

# Processing of Novel Electronic Health Data to Support Public Health Surveillance

Peter Hicks\*, Henry Rolka, Mark Wooster and Lynette Brammer

CDC, Atlanta, GA, USA

## Objective

To describe data management and analytic processes undertaken to rapidly acquire and use previously unavailable data during a public health emergency response.

## Introduction

Accurately gauging the health status of a population during an event of public health significance (e.g. hurricanes, H1N1 2009 pandemic) in support of emergency response and situation awareness efforts can be a challenge for established public health surveillance systems in terms of geographic and population coverage as well as the appropriateness of health indicators. The demand for timely, accurate, and event-specific data can require the rapid development of new data assets to “fill-in” existing information gaps to better characterize the scope, scale, magnitude, and population health impact of a given event within a very narrow time-window. Such new data assets may be concurrently under development and evaluation while being used to support response efforts. Recent examples include the “drop-in” surveillance processes deployed at evacuation centers following Hurricane Katrina and the illness and injury surveillance systems established for response workers during the Deepwater Horizon Oil spill response. During the 2009 H1N1 pandemic response, CDC acquired access to data from several national-level health information systems that previously had been un-vetted as public health information sources. These sources provided data extracts from massive administrative or electronic medical records (EMR) based in hospital and primary care settings. It was hoped that such data could supplement existing influenza surveillance systems and aid in the characterization of the pandemic. Few of these new data sources had formal documentation or concise information on the underlying populations and geographies represented.

## Methods

Throughout CDC’s H1N1 response; epidemiologists, data managers, and IT specialists collaborated to develop standardized methods to rapidly characterize, process, store, and provision these new data for analysis and reporting by subject matter experts. These new data were not part of a formally designed sample so each data source needed to undergo extensive empirical review to understand, representativeness, unique nuances, and facilitate the interpretation of analytic results and accurate reporting to public health decision makers.

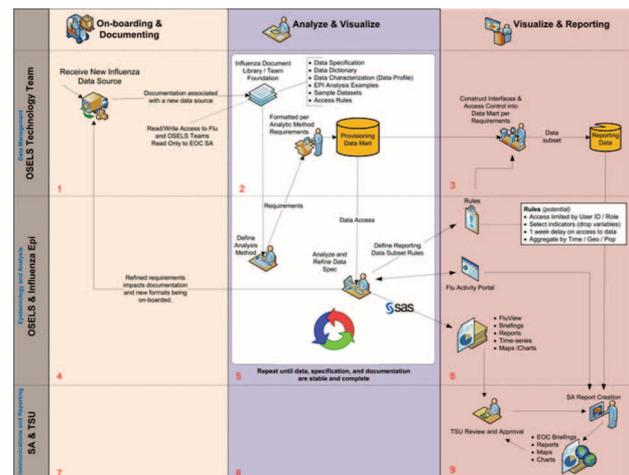
## Results

Such work requires a multi-disciplinary approach that cyclically reviews incoming data iteratively while concurrently documenting findings, modifying initial business rules (e.g. extraction, binning, or coding logic), and analytic techniques to produce the most interpretable and informative results. To elucidate the underlying complexity for these sequential and contingent activities occurring across information technology, informatics, and epidemiology domains, we

retrospectively described the intersection of the discrete tangible tasks and workforce roles via a TaskFlow diagram (Figure 1). Vertical “swim lanes” represent discrete tasks: On-boarding/Documentation, Analysis/Visualization, and Visualization/Reporting. Workforce roles such as Data management, Epidemiological Analysis, and Communications are broken into three horizontal “swim lanes” as each requires dramatically different skillsets and are accomplished by different individuals. Each of the steps (1-9) in the diagram were leveraged to produce supplemental artifacts (e.g. code books, extraction guides, defined analytic methods, etc.) to support ongoing analysis, interpretation, reporting, and over process improvement. The totality of all of these interrelated activities have an a priori purpose of characterizing population health during an event of public health significance to support disease prevention and control efforts in a timely fashion.

## Conclusions

This presentation describes the underlying business processes, activities, and roles used in transforming novel data sources, during the H1N1 response, into informative assets to support public health surveillance. By formally articulating and describing each of these steps, in a structured manner, we hope to contribute to the dialogue of developing useful practices for leveraging electronic health data to meet public health surveillance challenges.



## Keywords

informatics; surveillance; emergency response; h1n1; data management

\*Peter Hicks

E-mail: phicks@cdc.gov

