

Data sharing in the era of COVID-19

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) continues to test the capacity of world health systems. Since the outbreak started, the global community has learned about coronavirus disease 2019 (COVID-19), the disease resulting from SARS-CoV-2. In the first few weeks of the pandemic, knowledge about the disease and its treatment was generated from sharing of anecdotal observations and small case series. Although health-care professionals use modern technology to communicate, never before has the failure to build robust data-sharing systems for large-scale near real-time analysis in health care been more obvious.

In the era of electronic health records, physiologic, laboratory, imaging, decision-making, and treatment data are continuously recorded. Inferences drawn from these data can inform epidemiological inquiries and guide treatment protocols when clinical trial data do not exist or might be too slow to inform a rapidly evolving situation. While the number of trials increases, real-time treatment data accumulates, siloed within hospital systems. When considering COVID-19, the insight we could gain from a pooled, publicly available dataset analysed by researchers in academic institutes and industry is invaluable and necessary.

Unfortunately, patient-level COVID-19 data is not publicly available. These data also lack comprehensive information beyond typical registry resolution. In this interconnected world, we can imagine a unifying multinational COVID-19 electronic health record waiting for global researchers to apply their methodological and domain expertise. No such database exists, and this failing is not rooted in an absence of technology or precedent. Within intensive care, for instance, the Medical Information Mart for Intensive Care (MIMIC) has been a model

of publicly-available, deidentified, electronic health record data sharing since 1996.^{1,2} Containing approximately 50 000 patient admissions to the Beth Israel Deaconess Medical Center (BIDMC), MIMIC represents the most studied critical care cohort in the world, allowing clinicians and computer scientists to address research questions and build predictive models.³

MIMIC is evidence of the possibility of data sharing beyond BIDMC's critical care department and hospital. Although the academic community has embraced data monetisation, regulatory hurdles, funding apparatuses, and a publish-or-perish academia at the expense of open data sharing, this short-sightedness does not need to be an undoing. A dreadful unprecedented worldwide event deserves an appropriate response, and this response begins with an extraordinary joining of forces—and data—to best understand the event, and the successes and failures of different treatments. The clinical and academic community will learn many lessons during these turbulent times, and crucially, needs to learn that data on health and disease should be shared universally so that everyone can benefit.

LAC is part of the team behind the Medical Information Mart for Intensive Care database at the Massachusetts Institute of Technology. The other authors declare no competing interests.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

*Christopher V Cosgriff, Daniel K Ebner,
Leo Anthony Celi
lceli@mit.edu

MIT Critical Data, Laboratory for Computational Physiology, Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (CVC, DKE, LAC); Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA, USA (CVC); Alpert Medical School, Brown University, Providence, RI, USA (DKE); and Division of Pulmonary Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA (LAC)

1 Moody GB, Mark RG. A database to support development and evaluation of intelligent intensive care monitoring. *Comp Cardiol* 1996; **33**: 657–60.

- 2 Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; **3**: 160035.
- 3 Cosgriff CV, Celi LA, Stone DJ. Critical care, critical data. *Biomed Eng Comput Biol* 2019; **10**: 1179597219856564.

