.10.068
© 2011 WW Chapman *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Challenges in adapting an natural language processing system for real-time surveillance

WW Chapman, M Conway, JN Dowling, F-C Tsui, Q Li, LM Christensen, H Harkema, T Sriburadej, and JU Espino

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: wendy.w.chapman@gmail.com

Objective

Adapt an existing natural language processing (NLP) system to be a useful component in a system performing real-time surveillance.

Introduction

We are developing a Bayesian surveillance system for real-time surveillance and characterization of outbreaks that incorporates a variety of data elements, including free-text clinical reports. An existing NLP system called Topaz is being used to extract clinical data from the reports. Moving the NLP system from a research project to a real-time service has presented many challenges.

Methods

We describe the approaches we are taking to address the challenges, along with results of the current implementation on findings relevant for Shigellosis and influenza surveillance.

Modeling relevant knowledge

Which diseases should we monitor, what clinical information provides evidence for those diseases, how can we represent the knowledge so that collaborators in public health, clinical medicine, knowledge engineering, and software development have access to consistent information? We are developing an application for ontology/thesaurus-relating diseases of interest to public health practitioners to relevant findings potentially described in clinical reports.[1] The thesaurus provides a single point of reference used to manually design a Bayesian case detection model and represent target concepts the NLP system will extract.[2,3]

Creating an efficient web service for the NLP system

We are receiving real-time HL-7 messages from the University of Pittsburgh Medical Center and must process reports as they become available; however, the NLP system was much too slow for real-time processing. We increased efficiency by only loading a small portion of the UMLS Metathesaurus and by replacing the MetaMap indexing module with an implementation of IndexFinder, which is substantially faster. Moreover, we lacked a communication mechanism between the NLP system and the Bayesian case detector. We created an xml schema based on the concepts in the thesaurus for standardized NLP output that can be parsed by the case detector.

Building a framework for tracking performance and diagnosing errors

Assessing errors and updating the NLP system's lexicon and rules were initially accomplished through a painstaking process involving translating output to tab-delimited files, calculating outcome measures in Excel, identifying reports with errors from the Excel files, and examining sentences causing the errors in separate text documents. We are building a framework to simplify results review that automatically calculates agreement between two xml files (using the schema described previously). To quickly identify errors, a user can click on a cell in the contingency table and find the reports contributing to the cell's number. Clicking on a report brings up the text of the report and shows the annotations made by the human reference standard and by the NLP system.

Results

Using the results review analysis tool we have iteratively identified errors on 53 target concepts and made changes to the NLP system using a development set of 26 laboratory-verified Shigellosis cases and 20 laboratory-verified influenza cases. Positive predictive value (PPV) and sensitivity are 84 and 47% for Shigellosis concepts and 89 and 67% for influenza concepts. Sensitivity is consistently lower than PPV, indicating an incomplete lexicon. We are developing a lexicon-building tool to mine synonyms from clinical documents and the UMLS Metathesaurus.

Conclusions

Implementing an NLP system for a real-time surveillance application is more difficult than we expected it to be, and the

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

challenges are caused less by NLP technicalities than they are by the lack of a robust environment for NLP engineering, adaptation, and improvement. We are developing processes for knowledge engineering and modeling, communication standards for applications using NLP system output, and a dashboard-like interface for performing results review, error analysis, and results tracking. We hope the tools and standards will be generalizable to other surveillance projects and can be used for NLP system adaptation in other domain areas.

Acknowledgements

The project is funded by CDC grants P01 HK000086 and 1U38 HK000063-01. This paper was presented as an oral presentation at the 2010, International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Conway M, Dowling J, Chapman W. *Developing a Biosurveillance Application Ontology for Influenza-Like-Illness*. Proceedings of the OntoLex Workshop: Beijing, 2010.
2 Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001; 17–21. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243666/.
3 Zou Q, Chu WW, Morioka C, Leazer GH, Kangarloo H. IndexFinder: a method of extracting key concepts from clinical texts for indexing. *AMIA Annu Symp Proc* 2003; 763–7. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480259/.

Emerging Health Threats Journal 2011, **4**:s69. **doi:** 10.3134/ehtj.10.069

© 2011 L Chen *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Automated detection of data entry errors in a real time surveillance system

L Chen[1], A Dubrawski[1], N Waidyanatha[2], and C Weerasinghe[2]

[1]*Auton Lab, Carnegie Mellon University, Pittsburgh, PA, USA; and* [2]*LIRNEAsia, Colombo, Sri Lanka*
E-mail: lujiec@andrew.cmu.edu

Objective

We present a method for automated detection of systematic data entry errors in real time biosurveillance.

Introduction

Real-Time Biosurveillance Program (RTBP) introduces modern surveillance technology to health departments in Sri Lanka and Tamil Nadu, India.[1] Triage data from each patient visit (basic demographics, signs, symptoms, preliminary diagnoses) is recorded on paper at health facilities. Case records are transmitted daily to a central database using the RTBP mobile phone application. It is done by medical professionals in India, but in Sri Lanka, due to staffing constraints, the same duty is performed by lower cost personnel with limited domain knowledge. That results in noticeable differences in data entry error rates between the two locations. Most of such issues are due to systematic and subjective misinterpretations of the handwritten doctor notes by the data entry personnel. If not identified and remedied quickly, these errors can adversely affect accuracy and timeliness of health events detection. There is a need to support system managers in their efforts to maintain high reliability of data used for public health surveillance.

Methods

To address the need, we develop algorithms for automated detection of systematic data quality issues. They are used in automated, on-line screening of incoming data for potential discrepancies. Lists of potential issues are presented to human operators for evaluation. The operators may then correct the contents of the database, re-execute analyses, which might have been affected by errors, and follow-up with data entry personnel to prevent similar issues from happening in the future.

One of such algorithms relies on the assumption that most disease outbreaks do not affect population served by just one health facility. We expect neighboring facilities to show at least some level of correlated activity. We use entropy as a measure of uniformity of the geographic distribution of disease. Spatially isolated outbreaks will be characterized with low values of entropy, whereas widespread events will show high entropies. The data used in our experiments contains 62,000 disease cases from 13 locations in Sri Lanka covering 158 disease categories. Each data entry clerk is assigned to a specific health facility. Therefore, systematic subjective data entry errors tend to show as localized patterns ('spikes' or 'dips') with low spatial entropy. Our algorithm exhaustively searches data for such instances.

Results and conclusion

The Figure 1 below presents the results of analysis limited to notifiable diseases observed in Sri Lanka during a specific period of time. Horizontal axis denotes normalized entropy of disease geospatial distribution, vertical axis depicts relative frequency of diseases, and each point corresponds to one notifiable disease. Points in the upper left quadrant of the plot are likely the results of miscoding. These diseases show relatively high numbers of cases combined with a very limited spatial dispersion. Indeed, upon investigation, it was revealed that the spike of *Tetanus* resulted from mistakenly entering immunization records as if they were disease cases. The phantom outbreak of measles resulted from mistakenly recording lexically similar diagnoses of muscle pain. However, the unexpected spike of 150 cases of *fever greater than 7 days* observed at just one location in February and March of 2010, an ailment not typically seen in dry season, was a legitimate health event.

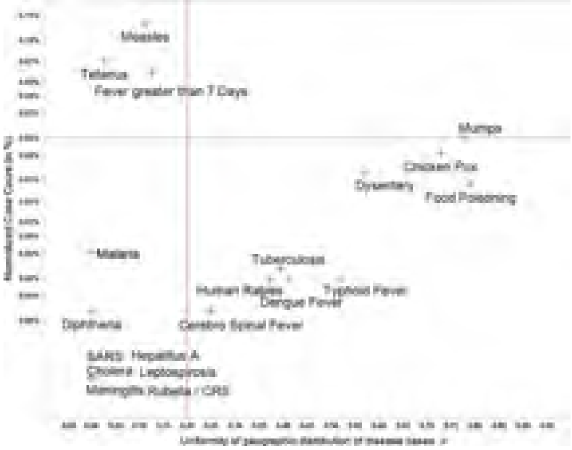


Figure 1 Entropy map of notifiable disease in Sri Lanka.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Low quality of data can easily invalidate analyses. Implementers of analytic systems must be prepared to face such challenges, especially in environments with limited resources. We have shown how to automatically detect certain types of avoidable systematic data entry errors in support of real time biosurveillance.

Acknowledgements

This work was supported in part by the International Development Research Centre of Canada (Award 105130) and National Science Foundation under grant number 0911032. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

Reference

1 Sabhnani M, Dubrawski A, Waidyanatha N. T-Cube Web Interface for Real-time Biosurveillance in Sri Lanka. Proceedings of the 2009 ISDS Annual Conference, Miami, Florida.

Emerging Health Threats Journal 2011, **4**:s124. **doi:** 10.3134/ehtj.10.124

© 2011 YE Cheng *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

PHIN-MS deployment acceptability survey for the data transfer of syndromic data between hospitals and public health

YE Cheng, H Dhotre, and D Drociuk

Division of Acute Disease Epidemiology (DADE), SC Department of Health and Environmental Control (SC DHEC), Columbia, SC, USA
E-mail: chengye@dhec.sc.gov

Objective

The objective of this presentation is to review and evaluate the use of Public Health Information Network (PHIN)[1] Messaging Service (PHIN-MS)[2] for the data transfer of the syndromic data between hospitals and public health. Included is an overview of the methodology used for PHIN-MS, and a review of the usage, adoption, benefits, and challenges within the hospitals and public health agencies in South Carolina. A formal survey is planned with results discussed during the presentation of this manuscript.

Introduction

PHIN-MS can send and receive data securely and automatically. It is used by many hospitals in the state(s) to send data to the South Carolina Department of Health and Environmental Control (SC DHEC) for both our National Electronic Disease Surveillance System and our South Carolina Early Aberration Alerting Network syndromic surveillance system.

DHEC has also implemented the PHIN-MS Route-not-Read (RnR) hub to allow hospitals to easily poll/receive syndromic data reports back from SC DHEC. This enables a two-way communication between the health partners. A basic process flow diagram is shown in Figure 1 below.

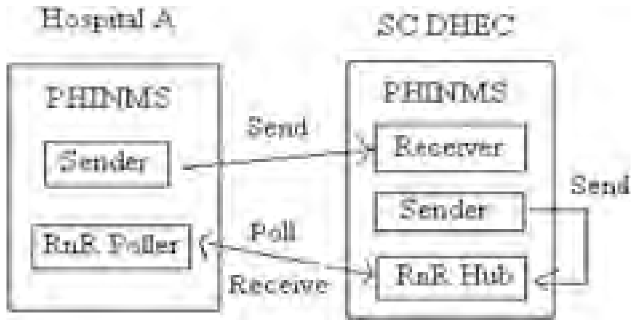


Figure 1 PHIN-MS data transfer diagram.

DHEC began using the PHIN-MS in 2006 for several systems, including the syndromic system beginning in 2008. We initially used PHIN-MS version 2.6, but upgraded to version 2.8.01.

As of August 2010, DHEC was receiving daily syndromic data from 14 emergency departments. These hospitals are receiving data from DHEC using PHIN-MS RnR. Many hospitals are expected to enroll. Positive feedback is normally received, but a formal survey is in process.

Methods

DHEC provides all of the PHIN-MS installation files and technical support to the hospitals, with additional technical support provided by the Center for Disease Control and Prevention in an as needed basis. Most of the intensive IT deployment work is done at SC DHEC rather than at the hospital level, especially with the PHIN-MS RnR hub.

Multiple anecdotal comments regarding the ease of deploying PHIN-MS have been received by SC DHEC. This prompted the creation of a formal survey tool in order to systematically assess the PHIN-MS deployment process. This survey will be provided to hospitals participating in both our National Electronic Disease Surveillance System-based system and the South Carolina Early Aberration Alerting Network syndromic surveillance system.

The survey will focus on the following areas: hospital adoption, satisfaction, challenges, and benefits of using PHIN-MS.

Some questions included in the survey are:

- How satisfied is the hospital using PHIN-MS?
- How easy is the hospital to adopt PHIN-MS as the messaging tool for the data exchange?
- How easy is the PHIN-MS installation process for the hospital?
- How satisfied is the hospital with the technical support provided by SC DHEC?

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- What are the top challenges the hospital has experienced during the implementation?

Results

A summary of the survey will be discussed. Barriers to recruiting hospitals into the system were mostly because of competing priorities within the healthcare setting. South Carolina does not mandate hospitals to report syndromic data. Once a hospital is recruited into the system, sending data through PHIN-MS at the hospital has proved successful so far.

Conclusions

PHIN-MS sends data securely, reliably, and automatically. The hospitals need only very little resources to implement and maintain, especially with the PHIN-MS RnR; health partners can exchange data in a two-way communication securely and easily. PHIN-MS has shown its potential to be the standard for data transfer of syndromic data, as well as several other systems like ELRs, cancer registry, and so on. PHIN-MS will be a valuable tool for increasing the state and local hospital's surveillance systems implementation.

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 The Centers for Disease Control and Prevention. Public Health Information Network. http://www.cdc.gov/phin/.
2 The Centers for Disease Control and Prevention. Public Health Information Network Messaging System. http://www.cdc.gov/phin/phinms/.
3 The Centers for Disease Control and Prevention. The Early Aberration Reporting System. http://emergency.cdc.gov/surveillance/ears/.

Emerging Health Threats Journal 2011, **4**:s72. **doi:** 10.3134/ehtj.10.072
© 2011 KF Comer *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Collaborative development of use cases for geospatially enabling a health information exchange

KF Comer[1], S Wiehe[2], B Dixon[3], R Gamache[3,4], J Wilson[5], and S Grannis[3,4]

[1] *The Polis Center, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA;* [2] *Children's Health Services Research, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA;* [3] *Regenstrief Institute, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA;* [4] *IU School of Medicine, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA; and* [5] *IUPUI Department of Geography, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA*
E-mail: kfrederi@iupui.edu

Objective

This presentation describes a collaborative approach for realizing the public health potential of a geospatially enabled statewide health information exchange.

Introduction

Given the clear relationship between spatial contexts and health,[1] the Indiana Center of Excellence in Public Health Informatics (ICEPHI) aims to serve both the needs of public health researchers and practitioners by contextualizing the health information of large populations. Specifically, ICEPHI will integrate one of the nation's largest health information exchanges, the Indiana Network for Patient Care (INPC),[2] with well-established community information systems that collect, geocode, organize, and present integrated data on communities in Indiana and surrounding states, including data on public safety, welfare, education, economics, and demographics.[3]

Methods

To integrate more than 3 billion clinical results for more than 12 million patients in the INPC with rich, locally available contextual data, an ongoing process for geocoding all clinical data is being developed that links the addresses associated with electronic medical records to geographical coordinates and other useful geographical identifiers including census block group data. To leverage this new data, a collaborative team of information scientists and public health researchers, practitioners, and decision makers is developing use cases that reflect diverse public health needs. These use cases in turn are being used to refine system requirements. ICEPHI anticipated that requirements would vary based on the spatial context relevant to the public health issues of interest[1,4] and targeted sources of contextual information.[5]

Results

The use case requirements differ across different dimensions. These dimensions include accuracy of geocoding results, type of geographic identifier (for example, county, census tract, neighborhood), and date of geographical identifier (for example, 1990 versus 2000 census tract ID). Some use cases, such as targeting neighborhoods for clinical interventions, do not require the same level of geocoding precision as others, such as assessing whether proximity to environmental health hazards relates to an individual's risk of a particular health outcome. Street centerline addressing geocoding is sufficient for the first use case example, while property parcel geocoding is more desirable for the later example. The most commonly needed contextual data and associated requirements for allowing proper linkage were identified for initial prototype development. The prototype geocoding service will return geographical coordinates and block group ID based on the use of a composite geocoding method that uses multiple spatial reference layers. Some use cases will require that geocoding be restricted to the use of a more limited set of reference layers. The developed metadata protocol will allow the source of generated spatial attributes to be tracked and reported.

Conclusions

Enabling geospatial data within health information exchanges has great potential for supporting and advancing public health research and practice. Multi-sector collaboration on the development and evaluation of associated uses cases allows informed decisions on system integration,

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

allowing spatially aware research and practice to be more quickly realized.

The functionality of the ICEPHI geocoding service will be expanded as use cases are further developed and prioritized. Potential uses of de-identified and aggregated health information by community-based organizations will be considered in the future.

Acknowledgements

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

References

1 Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010;**1186:**125–45.
2 Biondich PG, Grannis SJ. The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience. *J Public Health Manag Pract* 2004; (Suppl): S81–6.
3 Hayes RM. Informing our nation: Improving how to understand and assess the USA's position and progress. *Libr Info Sci Res* 2006;**28:**154–7.
4 Mobley LR, Kuo TM, Andrews L. How sensitive are multi-level regression findings to defined area of context?: a case study of mammography use in California. *Med Care Res Rev* 2008;**65:**315–37.
5 Schootman M, Sterling DA, Struthers J, Yan Y, Laboube T, Emo B, *et al.* Positional accuracy and geographic bias of four methods of Geocoding in Epidemiologic research. *Ann Epidemiol* 2007;**17:**464–70.

Emerging Health Threats Journal 2011, **4**:s73. **doi:** 10.3134/ehtj.10.073
© 2011 M Conway *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Developing an application ontology for mining clinical reports: the extended syndromic surveillance ontology

M Conway, J Dowling, R Tsui, and W Chapman

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: michaelambroseconway@gmail.com

Objective

To develop an application ontology—the extended syndromic surveillance ontology (ESSO)—to support text mining of ER and radiology reports for public health surveillance. The ontology encodes syndromes, diagnoses, symptoms, signs and radiology results relevant to syndromic surveillance (with a special focus on bioterrorism).

Introduction

Ontologies representing knowledge from the public health and surveillance domains currently exist. However, they focus on infectious diseases (infectious disease ontology), reportable diseases (PHSkb—retired) and internet surveillance from news text (BioCaster ontology), or are commercial products (OntoReason public health ontology). From the perspective of biosurveillance text mining, these ontologies do not adequately represent the kind of knowledge found in clinical reports. Our project aims to fill this gap by developing a stand-alone ontology for the public health/biosurveillance domain, which (1) provides a starting point for standard development, (2) is straightforward for public health professionals to use for text analysis, and (3) can be easily plugged into existing syndromic surveillance systems.

