Daily Visualization of Statewide COVID-19 Healthcare Data

Brian E. Dixon, PhD, MPA[1]
Regenstrief Institute, Inc
Indiana University – Fairbanks School of Public Health

Shaun J. Grannis, MD, MS[2]
Regenstrief Institute, Inc
Indiana University School of Medicine

Umberto Tachinardi, MD, MS[3]
Regenstrief Institute, Inc
Indiana University School of Medicine

Jennifer L. Williams, MPH[4]
Regenstrief Institute, Inc

Connor McAndrews, MS[5]
Regenstrief Institute, Inc

Peter J. Embí, MD, MS[6]
Regenstrief Institute, Inc
Indiana University School of Medicine

ABSTRACT

To manage a localized outbreak or global pandemic like COVID-19, Public Health agencies (PH) and health systems utilize a variety of information systems. Although existing PH information systems enable capture of data on laboratory-confirmed cases of COVID-19, the current pandemic has illuminated several deficits in the existing U.S. information infrastructure, including gaps in access to and visualization of near-real-time (daily) impacts to the healthcare system. To address these gaps, we leveraged our state-wide health information exchange-derived dataset that represents nearly all healthcare facilities in Indiana. The resultant dashboard has evolved to present data on hospitalization, emergency department utilization, and other metrics of interest to PH and a broader constituency across the state.

Index Terms and Keywords: COVID-19. Coronavirus. Dashboard. Health Information Exchange. Public Health Informatics. Visualization.COVID-19. Coronavirus. Data Visualization.

BACKGROUND

Public health (PH) agencies play a critical role in managing disease epidemics, such as COVID-19. To manage a localized outbreak or global pandemic, PH agencies utilize a variety of information systems. Public health laboratories, for example, are often the first entities to identify and document a novel virus or genetic strain of a disease in their management information systems (1).

[1]bedixon@regenstrief.org
[2]sgrannis@regenstrief.org
[3]utahina@iu.edu
[4]craigjen@regenstrief.org
[5]conmcand@regenstrief.org
[6]pembi@regenstrief.org (corresponding)

Public health agencies further use case management systems to capture data on infected populations, like details collected during contact tracing following laboratory confirmation of infection. Although existing PH information systems enable capture of data on laboratory-confirmed cases of COVID-19, the current pandemic has illuminated several deficits in the existing U.S. information infrastructure. The holes in the "swiss cheese" of PH infrastructures created challenges as demand for information quickly outstripped the capacity of most state and local PH agencies to produce data on the emerging outbreak. One important type of information that PH agencies typically do not receive but is critical for situational awareness of epidemic effects is near-real-time data on visits to emergency departments and hospitals (2). This is an exemplary case of the classic paradigm of transforming data into information that drives knowledge, in this case informed decision making in PH.

OBJECTIVE

Our objective was to support public health response to COVID-19 across a single state in the U.S by rapidly developing and implementing a novel visualization tool to present data on hospitalization, emergency department utilization, and other metrics of interest. The visualization tool leveraged data available through a regional health information exchange (HIE) and state public health laboratory testing sources.

METHODS

We developed, implemented, and refined two versions of a COVID-19 visualization tool to address the needs of various stakeholders and partners. These included the Regenstrief Institute (RI) COVID-19 Tracker (restricted access) and the RI COVID-19 Dashboard (public access). Both are Tableau©-based visualization tools that collate information on individuals tested for and infected with COVID-19, as well as COVID-like illnesses including influenza and pneumonia. The data that feeds the visualization platform is updated and pre-processed daily to inform situational awareness with respect to COVID-related testing, cases, hospitalizations, emergency department visits, intensive care utilization, mortality, recovery, co-morbidity, demographics, and geography.

In addition to descriptive representations of the data, we employ probabilistic algorithms (3-5) to uniquely identify each person across data sources. This ensures accurate reporting of information, because many persons receive multiple laboratory tests for COVID-19 due to a) uncertainty in test results; b) hospitals often test persons transferred from other hospitals and

nursing homes; and c) employers testing persons multiple times to determine eligibility to report for work.

The visualization tools leverage the robust information infrastructure available in Indiana, referred to as the Indiana Network for Patient Care (INPC), one of the oldest and largest health information networks in the U.S. (6) The INPC is a repository of clinical and administrative health data collected from over 38 health systems representing 117 hospitals, 18,486 physician practices, commercial laboratories, and public health departments across the state. Data are collected, harmonized, and stored by IHIE. Emergency department visits and hospitalizations are sourced from admission, discharge, and transfer (ADT) messages. In the INPC, ELRs and ADTs as well as encounter level diagnoses and co-morbidities are exchanged predominantly using HL7 version 2.5.1 messages, although some source systems use alternative message formats including prior versions of HL7.

Analysts from the Regenstrief Data Services team pull data once per day and, after performing data quality procedures, transform the data into an analytical dataset for the two visualization tools. The data pulls involve not only data extraction and deduplication functionality, but also descriptive statistical calculations. The per capita rate (e.g., rates per 100,000 population) is calculated by dividing the number of individuals meeting numerator criteria by the 2018 Census-based population of the geographic area selected by the user. Overall prevalence positivity and daily incidence positivity are calculated as well as other metrics, including a metric measuring the estimated active and recovered cases across the state.

