

# Accrued – An R Package for Visualizing Data Quality for Aggregate Surveillance Data

Julie Eaton<sup>1</sup>, Ian Painter\*<sup>2</sup> and William Lober<sup>2</sup>

<sup>1</sup>University of Washington - Tacoma, Tacoma, WA, USA; <sup>2</sup>University of Washington, Seattle, WA, USA

## Introduction

The utility of specific sources of data for surveillance, and the quality of those data, are an ongoing issue in public health(1). Syndromic surveillance is typically conducted as a secondary use of data collected as part of routine clinical practice, and as such the data can be of high quality for the clinical use but of lower quality for the purpose of surveillance. A major data quality issue with surveillance data is that of timeliness. Data used in surveillance typically arrive as a periodic process, inherently creating a delay in the availability of the data for surveillance purposes. Surveillance data are often collected from multiple sources, each with their own processes and delays, creating a situation where the data available for surveillance are accrued piecemeal. From 2006 to 2012 the ISDS ran Distribute (2), a surveillance system for monitoring influenza like illness (ILI) and gastroenteritis (GI) ED visits on a nationwide basis. This system collected counts for ILI, GI and total ED visits, aggregated to the level of jurisdiction. The primary data quality issue faced with the Distribute system was that of timeliness due to accrual lag; variable delays in the receipt of surveillance data from sources by jurisdictions together with variable delays in the reporting of aggregate data from jurisdictions to Distribute resulted in data which accrued over time(3).

## Methods

We have developed “accrued”, an R package for visualizing and analyzing data quality for accrued data, based on methods developed and used for data quality analysis in the Distribute project. R is an open source system for statistical analysis and programming, available on a multitude of platforms (including OS X, Windows and all major Linux variants), is used extensively in many research fields, and has a robust system for adding functionality through add-on packages (4).

The “accrued” package contains several visualization tools for understanding data quality issues, (upload pattern graphs, stacklag difference graphs - stacked time series graphs of counts by lag), summarizing the effects of timeliness on data quality (stacked lag histograms, timeliness (completion summary curves) and analyzing the effects of timeliness of accuracy (quantile error plots).

This package is freely available under the GPL3 open source license on the R project website(5), and includes extensive documentation using R’s built in help system as well as with a summary vignette showing how to use the package. We will discuss the motivation for the package, the underlying technical structure of the package and demonstrate use of the various functions in the package. We will also discuss progress on an additional package under development for monitoring data quality for individual level surveillance data.

## Keywords

Data Quality; Surveillance; Software

## References

1. Reynolds T, Painter I, Streichert L. (2013) Data Quality: A Systematic Review of the Biosurveillance Literature. *Online Journal of Public Health Informatics* 04/2013; 5(1).

2. Olson DR, Paladini M, Lober WB, Buckeridge D, ISDS Distribute-Working Group (2011), “Applying a New Model for Sharing Population Health Data to National Syndromic Surveillance for Influenza: The DiSTRIBuTE. Project Proof of Concept, 2006 to 2009,” *PLoS Currents: Influenza* 2011, Aug 2;3: RRN1251. <http://www.ncbi.nlm.nih.gov/pubmed/21894257>
3. Painter I, Eaton J, Olson D, Revere D, Lober W (2011) .” Visualizing data quality: tools and views” *Emerging Health Threats Journal* 01/2011; 4:11144. DOI:10.3402/ehth.v4i0.11144
4. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
5. <http://cran.r-project.org/web/packages/accrued/>

\*Ian Painter

E-mail: [ipainter@uw.edu](mailto:ipainter@uw.edu)