Table 1 Syndromes represented with example concepts

Syndrome	*Example concept*
Rash	Measles
Hemorrhagic	Hematuria
Botulic	Slurred speech
Neurological	Seizures
Constitutional	Lethargy
Respiratory, sensitive	Sore throat
Respiratory,specific	Pneumonia
Gastrointestinal, sensitive	Dehydration
Gastrointestinal, specific	Vomiting
Influenza-like-illness	Prostration

Note that respiratory and gastrointestinal syndrome are subdivided into *specific* and *sensitive* syndromes.

Table 2 Types of clinical concepts represented (with examples)

Clinical concept type	*Example concept*
Diagnosis	Emphysema
Syndrome	Reactive airways
Symptom	Abdominal pain
Chest X ray finding	X ray pneumonia
Bioterrorism disease	Cholera

Methods

The extended syndromic surveillance ontology (ESSO) uses the syndromic surveillance ontology (SSO) as a starting point. SSO is an ontology of chief complaints and consensus syndrome definitions developed by representatives from 10 currently functional surveillance systems.[1] In developing the ESSO, we tripled the number of concepts represented and added important new relations (for example, has_symptom). All new clinical concepts and relations were identified by an infectious disease physician (JD).

The ontology consists of 300 clinical concepts, each mapped to one (or more) of eight syndromes (see Table 1). Each concept is associated with a concept *type* (for example, diagnosis, bioterrorism disease—see Table 2) and, where possible, mapped to terms from other, related knowledge sources (for example, BioCaster, UMLS). Additionally, we include textual patterns (in the form of regular expressions), which help map clinical concepts to their textual instantiations (and *vice versa*). We encoded this structure as an OWL (web ontology language) file using the Protégé-OWL ontology editor (the current standard for ontology development).

Results

ESSO consists of 300 clinical concepts organized according to their clinical category (for example, *diagnosis*, *symptom*, or

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

radiology finding). Further, these 300 concepts are linked together through a series or relations (for example, DISEASE has_symptom Symptom, SYMPTOM isRelatedTo SYMPTOM). Preliminary evaluation in the influenza-like-illness domain has shown that the ontology has good domain coverage.[2]

Conclusion

The ESSO is a light weight, easily embedded ontology application designed to provide reasoning support for syndromic surveillance of clinical reports using text mining.

Acknowledgements

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

References

1 Okhmatovskaia A, Chapman W, Collier N, Espino J, Buckeridge DL. SSO: the syndromic surveillance ontology. 2009 *ISDS Annual Conference abstract* (unpublished).
2 Conway M, Dowling J, Chapman W. Developing a biosurveillance application ontology for influenza-like-illness. Proceedings of the OntoLex workshop: Beijing 2010.

Emerging Health Threats Journal 2011, **4**:s125. **doi:** 10.3134/ehtj.10.125
© 2011 K Davis *et al.*; licensee Emerging Health Threats Journal.
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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

Emerging Health Threats Journal 2011, **4**:s74. **doi:** 10.3134/ehtj.10.074
© 2011 GB Rattinger *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Clinical decision support at the time of an e-prescription can sustainably decrease unwarranted use of antibiotics for acute respiratory infections

GB Rattinger[1,2], CD Mullins[2], E Onukwugha[2], IH Zuckerman[2], LD Walker[2], B South[3,4], A Gundlapalli[3,4], M Samore[3,4], and S DeLisle[1,2]

[1]*Veterans Administration Maryland Health Care System, Baltimore, MD, USA;* [2]*Departments of Medicine and Pharmaceutical Health Services Research, University of Maryland, Baltimore, MD, USA;* [3]*VA Salt Lake City Health Care System, Salt Lake City, UT, USA; and* [4]*Department of Medicine, University of Utah, Salt Lake City, UT, USA*
E-mail: sdelisle@umaryland.edu

Introduction

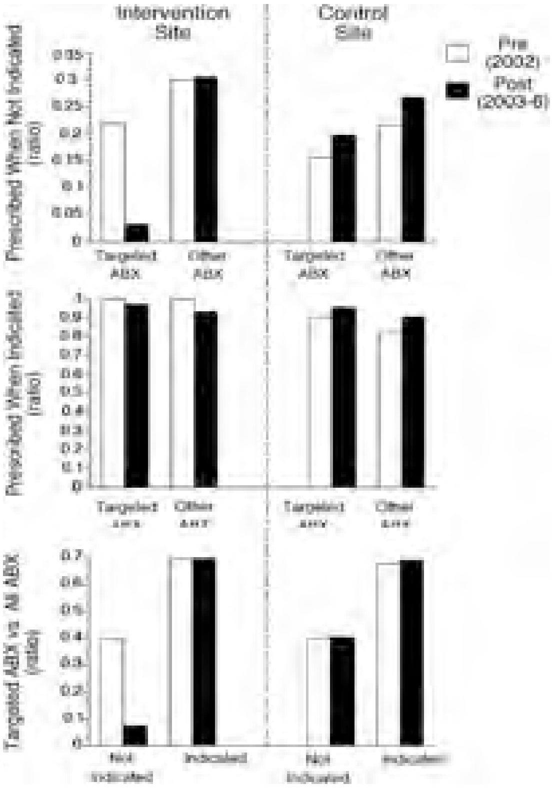
Microorganisms resistant to antibiotics (ABX) increase the mortality, morbidity and costs of infections. In the absence of a drug development pipeline that can keep pace with the emerging resistance mechanisms, these organisms are expected to threaten public health for years to come. Because exposure to ABX promotes the development of bacterial resistance, health care providers have long been urged to avoid using antibiotics to treat conditions that they are unlikely to improve, including many uncomplicated acute respiratory infections (ARI). We asked if interposing clinical decision support (CDS) software at the time of electronic order entry could adjust ABX utilization toward consensus guidelines for these conditions .[1]

Methods

The CDS was programmed in-house and introduced on January 2003 at the Maryland VA site (Intervention) but not at the Utah VA site (control). The CDS targeted two ABX, gatifloxacin and azithromycin (targeted ABX). Access to all other ABX was unrestricted (other ABX). To derive an enriched, population-based sample, we applied a previously validated[2] ARI surveillance algorithm ((one of 197 'respiratory' ICD-9 diagnostic code or new cough remedies or temp$\geqslant$38 °C) and (text analysis of the clinical note for non-negated ARI symptoms)) to EMR entries surrounding all outpatient visits ($n=4.1$ million) during our study period (January 2002–December 2006). Flagged records ($n=7000$) were then manually abstracted on the day of each index visit for all information required to establish guideline-defined ARI diagnoses and ABX treatment.

Results

A total of 3831 unique patients satisfied the consensus case definitions for pneumonias (537), bronchitis (2931), sinusitis (717) and non-specific acute respiratory infections (145). For the two targeted ABX, the proportion of unwarranted prescriptions for ARI decreased from 22 to 3%, pre- (2002) vs post-intervention period (2003–2006, $P<0.0001$). This proportion did not change for the other ABX at the intervention site (30 pre vs 31% post) or for both the targeted (16 vs 20%) and the other ABX (22 vs 27%) at the control site



OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Figure, *upper panel*). The CDS did not impede utilization of antibiotics when they were indicated (*middle panel*), including all cases with pneumonia. Azithromycin and gatifloxacin remained the predominant antibiotics prescribed when indicated for ARI (68–70% of all ABX at both sites, pre and post, *lower panel*).

Conclusion

CDS interposed during the e-prescription process nearly extinguished unwarranted use of targeted ABX for the treatment of ARI for 4 years, and did so without reducing indicated use or shunting misutilization toward unrestricted agents. If deployed at the appropriate scale, this approach could exert a positive impact on public health.

Acknowledgements

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

References

1 Gonzales R, Bartlett JG, Besser RE, Cooper RJ, Hickner JM, Hoffman JR, *et al*. Principles of appropriate antibiotic use for treatment of acute respiratory tract infections in adults: background, specific aims, and methods. (2001). *Ann Intern Med* 2001;**134**:479–86.

2 DeLisle S, South BR, Anthony J, Kalp E, Gundlapalli A, Curriero FC, *et al*. Combining free text and structured electronic medical record entries to detect acute respiratory infection. *PLoS ONE* 2010;**5**(10)**:** e13377.

Emerging Health Threats Journal 2011, **4**:s76. **doi:** 10.3134/ehtj.10.076
© 2011 G Divita *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Standardization to aid interoperability between NLP systems

G Divita[1,2], QZ Treitler[1,2], SM Meystre[1,2], B South[1,2], S Shen[1,2], R Cornia[1,2], J Garvin[1], M Samore[1,2], J Nebeker[1,2], S DuVall[1,2], J Potter[1,2], S Lee[2], P Haug[2], B Bray[2], W Chapman[3], H Harkema[4], L D'Avolio[5], P Elkin[6], M Tuttle[6], G Savova[7], A Coden[8], M Tanenblatt[8], I Sominsky[8], C Clark[9], J Friedlin[10], D Demner-Fushman[11], TC Rindflesch[11], A Aronson[11], AC Browne[11], M Fiszman[11], and O Bodenreider[11]

[1]*VA Salt Lake City Health Care System, Salt Lake City, UT, USA;* [2]*University of Utah, Salt Lake City, UT, USA;* [3]*University of California San Diego, San Diego, CA, USA;* [4]*University of Pittsburgh, Pittsburgh, PA, USA;* [5]*VA Boston Healthcare System, Boston, MA, USA;* [6]*Mount Sinai, New York, NY, USA;* [7]*Harvard Medical School, Boston, MA, USA;* [8]*IBM T.J. Watson Research Center, Hawthorne, NY, USA;* [9]*Mitre Corporation, Boston, MA, USA;* [10]*Regenstrief Institute, Indianapolis, IN, USA; and* [11]*NLM, Bethesda, MD, USA*
E-mail: guy.divita@hsc.utah.edu

Objective

The Consortium for Healthcare Informatics Research, a Department of Veterans Affairs (VA) Office of Research and Development is sponsoring the development of a standard ontology and information model for Natural Language Processing interoperability within the biomedical domain.

Introduction

There are a number of Natural Language Processing (NLP) annotation and Information Extraction (IE) systems and platforms that have been successfully used within the medical domain.[1] Although these groups share components of their systems, there has not been a successful effort[2] in the medical domain to codify and standardize either the syntax or semantics between systems to allow for interoperability between annotation tools, NLP tools, IE tools, corpus evaluation tools and encoded clinical documents. There are two components to a successful interoperability standard: an information and a semantic model.

Methods

Platform-specific information models, such as the UIMA CAS, GATE annotation graphs, and to a lesser degree, Protégé frames and Knowtator,[1] have been adapted to serve as the information model. However, each includes complexity and/or verbosity that has hindered wide adoption. A GATE-lite syntax is under development through this effort, with an *Annotation* at the center of the syntax.

Design principles have emerged around the representation of each *Annotation*: decouple the message syntax from the semantics; design for (space) efficiency; use a standoff annotation model. Create a model is simple enough to be adopted by a wide community, yet be expressive enough to encode a clinical document, a named entity, relationships between entities, and be able to represent temporal features. This model should be usable within annotation tasks, information extraction tasks, and document and corpus evaluation tasks.

Each *Annotation* is intrinsically typed with a tag name or category from a tag set. The naming convention for tags and their semantics are the focus of the other component of the interoperability standard. Semantics have been a stumbling block of other efforts. Under specification of entities such as *Sentence*, *Phrase* and *Token* has led to incomplete integrations at best. This effort includes definition and naming conventions from several established standards to define document structure and named entity semantics. Clinical Document Architecture[3] standard is being used to define clinical document structure, and to represent named entities at the annotation level. This is being augmented with tags from the Penn Treebank parsing and tagging guidelines[4] to define types of phrases, and part-of-speech tags. The model that has emerged includes components such as *Token* and *Term* that have otherwise not been defined within adoptable standards. Elements of these standards have been augmented to allow for relationships between named entities, and for post-coordinated concept coverage within a named entity.

Discussion

A draft should be available for discussion in September 2010. A reference version 0.0 should be available shortly thereafter. This effort is not just an academic exercise for Consortium for Healthcare Informatics Research. The Veteran's Informatics and Computing Infrastructure initiative within

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Veterans Affairs has a real need to have an interlingua between external NLP, annotations and IE systems, as well as for in-house development.

Acknowledgements

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

References

1 Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J of Biomed Inform. Biomedical Natural Language Processing* 2009;**42:**760–72.
2 Ide NM, Suderman K. Bridging the Gaps: Interoperability for GrAF, GATE, and UIMA. *Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP* 2009, pp 27–34. www.aclweb.org/anthology/W/W09/W09-3004.pdf.
3 Dolin R, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, *et al.* HL7 clinical document architecture, release 2. *J Am Med Inform Assoc* 2006;**13:**30–9.
4 Penn Treebank Guidelines. http://www.cis.upen.edu/~treebank.

Emerging Health Threats Journal 2011, **4**:s128. **doi:** 10.3134/ehtj.10.128
© 2011 J Duncan and T Grey.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Integrating medical examiner data in Utah

Jeffrey Duncan, and Todd Grey

Utah Department of Health, Salt Lake City, UT, USA
E-mail: jduncan@utah.gov

Objectives

The objectives of the Utah Medical Examiner Database (UMED) project are:

- To provide a single point of entry for medical examiner pathologists and staff to manage investigation information.
- To develop an operational system that links death certificate, medical examiner, and laboratory data in real time as a resource for epidemiology and public health surveillance.

Introduction

The Office of the Medical Examiner (OME) is a statewide system for investigation of sudden and unexpected death in Utah. OME, in the Utah Department of Health (UDOH), certified over 2000 of the 13,920 deaths in Utah in 2008.

Information from OME death investigations is currently stored in three separate UDOH data silos that have limited interoperability. These three electronic data systems include death certificates, medical examiner investigations, and laboratory results. Without interoperability, OME staff is required to enter the same data into multiple systems. In addition, the process of requesting laboratory analysis and receiving results is paper based, significantly slowing final cause of death determination in a majority of cases.

Epidemiological studies and surveillance activities are hindered by the lack of systems integration in UDOH and often require retrospective data linkage. As an example, in 2005, CDC and the UDOH reported that deaths in Utah caused by drug poisoning from non-illicit drugs had increased fivefold from 1991 to 2003.[1] This significant finding relied on retrospective linkage of death certificates, medical examiner records, and toxicology results to describe the problem.

In 2008, funding from a bioterrorism grant from the US Department of Homeland Security was secured to support development of a unique, integrated system for medical examiner and death certificate data.

Methods

The UMED was designed by staff from Utah's Office of Vital Records and Statistics, OME, and Department of Technology Services. UMED was designed with a three-tier architecture. In the data tier, information from historical medical examiner investigations has been linked with death certificate data to form a foundation database with combined death certificate and investigation information from 1991 to present. Moving forward, deaths entered into UMED will be intrinsically linked to death records in the data layer. Laboratory data is linked in the logic tier where UMED's unique person identifier is used to send and receive requests from the laboratory information system.

For deaths not under OME jurisdiction, funeral directors and non-OME physicians will interact through the web-based electronic death registration system interface to file death certificates. OME staff and physicians will interact through the web-based UMED interface that will provide a single point of entry for investigation, death certificate and laboratory data, eliminating multiple entry.

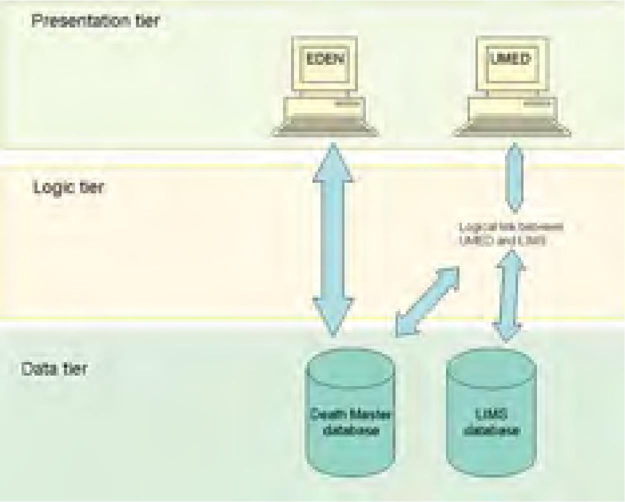


Figure 1 Architectural diagram of the UMED system.

This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The UMED system is currently undergoing internal testing and is scheduled for user acceptance testing in September 2010 (Figure 1).

Conclusions

The UMED system represents a major step forward for epidemiology and surveillance in Utah. The UMED database, containing linked medical examiner and death certificate data from 1991 to the present, is a rich resource for retrospective epidemiological research. Intrinsically linked operational data from UMED will facilitate real-time surveillance by public health programs in the UDOH.

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.

Reference

1 Caravati EM, Grey T, Nangle B, Rolfs RT, Peterson-Porucznik CA. Increase in poisoning deaths caused by non-illicit drugs–Utah, 1991–2003. *MMWR* 2005;**54**:33–6.

Emerging Health Threats Journal 2011, **4**:s77. **doi:** 10.3134/ehtj.10.077
© 2011 W Edwards *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Applying classification and anomaly detection techniques to real-world data

W Edwards[1], A Vaid[2], and I Brooks[1]

[1]*National Center for Supercomputing Applications, Urbana, IL, USA; and* [2]*Champaign-Urbana Public Health District, Champaign, IL, USA*
E-mail: wedwards@ncsa.uiuc.edu

Objective

This paper compares different approaches with classification and anomaly detection of data from an emergency department.

Introduction

Real-world public health data often provide numerous challenges. There may be a limited amount of background data, data dropouts, noise, and human error. The data from an emergency department (ED) in Urbana, IL includes a diagnosis field with multiple terms and notes separated by semicolons. There are over 7000 distinct terms, excluding the notes. Because it begins in April 2009, there is not yet adequate background data to use some of the regression-based alerting algorithms. Values for some days are missing, so we also needed an algorithm that would tolerate data dropouts.

INDICATOR[1] is a workflow-based biosurveillance system developed at the National Center for Supercomputing Applications (NCSA). One of the fundamental concepts of INDICATOR is that the burden of cleaning and processing incoming data should be on the software, rather than on the health care providers.

Methods

There were two major challenges to processing the ED data. First, we needed a way to reduce the vocabulary to a more manageable size. Second, we needed an algorithm that could tolerate a limited amount of baseline data and some data dropouts.

