

National Healthcare Safety Network Innovations

Creation of Synthetic Data Sets for Ventilator-Associated Events

National Center for Emerging and Zoonotic Infectious Disease

Project value

This project provided vendors and facilities with the means to check their ventilator associated events code, to determine if an event occurred. The validation process went from a manual, labor intensive process to an electronic process.

Program perspective

The National Healthcare Safety Network (NHSN) is a centralized web-based system that is used to monitor healthcare-associated infections (HAIs). NHSN supports national-level public surveillance, state reporting mandates, and aggregated data for CMS who provides reimbursement based on reporting of quality data.

Context of project

NHSN tracks a type of HAI called ventilator associated events (VAE) by asking facilities and vendors to report when events occur. NHSN provides the protocol through a definition of what an event is, employing a combination of factors (e.g., metrics, patient vital signs, and ventilator readings), but NHSN had no insight into how a specific facility implements their internal code to determine whether an event occurred.



Currently, healthcare facility workers enter most data manually into the NHSN system. However, over 25% of the 14,000 reporting facilities submit at least some data electronically via Clinical Document Architecture (CDA). The problem with the current reporting process is that validating the electronic data manually is a labor intensive process and can be limited in scope. The Centers for Disease Control and Prevention (CDC) needed a way to electronically validate that data providers at the healthcare facilities apply NHSN definitions properly and consistently in order that the multiple public health partners (including other CDC programs, healthcare facilities, state public health agencies, and CMS) can receive the quality, accurate, and complete data they depend on from NHSN. The goal is to increase the number of facilities reporting data electronically via CDA.

Stakeholders

The stakeholders were CDC epidemiologists, vendors and facilities that the vendors represent, staff within the Division for Healthcare Quality and Promotion who study the data, public health departments that have jurisdiction over the facility and want to be involved in the prevention efforts



to reduce VAE infections, subject matter experts at Brigham and Women's Hospital in Boston, and NHSN resources: an informatician, computer scientist, and an infection preventionist VAE expert.

Methodology

The project team received de-identified data from Brigham Women's Hospital in Boston to create synthetic Electronic Health Records (HER) patient datasets. It took approximately 2-3 months to build the synthetic data set, then an XML version of the data was created. The data was manipulated to cover typical cases, outlier cases, and edge cases. Then the data was amended to cover all conditions. They compared the data set against an existing VAE web service, and viewed the result set. The results included whether they had a VAE event, and the event type. Once a result set was done, an SME reviewed it to confirm that the data was correct, then they sent the data to Brigham and Women's Hospital for review.

Results

The final product of this project included a synthetic data set in an XML format with documentation on how to use the data set. The synthetic data set allowed vendors and other data providers to test their VAE efforts and validate the data within their systems. It doesn't matter as much how different vendors implement their code as long as they can identify the embedded events in the synthetic dataset. Before the project, vendors tested their own code by running it against a set of patients and someone would look over the data and do a validation based on their own understanding of conditions.

The electronic validation process overcomes several challenges including people making mistakes or misinterpreting the nuances of the data set, different facilities using variable means of checking for events, and privacy and sensitivity concerns with testing new code on a facility's patient data. As a result of this project, when a CDA message is sent to NHSN, it accurately represents an HAI event and provides a more reliable public health resource.

Opportunities for reuse in public health

Validating data is not new in other healthcare areas, but it is new for HAI reporting. The validation process for HAIs can be replicated with healthcare facilities, vendors, and other stakeholders. Because different vendors are comfortable doing different things. Don't write code for specific vendors. The synthetic data set allows vendors to code for themselves while validating their work.

The entire process involved:

- CDC providing the dataset to a vendor,
- the vendor imports the data, runs algorithms, produce CDAs, report to NHSN, then CDC will check the validity of the CDA format and ensure the results match the data the vendor provides; if the data matches, the data is certified, however, if messages fail to validate, vendors will receive an error messages in their email. Once CDC certifies the results, and finally,
- the information is made available on the NHSN website.

The next step is to create a formal EHR certification process and publish the data set along with instructions on the NHSN web site.

What is CDC's Health Information Innovation Consortium (CHIIC)?

CDC's Office of Public Health Scientific Services (OPHSS) launched the CDC Health Information Innovation Consortium (CHIIC), as part of the agency's Surveillance Strategy, to foster and promote creative solutions to surveillance challenges that are unique to public health. CHIIC serves as a channel for innovative projects in CDC programs and collaborates with State, Tribal, Local, and Territorial agencies.

More Information

Creation of Synthetic Data Sets for Ventilator-Associated Events - Project Details

<http://www.cdc.gov/ophss/chiic/forums/2015-02-24-synthetic-patient-dataset-.html>

National Healthcare Safety Network (NHSN)

<http://www.cdc.gov/nhsn/>

CDC Surveillance Innovation Project Descriptions

www.cdc.gov/ophss/chiic/project_portfolio.html

CDC's Health Information Innovation Consortium (CHIIC)

www.cdc.gov/ophss/chiic

CDC Surveillance Strategy

<http://www.cdc.gov/surveillance>

Email CHIIC for more information

chiic@cdc.gov

Disclaimer: Funding for this project was provided through the CDC Health Information Innovation Consortium (CHIIC) as part of the CDC Surveillance Strategy Innovation Project Portfolio.

