

Integrating data from disparate data systems for improved HIV reporting: Lessons learned

Kamran Ahmed*¹, Yvette Temate–Tiagueu¹, Joseph Amlung², Dennis L Stover², Philip J Peters¹, John T Brooks¹, Sridhar Papagari Sangareddy¹ and Jina J Dcruz¹

¹Centers for Disease Control and Prevention, Atlanta, GA, USA; ²Indiana State Department Of Health, Indianapolis, IN, USA

Objective

To assess the integration process of HIV data from disparate sources for reporting HIV prevention metrics in Scott County, Indiana

Introduction

In 2015, the Indiana State Department of Health (ISDH) responded to a large HIV outbreak among persons who inject drugs (PWID) in Scott County¹. Information to manage the public health response to this event and its aftermath included data from multiple sources such as surveillance, HIV testing, contact tracing, medical care, and HIV prevention activities. Each dataset was managed separately and had been tailored to the relevant HIV program area's needs, which is a typical practice for health departments. Currently, integrating these disparate data sources is managed manually, which makes this dataset susceptible to inconsistent and redundant data. During the outbreak investigation, access to data to monitor and report progress was difficult to obtain in a timely and accurate manner for local and state health department staff. ISDH initiated efforts to integrate these disparate HIV data sources to better track HIV prevention metrics statewide, to support decision making and policies, and to facilitate a more rapid response to future HIV-related investigations. The Centers for Disease Control and Prevention (CDC) through its Info-Aid mechanism is providing technical assistance to support assessment of the ISDH data integration process. The project is expected to lead to the development of a dashboard prototype that will aggregate and improve critical data reporting to monitor the status of HIV prevention in Scott County.

Methods

We assessed six different HIV-related datasets in addition to the state-level integrated HIV dataset developed to report HIV monitoring and prevention metrics. We conducted site visits to the ISDH and Scott County to assess the integration process. We also conducted key informant interviews and focus group discussions with data managers, analysts, program managers, and epidemiologists using HIV data systems at ISDH, Scott County and CDC. We also conducted a documentation review of summary reports of the HIV outbreak, workflow, a business process analysis, and information gathered during the site visit on operations, processes and attributes of HIV data sources. We, then, summarized the information flow, including the data collection process, reporting, and analysis at federal, state and county levels.

Results

We have developed a list of lessons learned that can be translated for use in any state-level jurisdiction engaged in HIV prevention monitoring and reporting:

Standardization of data formats: The disparate data sources storing HIV-related information were developed independently on different platforms using different architectures; they were not necessarily designed to link and exchange data. Hence, these systems could not seamlessly interact with each other, posing challenges when rapid data linkage was needed.

To better manage unstructured data coming from disparate data sources and improve data integration efforts, we recommend standardization of data formats, unique identifiers for registered individuals, and coding across data systems. Use of standard operating procedures can streamline data flow and facilitate automated creation of integrated datasets. This approach may be helpful for future integration efforts in other healthcare domains.

Data integration process: Manually integrating data is time intensive, increases workload, and poses significant risk of human error in data compilation. Hence, it may compromise data quality and the accuracy of HIV prevention metrics used by decision-makers.

We propose an automated integration process using an extract, transform and load (ETL) method to extract HIV-related data from disparate data sources, transforming it to fit the prevention metrics reporting needs and loading it into a state-level integrated HIV dataset or database. This approach can drastically decrease dependency on manual methods and help avoid data compilation errors.

Dashboard development: Major challenges in the process of integrating HIV-related data included disparate data sources, compromised data quality, and the lack of standard metrics for some of the HIV-related metrics of interest. Despite these challenges to data integration, creation of a dashboard to track HIV prevention metrics is feasible. Integrating data is a critical part of developing an HIV dashboard that can generate real-time metrics without creating additional burden for the health department staff, if manual integration is no longer needed.

Stakeholder participation: Due to the immediate need for outbreak response, involvement of stakeholders at all levels was limited. Active stakeholder engagement in this process is essential. The stakeholders' interest and participation can be improved by helping them understand the value of each other's data, and providing regular feedback about their data and its best use in public health interventions.

Conclusions

This assessment highlighted the importance of standardizing data formats, coding across systems for HIV data, and the use of unique identifiers to store individuals' information across data systems. Promoting stakeholder understanding of the value and best use of their data is also essential in improving data integration efforts. The results of this assessment offer an opportunity to learn and apply these lessons to improve future public health informatics initiatives, including HIV (but not limited to HIV), at any state-level jurisdiction.

Keywords

Integration; HIV; Dashboard; Data-Systems

References

1. Peters PJ, Pontones P, Hoover KW, Patel MR, Galang RR, Shields J, et al. HIV Infection Linked to Injection Use of Oxycodone in Indiana, 2014-2015. *N Engl J Med*. 2016;375(3):229-39.

*Kamran Ahmed

E-mail: drkamranrajput@gmail.com