We grouped the terms into six syndromic groups: 'GI-Sensitive', 'GI-Specific', 'Respiratory-Sensitive', 'Respiratory-Specific', 'Flu-Like Illness', and 'Constitutional.'[2] The 'sensitive' groups include a larger set of symptoms than the 'specific' groups. Generated graphs suggested elevated activity in the Respiratory and Flu-like Illness groups around the time of the H1N1 flu outbreak in Fall 2009.

To generate alerts, we used a modified version of CDC's Early Aberration and Reporting System (EARS).[3] EARS uses an Estimated Weight Moving Algorithm (EMWA) to generate alerts. The modified approaches expand the baseline period to 28 days, separate data into weekdays and weekends, and also adjust for the total number of ED visits on a particular day.

For comparison, we grouped the data into Flu-like Illness, Respiratory-Sensitive, Respiratory-Specific, GI-Sensitive, and GI-Specific, and ran the EARS algorithm on the raw data, segregated into weekends and weekdays, and adjusted for total number of visits.

Results

For the Flu-Like Illness Syndrome, the algorithm that separated weekend from weekday data and adjusted for the total number of ED visits was the most sensitive, generating alerts on 19 days (~6% of the total). This pattern also held for the respiratory-sensitive and respiratory-specific groups Figure 1.

Conclusions

This approach has yielded promising results, and in the future, we plan to expand the number of syndromic groups to explore rates of ED activity related to substance abuse and

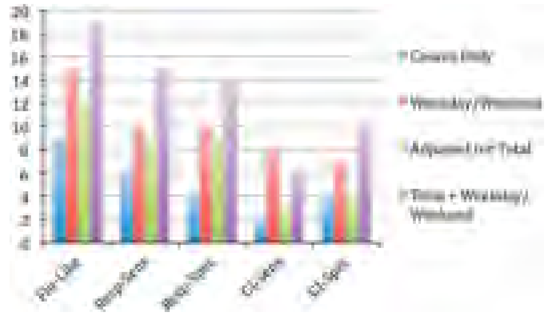


Figure 1 Number of alerts generated for each syndrome group using different modifications of the EARS algorithm. The blue color analyzes only the daily counts for a particular syndrome. The red color separates the counts into weekdays and weekends. The green color adjusts for the total number of ED visits that day. The purple color both separates weekday/weekend data and adjusts for ED totals.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

West Nile virus. The modified EARS algorithms also worked well for us, and we plan to apply them to the school absence data, where enrollment, as well as absence figures have fluctuated.

Acknowledgements

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

References

1 Brooks I, Edwards W. INDICATOR: A Cyberenvironment for biosurveillance and response. *Syndromic* 2009, 8th Annual Conference of the International Society for Disease Surveillance, 2009.
2 Chapman W. Developing Syndrome Definitions based on Consensus and Current Use. *J Am Med Inform Assoc* 2010;**17:** 595–601.
3 Tokars J, Burkom H, Xing J, English R, Bloom S, Cox K, *et al*. Enhancing Time-Series Detection Algorithms for Automated Biosurveillance. *Emerging Infectious Diseases* 2009; **15**:4.

Emerging Health Threats Journal 2011, **4**:s129. **doi:** 10.3134/ehtj.10.129

© 2011 JU Espino and K Bui.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

EpiScape: a map generation service for spatial temporal visualization

JU Espino, and K Bui

Real-time Outbreak and Disease Surveillance Laboratory, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: juest4@pitt.edu

Objective

This paper describes EpiScape, our map generation service. It generates three-dimensional static or animated maps as Keyhole Markup Language (KML) files that can be used to display epidemiologic data over time and space using Google Earth or Google Maps software.

Introduction

The KML format has become a recognized standard for the distribution of geographic information system data.[1] In most recent versions of the Real-Time and Outbreak Disease Surveillance (RODS) system, we standardized on KML as our mapping solution. This decision obviates the need for commercial GIS servers and clients, and permits users to easily overlay RODS map output with other websites and software that output KML, for example, EPA, NASA, and NOAA.

We quickly recognized that the mapping tools in RODS have broad applicability in public health and other domains where there is a requirement to display spatial temporal data as it relates to state, county, and zip code geographies. To facilitate these needs, we created the EpiScape map generation service for public use.

Methods

EpiScape comprises a spatial database, map generation server, and Google Earth. We utilize the open source PostgreSQL database to store the spatial data. The map generation server is implemented in Java Enterprise Edition and makes significant use of the PostGIS Java libraries. The

Figure 1 Stores monitored by the National Retail Data Monitor for over the counter medication sales aggregated by county as of 22 April 1999.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

client for viewing the output of the EpiScape service is Google Earth. Both free and commercial versions of Google Earth are compatible with the EpiScape output.

We programmed the map generation server so that users can define the:

- title for their map
- spatial or spatial temporal data
- labels for each geographic area
- the number for bins and binning method
- a color scheme
- a map type
- transparency levels
- the magnitude of the three-dimensional effect
- the degree of detail for polygons

The service is accessible from a web page or through an http-based application programming interface (API) (API documentation is available at http://betaweb.rods.pitt.edu/wiki/index.php/KML_Generator). For confidentiality, the service deletes map data immediately after the users download their file. In addition, the source code for the service is available from the RODS Open Source Project.

We deployed the EpiScape service on the Apache Tomcat Servlet Engine and Apache web server. The service utilizes SSL encryption when transferring data to and from the user. We loaded the PostgreSQL database with state, county, and zip code tabulation area polygon data from the United States Census.

Results

We use the EpiScape service as the map generation tool for multiple software projects at our laboratory. These projects include the RODS system, National Retail Data Monitor, and Allegheny County influenza monitoring system. Figure 1 is an example map from the EpiScape service showing the number of stores monitored by the National Retail Data Monitor by county.

Conclusions

The EpiScape map generation service has become an indispensable tool at our laboratory. Its web accessibility makes it easy for users to create high quality maps with minimal effort. As it has a web-based API, we have been able to easily incorporate it into other software projects. We hope that users outside our laboratory will find it useful.

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.

Reference

1 http://www.opengeospatial.org/standards/kml/.

Emerging Health Threats Journal 2011, **4**:s79. **doi:** 10.3134/ehtj.10.079

© 2011 DM Falls *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Securing protected health information in NC DETECT

DM Falls, M Li, and A Waller

Carolina Center for Health Informatics, Department of Emergency Medicine, UNC, Chapel Hill, NC, USA
E-mail: dfalls@med.unc.edu

Objective

This paper describes how the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) utilizes various methods of encryption and access control to protect sensitive patient data during both integration and reporting.

Introduction

NC DETECT receives daily data files from emergency departments (ED), the statewide EMS data collection system, the statewide poison center, and veterinary laboratory test results. Included in these data are elements, which may contain Protected Health Information (PHI). It is the responsibility of NC DETECT to ensure that security of these data is managed during their entire life cycle, including receiving, loading, cleaning, storage, managing, reporting, user access, archiving, and destruction. A web interface is provided for users at state, regional and local levels to access syndromic surveillance reports, as well as reports for broader public health surveillance such as injury, occupational health, and disaster management.

Methods

Data files are downloaded via the internet using both Secure File Transfer Protocol (SFTP) and Secure Hyper Text Transfer Protocol (HTTPS). An off-the-shelf Extraction, Transformation and Loading (ETL) tool, capable of receiving data from any of nearly 200 data types including ASCII, HL7 and XML, allows for easy database loading and data encryption. A combination of secure hash algorithm (SHA-1) and triple DES encryption algorithm are used to secure PHI upon database loading. Role based access with multiple tiers controls, data source, geography, aggregate data, line listing data, PHI, and annotation privileges, functionality which allows authorized users to document signals and keep track of signal investigations.

	Hospital Emergency Department	Poison Center	PreMIS
Medical Record Number	Hash & Encrypt	N/A	N/A
Account Number	Hash & Encrypt	N/A	N/A
Notes	N/A	Encrypt	N/A
Last Name	N/A	N/A	Encrypt
First Name	N/A	N/A	Encrypt
Middle Name	N/A	N/A	Encrypt
Incident Address	N/A	N/A	Encrypt
Patient Address	N/A	N/A	Encrypt

Figure 1 Security methods per data source/data element.

Data files are retained for 14 days in a location secured using Windows Encrypted File System (EFS).

Results

NC DETECT currently receives data from 120 hospital-based emergency departments, 100 EMS systems, and the state poison control center. Eight various data elements are either encrypted, hashed or both (Figure 1).

There are 356 active web interface users; 276 are authorized to access limited PHI based on professional role and geographical location.

Conclusion

PHI must be secured for both storage and transportation. NC DETECT's data processing system provides the functionality to meet HIPAA standards for data storage encryption[1] and our Role Based Web Interface provides protection of data being transmitted across the internet.[2] NC DETECT 4.0 provides users at all levels with secure and tailored access to syndromic, injury, post-disaster, occupational health and other types of public health surveillance reports. Role-based

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

access designs must be flexible enough to accommodate changing user needs as well as state and federal privacy and security regulations.

Acknowledgements

Funding for NC DETECT is provided by the North Carolina Division of Public Health. 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 45 CFR Part 142, § 142.308 (c). 'Technical security services to guard data integrity, confidentiality and availability.'
2 45 CFR Part 142, § 142.308 (d). 'Technical security mechanisms.'

Emerging Health Threats Journal 2011, **4**:s80. **doi:** 10.3134/ehtj.10.080
© 2011 AM Fine *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Improved diagnosis of group A streptococcal pharyngitis using real-time biosurveillance

AM Fine[1,2], V Nizet[1,2], and KD Mandl[1,2]

[1]*Informatics and Emergency Medicine, Children's Hospital Boston, Boston, MA, USA; and* [2]*Department of Pediatrics, University of California San Diego, San Diego, CA, USA*
E-mail: Andrew.Fine@childrens.harvard.edu

Objective

The objective of this study was to measure the value of integrating real-time contemporaneous local disease incidence (biosurveillance) data with a clinical score, to more accurately identify patients with Group A Streptococcal (GAS) pharyngitis.

Introduction

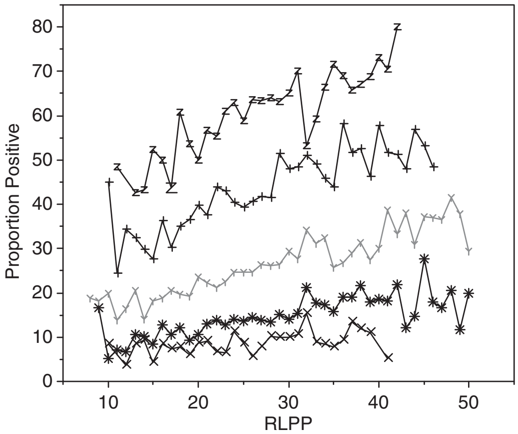
Group A Streptococcal (GAS) pharyngitis, the most common bacterial cause of acute pharyngitis, causes more than half a billion cases annually worldwide. Treatment with antibiotics provides symptomatic benefit and reduces complications, missed work days and transmission. Physical examination alone is an unreliable way to distinguish GAS from other causes of pharyngitis, so the 4-point Centor score, based on history and physical, is used to classify GAS risk. Still, patients with pharyngitis are often misclassified, leading to inappropriate antibiotic treatment of those with viral disease and to under-treatment of those with *bone fide* GAS. One key problem, even when clinical guidelines are followed, is that diagnostic accuracy for GAS pharyngitis is affected by earlier probability of disease, which in turn is related to exposure. Point-of-care clinicians rarely have access to valuable biosurveillance-derived contextualizing information when making clinical management decisions.

Methods

We analyzed data from patients tested for GAS, who presented with pharyngitis from 2006 to 2008 to MinuteClinic, a large national retail health chain. Analysis was restricted to nine markets with >7000 patient visits for pharyngitis, for a total of 138,910 patient encounters across six states. Anonymized extracted data included visit date, location, signs and symptoms included in the Centor score, and pharyngitis test results. To enable integration of contemporaneous, local GAS data with clinical data, we created a biosurveillance variable called the recent local proportion positive (RLPP), a moving window reflecting the proportion of local patients testing positive in the previous week. Patients were grouped by Centor score (0–4) and further categorized by RLPP. We calculated the percent of patients who tested positive for GAS for all combinations of Centor score and RLPP. Using standard metrics (sensitivity, specificity, AUC), we compared the accuracy of the Centor score alone and RLPP alone with the accuracy of a biosurveillance-responsive score that integrated the Centor score and the RLPP to predict which patients tested positive for GAS. We examined the public health effects of subtracting 1 point from the Centor score when the RLPP was below certain thresholds, and adding 1 point when the RLPP exceeded thresholds.

Results

There was no distinct seasonal GAS pattern. For patients with Centor scores of 1–4, represented by the top 4 lines in the figure below, the percent of patients testing positive increases as the RLPP increases ($P<0.0001$). When RLPP >0.30, managing patients with Centor scores of 1 (where the American College of Physicians recommendation is neither



OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

test nor treat) as if their scores were 2 would identify 114,850 previously missed patients who would test positive for GAS each year in the United States while misclassifying 33,161 patients who tested negative. When RLPP <0.20, approaching patients with Centor scores of 3 (where one guideline suggests empiric treatment) as if their scores were 2 would spare unnecessary antibiotics for 78,367 patients while missing 8,195 positives. The AUC is best for the biosurveillance-responsive model incorporating RLPP with the Centor score (0.72), followed by Centor score alone (0.70), and then by RLPP alone (0.57).

Conclusions

Incorporating live epidemiological data into clinical guidelines for GAS should be considered to reduce missed cases when the contemporaneous incidence is elevated, and to spare unnecessary antibiotics when the contemporaneous incidence is low.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s81. **doi:** 10.3134/ehtj.10.081

© 2011 P Gesteland *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Population-based, pathogen-specific surveillance in Utah

P Gesteland[1,2], K Korgenski[2], K Peterson[2], and M Samore[1]

[1]*University of Utah School of Medicine, Salt Lake City, UT, USA; and* [2]*Intermountain Healthcare, Salt Lake City, UT, USA*
E-mail: Per.Gesteland@hsc.utah.edu

Objective

The objective of this study is to describe a system—Germ Watch—that provides information about the regional activity of common communicable infectious diseases.

Introduction

Epidemiological information realized by modern disease-surveillance systems offers great potential for supporting clinical decision-making. Providing health practitioners with population-based, pathogen-specific information about regional communicable infectious disease epidemiology can engender enhanced knowledge about specific pathogens, which may, in turn, lead to improved clinical performance. To enhance the pathogen-specificity of Utah's surveillance system, which includes tracking syndromes[1] and notifiable diseases, we developed a system that tracks microbiologic testing in Utah's largest health care delivery system.

Methods

Setting

Intermountain Healthcare operates 21 hospitals and >100 outpatient clinics. The Intermountain Healthcare Data Warehouse is a seven terabyte (TB) database containing clinical (lab, radiology, meds, vitals) and administrative data from inpatient and outpatient settings.

Data source

Microbiological testing, including molecular diagnostic testing (DFA, PCR, EIA), and cultures ordered during routine clinical care and performed in one of Intermountain Healthcare's microbiology laboratories.

Pathogens currently tracked

Nine respiratory pathogens and seven enteric pathogens for which testing is available and routinely performed (Table 1).

Data warehousing

Lab results are uploaded daily into a dimensional (star) schema in the Intermountain Healthcare Data Warehouse to support *ad-hoc* multidimensional online analytical processing analysis of daily surveillance data. Additional normalized tables are utilized to enable fast web-query performance of the web applications. Data is available at patient, region and organism granularity.

Table 1 Pathogens tracked by Germ Watch and current count of encounters in the database (2002 through Sept. 2010)

Pathogen	*Encounters*
Respiratory	
Influenza A[a]	13,958
Influenza B[a]	2108
Respiratory Syncytial Virus (RSV)	16,072
Human Metapneumovirus (hMPV)	3025
Adenovirus	2517
Parainfluenza (types 1, 2 and 3)	4130
Rhinovirus	7031
Enterovirus	914
Bordetella pertussis[a]	712
Enteric	
Rotavirus	2898
E. Coli 0157H7[a]	973
Giardia lamblia[a]	1503
Cryptosporidium[a]	1286
Salmonella[a]	1160
Shigella[a]	210
Campylobacter[a]	1563

[a]Reportable disease in Utah.

Reporting

We use the IBM COGNOS (International Business Machines Corp. Armonk, New York, USA; http://www-01.ibm.com/software/data/cognos/) platform to automatically generate and disseminate dashboard-style summary reports, and NetCharts (Visual Mining, Inc., Rockville, MD, USA; http://www.visualmining.com/) to provide interactive, web-based graphing. GIS visualizations are made available using Google Maps (Google Inc. 2011. Map data—Europa Technologies, INEGI).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

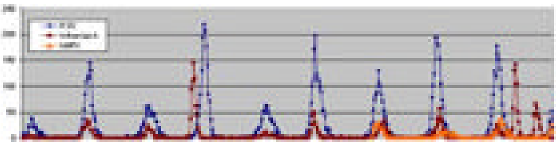


Figure 1 Time-series graph of Germ-Watch data showing yearly epidemics of RSV (blue), Influenza (dark red) and hMPV (orange).

Results

Presently, the Germ-Watch schema contains over 60,000 encounters associated with a lab-detected infection (Table 1). Time-series graphs of these data provide meaningful information about epidemic and endemic activity of various pathogens (Figure 1).[2] Summary reports, along with a bulleted text summary of the week's activity, are e-mailed weekly to >300 physicians and made available on the web. The system has proven sustainable and has been well-received by Utah providers who believe that it helps their clinical performance.[3]

Conclusions

Pathogen-specific information derived from routine diagnostic testing can provide an important signal to population health, when data are available at a population level. Systems that generate and disseminate this information to health practitioners can, and will have an important role in supporting public health practice and clinical decision-making.

Acknowledgements

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

References

1 Gesteland PH, Gardner RM, Tsui FC, Espino JU, Rolfs RT, James BC, *et al.* Automated syndromic surveillance for the 2002 Winter Olympics. *JAMIA* 2003;**10**:547–54.
2 Gesteland PH, Samore MH, Pavia AT, Srivastava R, Korgenski K, Gerber K, *et al.* Informing the front line about common respiratory viral epidemics. *Proc AMIA Symp* 2007; 274–8.
3 Gesteland PH, Allison MA, Staes CJ, Samore MH, Rubin MA, Carter ME, *et al.* Clinician use and acceptance of population-based data about respiratory pathogens: implications for enhancing population-based clinical practice. *Proc AMIA Symp* 2008; 232–6.