RESULTS

We launched the RI COVID-19 Tracker on March 30, 2020, and the public-facing Dashboard approximately 3 weeks later. The original tracker debuted with a series of bar charts reporting individuals tested for COVID-19, influenza, or other coronaviruses. (**Figure 1)** Demographics for each person were displayed using stacked bar charts, and line graphs depicted cumulative hospitalizations since the start of the outbreak. While useful, we obtained user-feedback over-time that indicated the early visuals were increasingly complex and cognitively burdensome for some. We also received feedback that many more users across both public and private sectors wanted to view key elements of the Tracker. Therefore, we developed a public version of our dashboard, that preserved individual privacy and changed the visualizations to reduce cognitive burden. To achieve the latter, we used visual design techniques including creating headline graphics for key daily numerical figures, a consistent color palette to call out different portions of the dashboard, migrating to the use of pyramid charts trend lines that communicated more information in a more intuitive fashion. As the project progressed, we continued to reduce burden by removing redundant or less important information through iterative cycles of development that involved end-users working on the state's COVID-19 response (**Figure 2**).

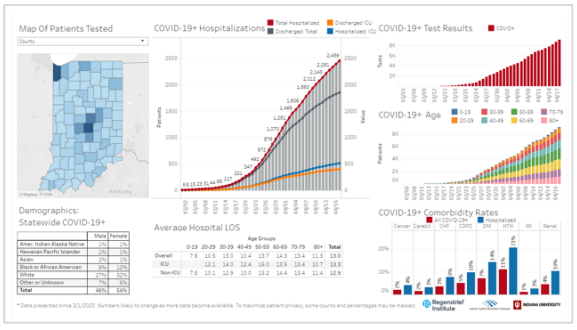

Figure 1: Early version of the Regenstrief Institute COVID-19 Tracker – private dashboard

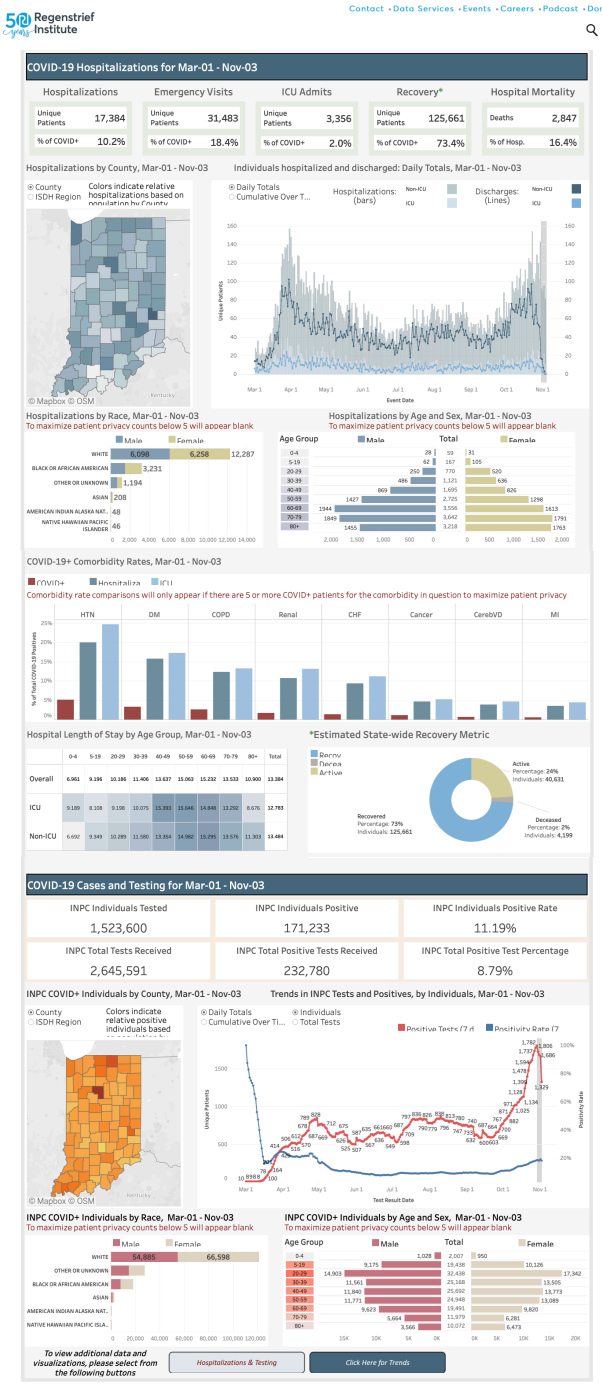

Figure 2: Regenstrief Institute COVID-19 public-facing dashboard: https://www.regenstrief.org/covid-dashboard/

The visualizations also feature a number of functions enabling users to drill down into details useful for examining trends at the local level. Users can filter results by selecting a geographic area of interest (e.g., county, zip code, school district) or use a radio button at the top to filter results to a specific diagnosis or laboratory test of interest. Users can also change the time period of

interest, allowing examination of data before or after certain events such as the Governor's stay-at-home order or a major holiday (e.g., Fourth of July).

