Using medical records of patients with COVID-19, researchers create agile analytic tool for rapid disease insights
19 August 2020

More than a decade ago, electronic medical records were all the rage, promising to transform health care and help guide clinical decisions and public health response.

With the arrival of COVID-19, researchers quickly realized that electronic medical records (EMRs) had not lived up to their full potential—largely due to widespread decentralization of records and clinical systems that cannot "talk" to one another.

Now, in an effort to circumvent these impediments, an international group of researchers has successfully created a centralized medical records repository that, in addition to rapid data collection, can perform data analysis and visualization.

The platform, described Aug.19 in Nature Digital Medicine, contains data from 96 hospitals in five countries and has yielded intriguing, albeit preliminary, clinical clues about how the disease presents, evolves and affects different organ systems across different categories of patients COVID-19.

For now, the platform represents more of a proof-of-concept than a fully evolved tool, the research team cautions, adding that the initial observations enabled by the data raise more questions than they answer.

However, as data collection grows and more institutions begin to contribute such information, the utility of the platform will evolve accordingly, the team said.

"COVID-19 caught the world off guard and has exposed important deficiencies in our ability to use electronic medical records to glean telltale insights that could inform response during a shapeshifting pandemic," said Isaac Kohane, senior author on the research and chair of the Department of Biomedical Informatics in the Blavatnik Institute at Harvard Medical School. "The new platform we have created shows that we can, in fact, overcome some of these challenges and rapidly collect critical data that can help us confront the disease at the bedside and beyond."

In its report, the Harvard Medical School-led multi-institutional research team provides insights from early analysis of records from 27,584 patients and 187,802 lab tests collected in the early days of epidemic, from Jan. 1 to April 11. The data came from 96 hospitals in the United States, France, Italy, Germany and Singapore, as part of the 4CE Consortium, an international research repository of electronic medical records used to inform studies of the COVID-19 pandemic.
"Our work demonstrates that hospital systems can organize quickly to collaborate across borders, languages and different coding systems," said study first author Gabriel Brat, HMS assistant professor of surgery at Beth Israel Deaconess Medical Center and a member of the Department of Biomedical Informatics. "I hope that our ongoing efforts to generate insights about COVID-19 and improve treatment will encourage others from around the world to join in and share data."

The new platform underscores the value of such agile analytics in the rapid generation of knowledge, particularly during a pandemic that places extra urgency on answering key questions, but such tools must also be approached with caution and be subject to scientific rigor, according to an accompanying editorial penned by leading experts in biomedical data science.

"The bar for this work needs to be set high, but we must also be able to move quickly. Examples such as the 4CE Collaborative show that both can be achieved," writes Harlan Krumholz, senior author on the accompanying editorial and professor of medicine and cardiology and director of the Center for Outcomes Research and Evaluation at Yale-New Haven Hospital.

What kind of intel can EMRs provide?

In a pandemic, particularly one involving a new pathogen, rapid assessment of clinical records can provide information not only about the rate of new infections and the prevalence of disease, but also about key clinical features that can portend good or bad outcomes, disease severity and the need for further testing or certain interventions.

These data can also yield clues about differences in disease course across various demographic groups and indicative fluctuations in biomarkers associated with the function of the heart, kidney, liver, immune system and more. Such insights are especially critical in the early weeks and months after a novel disease emerges and public health experts, physicians and policymakers are flying blind. Such data could prove critical later: Indicative patterns can tell researchers how to design clinical trials to better understand the underlying drivers that influence observed outcomes. For example, if records are showing consistent changes in the footprints of a protein that heralds aberrant blood clotting, the researchers can choose to focus their monitoring, treatments on organ systems whose dysfunction is associated with these abnormalities or focus on organs that could be damaged by clots, notably the brain, heart and lungs.

The analysis of the data collected in March demonstrates that it is possible to quickly create a clinical sketch of the disease that can later be filled in as more granular details emerge, the researchers said.

In the current study, researchers tracked the following data:

- Total number of COVID-19 patients
- Number of intensive care unit admissions and discharges
- Seven-day average of new cases per 100,000 people by country
- Daily death toll
- Demographic breakdown of patients
- Laboratory tests to assess cardiac, immune and kidney and liver function, measure red and white blood cell counts, inflammatory markers such as C-reactive protein, as well as two proteins related to blood clotting (D-dimer) and cardiac muscle injury (troponin)

Tell tale patterns

The report's observations included:

- Demographic analyses by country showed variations in the age of hospitalized patients, with Italy having the largest proportion of elderly patients (over 70 years) diagnosed with COVID-19.
- At initial presentation to the hospital, patients showed remarkable consistency in lab tests measuring cardiac, immune, blood-clotting and kidney and liver function.
- On day one of admission, most patients had relatively moderate disease as measured by lab tests, with initial tests showing moderate abnormalities but no indication of organ failure.
Major abnormalities were evident on day one of diagnosis for C-reactive protein—a measure of inflammation—and D-dimer protein, a chemical that measures blood clotting with test results progressively worsening in patients who went on to develop more severe disease or died.

Levels of the liver enzyme bilirubin, which indicate liver function, were initially normal across hospitals but worsened among persistently hospitalized patients, a finding suggesting that most patients did not have liver impairment on initial presentation.

Creatinine levels—which measure how well the kidneys are filtering waste—showed wide variations across hospitals, a finding that may reflect cross-country variations in testing, in the use of fluids to manage kidney function or differences in timing of patient presentation at various stages of the disease.

On average, white blood cell counts—a measure of immune response—were within normal ranges for most patients but showed elevations among those who had severe disease and remained hospitalized longer.

Even though the findings of the report are observations and cannot be used to draw conclusions, the trends they point to could provide a foundation for more focused and in-depth studies that get to the root of these observations, the team said.

"It's clear that amid an emerging pathogen, uncertainty far outstrips knowledge," Kohane said. "Our efforts establish a framework to monitor the trajectory of COVID-19 across different categories of patients and help us understand response to different clinical interventions."

More information: Nature Digital Medicine, DOI: 10.1038/s41746-020-00308-0