Emerging Health Threats Journal 2011, **4**:s82. **doi:** 10.3134/ehtj.10.082

© 2011 PJ Gibson and F Theadore.; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Placing surveillance in a preparedness business process framework

PJ Gibson[1], and F Theadore[2]

[1]*Marion County Health Department, Ocala, FL, USA; and* [2]*CMA Consulting, Kansas City, MO, USA*
E-mail: jgibson@hhcorp.org

Objective

The Public Health Emergency Preparedness Business Process Interdependencies Framework was developed through a collaboration of state and local health departments, brought together to define public health's business processes related to preparedness. This session will explain the framework, the role of surveillance within it, and the methodologies used to develop it.

Introduction

Most public health workers could reach agreement on the general functions and operations regarding public health's traditional operations, such as food safety inspections, outbreak investigations, or immunization clinics. However, there is much less shared understanding about public health's emergency preparedness activities.

A consortium of state and local health departments was brought together to define public health's business processes related to preparedness. That consortium developed the Public Health Emergency Preparedness Business Process Interdependencies Framework, a concise yet comprehensive description of what public health agencies do to prepare for, respond to, and recover from public health emergencies. The framework's 'monitor' group of processes includes *Conduct Syndromic Surveillance*, placing it in context of other emergency preparedness processes.

Methods

This session will present the framework, explain the methodologies used to develop it, and describe how syndromic surveillance and other surveillance processes interact with the rest of the framework. The framework's applicability will be demonstrated through two scenarios—a syphilis outbreak and a health agency's response to H1N1. We will also compare this framework with four existing emergency response frameworks—the National Response Framework,[1] the National Health Security Strategy,[2] the CDC goals for public health emergency preparedness (PHEP) activities, and the PREPARE for Pandemic Influenza Quality Improvement toolkit[3]—and discuss their relative advantages and uses.

We will show how this framework can accommodate local variations in structure, be used as a quality improvement tool, and aid in communicating public health's role to its emergency response partners.

Results

At the end of this session, participants will be able to:

1. Describe a framework that shows how public health agencies operate in public health emergency preparedness;
2. Understand the relative value of this framework in comparison with other emergency response frameworks; and
3. Describe how the disease surveillance process interacts with the other public health processes used in response to public health emergencies.

Conclusion

The framework generalizes well-to-health agencies nationwide, despite local variations in structure. Within the framework, the surveillance processes have an important role in triggering the initial response, as well as in informing continuing response activities. Public health leadership may use the framework to better organize, monitor, prioritize,

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and fund public health preparedness activities. It can also be used as a quality improvement tool to understand and enhance preparedness processes for improved outcomes.

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 National Response Framework. http://www.dhs.gov/files/programs/editorial_0566.shtm.
2 National Health Security Strategy. US Department of Health & Human Services/Assistant Secretary for Preparedness and Response; 2009:43.
3 Lotstein D, Leuschner KJ, Ricci KA, Ringel JS, Lurie N. PREPARE for Pandemic Influenza: A Quality Improvement Toolkit; RAND Health; 2008:217.

Emerging Health Threats Journal 2011, **4**:s133. **doi:** 10.3134/ehtj.10.133
© 2011 KS Fiske *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

EHR-based syndromic surveillance during the 2009–2010 H1N1 pandemic

KS Fiske, M Siefert, and M Hoffman

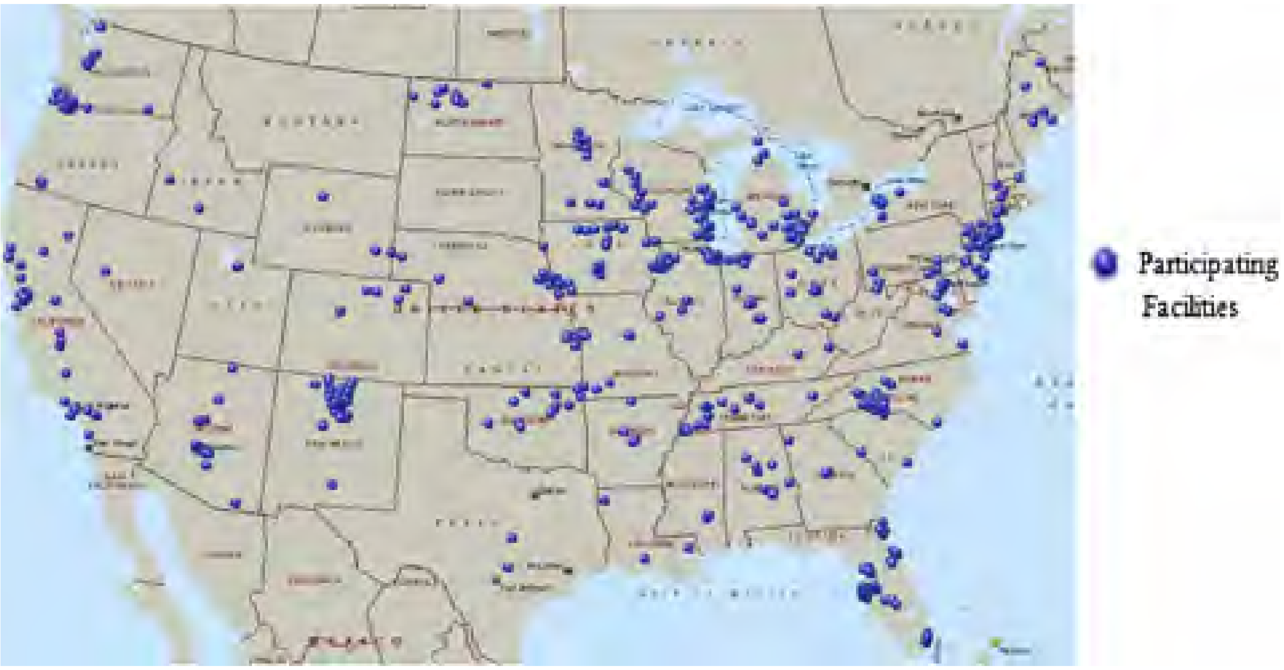
Cerner® Corporation, Kansas City, MO, USA
E-mail: mhoffman1.cerner.com

Objective

This paper describes a national initiative connecting 850 non-affiliated healthcare provider organizations throughout the United States in order to provide situational awareness during the 2009–2010 H1N1 influenza A pandemic. We addressed the challenge of semantic variability between organizations through a centralized data-mapping approach.

Introduction

The global H1N1 influenza A pandemic in 2009 heightened the need for automated disease surveillance capabilities. After an initial surge in confirmatory testing, clinicians moved to diagnosis based on patient assessment for fever combined with cough or sore throat, the influenza-like indicators (ILI). Although some organizations used automated data capture[1] or national systems with manual data entry (www.cdc.gov/flu/weekly/fluactivity.htm), there was not a turnkey national automated system in place to support syndromic surveillance for ILI among non-affiliated organizations. Semantic interoperability through standards utilization is widely expected to simplify large-scale data initiatives but is challenging with widely disparate uses of terminology.



OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Methods

Organizations utilizing Cerner electronic health record (EHR) systems were invited to participate in the HealthAware Flu Initiative. 166 organizations in 48 states representing 850 healthcare provider facilities are currently participating. Local terminologies from all participating organizations were reviewed by Cerner terminology curators. Data elements associated with ILI were mapped to a core vocabulary. Participating organizations installed an automated query that ran every 24 h and summarized the prevalence of patients with ILI, diagnostic test orders and positive influenza A results. These data, with limited demographical information and no personal health identifiers, were uploaded to the Cerner data operations team, aggregated and published to participating health organizations daily. Local (34), state (33) and national public health organizations (CDC) also received access to the system.

Results

The reference data models from the 166 organizations demonstrated widely disparate terminologies for the key ILI data elements. For example, 88 participating systems captured 'cough' using 101 potential prompt labels, with one organization using 17 different prompts. Fever was captured by 75 organizations using 72 potential prompts, whereas sore throat was captured by 26 organizations with only nine possible prompts. The responses to these prompts varied widely. By 1 September 2009, 57 million patient encounters were monitored by this system for either syndromic or laboratory indications.

Conclusions

The HealthAware initiative implemented a large-scale nationwide network to monitor indicators of influenza. Participating organizations included ambulatory clinics, private hospitals, pediatric hospitals, academic medical centers and smaller regional medical centers. We found limited use of standard terminology amongst these organizations, highlighting the value of human-mediated curation until standards become more widely adopted.

Acknowledgements

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

Reference

1 Schirmer P, Lucero C, Oda G, Lopez J, Holodniy M. Effective detection of the 2009 H1N1 influenza pandemic in the US Veterans Affairs medical centers using a national electronic biosurveillance system. *PLoS One* 2010;**5**:e9533.

Emerging Health Threats Journal 2011, **4**:s134. **doi:** 10.3134/ehtj.10.134
© 2011 Y Jiang and CJ Staes.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Virtual public health informatics training: development and delivery of an online public health informatics course

Y Jiang, and CJ Staes

University of Utah, Salt Lake City, UT, USA
E-mail: yuling.jiang@utah.edu

Objective

This poster describes the development and delivery of an online American Medical Informatics Association (AMIA) 10 × 10 Public Health Informatics course at the University of Utah.

Introduction

Public health informatics is an emerging interdisciplinary field that uses information technology and informatics methods to meet public health goals. To achieve these goals, education and training of a new generation of public health informaticians is one of the essential components. AMIA's 10 × 10 program aims to realize the goal of training 10,000 health care professionals in applied health and medical informatics by the year 2010.[1] The Department of Biomedical Informatics of the University of Utah was established in 1964. As one of the largest biomedical informatics training programs in the world, the department is internationally recognized as a leader in biomedical informatics research and education.[2] The poster hereby describes the collaborative effort between Utah and AMIA to develop a public health informatics online course.

Methods

The course was developed based on the classroom version of the Public Health Informatics course taught in the Biomedical Informatics Department, with input from practitioners at the Utah Department of Health. The course contains six modules:

- Describe the mission and practice of public health and identify opportunities using informatics methods and tools;
- Describe fundamental informatics principles and their application to public health, including database design and process diagramming;
- Examine standards relevant to public health and create design artifacts to enable system interoperability;
- Describe the current and evolving relationship between clinical and public health systems;
- Examine roles required to develop and manage public health informatics projects and systems.

Each module contains a guidance file, one or more voice-over-power-point lectures, readings and resources, a short quiz and/or discussion. A student project was also developed for the students to look into a current public health information system of their own choice. An in-person session was planned at the end of course for the students to present their projects. Feedback was collected by distributing a student evaluation questionnaire to each student at the in-person session.

Results

The course was delivered from 18 January to 10 May 2010 using Blackboard Vista, the University of Utah's online

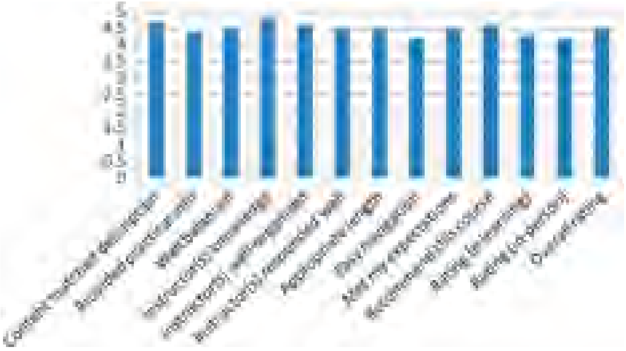


Figure 1 Average scores of students' evaluation of the Utah-AMIA 10 × 10 Public Health Informatics online course.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

teaching resource. Fifteen students from Puerto Rico to California had various backgrounds from clinician, public health practitioner, to business development manager. Twelve students described surveillance workflows and information process at the in-person session before the AMIA NOW! conference. The students highly scored the course (Figure 1) and made comments such as: 'It covered a broad range of skills and knowledge relevant to public health informatics.' 'A new professional field, I would like to continue in this field (public health informatics).'

Conclusions

The Utah-AMIA 10 × 10 Public Health Informatics online course successfully introduced clinicians and public health practitioners to informatics principles and their application to public health problems. With a high level of satisfaction, students developed basic analysis and lifelong-learning skills to engage in the evolving field of public health informatics.

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 AMIA 10 × 10. Available at https://www.amia.org/10 × 10.
2 About the Deparment. Available at http://www.bmi.utah.edu/?pageId = 2182.

Emerging Health Threats Journal 2011, **4**:s136. **doi:** 10.3134/ehtj.10.136
© 2011 E Lanktree *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Assessing address data quality for public health surveillance in Montreal

E Lanktree[1], C Jauvin[1,2], L Bédard[1], and D Buckeridge[1,2]

[1]*Bureau de surveillance épidémiologique, Direction de santé publique, Agence de santé et des services sociaux de Montréal, Montréal, Canada; and* [2]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada*
E-mail: elanktre@santepub-mtl.qc.ca

Objective

To (1) validate an address verification algorithm (*Dracones qualité*, DQ, described by Zinszer *et al.*[1]) developed to improve data quality for public health mapping and (2) identify the origin of address errors.

Introduction

In Montreal, notifiable diseases are reported to the Public Health Department (PHD). Of 44, 250 disease notifications received in 2009, up to 25% had potential address errors. These can be introduced during transcription, handwriting interpretation and typing at various stages of the process, from patients, labs and/or physicians, and at the PHD. Reports received by the PHD are entered manually (initial entry) into a database. The archive personnel attempts to correct omissions by calling reporting laboratories or physicians. Investigators verify real addresses with patients or physicians for investigated episodes (40–60%).

The DQ address verification algorithm compares the number, street and postal code against the 2009 Canada Post database. If the reported address is not consistent with a valid address in the Canada Post database, DQ suggests a valid alternative address.

Methods

Individual disease episodes for Montreal residents reported to the PHD between 26 January and 9 May 2009 and followed by PH nurses were included ($n=798$). Exclusion criteria were H1N1 episodes, those whose investigator address was completely different from the initially reported address, invalidated or non-nominal episodes, for a sample of 408.

Complete investigator addresses treatable by DQ ($n=342$) were verified manually by a member of the research team (EL) using the Canada Post website.[2] These addresses were also scanned by DQ. DQ and Canada Post results were compared to determine whether DQ correctly identified valid and invalid addresses and proposed valid suggestions.

Calls were made to labs, hospitals or clinics, between 9 April and 9 July 2010, to confirm reported addresses. A list was compiled by selecting all cases that had a reported address deemed invalid by DQ, ($n=66$) and a 20% random sample[3] of the 181 cases that had reports with hand-written addresses.

Results

DQ detected an error rate of 11.8% (47 of 398) upon initial data entry, 10.0% (40 of 401) at final entry, after archive verification, 18.0% from investigator addresses (60 of 333), and 36.5% from calls to reporting labs/physicians (23 of 63). All addresses corrected at data entry corresponded to DQ suggestions.

Of the 336 episodes with investigator addresses corresponding exactly to initial entry, 334 (99.4%) DQ suggestions correspond to Canada Post. For the two others, both numbered avenues (that is, ninth), DQ proposed one suggestion and Canada Post proposed two.

The investigator address was chosen as the gold standard, as it was closest to the case's actual address. Among initial entries, DQ detected three types of errors: 27 postal code, five street number and six street number or postal code errors. Most corresponding investigator addresses were identical. All other addresses (11 of 27, three of five, and one of six, respectively) that were corrected during the investigator calls corresponded to DQ suggestions.

Conclusions

The DQ algorithm is valid but short street names seem problematic. The algorithm will be refined using a street name length and edit distance solution.

Approximately two percent of errors detected at initial data entry were corrected through data entry and/or archive team efforts, but could be corrected by DQ if integrated into the electronic chart database, saving manpower and time. Labs and physicians should update patient addresses regularly to decrease error rates. Surveillance and intervention

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

would benefit as reliable data would improve disease cluster identification and communication with cases.

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 Zinszer K, Jauvin C, Verma A, Bedard L, Allard R, Schwartzman K, *et al*. Residential address errors in public health surveillance data: a description and analysis of the impact on geocoding. *Spat and Spatio-temporal Epidemiolo* 2010;**1**:163–8.
2 Canada Post Corporation 2010. http://www.canadapost.ca/cpotools/apps/fpc/personal/findByCity?execution = e1s1.
3 StataCorp 2009. STATA 10.1.

Emerging Health Threats Journal 2011, **4**:s93. **doi:** 10.3134/ehtj.10.093

© 2011 SL Lewis *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

SAGES: a suite of freely available software tools for electronic disease surveillance in resource-limited settings

SL Lewis[1], BH Feighner[1], WA Loschen[1], RA Wojcik[1], JF Skora[1], JS Coberly[1], and DL Blazes[2]

[1] *The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; and* [2] *Division of GEIS Operations, Armed Forces Health Surveillance Center, Silver Spring, MD, USA*
E-mail: Sheri.Lewis@jhuapl.edu

Objective

This paper describes the development of the Suite for Automated Global bioSurveillance (SAGES), a collection of freely available software tools intended to enhance electronic disease surveillance in resource-limited settings around the world.

Introduction

Emerging and re-emerging infectious diseases are a serious threat to global public health.[1,2] The World Health Organization (WHO) has identified more than 1100 epidemic events worldwide in the last 5 years alone.[3] Recently, the emergence of the novel 2009 influenza A (H1N1) virus and the SARS coronavirus has demonstrated how rapidly pathogens can spread worldwide. This infectious disease threat, combined with a concern over man-made biological or chemical events, spurred WHO to update their International Health Regulations (IHR) in 2005.[4] The new 2005 IHR, a legally binding instrument for all 194 WHO member countries, significantly expanded the scope of reportable conditions, and are intended to help prevent and respond to global public health threats. SAGES aims to improve local public health surveillance and IHR compliance, with particular emphasis on resource-limited settings.

Methods

More than a decade ago, in collaboration with the US Department of Defense (DoD), the Johns Hopkins University Applied Physics Laboratory (JHU/APL) developed the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE). The current SAGES initiative leverages the experience gained in the development of ESSENCE; the analysis and visualization components of SAGES are built with the same functionalities in mind. Cognizant of work underway on individual surveillance systems components, for example, collection of data by cell phones, we have focused our efforts on the integration of inexpensive, interoperable software tools that facilitate regional public health collaborations.

