

Monitoring and Improving Syndromic Surveillance Data Quality

MisChele A. Vickers

Infectious Diseases and Outbreaks Division, Alabama Department of Public Health, Montgomery, Alabama, United States

Objective

To monitor and improve the data quality captured in syndromic surveillance for Alabama Department of Public Health Syndromic Surveillance (AlaSyS).

Introduction

The public health problem identified by Alabama Department of Public Health Syndromic Surveillance (AlaSyS) was that the data reflected in the user application of ESSENCE (Electronic Surveillance System for the Early Notification of Community-based Epidemics) was underestimating occurrences of syndromic alerts preventing Alabama Department of Public Health (ADPH) from timely recognition of potential public health threats. Syndromic surveillance (SyS) data in ESSENCE were not reliable for up to a week after the visit date due to slow processing, server downtime, and untimely data submission from the facilities. For AlaSyS, 95 percent of data should be submitted within 24 hours from time of visit, for near real time results. The slow data processing caused latency in the data deeming it less useful for surveillance purposes, consequently the data was not meaningful for daily alerts. For example, if a user ran a report to assess the number of Emergency Department (ED) visits that mentioned heroin in the chief complaint (CC), depending on the status of the data coming from the facility (processing, sending, or offline), the number of visits visible to the user could vary from one to several days. With the opioid epidemic Alabama is currently facing, this delay poses a major public health problem.

Methods

During the data quality improvement review, AlaSyS PHRA addresses all three data quality metrics: (1) completeness – any data element of interest that is less than 95 percent complete, that field is reported to the facility along with some guidance to reference the Public Health Information Network (PHIN) messaging guide for assistance on how to correct; (2) validity – any data element of interest that is less than 95 percent conforming is flagged for corrections (this is not a very big data quality issue for production or onboarding); and (3) timeliness – data are requested to be sent in a timely manner (i.e., at least once every 24 hours); anything sent more than 24 hours after the visit is highlighted and sent to the facility in a report. Now that quality data is coming into ESSENCE, AlaSyS staff had to address the issue of latency to improve representativeness. AlaSyS has approximately 82 facilities sending data to production. When updates occur from the NSSP or a facility was not sending data in a timely manner, facilities in ESSENCE would appear to be offline. This bottleneck of data being processed caused a backlog of data sometimes in excess of 3 days. For example, the data coming to the ESSENCE platform would, in some cases, appear 7 days after the patient visit. These occurrences led to the development of the AlaSyS “Current Production” spreadsheet. This allows the AlaSyS Team to record the status of each facility in the event data is not current, e.g., a facility temporarily drops from production due to a vendor change or upgrade. At any given moment, AlaSyS staff know the count of facilities in production, regardless of the overall general status. AlaSyS PHRA has developed queries in R Studio to help monitor the data flow status. If the data drops, this is noted on the Current Production spread sheet and AlaSyS staff is aware, even before the disruption of the data flow is reflected in ESSENCE. The query returns the name of facilities that are sending data on a particular date. This has allowed AlaSyS staff to identify data drops earlier.

Results

After the implementation of the Current Production spread sheet, monitoring of the timeliness metric in syndromic surveillance data has improved. By analyzing the NSSP data validation reports for completeness and validity, and providing feed back to the vendors and facilities, the data quality of what is captured in ESSENCE has also improved.

The data quality reports that target the onboarding facilities were used to transition seven facilities (six hospitals and one urgent care) from onboarding to production during the period. The completeness data quality reports were used to validate the completeness metric in order to support the transition to production. The data quality reports that targeted the production data



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

generated conversations between the AlaSyS Team and the facilities regarding barriers that impeded their improvement. The timeliness metric for example, some facilities are set up to send data once every 24 hours. This results in a lag time to ESSENCE of up to almost 48 hours. Facilities may not be able to improve their timeliness measure without incurring a cost from the vendor for an upgrade. In other instances, facilities are able to send in real time. However, at the time of this document, the BioSense platform is only capable of accepting data in 15minute increments for ESSENCE. AlaSyS staff were able to improve representativeness using the Current Production Status spreadsheet. This communication tool allows users more reliability of the data by knowing the status of the facilities in the catchment area on any given day. For example, the user is able to know if a feed is down on a day they want to run a report. That same report, run again later, (when the feed is active) may show up as a fluctuation in the data. Understanding the nature of the data in this scope will help improve and support reliability. AlaSyS PHRA will incorporate the spreadsheet into an Access database to be displayed on the AlaSyS data management internal website. The Current Production Status spreadsheet as a validation tool, also supports the monitoring of the timeliness metric.

An internal website for syndromic surveillance data management has been developed so users can be informed of the facility status and the latest data quality reports of each facility. The development of the “AlaSyS Data Intranet” web site will be available for registered ADPH staff who use ESSENCE. The AlaSyS data analysis and reporting (A&R) will be available for users to check the status of the data feed for facilities before making a report. This intranet will provide back door information so the ESSENCE user can have a window of confidence with the data prior to creating a report.

Conclusions

By engaging the facilities with the data quality reports, AlaSyS staff was able to find out some of the barriers such as the facility not having funds to upgrade to a more timely system. Also uncovered was acknowledgment the capability (including limitations) of how quickly the NSSP server can process data. For example, while a facility may be able to submit in real time, (as opposed to near-real time) the ability to process data in real time is not option at the time of this document. AlaSyS staff also learned when data is not appearing in ESSENCE, this absence does not necessarily mean that the data is not being sent from the facility. Point to consider also, is the understanding, nature and behavior of the data helps to improve reliability.

When reporting using AlaSyS data, it is important to be mindful of the limitations. Developing a check and balance system for data validation to find root cause for proper evaluation and resolve will support improvement to data quality. Utilizing properly calibrated measuring tools helps to ensure that data quality metrics are effective and measuring as intended.

Acknowledgement

Alabama Department of Public Health Centers for Disease Control and Prevention National Syndromic Surveillance Program



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.