4CE: Extracting COVID-19 Data from EHRs

As the respiratory disease COVID-19 is caused by a new pathogen, SARS-CoV-2, there is so much uncertainty that makes the battle against the pandemic even more challenging. Experts continue with their research work to have a better understanding on COVID-19's underlying physiology and its interactions with the body's different system processes.

You might also like: [EHR Systems: From Record to Plan](#)

A global community of researchers have formed the Consortium for Clinical Characterization of COVID-19 by EHR (4CE) – pronounced "foresee" – to help answer some of the clinical and epidemiological questions around COVID-19 through harmonisation and analysis of data stored in electronic health records (EHRs).

In a recent paper (Brat et al. 2020) 4CE, comprising 96 hospitals across five countries, reports that the largely untapped EHR data is helpful in bridging the knowledge gaps about COVID-19. This includes data on incidence, prevalence, case-fatality rates, and clinical predictors of disease severity and outcomes.

The consortium's initial efforts seek to consolidate, share and interpret data about the clinical trajectories of the infection in patients with a first focus on laboratory values and comorbidities. Within a three-week period, 96 hospitals in the United States (45), France (42), Italy (5), Germany (3) and Singapore (1) contributed data to the consortium.

Overall, the consortium's dataset covers 27,584 patients with COVID-19 diagnosis. 4CE researchers – most of whom are or have been members of the [i2b2](#) Academic Users Group – also collected 187,802 laboratory values and harmonised them across sites.

Using automated data extraction methods, the consortium was able to show results consistent with country-level demographic and epidemiological differences identified in the literature. These findings include:

- Demographic breakdown by age and sex. Age distribution was different across countries and consistent with previously identified patterns. In particular, patients from Italy were more commonly over the age of 70 relative to other countries.
- Rates of total case rise consistent with international tracking sites. To normalise across sites and countries with varying sizes, 4CE researchers reported 7-day average new case rate per 100K over time for each country normalised by the ratio between the inpatient discharge rate for each country and inpatient discharge rate for the 4CE sites in that country.
- Results varied for each laboratory value and site, with no obvious country-level pattern. Notably, there was greater between-hospital variation for laboratory test performance than between-country variation.

"Our initial data extraction included 14 laboratory markers of cardiac, renal, hepatic and immune dysfunction that have been strongly associated with poor outcomes in COVID-19 patients in previous publications," 4CE researchers wrote.

The consortium said that aside from interoperability issues, i.e. wide differences in units and data presentation, variations in international clinical diagnosis (ICD) coding and inclusion made harmonisation difficult. Despite these limitations, 4CE was able to create a framework to capture the trajectory of COVID-19 disease in patients and their response to interventions. This framework, designed to be a highly scalable system, is now

implemented at 23 sites.

"We recognise that these early data are incomplete and are subject to many biases and limitations, which constrain the conclusions we can currently draw," the researchers point out. "However, we believe the sources of our data and the mechanism we have established for sharing them are sound, reproducible, and scalable."

The consortium hopes that this collaborative project will encourage other sites to share data and contribute to this important research effort.

Source: Nature

Image credit: a-image via iStock

Published on : Tue, 8 Sep 2020