Results

SAGES tools are organized into four categories: (1) data collection, (2) analysis and visualization, (3) communications, and (4) modeling/simulation/evaluation. Within each category, SAGES offers a variety of tools compatible with surveillance needs and different types or levels of information technology infrastructure. In addition to the flexibility of tool selection, there is flexibility in the sense that the analysis tools do not require a fixed database format. For example, rather than requiring an existing database to adapt to the tool, the SAGES database tools adapt to the format of all Java database compliant formats. Lastly, the SAGES tools are modular in nature, allowing the user to select one or more tools to enhance an existing surveillance system, or use the tools *en masse* for an end-to-end electronic disease surveillance capability. Thus, each locality can select tools from SAGES based upon their needs, capabilities, and existing systems to create a customized electronic disease surveillance system.

Conclusions

We have combined electronic disease surveillance tools developed at the Johns Hopkins University Applied Physics Laboratory with other freely available, interoperable software products to create SAGES. We believe this suite of tools will facilitate local electronic disease surveillance, regional public health collaborations, and international disease reporting. The Armed Forces Health Surveillance Center

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

welcomes inquiries on the SAGES tools from interested WHO member countries.

Acknowledgements

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

References

1 Binder S, Levitt AM, Sacks JJ, Hughes JM. Emerging infectious diseases: public health issues for the 21st century. *Science* 1999; **284:**1311–13.
2 Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature* 2004;**430:**242–9.
3 WHO. *The World Health Report 2007—A Safer Future: Global Public Health Security in the 21st Century.* World Health Organization: Geneva, 2007.
4 WHO. *International Health Regulations (2005).* World Health Organization: Geneva, 2005.

Emerging Health Threats Journal 2011, **4**:s94. **doi:** 10.3134/ehtj.10.094
© 2011 N Lipskiy *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Health IT meaningful use: how the EHR certification may improve the electronic laboratory reporting (ELR)

N Lipskiy, N Garrett, and S Tandon

Office of Surveillance, Epidemiology and Laboratory Science (OSELS), Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA
E-mail: dgz1@cdc.gov

Objective

The objective of this presentation is to evaluate the potential impact of Stage 1 meaningful use (MU) health IT certification (MUC), on development of national electronic laboratory reporting (ELR) capacities.

Introduction

The American Recovery and Reinvestment Act (ARRA) brought significant incentives to providers for implementing certified EHR technologies. It specifically requires utilization of certified electronic health records (EHRs) for electronic exchange of health information and for submission of clinical quality and other measures to the federal agencies. The most important barriers in the ELR implementation are a lack of funding at health departments, shortage of staff at health departments, and the variable content and format of ELR messages.[1] The MU is a new factor that may foster ELR technologies through implementation incentives and through standardization of EHRs.

Methods

(1) Analysis of ELR data structure and data exchange requirements (HL7) (a) for a communication between states and CDC programs and (b) across all MU objectives in comparison with the ELR public health reporting objective. (2) Impact assessment of EHR-related and ELR-related objectives on development of state and national ELR capacities. An assessment of states' ELR capacities was based on the 2008 ELR survey[2] and official web sites of states' department of health. We analyzed ELR messaging requirements using documents that were published through official HL7 and CDC web sites.

Results

A baseline assessment of ELR status before enacting of the ARRA demonstrates that most of the state legislations did not require ELR. The MUC fosters an implementation of ELR for notifiable conditions. Core MU objectives[3] require an inclusion in certified EHRs data elements that describe patients' demographics, clinical diagnoses, and test results. It aims to harmonize data elements in health IT systems that are relevant to ELR. For example, a MUC requirement on use of Logical Observation Identifiers and Codes (LOINC) for test procedures in hospital systems will foster using LOINC in ELR. The MUC of EHRs leverages national activities on creation, maintenance, and access relevant to ELR standards for content, messages exchange and their security, and patient privacy.

For example, because the MUC requires use of the HL7 version 2.5.1 for submitting of laboratory results to public health (PH) agencies, it may stimulate PH for using the same HL7 version for ELR. If an expected inclusion of certification requirements for a bi-directional communication between clinical care and PH is added to the MUC at Stage 2, then the MUC impact on ELR will be significantly increased.

Conclusions

Even though EHRs and ELR operate in different health IT systems (respectively, in clinical care and PH), the MUC of EHRs is a very important factor that may help in further ELR implementation, improvement of interoperability of state and national PH IT systems, and timeliness of PH emergency response.

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 Council of State and territorial Epidemiologists. About ELR. Available at: http://www.cste.org/dnn/ProgramsandActivities/PublicHealthInformatics/tabid/346/Default.aspx.
2 The 2008 National Electronic Laboratory Reporting (ELR) Snapshot Survey. Available at: http://www.cste.org/dnn/Portals/0/2008NationalELRSurveySum.pdf.
3 42 CFR Part 170, Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, Final Rule, 2010.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s95. **doi:** 10.3134/ehtj.10.095
© 2011 S Ganesan *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Reporting nationally notifiable conditions (NNC): vocabulary aspects

S Ganesan[1], N Lipskiy[2], R Merrick[3], and C Vinion

[1]*Northrop Grumman Consultant to CDC, Atlanta GA, USA;* [2]*Office of Surveillance, Epidemiology and Laboratory Science (OSELS), Centers for Disease Control and Prevention (CDC), Atlanta GA, USA; and* [3]*Independent Consultant to CDC, Atlanta GA, USA*
E-mail: dsg3@cdc.gov

Objective

Electronic laboratory reporting (ELR) has a key role in public health case reporting and case notification. This paper will discuss the current status, problems, and solutions in a vocabulary support of nationally notifiable conditions (NNC) reporting.

Introduction

Standard vocabulary facilitates the routing and filtering of laboratory data to various public health programs. In 2008, Council of State and Territorial Epidemiologists (CSTE) developed 67 Technical Implementation Guides (TIGs) that accompany each condition and contain standard codes for NNC reporting. Those TIGs were reviewed by a public health subject matter expert panel (SMEP), in May 2010, consisting of members of the CDC CSTE Laboratory and PHIN Vocabulary and Messaging Communities of Practice Program, and representatives from the Regenstrief Institute and the International Health Terminology Standards Development Organization.

Methods

The SMEP reviewed CSTE position statements for NNC[1] and content of Table 2 of the 67 TIGs that were provided by CSTE. All laboratory criteria that exist in current position statements were compared with a list of general microbiology laboratory methods (that is, microscopy, antigen detection, antibody detection, DNA methods, and so on). Also, a list of existing Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) codes for notifiable conditions were compared with those that were listed in the TIGs. The Regenstrief LOINC Mapping Assistant, RELMA,[2] was used for retrieving LOINC codes sorted by each NNC and laboratory criterion. SNOMED-CT codes, as of May 2010, were used for a comparison with codes that were included in the TIGs.

Results

The SMEP concluded that laboratory vocabulary, after minor updates to current codes, in TIGs is ready to be implemented. The SMEP also suggested to CSTE a change in the format of the CSTE position statements. Specifically, the recommendations are to (1) add a section on 'Microorganisms/Agents' to applicable CSTE position statements,[2] (2) define a CSTE process for new development and maintenance TIG laboratory vocabulary, (3) clarify the use of paired serology tests, (4) add quantitative results in reports, and (5) develop additional value sets (a qualitative laboratory test finding value set based on SNOMED-CT evaluation finding domain and a specimen value set based on SNOMED-CT and HL7 as specified in the ELR v2.5.1 implementation guide).

Conclusions

There is a notable progression in the standardization of vocabulary for NNC reporting. The SMEP found existing problems (development of TIGs, vocabulary gaps, evaluation of laboratory results, and so on.) that may be eliminated by a defined process for collaborative work of public health and standard development organizations.

Acknowledgements

The authors would like to acknowledge the excellent participation of the CSTE Case Report Standardization Workgroup and the PHIN Laboratory Messaging Community of Practice. 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 CSTE. Position Statements at: http://www.cste.org/dnn/AnnualConference/PositionStatements/09PositionStatements/tabid/333/Default.aspx 2009.
2 Regenstrief LOINC Mapping Assistant, RELMA. Available at: http://loinc.org/relma.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s137. **doi:** 10.3134/ehtj.10.137
© 2011 N Lipskiy *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Improving quality and interoperability of public health messaging guides: CDC messaging quality framework (MQF)

N Lipskiy[1], G Kesarinath[1], M Pray[2], and V Nwadiogbu[2]

[1]*Office of Surveillance, Epidemiology and Laboratory Science (OSELS), Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA; and* [2]*Northrop Grumman Consultant to CDC, Atlanta, GA, USA*
E-mail: dgz1@cdc.gov

Objective

This document describes the Public Health Information Network (PHIN) efforts on the development of the MQF, a flexible framework of services and utilities designed to assist public health partners with preparing and communicating quality, standard electronic messages.

Introduction

The HL7 messaging standard, version two[1] that was implemented by most vendors and public health agencies did not resolve all systems' interoperability problems. Design and tool implementation for automated machine-testing messages may resolve many of those problems. This task also has critical importance for rapid deployment of electronic public health systems.

Methods

(1) Assessment of messaging structures that support reporting of nationally notifiable conditions, NND[2] and Public Health Lab Interoperability Project, PHLIP.[3] (2) Analysis of best practices on a structural validation of messages. (3) Analysis of CDC MQF[4] capabilities for conformance testing and support rapid implementation of biosurveillance applications.

Results

Released in March 2010, a web-hosted MQF application (version 1.0) allows several options for loading messages. It contains a message validation tool based on a current messaging standard for a NND case notification,[2] which is constrained to a Generic case Notification Messaging Guide (MG) and MGs for Varicella and Tuberculosis. Another MQF option allows validating messages against a PHLIP MG for Influenza test reporting.[3] At the first step of a message validation process the MQF application provides a structural validation of messages on message, segment, field and component levels (that is, cardinality of segments groups, existence of required segments and their cardinality, existence and cardinality of segments and so on). The MQF then provides a message-constrained validation against messaging specifications (that is, validation of data types within OBX segments, lengths of specific data elements, verification that data elements are consist of only supported values and so on). The offline version adds a capability for synchronizing MQF with all NND MGs.

Conclusions

Even at the initial stage of development, the CDC MQF represents a flexible framework of services available to systems and public health partners inside and outside of CDC to prepare, test, and validate electronic messages against the relevant messaging, vocabulary standards, and program specific business rules.

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 Health Level Seven, Inc. *HL7 Messaging Standard Version 2.5: An Application Protocol for Data Exchange in Healthcare Environments*. Ann Arbor, MI, 2007.
2 PHIN HL7 Version 2.5. Messaging Standard National Condition Reporting Case Notification. ORU^R01 Message Structure Specification/Profile Version 2.0, October 23, 2008. Available at: http://www.cdc.gov/phin/library/documents/pdf/PHIN%20Notification-MessageSpecificationProfile%20v2%200.pdf.
3 APHL/CDC PHLIP Messaging Guide for Influenza Test Result Reporting by Public Health Laboratories, ORU R01 HL7 v2.3.1, Document version 1.0.2, Sept. 15, 2009. Available at: https://phinmqf.cdc.gov/HL7v2_3_1_APHL_Influenza_Msg_Guide_ORU_v1%200%202.pdf.
4 PHIN Message quality Framework. Available at: https://phinmqf.cdc.gov/.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s96. **doi:** 10.3134/ehtj.10.096
© 2011 J Lipsky *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Influenza messaging and interoperability: the PHLIP assistance team approach

J Lipsky[1], L Cohen[1], M Meigs[2], and P Zarcone[2]

[1]*The St John Group, LLC, Atlanta, GA, USA; and* [2]*Association of Public Health Laboratories (APHL), Silver Spring, MD, USA*
E-mail: jlipsky@TSJG.com

Objective

This paper describes the Public Health Laboratory Interoperability Project (PHLIP) assistance team (PAT) approach and the collaborative efforts between the Association of Public Health Laboratories (APHL) and the Centers for Disease Control and Prevention (CDC) to achieve electronic laboratory surveillance messaging (ELSM) for Influenza. The knowledge transfer and experience gained by state public health laboratories (PHLs) participating in PHLIP could serve as an interoperability model for other data messaging and surveillance initiatives.

Introduction

An essential theme of the US Federal Health Information Technology Strategic Plan is interoperability and the ability to effectively exchange information using specific data and technical standards.[1] In 2005, in an effort to accelerate the development of a national laboratory standards-based electronic data-sharing network, APHL and CDC collaborated to launch PHLIP.[2] The goals of PHLIP include, but are not limited to, improving the quality of data exchanged, piloting sustainable architecture for laboratory data exchange, sending and receiving HL7 test results from states to CDC programs (v2.3.1), increasing the use of Route-not-Read hubs for regional data exchange, and expanding these efforts beyond National Notifiable Diseases (NNDs). In an effort to achieve these goals, APHL solicited input directly from the PHL community to understand what assistance was necessary to achieve success with ELSM; in this case, Influenza as a prototype. After receiving feedback from PHLs responsible for reporting NNDs, the concept of technical assistance teams was formulated. In early 2010, APHL initiated an effort to send out the PATs to implement the ELSM message for Influenza in as many PHLs as possible by December 2010.

Methods

A detailed retrospective review of the experiences of the PATs experience will be presented. PHLs signed up for PAT support and were prioritized according to input from the CDC's Influenza Division, as well as their own timeline for assistance. Deployed in spring 2010, two technical assistance teams were tasked with visiting PHLs to provide 'hands-on' assistance in areas including technical architecture, vocabulary, and project management expertise. States that were not available to host an onsite PAT visit were given the option of receiving 'virtual' support. Over the last 6 months, the process from initial state engagement until going 'live' with the PHLIP ELSM has evolved and best practices for achieving interoperable electronic data exchange have been captured.

Results

Presently, over 30 PHLs are working in some capacity on PHLIP, and 11 PHLs have been visited by PATs. Since March 2010, approximately 10 PHLs receiving either an onsite visit or virtual support by the PAT are now sending automated Influenza messages to CDC. Next states to be visited before December 2010 include NH, KY, NY, and WA, as well as others. (Table 1).

Conclusions

Over the last 6 months, PHLs have increased their interoperability functionality as well as enhanced their surge capacity for Influenza. These improvements are because of the PAT approach and the foresight of the collaborating organizations; APHL and CDC. Given the changing climate of infectious disease transmission, it is important to heed the lessons learned for effective laboratory information management processes, while keeping an eye toward interoperability in the public health domain.

Table 1 Process steps from initial PHL engagement until electronic message validation by state model type

Model	*Time to validation step*
Original, non-PAT states	6–12 Months
States with PAT support	Under a month

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

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

References

1 Benson T. *Principles of Health Interoperability HL7 and SNOMED*. Springer-Verlag London Limited: London, UK, 2010.
2 Zarcone P, Nordenberg D, Meigs M, Merrick U, Jernigan D, Hinrichs SH. Community-driven standards-based electronic laboratory data-sharing networks. *Public Health Rep* 2010;**125** (2): 47–56.

Emerging Health Threats Journal 2011, **4**:s103. **doi:** 10.3134/ehtj.10.103

© 2011 P Mirhaji *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Real-time, reusable, dynamic public health surveillance

P Mirhaji, R Vedam, and MM Vagnoni

School of Biomedical Informatics, University of Texas, Houston, TX, USA
E-mail: Parsa.Mirhaji@uth.tmc.edu

Objective

This paper describes use of semantic technologies in combination with Services Oriented Architecture (SOA) to construct dynamic public health surveillance systems[1] used for just-in-time monitoring of emerging infectious disease outbreaks. The system was used for surveillance of schools in the third largest population center, Harris County.

Introduction

The resources available in most public health departments are limited. Access to trained technical personnel and state-of-the-art computing resources are also lacking. Customizable off-the-shelf systems contribute only to creation of information silos, are expensive, and not affordable by the limited budget available to the departments of health (only growing worse with the recession). The one thing that has increased is the need for surveillance in more areas, from diseases to environmental exposures to unexpected disasters. One solution would be an adaptable system able to cope with changing requirements while reusing or eliminating infrastructure from both computing hardware and technical personnel.[2] We report in this paper an instance of such system as used to perform disease surveillance across the Harris County school system. The system is designed to be customizable for surveillance of any disease, while simultaneously accommodating other use cases like disaster response and registries.

Methods

The Survey-On-Demand System (SODS) enables epidemiologist to create all necessary data capture for an area to be surveilled. This eliminates the need for personnel to design a web page, desktop client, or the backend data store. In addition, the system ties directly into a surveillance dashboard that allows epidemiologist to view trends, set alerts, and monitor multiple surveillance projects simultaneously.

The system is designed to abstract the format of the data captured so that it can be seamlessly exported to other formats (that is, Excel or SAS). All the data captured is backed by controlled and/or colloquial terminologies, which enable integration across multiple surveillance projects (past, present, and future) or information exchange with remote collaborators. Data is either mapped manually or natural language processing suggests appropriate terms from formal terminologies.

In order to cope with partial Internet connection (online somewhere, offline most places), an interaction model enables offline use and then synchronization when Internet is available. The server manages the state and macroscopic provenance of the forms. The client software logs every user interaction providing audit trails and item level rollback. All data is strongly encrypted. The system effectively manages the dynamic changes in the structure of the survey; no data is lost. The client keeps past data coherent with current data, as data collection needs change.

Results

More than 200 schools from Harris County were surveilled for absenteeism due to influenza-like illness through the academic year 2009–2010. The system is being evaluated to be used in other surveillance projects, including food-borne illness, environmental services, and children learning assessment. The SOA architecture enables a cloud-based information processing and data storage.[3]

Conclusions

The SODS system has been successful at both disaster and disease surveillance. It is being applied to multi-institutional clinical research and learning assessment in primary schools. The dynamic, model driven, information processing backend services support context independent, disparate use cases harnessing the same system and investment without creating another information/systems silo. The warehouse is a hybrid containing both relational tables for Online Analytics Processing (OLAP) and a semantic repository for flexible, extensible representation of highly variable data. In addition, the surveys can be shared across organizations and projects, enabling reuse and further minimizing resources required to do surveillance. There are many interesting possibilities to expand this surveillance toolkit.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

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

References

1 Mirhaji P, Casscells SW, Srinivasan A, Kunapareddy N, Byrne S, Richards DM, *et al*. Services oriented architectures and rapid deployment of *ad-hoc* health surveillance systems: lessons from Katrina relief efforts. *AMIA Annu Symp Proc* 2006; 569–573.

