Inconsistencies in data presentation could harm efforts against COVID-19

August 20 2020

Credit: Pixabay/CC0 Public Domain

Since COVID-19 emerged late last year, there's been an enormous amount of research produced on this novel coronavirus disease. But the content publicly available for this data and the format in which it's presented lack consistency across different countries' national public health institutes, greatly limiting its usefulness, Children's National Hospital scientists report in a new study. Their findings and suggestions,

published online August 19 in *Science & Diplomacy*, could eventually help countries optimize their COVID-19-related data—and data for future outbreaks of other diseases—to help further new research, clinical decisions and policy-making around the world.

Recently, explains study senior author Emmanuèle Délot, Ph.D., research faculty at Children's National Research Institute, she and her colleagues sought [data](link) on sex differences between COVID-19 patients around the world for a new study. However, she says, when they checked the information available about different countries, they found a startling lack of consistency, not only for sex-disaggregated data, but also for any type of clinical or demographic information.

"The prospects of finding the same types of formats that would allow us to aggregate information, or even the same types of information across different sites, was pretty dismal," says Dr. Délot.

To determine how deep this problem ran, she and colleagues at Children's National, including Eric Vilain, M.D., Ph.D., the James A. Clark Distinguished Professor of Molecular Genetics and the director of the Center for Genetic Medicine Research at Children's National, and Jonathan LoTempio, a doctoral candidate in a joint program with Children's National and George Washington University, surveyed and analyzed the data on COVID-19.

The research spanned data reported by public health agencies from highly COVID-19 burdened countries, viral genome sequence data sharing efforts, and data presented in publications and preprints.

At the time of study, the 15 countries with the highest COVID-19 burden at the time included the US, Spain, Italy, France, Germany, the United Kingdom, Turkey, Iran, China, Russia, Brazil, Belgium, Canada, the Netherlands and Switzerland. Together, these countries represented

more than 75% of the reported global cases. The research team combed through COVID-19 data presented on each country's public health institute website, looking first at the dashboards many provided for a quick glimpse into key data, then did a deeper dive into other data on this disease presented in other ways.

The data content they found, says LoTempio, was extremely heterogenous. For example, while most countries kept running totals on confirmed cases and deaths, the availability of other types of data—such as the number of tests run, clinical aspects of the disease such as comorbidities, symptoms, or admission to intensive care, or demographic information on patients, such as age or sex—differed widely among countries.

Similarly, the format in which data was presented lacked any consistency among these institutes. Among the 15 countries, data was presented in plain text, HTML or PDF. Eleven offered an interactive web-based data dashboard, and seven had comma-separated data available for download. These formats aren't compatible with each other, LoTempio explains, and there was little to no documentation about where the data that supplies some formats—such as continually updated web-based dashboards—was archived.

Dr. Vilain says that a robust system is already in place to allow uniform sharing of data on flu genomes—the World Health Organization's (WHO) Global Initiative on Sharing All Influenza Data (GISAID)—which has been readily adapted for the virus that causes COVID-19 and has already helped advance some types of research. However, he says, countries need to work together to develop a similar system for harmonized sharing other types of data for COVID-19. The study authors recommend that COVID-19 data should be shared among countries using a standardized format and standardized content, informed by the success of GISAID and under the backing of the WHO.

In addition, the authors say, the explosion of research on COVID-19 should be curated by experts who can wade through the thousands of papers published on this disease since the pandemic began to identify research of merit and help merge clinical and basic science.

"Identifying the most useful science and sharing it in a way that's usable to most researchers, clinicians and policymakers, will not only help us emerge from COVID-19 but could help us prepare for the next pandemic," Dr. Vilain says.

More information: Jonathan LoTempio et al. We Can Do Better: Lessons Learned on Data Sharing in COVID-19 Pandemic Can Inform Future Outbreak Preparedness and Response. *Science and Diplomacy.* www.sciencediplomacy.org/artic … ic-can-inform-future

Provided by Children's National Hospital