Minor changes and additional features have been added on approximately a bi-weekly update cycle since launch, in response to changes in the pandemic's effects and end user requests. For instance, in May of 2020, we created a calculation to represent estimated recovery statistic and represent that on the dashboard (**Figure 2**). Occasionally, we have also added major new updates. Once such example was the development and incorporation in June, 2020, of a Trending section to the public-facing dashboard (**Figure 3**).

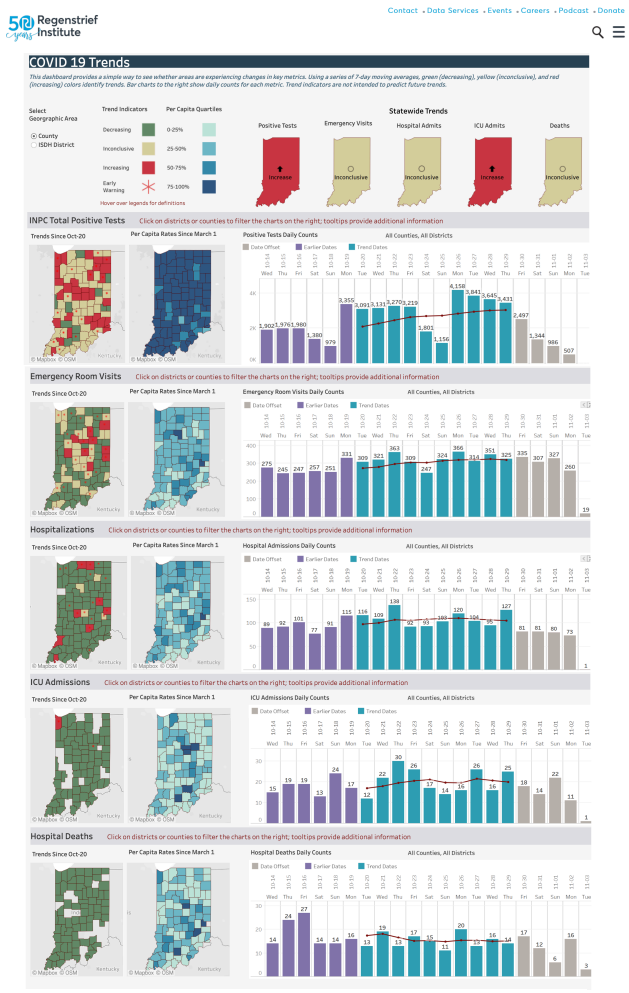

Figure 3: Top of the Trends section of the Regenstrief Institute COVID-19 Dashboard:

CONCLUSION

The development by the Regenstrief Institute of novel data visualization tools for COVID-19-related health events has been critical our State's response to the pandemic. Led by Regenstrief Institute scientists, working in close collaboration with statewide partners across multiple sectors, these tools were enabled by the existence of our state-wide HIE and the integration of myriad data sources in near-real-time to support decision making among multiple stakeholder groups. Indeed, as reported by our state partners and demonstrated by the regular referencing of the data they display, these tools have provided essential and otherwise unavailable health-system level data to inform key decision makers throughout the pandemic and enable a data-driven strategy to a state-wide response.

Since their creation, we continue to add new functions as deemed necessary by users with input from informatics faculty at the Regenstrief Institute as well as state and local public health officials. As flu season approaches, we are working to further enhance visualizations of syndromic surveillance data to distinguish between influenza and COVID-19 cases. The lessons learned from these data visualization efforts are already illuminating data collection, standardization, and integration challenges, as well as providing evidence for optimizing data visualization. In addition to continuing to use these tools for the current pandemic, we are also working to further enhance our data infrastructure and capabilities to track and respond to other diseases and health conditions at a State-level.

REFERENCES

[1] Meigs M, Merrick R, Hinrichs S. Public Health Laboratories. In: Magnuson JA, Dixon BE, editors. Public Health Informatics and Information Systems. Health Informatics. 3 ed. Switzerland AG: Springer Nature; 2020.

[2] Andrews RM. Statewide Hospital Discharge Data: Collection, Use, Limitations, and Improvements. Health Serv Res. 2015;50 Suppl 1(Suppl 1):1273-99.

[3] Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. AMIA Annu Symp Proc. 2003:259-63.

[4] Grannis SJ, Xu H, Vest JR, Kasthurirathne S, Bo N, Moscovitch B, et al. Evaluating the effect of data standardization and validation on patient matching accuracy. J Am Med Inform Assoc. 2019;26(5):447-56.

[5] Li X, Xu H, Shen C, Grannis S. Automated linkage of patient records from disparate sources. Stat Methods Med Res. 2018;27(1):172-84.

[6] Overhage JM. The Indiana Health Information Exchange. In: Dixon BE, editor. Health Information Exchange: Navigating and Managing a Network of Health Information Systems. 1 ed. Waltham, MA: Academic Press; 2016. p. 267-79.