2 Mirhaji P, Zhu M, Vagnoni M, Bernstam EV, Zhang J, Smith JW, *et al*. Ontology driven integration platform for clinical and translational research. *BMC Bioinformatics* 2009;**10** (Suppl 2): S2.

3 Bechhofer SK, Stevens RD, Lord PW. GOHSE: ontology driven linking of biology resources. *Web Semantics: Science, Services and Agents on the World Wide Web* 2006;**4**:155–63.

Emerging Health Threats Journal 2011, **4**:s140. **doi:** 10.3134/ehtj.10.140
© 2011 A Mostaghimi *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

HAIISS Data Warehouse (HDW)—A new data access architecture for ESSENCE in the VA

A Mostaghimi[1], G Oda[1], C Lucero[1], P Schirmer[1], J Lombardo[2], R Wojcik[2], and M Holodniy[1]

[1]*Department of Veterans Affairs, Office of Public Health Surveillance and Research, Palo Alto, CA, USA; and* [2]*Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA*
E-mail: anoshiravan.moshtaghimi@va.gov

Objective

To describe the new data warehouse, HAIISS Data Warehouse (HDW) architecture whereby VA's Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE) will receive its required data elements from VA's 128 VistA systems in a more accurate, robust and time sensitive manner.[1]

Introduction

The data elements required for the proper functionality of VA's ESSENCE system are all currently available within VA's 128 VistA systems. These data are made available to VA's ESSENCE system via a series of complicated MUMPS extraction routines, multiple data transformations crossing multiple servers, networks, operating systems and HL7-parsing

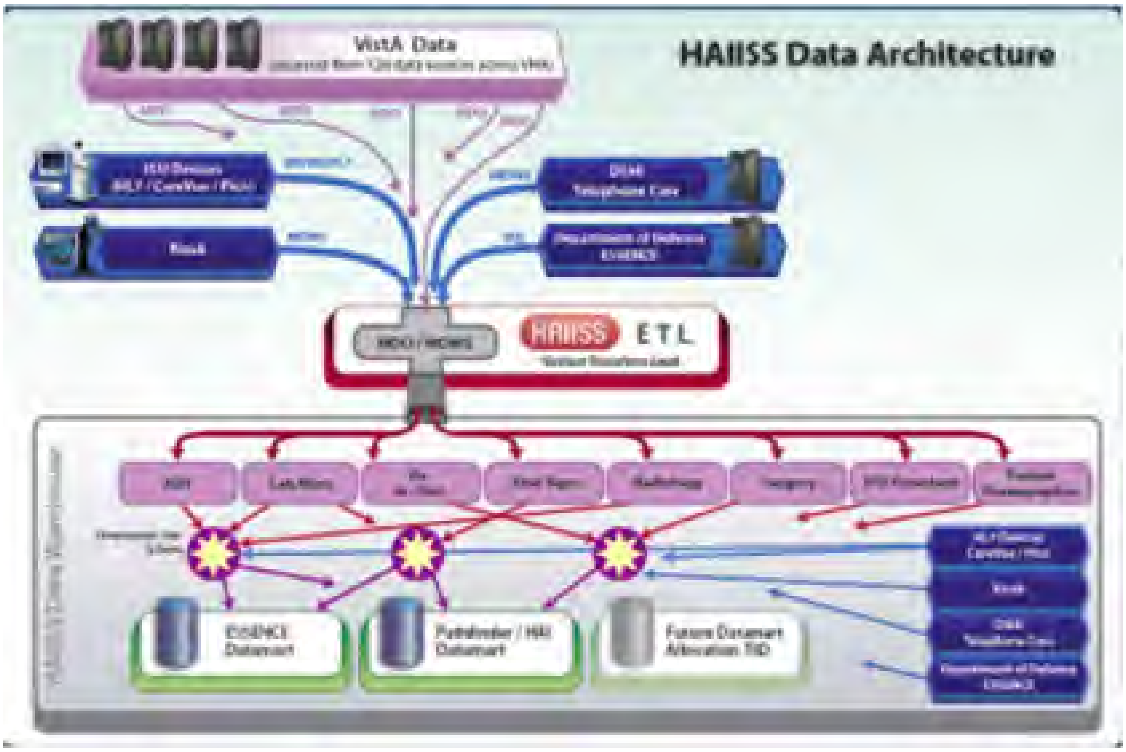


Figure 1 Required data elements are collected via a MDO-based ETL process and populate an ESSENCE data mart.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

routines on a daily interval. With recent changes emerging in VA's information technology infrastructure, a new data architecture supporting ESSENCE's surveillance capabilities is becoming possible.

Methods

Healthcare Associated Infection and Influenza Surveillance System (HAIISS) is a VA project tasked with the operation and deployment of VA's biosurviellance system using ESSENCE, and a nosocomial surveillance system. In order for HAIISS to successfully untangle the current data access challenges supporting these systems, a new data access architecture has been developed and is being embedded as part of Extract-Transform-Load (ETL) layer of a data warehouse in support of HAIISS Data Warehouse (HDW) project. The Extraction methodology is based on the Remote Procedure Call (RPC) capabilities of VistA. This VistA RPC technology is in turn abstracted into a Data Access Object (DAO) design pattern written in C# hence, an Application Programming Interface (API) called Medical Domain Objects (MDO) is written whereby targeted, inexpensive extraction calls are made against VistA systems. The resulting returning object in turn is passed along to a Business Processing Engine of the HDW whereby the Transformation-Load steps associated with populating the data warehouse and consequently the ESSENCE data mart are accomplished (Figure 1).

Conclusions

Utilization of MDO as part of the data extraction capabilities of HDW has offered us the ability to query for the required data elements using an extremely efficient and low overhead process. The required surveillance data are now made accessible to ESSENCE as queriable data inside a relational database hence, alleviating the expensive and often error-prone HL7-based flat-file parsing process. Finally, the Business Processing Engine associated with HDW allows the data warehouse manger the ability to poll the VistA data sources in a situationally specific interval, such that the periodicity of critical data elements used by ESSENCE-alerting mechanism is much shorter than a daily interval.

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.

Reference

1 Gillon, Joseph—VA SharePoint. http://medora.sharepoint.med.va.gov/sites/mdws/default.aspx.

Emerging Health Threats Journal 2011, **4**:s160. **doi:** 10.3134/ehtj.10.160
© 2011 S Mottice *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Implementing an outbreak/event tracking monitor for public health

S Mottice[1], J Reid[1,2], T Jeppson[1], and R Rolfs[1]

[1]*Utah Department of Health, Salt Lake City, UT, USA; and* [2]*Department of Clinical Epidemiology, University of Utah, Salt Lake City, UT, USA*
E-mail: smottice@utah.gov

Objective

The aim of this project was to create a secured web-based application that would run within the PH Access communication framework at the Utah Department of Health. This application—Epi Issue Tracker—would provide State and Local Health Departments with an environment that can be used to monitor actions during an event. We created a tracking system that functions as a combination dashboard/notification system to permit timely and effective communication of epidemiology events. This tracking system, Epi Issue Tracker, is used by all 12 Local Health Departments and the State Department of Health to share information across the state. There have been 830 issues/outbreaks entered into Epi Issue Tracker since January 2009, with 647 updates posted for those issues/outbreaks.

Introduction

Utah has a centralized State Health Department and 12 Local Health Departments situated throughout the state. Coordination of outbreaks or events that crosses jurisdictions has been historically difficult. Utah has not had a functional NEDSS-compliant database until 2009 and still does not have an Outbreak Management System (OMS). A survey was sent to Local Health Departments to assess their perception of need for real-time knowledge of current outbreak/events, with the majority indicating that current processes were inadequate.

Methods

We developed this application using open-source tools from PHAccess written in PHP and subject matter experts to provide initial application requirements and feedback. The development process used a series of iterative development cycles that included development, testing and feedback. Through these cycles we are able to add product enhancements and fixes quickly and efficiently.

Results

We created Epi Issue Tracker and provided it to Local Health Departments during a sequential rollout. The application was designed to provide a dashboard detailing all ongoing events within the state. Events are individually managed and activities can be added and maintained in chronological sequence, thus expediting the creation of event reports. Completed events are moved to resolved status and can be maintained in a database that can be exported via CSV for tracking of outbreaks over time. There have been 830 events created since Epi Issue Tracker was rolled out to Local Health Departments, with 647 event updates posted to those events. Since April 2009 the system has been accessed 3099 times by both state and local epidemiologists.

Conclusion

This program is an easy-to-use issue tracker that facilitates cross-jurisdictional management of outbreaks and events.

Through Epi Issue Tracker, Local Health Departments are able to be informed of cross-jurisdictional outbreak events and post updates to those events for other Local Health Departments or the State Health Department. This application has helped increase communication within the state and improve relationships between the State and the Local Health Departments.

Acknowledgements

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

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s142. **doi:** 10.3134/ehtj.10.142

© 2011 A Pérez *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Classification of errors for quality assurance in the emerging infections program influenza hospitalizations surveillance system

A Pérez, T D'Mello, L Kamimoto, L Finelli, and MH Torres-Urquidy

Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA
E-mail: hvv9@cdc.gov

Objective

Introducing data quality checks can be used to generate feedback that remediates and/or reduces error generation at the source.[1] In this report, we introduce a classification of errors generated as part of the data collection process for the Emerging Infections Program (EIP)'s Influenza Hospitalization Surveillance Project at the Centers for Disease Control and Prevention (CDC). We also describe a set of mechanisms intended to minimize and correct these errors via feedback, with the collection sites.

Introduction

The CDC's Emerging Infections Program monitors and studies many infectious diseases, including influenza.[2] In 10 states in the US, information is collected for hospitalized patients with laboratory-confirmed influenza. Data are extracted manually by EIP personnel at each site, stripped of personal identifiers and sent to the CDC. The anonymized data are received and reviewed for consistency at the CDC before they are incorporated into further analyses. This includes identifying errors, which are used for classification.

Methods

We evaluated the most current dataset as of 24 August 2010, containing records for 6521 persons with influenza-associated hospitalizations from 1 September 2009 through 30 April 2010. We built fully automated software routines using SAS version 9.2 (SAS Institute Inc., Cary, NC, USA) to conduct quality assurance. For instance: when data about the patient age are not provided, our software identifies the missing information as an error. We generated our classification based on the characteristics of these errors using a data-driven approach (that is, clustering errors with similar properties). The classification was then discussed internally. Based on the common characteristics of the clusters we developed common definitions for each category in the classification. Finally, we measured the actual number of errors in the most current collected dataset and categorized the classification by type, with the most prominent ratio (predefined errors/errors in reports).

Table 1 Error classification

Name	*Predefined errors*	*Errors in reports*	*Ratio*
Data entry errors	43	761	18.12
Missing data	19	263	13.84
Integrity	18	129	7.17
Failure to meet case definition	4	28	7.00
Chronology	32	113	5.35

Results

The implementation of the error classification occurred during the preparation of the monthly report submitted to sites. The error classification was generated (Table 1). We found that 'Data Entry Errors' were the most prominent followed by 'Missing Data.' Other types of errors were identified as well.

Conclusions

Classification of errors allows for easier identification and prompt correction. In addition, it will allow us to improve subsequent versions of the software used to capture information and possibly minimize errors during capture.

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 Mullooly JP. The effects of data entry error: an analysis of partial verification. *Comput Biomed Res* 1990;**23**:259–67.
2 Dawood FS, Fiore A, Kamimoto L, Nowell M, Reingold A, Gershman K, *et al*. Influenza-associated pneumonia in children hospitalized with laboratory-confirmed influenza, 2003–2008. *Pediatr Infect Dis J* 2010;**29**:585–90.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s106. **doi:** 10.3134/ehtj.10.106
© 2011 P Picciano *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

BioSINE: an intuitive visualization tool to enhance collaboration between research and practice in disease surveillance

P Picciano[1], J Beaubien[1], G Mizrahi[1], J Owen[1], F Grant[2], and AM Hegedus[2]

[1]*Aptima, Inc., Woburn, MA, USA; and* [2]*Northrop Grumman Corp., Atlanta, GA, USA*
E-mail: ppicciano@aptima.com

Objective

BioSINE strives to improve situational awareness by making data visualization and collaboration capabilities intuitive and readily available for a wide range of public health (PH) stakeholders.

Introduction

Funded by the Army's Telemedicine and Advanced Technology Research Center (TATRC), we developed the BioSINE toolset to provide visualization and collaboration capabilities to improve the accessibility and utility of health surveillance data. Investigation of PH practitioners' needs with cognitive engineering methods revealed two key objectives:

1. To provide analysts and decision makers with an intuitive, visually driven workspace.
2. To support a web presence to provide rapid updating and facilitate greater interaction with data analysis in the PH community.

To better serve under-resourced PH organizations, both domestic and abroad, it is necessary to minimize information technology (IT) requirements and expertise in complex analytic tools.

BioSINE provides decision makers with the ability to create customized visualizations, focus on specific aspects of the data, or conduct hypothesis testing. Users can also view or hide variables, specify data ranges, and filter data relevant to their interests. Figure 1 shows a display in which a user investigated seasonal effects by narrowing the analysis to the summer months. Intuitive filtering is a key characteristic of the application to quickly produce snapshots of local interests.

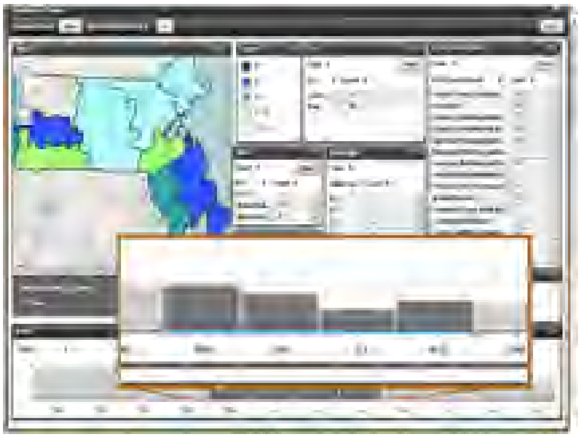


Figure 1 Screenshot of consumer probing for seasonal trends by customizing date range.

Methods

We conducted usability tests with a sample of six participants, including three (military) public health experts and three (civilian) naïve users. Each performed a series of standardized tasks that required probing the dataset to identify morbidity patterns and answer specific questions. Morae© usability testing software was used to record the entire session, including the number of errors made, real-time verbal commentary, and survey items measuring satisfaction and usability opinions. Each questionnaire was scored on a five-point Likert scale with anchors ranging from *Strongly Disagree* (1) to *Strongly Agree* (5). After each testing session was completed, users participated in a semi-structured interview to suggest additional software features and functionality that might be useful to the larger PH community. Each user session lasted approximately 90 min.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Results

Users were quite successful in completing their tasks. Military and civilian users performed similarly. The mean number of errors per user was 4.17 (s.d. = 2.04), none of which prevented completion. Users also reported they enjoyed using the BioSINE software. The mean satisfaction score was 4.07 (s.d. = 0.59). Of the narrative comments collected, users reported that the BioSINE interface was easy to use, and provided a unique approach to interacting with and extracting information from health surveillance data.

Conclusions

Our findings over the course of this project suggest the ability to *interact with data,* and not just view reports, leads to improved understanding, situational awareness, and collaboration. Our testing results indicate that even first-time users without PH expertise can successfully complete complex tasks and correctly answer health surveillance queries. Collected user opinions showed BioSINE to be useful and intuitive. Our intent is to harden the BioSINE prototype and make it a freely available web tool to the PH community.

Acknowledgements

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

Emerging Health Threats Journal 2011, **4**:s107. **doi:** 10.3134/ehtj.10.107
© 2011 A Poku *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Mitigating data collection challenges with adaptive frameworks

A Poku[1], RA Wojcik[1], J Skora[1], S Lewis[1], and DL Blazes[2]

[1]*The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; and* [2]*Division of GEIS Operations, Armed Forces Health Surveillance Center, Silver Spring, MD, USA*
E-mail: adjoa.poku@jhuapl.edu

Objective

This paper describes the common challenges of data collection and presents a variety of adaptable frameworks that succeed in overcoming obstacles in applications of public health and electronic disease surveillance systems and/or processes, particularly in resource-limited settings.

Introduction

Electronic disease surveillance systems can be extremely valuable tools; however, a critical step in system implementation is collection of data. Without accurate and complete data, statistical anomalies that are detected hold little meaning. Many people who have established successful surveillance systems acknowledge the initial data collection process to be one of the most challenging aspects of system implementation.[1] These challenges manifest from varying degrees of economical, infrastructural, environmental, cultural, and political factors. Although some factors are not controllable, selecting a suitable collection framework can mitigate many of these obstacles. JHU/APL, with support from the Armed Forces Health Surveillance Center, has developed a suite of tools, Suite for Automated Global bioSurveillance (SAGES), that is adaptable for a particular deployment's environment and takes the above factors into account. These subsystems span communication systems such as telephone lines, mobile devices, internet applications, and desktop solutions—each has compelling advantages and disadvantages depending on the environment in which they are deployed. When these subsystems are appropriately configured and implemented, the data are collected with more accuracy and timeliness.

Methods

With SAGES we piloted multiple data collection methods against disease surveillance use cases. These pilots allowed real-time assessment of initial requirements and evaluation of performance. The discussion describes various challenges encountered by users and implementers of data collection tools for disease surveillance systems: availability, privacy regulations, timeliness, latency, coverage, digitalization, automaticity, reliability, centralization, usability, sustainability, and cost.[2] Solutions for these obstacles are presented along with a comparative analysis of the advantages and disadvantages of the various frameworks. Field notes from data collection initiatives with collaborators in Peru (US Naval Medical Research Center Detachment) and the Republic of the Philippines (Philippines-Armed Forces Research Institute of Medical Services Virology Research Unit), as well as relevant analysis of popular frameworks, supplement the discussion and provide real-life examples. Emphasis is placed on a process of understanding the targeted locale and then adapting the collection framework to operate within that locale's environment and unique impacting factors to capture relevant, accurate, and timely information.

Results

The success of any data collection initiative is dependent on the type of framework put in place. The negative impact and long-term effects resulting from using inappropriate frameworks are explored during the discussion. Furthermore, the discussion defines a process for mitigating data collection challenges by leveraging adaptable tools (Figure 1).

