

# Measuring and Improving the Quality of Data Used for Syndromic Surveillance

Brian E. Dixon<sup>\*1, 2, 3</sup>, Jon Duke<sup>4, 1</sup> and Shaun Grannis<sup>1, 5</sup>

<sup>1</sup>Center for Biomedical Informatics, Regenstrief Institute, Inc., Indianapolis, IN, USA; <sup>2</sup>Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, IN, USA; <sup>3</sup>Department of Veterans Affairs, Health Services Research & Development Service, Indianapolis, IN, USA; <sup>4</sup>Georgia Tech Research Institute, Atlanta, GA, USA; <sup>5</sup>Indiana University School of Medicine, Indianapolis, IN, USA

## Objective

To extend an open source platform for measuring the quality of electronic health data by adding functions useful for syndromic surveillance.

## Introduction

Nearly all of the myriad activities (or use cases) in clinical and public health (e.g., patient care, surveillance, community health assessment, policy) involve generating, collecting, storing, analyzing, or sharing data about individual patients or populations. Effective clinical and public health practice in the twenty-first century requires access to data from an increasing array of information systems, including but not limited to electronic health records. However, the quality of data in electronic health record systems can be poor or “unfit for use.” Therefore measuring and monitoring data quality is an essential activity for clinical and public health professionals as well as researchers.

## Methods

Using the Health Data Stewardship Framework<sup>1</sup>, we will extend Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems (ACHILLES), a software package published open-source by the Observational Health Data Sciences and Informatics collaborative (OHDSI; [www.ohdsi.org](http://www.ohdsi.org)) to measure the quality of data electronically reported from disparate information systems. Our extensions will focus on analysis of data reported electronically to public health agencies for disease surveillance. Next we will apply the ACHILLES extensions to explore the quality of data captured from multiple real-world health systems, hospitals, laboratories, and clinics. We will further demonstrate the extended software to public health professionals, gathering feedback on the ability of the methods and software tool to support public health agencies’ efforts to routinely monitor the quality of data received for surveillance of disease prevalence and burden.

## Results

To date we have mapped key surveillance data fields into the OHDSI common data model, and we have transformed 111 million syndromic surveillance message segments pertaining to 16.4 million emergency department encounters representing 6 million patients for importation into ACHILLES. Using these data, we are exploring the existing 167 metrics across 16 categories available within ACHILLES, including a person (e.g., number of unique persons); and observation period (e.g., Distribution of age at first observation period). Syndromic surveillance (SS), however, is driven largely by monitoring patient stated chief complaints (non-standard free text clinical data) in addition to coded diagnoses. Consequently, ACHILLES must be extended to maximally support use in analyzing SS datasets.

## Conclusions

This work remains a work-in-progress. Over the coming year, we will not only explore existing ACHILLES constructs using real-world

public health data but also introduce new functionality to explore 1) patient demographics; 2) facility and location (e.g., emergency department where care was delivered); and 3) clinical observations (e.g., chief complaint). The design and methods for examining these aspects of surveillance data will be included on the poster, and they will be made freely available for distribution with a future instance of the ACHILLES software. We ultimately envision these tools being available for use on platforms such as the CDC’s Biosense – open to all local and state health agencies as a one-stop portal for surveillance data analysis – or research environments where they can be used to examine and improve the quality of data output from informatics systems.

## Keywords

Data quality; Syndromic surveillance; Electronic health records

## Acknowledgments

Research reported in this abstract was supported by the National Library Of Medicine of the National Institutes of Health under Award Number R21LM012219. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

1. Dixon BE, Rosenman M, Xia Y, Grannis SJ. A vision for the systematic monitoring and improvement of the quality of electronic health data. *Studies in health technology and informatics*. 2013;192:884-8.

**\*Brian E. Dixon**

E-mail: [bedixon@regenstrief.org](mailto:bedixon@regenstrief.org)

