

# Socio-environmental and measurement factors drive variation in influenza-like illness

Elizabeth Lee\*<sup>1</sup> and Shweta Bansal<sup>1,2</sup>

<sup>1</sup>Biology, Georgetown University, Washington, DC, USA; <sup>2</sup>Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

## Objective

To assess the use of medical claims records for surveillance and epidemiological inference through a case study that examines how ecological and social determinants and measurement error contribute to spatial heterogeneity in reports of influenza-like illness across the United States.

## Introduction

Traditional infectious disease epidemiology is built on the foundation of high quality and high accuracy data on disease and behavior. Digital infectious disease epidemiology, on the other hand, uses existing digital traces, re-purposing them to identify patterns in health-related processes. Medical claims are an emerging digital data source in surveillance; they capture patient-level data across an entire population of healthcare seekers, and have the benefits of medical accuracy through physician diagnoses, and fine spatial and temporal resolution in near real-time.

Our work harnesses the large volume and high specificity of diagnosis codes in medical claims to improve our understanding of the mechanisms driving spatial variation in reported influenza activity each year. The mechanisms hypothesized to drive these patterns are as varied as: environmental factors affecting transmission or virus survival, travel flows between different populations, population age structure, and socioeconomic factors linked to healthcare access and quality of life. Beyond process mechanisms, the nature of surveillance data collection may affect our interpretation of spatial epidemiological patterns [1], particularly since influenza is a non-reportable disease with non-specific symptoms ranging from asymptomatic to severe. Considering the ways in which medical claims are generated, biases may arise from healthcare-seeking behavior, insurance coverage, and medical claims database coverage in study populations.

## Methods

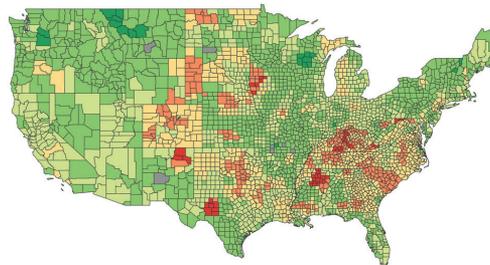
Using aggregated U.S. medical claims for influenza-like illness (ILI) from the 2001-2002 through 2008-2009 flu seasons [2], we developed a Bayesian hierarchical modeling framework to estimate the importance of both ecological and social determinants and measurement-related factors on observed county-level variation of influenza disease burden across the United States. Integrated Nested Laplace Approximation (INLA) techniques for Bayesian inference were used to render our questions computationally tractable due to the high spatial resolution of our data (Figure 1) and the multiplicity of models in our analysis [3]. Linking data from a variety of publicly available sources, we determined the strength, directionality, and consistency of these factors over multiple flu seasons.

## Results

We found that measurement-related factors – healthcare-seeking behavior, insurance coverage, and medical claims database coverage – were strong predictors of greater ILI intensity across seasons. Secondly, poverty and specific humidity were negatively associated with ILI intensity for several seasons. Finally, by incorporating mechanistic and measurement factors into our model, our model predictions present an improved map of influenza-like illness in the United States for the flu seasons in our study period.

## Conclusions

We present a flexible modeling approach that applies to different medical claims diagnosis codes and disease surveillance data and demonstrates the utility of Bayesian hierarchical models for large-scale ecological analyses. Our results increase our knowledge of the spatial distribution of influenza and the underlying processes that drive these patterns, promote finer spatial targeting for different types of interventions, and enable the interpolation of burden in areas difficult to surveil through traditional public health. Moreover, they highlight the relative contributions of surveillance data collection and ecological processes to spatial variation in disease, and highlight the importance of considering measurement biases when using surveillance data for epidemiological inference.



Seasonal Flu Intensity (150.2-300] (88.7-150.2] (49-88.7] (21.1-49] (0-21.1] (0-0]

Figure 1. Observed seasonal ILI intensity for United States counties during the 2007-2008 flu season demonstrates the extent and resolution of medical claims data coverage. Greater values indicate larger disease burden and grey areas had no reported data.

## Keywords

influenza; medical claims; ecological analysis; Bayesian; United States

## Acknowledgments

This work was supported by the Jayne Koskinas Ted Giovanis Foundation for Health and Policy (dissertation support grant to ECL), the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

## References

- Lee EC, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference. *J Infect Dis*. In press.
- Viboud C, Charu V, Olson D, et al. Demonstrating the use of high-volume electronic medical claims data to monitor local and regional influenza activity in the US. *PLoS One*. 2014; 9(7):e102429.
- Rue H, Martino S, Chopin N. Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. *J R Stat Soc Ser B*. 2009; 71(2):319–392.

\*Elizabeth Lee

E-mail: ecl48@georgetown.edu