Conclusions

With understanding of the targeted environment, the success of surveillance systems is achievable even with inherent challenges being present. The discussion offers a method for decomposing the environment, identifying potential obstacles, and comparing suitable solutions

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 1 Data collection quadrant of APL's SAGES suite is customizable for a deployment's needs.

for those challenges to define the data collection framework that will best meet the needs of stakeholders and end users.

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 Lewis SH, Wojcik R. Methodologies for data collection. *BMC Proc* 2008;**2** (Suppl 3):SS. http://www.biomedcentral.com/1753-6561/2/S3/S5.

2 Ashar R, Lewis S, Blazes DL, Chretien JP. Applying information and communications technologies to collect health data from remote settings: a systematic assessment of current technologies. *J Biomed Inform* 2010;**43**:332–41.

Emerging Health Threats Journal 2011, **4**:s143. **doi:** 10.3134/ehtj.10.143
© 2011 H Park *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

The National Biosurveillance Integration System

H Park[1], E Pedersen[1], D Stetson[1], D Carr[1], D Bailey[1], M Stephens[1], J Best[1], A Marouf, and T Quitugua

National Biosurveillance Integration Center, Office of Health Affairs, Department of Homeland Security, Washington, DC, USA
E-mail: teresa.quitugua@dhs.gov

Objective

The National Biosurveillance Integration System (NBIS) is a consortium of federal agencies, whose joint objective is to enhance the identification, location, characterization, and tracking of biological events potentially impacting homeland security.[1–6] Together, the consortium members benefit from a joint awareness of potentially significant biological events that are unfolding or imminent, based on information shared among the group. This presentation describes the framework, activities and benefits for NBIS participants, and invites participation by other agencies.

Methods

NBIS integrates worldwide biosurveillance information across the domains of human health, animal and plant health, food and pharmaceutical issues, and relevant environmental factors. It leverages unique interagency communications and relationships to identify potential biological events of national concern. NBIS also supports prevention and mitigation of such events by providing timely notifications and ongoing situational awareness to enhance response. Biosurveillance information is gathered from a wide variety of government, scientific, restricted access, and open sources. Information regarding biological events and associated contextual information is then analyzed by NBIS subject matter experts who identify, characterize, and report events of concern in addition to their significance and potential impacts.

Results

Participating agencies report relevant information on the federal Biosurveillance Common Operating Picture (BCOP). The BCOP is a geo-temporal display of events of interest to the NBIS community, and is maintained by the Department of Homeland Security. The BCOP is currently accessible to individuals within federal NBIS agencies. NBIS is developing a companion state, local, tribal, and territorial BCOP, which will share available information with non-federal government organizations. Expansion of this service to academic and private entities is under consideration.

Conclusions

NBIS and BCOP provide a robust information-sharing framework and communication pathway, facilitating timely and accurate notifications of biological events of concern to decision makers. This enhances the nation's ability to prevent, mitigate, and respond to significant events, efficiently and effectively.

Acknowledgements

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

References

1 Government Accountability Office (GAO)-10-171, Biosurveillance: Developing a Collaboration Strategy is Essential to Fostering Interagency Data and Resource Sharing. December 2009.
2 Government Accountability Office (GAO)-10-645, Biosurveillance: Efforts to Develop a National Biosurveillance Capability Need a National Strategy and a Designated Leader. June 2010.
3 Homeland Security Presidential Directive (HSPD)-9, Defense of United States Agriculture and Food, January 2004.
4 Homeland Security Presidential Directive (HSPD)-10, Biodefense for the 21st Century. June 2002.
5 Homeland Security Presidential Directive (HSPD)-21, Public Health and Medical Preparedness. October 2007.
6 Public Law 110.53, Implementing Recommendations of the 9/11 Commission Act of 2007, August 2007.

---

[1] Contractors supporting the National Biosurveillance Integration Center.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s161. **doi:** 10.3134/ehtj.10.161
© 2011 J Reid *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Development of public health communication tools using open source methods

Jonathan Reid[1,2], Susan Mottice[1], Wu Xu[1], and Matthew Samore[2]

[1]*Utah Department of Health, University of Utah School of Medicine, Salt Lake City, UT, USA; and* [2]*Departments of Internal Medicine and Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA*
E-mail: jreid@utah.gov

Objective

There is an urgent need for improved communication between stakeholders involved in outbreak investigations, public health reporting and events of interest occurring between different jurisdictions within the same state. Currently, state and local public health agency personnel rely on personal communications involving phone, fax and snail mail. The Utah Department of Health (UDOH) sought to develop and encourage the use of a secured web portal that allows access to a variety of applications using a single sign-on. This was achieved by developing a secured communications framework called PHAccess[1] that allows tools and applications to be implemented within a secure web environment, using open source software and Agile methodology techniques. The user-centric design currently hosts an electronic report-staging area, ELR/EMR reporting, web-based reporting, secure messaging between stakeholders and a state laboratory result look-up feature. Currently, there are over 700 registered users; 3693 secure messages that have been exchanged and the site has been accessed over 12,205 times since January 2009. Informal feedback from users has been encouraging and formal evaluation is planned, along with expansion and integration with state level health information exchange projects.

Introduction

Secure and confidential exchange of information is the cornerstone of public health practice. Often, this exchange has to occur between public health agencies across jurisdictions. Examples include notification of reportable diseases when the testing and residence of the patient are in different counties. The cross-jurisdictional issues become exaggerated in times of communicable disease outbreaks or events of interest that are not yet classified as outbreaks. Currently, such communication occurs between state and local agencies and between agencies and community clinicians on a personal level, with phone, fax and snail mail. There are a multitude of secured websites hosted by UDOH that offer access to single applications requiring approved users to remember multiple sites and logins/passwords. The goal of this project was to develop a centralized, single sign-on secure web portal, from which users could access multiple applications and communicate securely with each other.

Method

The framework for PHAccess was developed using open source software, PHP and My SQL. The novel concept adopted by Apple iTunes App Store was used to develop our applications. This was supported by Agile methodology techniques with frequent iterative development cycles that were driven by constant feedback from users and stakeholders. A central premise was employing a user-centric design philosophy that allows users to influence the development of as well as manage and maintain the product.

Results

PHAccess was commissioned on 1 January 2009 after 4 months of development and iteration. The site started with three applications (Issue Tracker, Secure Messaging, Admin Functions) and has since grown to over 40. Since then, there are 744 registered users (54% state health, 14% local health departments and 32% community clinicians). The site has been accessed 12,205 times and 3693 secure messages have been exchanged. The current suite of applications include an EDI (Electronic Data Interchange) reporting suite of services that include electronic report-staging area, ELR/EMR reporting, web-based reporting, secure messaging between stakeholders and a state laboratory result look-up feature. There is also a query agent that allows the user to securely query and display results from external sources such as the Utah NEDSS or State laboratory information systems.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusions

An integrated suite of services has been developed using open source tools and modest resources to provide public health agencies and clinicians a secure single sign-on access to relevant information. PHAccess is currently in use and further expansion is planned to integrate this service into state-wide clinical health information exchange projects.

Acknowledgements

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

Reference

1 Program demonstration. http://www.rockymountaincoe.org/phaccess.

Emerging Health Threats Journal 2011, **4**:s112. **doi:** 10.3134/ehtj.10.112
© 2011 S Shen *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Identification of features for detection and prediction of homelessness from VA clinical documents

S Shen[1,2], BR South[1,2], M Palmer[1], S Duvall[1], M Samore[1,2], and AV Gundlapalli[1,2]

[1]*Salt Lake City VA Health Care System, Salt Lake City, UT, USA; and* [2]*University of Utah, Salt Lake City, UT, USA*
E-mail: shuying.shen@hsc.utah.edu

Objectives

We demonstrate a semi-automated approach to induce and curate lexical domain knowledge for identification of evidence and risk factors for homelessness found in VA clinical documents. This domain knowledge can be used to support training and evaluation of automated methods such as Natural Language Processing (NLP) systems for detection and prediction of homelessness among veterans. This could serve as a proxy for public health and other surveillance involving homeless individuals. Similar methods could be used to identify other conditions of interest.[1]

Introduction

Homelessness in general is a major issue in the US today. The risk factors of homelessness are myriad, including inadequate income, lack of affordable housing, mental health and substance abuse issues, lack of social support, and non-adherence to treatment/follow-up appointments. Early identification of these factors from clinical documents may help detect or even predict homelessness cases, allowing adequate intervention and prevention measures.

Methods

Using a think out loud approach, we developed an initial lexicon of features related to homelessness using expert inputs and available literature sources. This lexicon consists of social stressors (that is, recent divorce, unemployment), behavioral factors (that is, drug abuse), evidence (that is, lives in shelter, no housing), other risk factors (that is, exposure to war-related trauma) and direct mention of homelessness in the medical record (that is, homeless patient). This initial list was applied as pre-annotations to 600 VA clinical documents extracted from the VA Region one and four Data Warehouse for the time period 1/1/200-12/31/2009. Documents were pre-annotated using a prototype system that supports interactive annotation and semi-automated curation of user-defined information classes.

Refining the Lexicon

Domain experts reviewed pre-annotated documents to determine if information was correctly identified, make modifications to annotations, add missing annotations, or reject annotations found to be incorrect or irrelevant. We applied an iterative process of revising the lexicon until further refinements were exhausted.

Results

Our initial lexicon had 83 entries. After two rounds of semi-automated curation on 75 documents, 38 concepts were added. Pre-annotations were helpful for reviewers to focus attention around the surrounding context, revealing important cues and textual patterns that would inform guideline development and creation of a reference standard for NLP system implementation.

Conclusions

Our approach can effectively generate lexical domain knowledge combining information from literature and expert feedback via iterative refinement. This method could be easily adapted to other surveillance efforts for case identification and prediction.

Acknowledgements

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

Reference

1 Rosenheck R, Fontana A. A model of homelessness among male veterans of the Vietnam War generation. *Am J Psychiatry* 1994;**151:** 421–7.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Emerging Health Threats Journal 2011, **4**:s115. **doi:** 10.3134/ehtj.10.115
www.eht-journal.org

© 2011 BR South *et al.*; licensee Emerging Health Threats Journal.

ABSTRACT

Using clinician mental models to guide annotation of medically unexplained symptoms and syndromes found in VA clinical documents

BR South[1,2], M Palmer[1], S Shen[1,2], G Divita[1,2], S Duvall[1,2], M Samore[1,2], and AV Gundlapalli[1,2]

[1]*Salt Lake City VA Health Care System, Salt Lake City, UT, USA; and* [2]*University of Utah, Salt Lake City, UT, USA*
E-mail: brett.south@hsc.utah.edu

Objectives

We sought to develop a guideline and annotation schema that can be consistently applied to identify medically unexplained syndromes (MUS) found in VA clinical documents. These efforts will support building a reference standard used for training and evaluation of a Natural Language Processing system developed for automated symptom extraction. Our overarching goal is to characterize the occurrence of MUS in Operation Enduring Freedom (OEF)/ Operation Iraqi Freedom (OIF) veterans.

Introduction

Medically unexplained syndromes are conditions that are diagnosed on the basis of symptom constellations and are characterized by a lack of well-defined pathogenic pathways.[1] The three most common MUS are chronic fatigue syndrome (CFS), irritable bowel syndrome (IBS), and fibromyalgia. Different types of persistent symptoms, originating from different organ systems, characterize these syndromes. Patients often meet the criteria for more than one MUS.

Methods

Identification of MUS requires some level of inference on the part of annotators. For this reason we integrate clinician mental models with these review tasks. We demonstrate a practical approach that can be used to generate labeled data upon which additional layers of annotation can be added. Using this approach, annotated spans of text represent the most granular level of information necessary to identify symptoms and contextual features representing assertional information, and symptom duration. We developed an annotation guideline that provides definitions and examples of spans of text to annotate. Using a Protégé plugin called Knowtator,[2] we also developed an annotation schema used to mark spans of text representing symptoms and contextual features, along with the ability to link contextual features with an annotated symptom. Once annotations have been created at the most granular level, additional assessments of medical explainability can be obtained from subsequent clinician review. We randomly sampled 492 clinical encounter documents from a cohort of OEF/OIF veterans, from the VA Region one and four Data Warehouse, who received health care services between 1 January 2007 and 12 December 2010. As part of tasks related to refining our annotation guideline, four reviewers annotated a random sample of 15 of these documents using the annotation guideline and schema. We report the estimated number of annotated symptoms and the number of symptoms that could potentially be included in non-mutually exclusive symptom constellations for the three most common MUS.

Results

Preliminary testing shows promising results for these methods. The number of words in the 15 annotated documents varied from 554 to 3314, with an average of 17 symptom annotations per document. The number of annotations (unique mentions) for all four annotators was 1358 (583) overall with symptoms 1010 (477), followed by assertions 262 (53), and symptom duration 86 (53). Clinician *post hoc* review of the 477 unique symptom annotations revealed non-mutually exclusive symptom clusters of which 274 (57%) could potentially describe CFS, IBS 29 (6%), and 155 (32%) fibromyalgia.

Conclusions

Our methods demonstrate a practical way of conducting annotation tasks that allow reuse of labels for other symptom surveillance efforts. Once symptoms have been annotated, additional layers of review can be done by another group of clinician annotators. Annotated data can be used to train NLP tools for related surveillance tasks including extracting symptoms representing known and

This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

unknown syndromes of interest to clinicians, epidemiologists and, public health agencies.

Acknowledgements

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

References

1 Fink P, Toft T, Hansen MS, Ørnbøl E, Olesen F. Symptoms and syndromes of bodily distress: an exploratory study of 978 internal medical, neurological, and primary care patients. *Psychosom Med* 2007;**69**:30–9.
2 Ogren PV. *Knowtator: a Protégé Plug-in for Annotated Corpus Construction. Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology.* Association for Computational Linguistics: Morristown, NJ, USA, 2006. pp. 273–275.

Emerging Health Threats Journal 2011, **4**:s119. **doi:** 10.3134/ehtj.10.0119

© 2011 M Torii *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Document classification toward efficient event-based biosurveillance

M Torii, N Nelson[1], and D Hartley[1]

Division of Integrated Biodefense, ISIS Center, Georgetown University Medical Center, Washington, DC, USA
E-mail: torii@isis.georgetown.edu

Objective

This paper describes ongoing efforts in enhancing automated document classification toward efficient event-based biosurveillance.

Introduction

Event-based biosurveillance is a practice of monitoring diverse information sources for the detection of events pertaining to human health.[1–3] Online documents, such as news articles on the Internet, have commonly been the primary information sources in event-based biosurveillance.[4–8] With the large number of online publications as well as with the language diversity, thorough monitoring of online documents is challenging. Automated document classification is an important step toward efficient event-based biosurveillance. In Project Argus, a biosurveillance program hosted at Georgetown University Medical Center, supervised and unsupervised approaches to document classification are considered for event-based biosurveillance.

Methods

In Argus operations, analysts are requested to label online documents that they read in their regular surveillance work. Currently, two document classes, relevant and irrelevant, are assumed. With such labeled articles, a customized classifier is trained for target geographic regions/languages using a machine-learning algorithm. Documents retrieved from a Boolean keyword search can be classified (filtered) or ranked according to the relevancy scores assigned. In addition, we considered dynamic grouping of documents, in contrast to classification into predefined classes. To reflect analysts' perspective in clustering documents, we try to weight features (for example, keywords) based on information extracted from class-labeled documents and/or past event reports.

Results

We have tested the proposed framework to facilitate supervised machine-learning classifiers on past data. The framework has been implemented in the Argus surveillance workflow. We are in the process of evaluating the performance of trained classifiers in operational settings. Meanwhile, we became aware of inherent challenges in the framework that could affect performance of classifiers, which include class-imbalance in training data sets, that is, few labeled irrelevant (or relevant) articles may be available because of labeling bias (or to the inherent class-imbalance), and dominant topics in labeled examples, for example, articles on seasonal influenza.

Using past data, we observed that informative subsets could be derived using document clustering, for example, k-means clustering.

Conclusions

We observed promising results on automated document classification in our preliminary experiments. Previously good results have been reported by other related studies in this domain.[4–8] Meanwhile, Boolean queries created and maintained by expert analysts have also been found effective in Project Argus. The utility of automated document classifiers in contrast to the Boolean keyword search should be evaluated in real-life surveillance settings in the future.

Acknowledgements

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

References

1 Hartley DM, Nelson NP, Walters R, Arthur R, Yangarber R, Madoff L, *et al.* The landscape of international event-based biosurveillance. *Emerg Health Threats J* 2010;**3:**e3.

---

[1] These authors contributed equally to this paper.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2 Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, *et al*. Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis* 2009;**15**:689–95.

3 Walters R, Harlan P, Nelson NP, Hartley DM. Data sources for biosurveillance. In: Voeller JG (ed) *Wiley Handbook of Science and Technology for Homeland Security: Risk Analysis*. Hoboken, NJ, USA, 2009.

4 Nelson NP, Brownstein JS, Hartley DM. Event-based biosurveillance of respiratory disease in Mexico, 2007–2009: connection to the 2009 inuenza A(H1N1) pandemic? *Euro Surveill* 2010;**15**, pii = 19626. Available online http://www.eurosurveillance.org/ViewArticle.aspx? ArticleId = 19626.

5 Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveill* 2007;**12**:E071129–071125.

6 Mawudeku A. Mining the Internet: GPHIN. In *The International Meeting on Emerging Diseases and Surveillance (IMED 2007)*. Vienna, Austria, 2007.

7 Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, *et al*. BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics* 2008;**24:** 2940–1.

8 Steinberger R, Fuart F, Groot Evd, Best C, Etter Pv, Yangarber R. *Text Mining from the Web for Medical Intelligence*. OIS Press, The Netherlands, 2008.

Emerging Health Threats Journal 2011, **4**:s101. **doi:** 10.3134/ehtj.10.101
© 2011 F-C Tsui *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Building an automated Bayesian case detection system

F-C Tsui, J Espino, T Sriburadej, H Su, and J Dowling

RODS Laboratory, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: tsui2@pitt.edu

Objective

This paper describes the architecture and evaluation of our recently developed automated Bayesian case detection (BCD) system.

Introduction

Current practices of automated case detection fall into the extremes of diagnostic accuracy and timeliness. In regards to diagnostic accuracy, electronic laboratory reporting (ELR) is at one extreme and syndromic surveillance is at the other. In regards to timeliness, syndromic surveillance can be immediate, and ELR is delayed 7 days from initial patient visit.[1]

A plausible solution, a middle way, to the extremes of diagnostic precision and timeliness in current case detection practices is an automated Bayesian diagnostic system that uses all available data types, for example, freetext ED reports, radiology reports, and laboratory reports. We have built such a solution—BCD. As a probabilistic system, BCD operates across the spectrum of diagnostic accuracy, that is, it outputs the degree of certainty for every diagnosis. In addition, BCD incorporates multiple data types as they appear during the course of a patient encounter or lifetime, with no degradation in the ability to perform diagnosis.

