OJPHI

ISDS 2015 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Data Blindspots: High-Tech Disease Surveillance Misses the Poor

Samuel V. Scarpino*[1], James G. Scott[2], Rosalind Eggo[2], Nedialko B. Dimitrov[2] and Lauren A. Meyers[2, 1]

[1]Santa Fe Institute, Santa Fe, NM, USA; [2]The University of Texas at Austin, Austin, TX, USA

Objective

Improve situational awareness for influenza by combining multiple data sources to predict influenza outbreaks in at-risk populations.

Introduction

Evidence from over 100 years of epidemiological study demonstrates a consistent, negative association between health and economic prosperity (Farmer 2001; Marmot 2005). In many settings, it is clear that causal links exist between lower socioeconomic status and both reduced access to healthcare and increased disease burden (Shi et al. 1999; Liao et al. 2004). However, our study is the first to demonstrate that the increased disease burden in at-risk populations interacts with their reduced access to healthcare to hinder surveillance.

Despite overwhelming evidence for a causative relationship between poverty and disease, critical gaps exist in our understanding of how to design surveillance systems for these at-risk communities. Past work on infectious disease surveillance has focused at the state-level (Polgreen et al. 2009; Scarpino et al. 2012) or assumed that risk was evenly spread across well-mixed populations (Pelat et al. 2014). Surveillance studies focused on broader definitions of health and on chronic diseases have found similar disparities to the ones presented here (Liao et al. 2004; Kandula et al. 2007).

Methods

As a measure of situational awareness, we focus on a surveillance system's ability to predict hospitalizations. To achieve this goal, we constructed generalized additive models. First, zip codes are partitioned into poverty quartiles. We then expanded each predictor in a third-order B-spline basis with six degrees of freedom to allow for non-linear effects. To avoid overfitting, we regularize the spline coefficients using a lasso penalty, with the regularization parameter chosen by cross-validation. We also evaluated out-of-sample Poisson log-likelihoods and performed least squares regressions. To test for the coherence of each poverty quartile, we calculated the pairwise correlation coefficient between all zip codes within a grouping. We confirmed these results using a principle component analysis. To determine significance, both for the correlation analysis and predictive performance, we randomly assigned zip codes to poverty groups 5000 times and repeated the analyses.

Results

We analyzed the effectiveness of an integrated surveillance system—-one that combines data from Biosense 2.0, ILINet, Hospital Discharge Records, and Google Flu Trends. At higher levels of aggregation—-e.g at the state-level, or multiple counties within a state-level—-we find that these data sources correlate well with seasonal influenza. We find strong evidence that these data sources work significantly better for affluent populations than for less affluent, at-risk populations. Furthermore, we find that these most at-risk zip codes are more synchronous with each other and have higher hospitalization rates for influenza.

Conclusions

Populations with lower socioeconomic status often experience higher hospitalization rates across a range of diseases. One causative mechanism for this increased burden is reduced access to healthcare (Shi et al. 1999). Our results suggest that this reduced access may also have profound public health consequences, by impairing situational awareness. Specifically, we find that an integrated, data-driven surveillance system can accurately predict one-week ahead inpatient influenza hospitalizations in wealthier, but not poorer, more at-risk zip codes. This indicates that the high-tech, integrated surveillance systems of recent focus in the literature have a data blindspot—-these technology-driven systems miss at-risk populations.

Keywords

Disease Surveillance; Poverty; Influenza; Forecasting

References

Farmer (2001). Infections and inequalities: The modern plagues.

Kandula et al. (2007). Differences in self-reported health among asians, latinos, and non-hispanic whites: the role of language and nativity. Annals of Epidemiology.

Liao et al. (2004). Reach 2010 surveillance for health status in minortty communities—united states, 2001–2002. Morbidity and mortality weekly report.

Pelat et al. (2014). Optimizing the precision of case fatality ratio estimates under the surveillance pyramid approach. American journal of epidemiology.

Polgreen et al. (2009). Optimizing influenza sentinel surveillance at the state level. American journal of epidemiology.

Scarpino et al. (2012). Optimizing provider recruitment for influenza surveillance networks. PLoS Comput Biol.

Shi et al. (1999). Income inequality, primary care, and health indicators. The Journal of family practice.

***Samuel V. Scarpino**
E-mail: scarpino@santafe.edu