Methods

The BCD system that we built has five components: real-time HL7 parsers, natural language processing (NLP) tools, Bayesian inference engine, a Bayesian network, and a database. The HL7 message parsers extract different data types from HL7 messages. Then, NLP tools, MedLEE,[2] and Topaz (a homegrown tool) find medical terms contained in each freetext report, including significant negative findings. We store the NLP results in a database. For non-freetext reports such as laboratory reports, we store the coded data directly to the database.

We built a Bayesian network with 57 nodes for detection of three diseases: flu, shigellosis, and measles. We formed the network structure and conditional probabilities by consulting a physician board certified in infectious diseases. We implemented the Bayesian inference engine in Java using the junction-tree algorithm. To make the BN portable, each node in the network is represented in an Unified Medical Language System (UMLS) Concept Unique Identifier.

We performed a preliminary evaluation of the BCD system using only freetext ED reports as input to detect influenza cases. The gold standard was laboratory-confirmed positive and negative reports.

Results

Our preliminary evaluation used 363 randomly selected reports (181 positive) from 12 January 2005 to 31 August 2007. We found an area under ROC curve of about 0.8 (95%CI: 0.76–0.85). When the posterior probability threshold was set to 0.8 (given $P(\text{flu}) = 0.1$), we found a sensitivity of 63.5% (95%CI: 56.5–70.5%), a specificity of 82.4% (95%CI: 76.9–87.9%), and a positive predictive value of 78.2% (95%CI: 71.5–84.9%).

We used BCD to estimate daily expected counts of flu cases presented in seven EDs of UPMC health system during H1N1 outbreak by summing the posterior probability for flu for each visit. Figure 1 shows chart of percent daily expected ED flu visits from July to December of 2009. Average daily ED visits was 569.

A demo web page of BCD is available at https://betaweb.rods.pitt.edu/casedetection-rest/demoPage.jsp. It demonstrates multiple data types' input for computing posterior probabilities of three diseases.

Conclusions

Our BCD system has good performance characteristics and is a solution to low diagnostic accuracy and timeliness in existing automated surveillance systems.

Figure 1 A flu chart showing daily percentage of ED visits with flu (green) and its 5-day moving average (black) between 7/1/2009 and 12/1/2009.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

The project is funded by CDC grants P01 HK000086 and 1U38 HK000063-01. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

1 Que J, Tsui FC, Wagner MM. Timeliness study of radiology and microbiology reports in a healthcare system for biosurveillance. *AMIA Annu Symp Proc* 2006; 1068.
2 Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004;**11**:392–402.

Emerging Health Threats Journal 2011, **4**:s102. **doi:** 10.3134/ehtj.10.102
© 2011 W Xing *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Epibasket: a prototype web tool aimed to speed up and standardize the epidemiological investigation of an outbreak

W Xing[1], G Hejblum[1], GM Leung[2], and A-J Valleron[1]

[1]*Ecole Pierre Louis de Santé Publique, Université P. & M. Curie & Inserm, Paris, France; and* [2]*School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China*
E-mail: Alain-jacques.valleron@upmc.fr

Objective

This paper describes how the ideas and tools of e-commerce can be translated to the investigation of outbreaks: epidemiologists will 'shop' the best available items for their questionnaire, enhance the chances of producing interoperable questionnaires, and speed up the whole process.

Introduction

The detailed analysis of the epidemiological literature on the 2003 SARS epidemic published in peer reviewed journals has shown that a majority (78%) of the epidemiological articles were submitted after the epidemic had ended, although the corresponding studies had relevance to public health authorities during the epidemic. The conclusion was that to minimize the lag between research and the exigency of public health practice in the future, researchers should consider adopting common, predefined protocols and ready-to-use instruments to improve timeliness, and thus, relevance, in addition to standardizing comparability across studies.[1]

Methods

The basic idea of the method is to avoid the epidemiologists faced to an outbreak to reinvent the wheel, and to provide them an online 'catalogue' of the 'best' items to include in their questionnaire. At the present time, the prototype concerns the psychobehavioral studies performed during an outbreak. We have shown[1] that they accounted for 19% of the studies made during the 2003 SARS epidemic. Questionnaires concern knowledge, attitudes, behaviors, and psychological and social impacts. The catalogue of possible questions was constituted from the comprehensive review of the relevant papers ($n=59$) published during the SARS epidemic. The 1962 variables collected in these papers were standardized with the aid of usual medical terminologies (for example, MeSH, UMLS, SNOMED, CDISC standards and so on). The user interface was built by adapting a freely available software used for e-commerce (Prestashop[2]). The queXML tool[2] was used to generate the questionnaire with the selected items in pdf format, and/or to generate a file that could be used with online survey development tools (for example, LimeSurvey, an open source free software[2]) to create an online questionnaire.

Results

In the present prototype, the epidemiologist 'shops' the items and puts them in a 'shopping basket'. At the end, he gets a paper questionnaire or a file of questions. Each item is documented (Figure 1). The parallel with e-shopping can be pursued to provide a 'value' for each item (for example, the

Figure 1 Description of an item from the online catalogue (users get the definition, MESH terms, and list of papers having used it).

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

value might include the Impact Factor of the journals that have published the papers having previously used the variable). This may be useful, for example, at the time of choosing between two multidimensional scales of risk perception.

Conclusions

The concept of EpiBasket could increase the rapidity of the data collection in the case of an outbreak, and help to standardize the data collected. Moreover, the constitution of a catalogue of 'important' items for investigative epidemiologists during an outbreak could be a collective work of ISDS.

Acknowledgements

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

References

1 Xing W, Hejblum G, Leung GM, Valleron AJ. Anatomy of the epidemiological literature on the 2003 SARS outbreaks in Hong Kong and Toronto: a time-stratified review. *PLoS Med* 2010;**7**:e1000272.
2 See www.prestashop.com, quexml.sourceforge.net/, and www.limesurvey.org.

Emerging Health Threats Journal 2011, **4**:s156. **doi:** 10.3134/ehtj.10.156

© 2011 C Weng *et al.*; licensee Emerging Health Threats Journal.

www.eht-journal.org

ABSTRACT

Map application to the New York state electronic syndromic surveillance system

C Weng, J-H Chen, C DiDonato, and H-G Chang

Statistical Unit, Division of Epidemiology, New York State Department of Health, Albany, NY, USA
E-mail: cxw14@health.state.ny.us

Objective

The objective of this paper is to describe a map application added to the New York state Electronic Syndromic Surveillance system (ESSS). The application allows system users to display the geographic distributions, and trends of fever syndrome that was used to monitor seasonal and H1N1 influenza activities.

Introduction

The primary goal of the Electronic Syndromic Surveillance system (ESSS) is to monitor trends in non-specific symptoms of illness at the community level in real time. The ESSS includes emergency department chief complaint data that are categorized into eight syndromes: respiratory, gastrointestinal, fever, asthma, neurological, rash, carbon monoxide, and hypothermia. Since the onset of H1N1, fever syndrome has been used to monitor flu activity. As H1N1 spread nationwide, the need of visualizing flu activity geographically became clear, and urgent.

Methods

SAS software (SAS Institute Inc., Cary, NC, USA) is used to generate the maps. Once the users log on to the ESSS, enter map-generating parameters via an HTML screen, a Java-based web application is developed to run SAS codes to produce the maps. As this is a web application, the backend systems will respond with the data back to the client within five minutes, which allows for on demand display of the resultant maps.

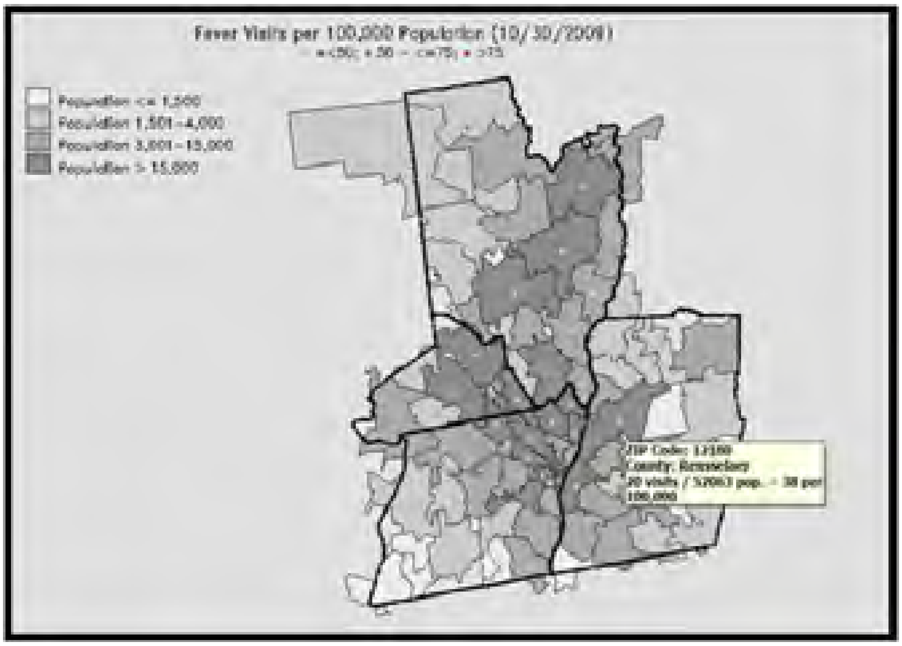


Figure 1 A snapshot of the user-generated map with pop-up label.

This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

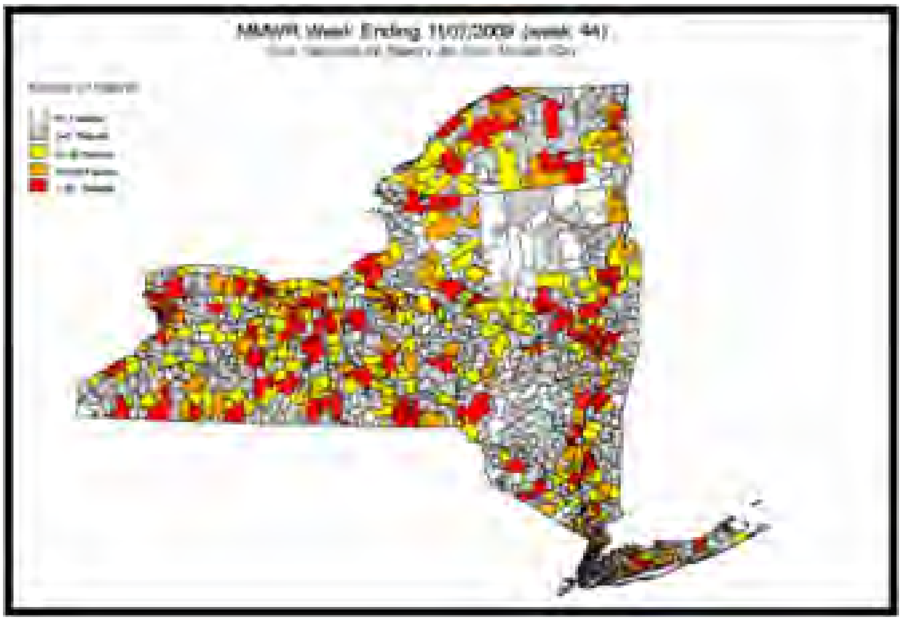


Figure 2 A snapshot of the statewide fever syndrome map.

To reduce the SAS running time for the end users, a separate SAS program was scheduled to run every morning before business hours that creates a small-sized sub-dataset for query. The SAS output is constrained to HTML and GIF images to meet the system requirement.

Results

The map application was implemented in August 2010. It contains two components. The first component is a user-generated geographic syndrome display. With this component, users input three parameters, which are where (that is, the geographic level by county, or counties), when (that is, one day in the past seven days that the users want to study), and what to generate (fever syndrome). On each resultant map, counts of emergency department fever syndrome visits are aggregated, and displayed as a rate (per 100,000 population) by patients' ZIP Codes. The ZIP Code population is represented with color shading in the resultant maps; the darker the color, the larger the population in that ZIP Code. The fever syndrome visit rates are denoted by three levels of colored dots. In addition, if users hover their cursor over a ZIP Code, a pop-up label will also appear, identifying the ZIP Code, county name, number of emergency department fever syndrome visits, population in that ZIP Code, and the fever syndrome visit rate (Figure 1). The second component of the application is a weekly statewide display of the fever syndrome visit counts. When requested, statewide fever syndrome maps of the most current four weeks will display to show the trend. Each map represents one of the CDC Morbidity and Mortality Weekly Report (MMWR) week. On the maps, emergency department fever syndrome visits are aggregated by patients' ZIP Codes. The maps are shaded by fever syndrome visit counts; the darker the color, the higher the fever syndrome visit counts (Figure 2). Users can also zoom in the maps or view the historical weekly maps for the past twelve months.

Conclusions

The map application is a new data visualization tool for the ESSS users. It provides hospital, local health department, region, and state users a user-friendly platform to monitor H1N1 and seasonal flu activities of their own jurisdiction, and the neighboring areas. The two components of the application enable users to generate informative local maps of fever syndrome visits, as well as to see the statewide trends over time. The map application does not require user-end software, or risk data security. It can be easily adapted for surveillance of other syndromic indicators.

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.

Emerging Health Threats Journal 2011, **4**:s91. **doi:** 10.3134/ehtj.10.091
© 2011 W Xu *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Enhance public health data capacity through statewide health information exchange of immunization records

W Xu, F Lanier, and N McConnell

Utah Department of Health, Salt Lake City, UT, USA
E-mail: wxu@utah.gov

Objective

The goal of the immunization information exchange is to expand the Utah Statewide Immunization Information System (USIIS) interoperability to all private providers, especially those who are not a USIIS' user but participate in clinical Health Information Exchange (cHIE).

Introduction

Utah Statewide Immunization Information System (USIIS) is the state immunization registry. It is connected to about 700 organizations, including 100% public health clinics, 60% of private providers, many schools, daycares, pharmacies, and Indian Health Services. Data exchange methods range from web data entry, batch file transfer, proprietary or HL7 data interfaces, and real-time exchange with Intermountain Healthcare's electronic health records (EHR). Clinicians in Utah ranked immunization data as one of the first five use cases for developing the statewide clinical Health Information Exchange (cHIE) in Utah Health Information Network. Utah Department of Health (UDOH) has collaborated with Utah Health Information Network to develop the immunization information exchange.

Method

We used the agile project management method and developed a roadmap with two developmental tracks and two implementation phases. The two tracks are parallel efforts to develop (1) policies and standards, and (2) technical infrastructure. The first phase is to 'push' the USIIS consolidated immunization records to the cHIE Virtual Health Records (VHR) for provider to query. The second phase aims to 'pull' new immunization information from the cHIE Virtual Health Records into USIIS.

Results

The policy track has completed following governance milestones: *The Electronic Commerce Agreement* is signed among all Utah cHIE partners. This is a statewide agreement covering all Utah Department of Health (UDOH) programs' exchange with cHIE providers.

Utah Department of Health enterprise-wide management policy and procedures are developed on the basis of USIIS and Medicaid regulations and management protocols. Roles and responsibilities for operating and maintaining the interoperability among independent information systems and cHIE are defined.

HL7 2.5.1 Immunization Standards with CDC CVX codes are formally adopted for cHIE. USIIS has worked with UHIN Standards Committee to complete the implementation specifications for cHIE users.

The technical track for phase I has developed or is developing:

The UDOH–HIE gateway

This gateway will serve all public health clinical information exchanges with cHIE. USIIS has loaded records to cHIE, including demographics, consolidated immunizations, contraindication, and exemption information. Weekly updates capacity is established as well.

CVX codes mapping

Some EHRs use national drug codes (NDC) codes to store vaccine records in drug tables. According to the standards, cHIE is mapping USIIS' CVX codes with NDC codes in the cHIE Virtual Health Records.

Displaying consolidated immunization information in patient summary has been the cHIE's latest, hopefully the last challenge to accurately present immunization records from multiple sources for cHIE users.

Conclusion

New policy and IT infrastructure are needed to enable public health to exchange clinical information through statewide HIEs. Immunization registry can be a good pilot to develop a roadmap for this endeavor.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgements

Acknowledgement to the CDC Rocky Mountain Center for Translational Research on Public Health Informatics (PI Matt Samore) and ONC funds for State HIE Cooperative Agreement Program (State HIT Coordinator David N Sundwall). This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Emerging Health Threats Journal 2011, **4**:s159. **doi:** 10.3134/ehtj.10.159
© 2011 Y Zhang and Q Chen.; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Public health data sharing policy and informatics initiatives at China CDC

Y Zhang, and Q Chen

Chinese Center for Disease Control and Prevention (China CDC), Beijing, Peoples Republic of China
E-mail: yingjie924@yahoo.com.cn

Objective

The objective of this study is to describe public health data-sharing policy, and informatics initiatives at China Center for Disease Control and Prevention (China CDC).

Introduction

Construction of data-sharing network for public health is one of the national scientific data-sharing projects, based on the data resource that distributed at China Center for Disease Control and Prevention (China CDC), universities, research institutes, and scientists, as well as the data from research projects were integrated.

Data sharing policy

China CDC believes that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The policy reaffirmed the principle that data should be made as widely, and freely available as possible while safeguarding the privacy of research participants, and protecting confidential and proprietary data.

Informatics initiatives

From 2004, China CDC has launched many informatics initiatives, including: to establish the technical platform releasing to the entire society, to push forward China's scientific data sharing, to assemble more valuable and available databases especially for epidemiology of the serious diseases, to build database integration for special research field in order to give more convenience to users, to set up the rules for technical standards data sharing, management, quality control of the databases, and service guide. The goal of these initiatives is to build infrastructure and networks to facilitate data sharing, integration, and interoperability.

Conclusions

Whole public health information, when combined with clinical and other phenotype data, offers the potential for increased understanding of disease processes affecting human health, improvement in the disease control, and prevention. For these reasons, China CDC believes that the full value of infrastructure and networks can be realized only if the public health datasets are made available as rapidly as possible to a wide range of scientific investigators.

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.

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.