EBioMedicine 74 (2021) 103738

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Commentary

Scalable data systems require creating a culture of continuous learning

James Brogan[a], Maria del Pilar Arias López[b], Harumi Tokashiki[c,*], Leo Anthony Celi[de,f]

[a] *Albert Einstein College of Medicine, Bronx, NY, United States*
[b] *Sociedad Argentina de Terapia Intensiva. SATIQ Program, Hospital de Niños Ricardo Gutierrez, Argentina*
[c] *Tufts University School of Medicine, Carney Hospital, Boston, MA, United States*
[d] *Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States 02139*
[e] *Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States 02215*
[f] *Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States, 02115*

A R T I C L E I N F O

Article History:
Received 22 November 2021
Accepted 23 November 2021
Available online 16 December 2021

Establishing the data infrastructure necessary to enable health research is essential for improving healthcare equity, global health, and economic prosperity. Research has played a pivotal role in the improvement of human health, but the gains are more pronounced in countries where most of the research takes place. Despite a higher burden of disease, only 2% of scientific publications in indexed journals come from low-income and middle-income countries (LMICs).[1]

Evidence-based guidelines derived from studies performed in high-income countries (HICs) are implemented in LMICs that have different patient populations, disease prevalence, resources, and needs. For instance, despite a higher burden of sepsis in LMICs, with 90% of 48·9 million cases and 11 million deaths in 2017 occurring in LMICs, guidelines provided by The Surviving Sepsis Campaign for management of septic shock and sepsis-associated organ dysfunction in children were formulated by forty-nine collaborators, only three of whom were from LMICs.[2] The potential to digitise medical data across the world could help LMICs build their own clinical practice guidelines and redraw the unequal map of medical knowledge generation and validation.[3]

Over the past decade, there has been a global movement towards digitisation of healthcare data leading to modernisation of healthcare data standards that integrate existing and emerging systems through programming interfaces to ensure safe and timely flow of data.[4,5] The Asia eHealth Information Network, for example, brings together national officers responsible for digital health and non-government stakeholders to build an approach towards data governance, program management and standards across Asia.[6] At the same time, decreased cost of computational capacity, the proliferation of smartphones and tablets, and migration of infrastructure to the cloud has shifted the barrier for building and deploying scalable healthcare data systems from one of hardware and software to one of people and culture.

As data standards such as Fast Healthcare Interoperability Resources and Logical Observation Identifiers Names and Codes gain traction across the world, LMICs will have the potential to collect, store, and exchange their healthcare data during patient encounters using universal data standards. When technological innovation no longer serves as the principal barrier, each country could build its own medical knowledge system specifically based on its patient population, and thus leverage and learn from what is otherwise the "digital exhaust" of care. In this regard, there has been limited success in the creation of knowledge systems in both HICs and LMICs to date. The arduous task of data curation, a requisite step before any analysis can be performed, does not scale in the absence of a collaborative research ecosystem. Such an ecosystem remains elusive across the globe even in countries with significant investments in digital health and artificial intelligence.

Investing in people and culture will enable health organisations to build the kinds of scalable data systems that will fuel quality improvement and continuous learning based on data from their own populations. International collaborations are important to enable and expedite knowledge discovery from local healthcare data systems with less resources while also facilitating their representation in research studies. With this vision in mind, MIT Critical Data works to break down silos across disciplines, institutions, regions, and stakeholders to advance health data science.[7] The consortium consists of healthcare practitioners, computer scientists, engineers, public health providers, and social scientists who believe that data and learning are the best medicine for population health. We build communities of practice with worldwide representation to derive knowledge from data routinely collected in the process of care to better understand health and disease, particularly in the local context.

The MIT Laboratory for Computational Physiology (LCP), which leads MIT Critical Data, developed and maintains the publicly

* Corresponding author at: Tufts University School of Medicine, Carney Hospital, 2100 dorchester ave, Boston, MA 02124, United States

E-mail addresses: james.brogan@einsteinmed.org (J. Brogan), harumi.tokashiki@tufts.edu (H. Tokashiki), lceli@mit.edu (L.A. Celi).

https://doi.org/10.1016/j.ebiom.2021.103738

available Medical Information Mart for Intensive Care (MIMIC). MIMIC II, III, and IV are the most widely-used electronic health record (EHR) databases with open-source code repositories (MIMIC Code Repository, http://github.com/LCP/mimic-code).[8] These resources comprise a large set of open-access educational materials that have been developed by the LCP and are freely shared with the global research community. Over 25,000 credentialed users in academia and industry from over 110 countries utilise the resource for clinical research studies, exploratory analyses, and the development of clinical algorithms and decision support tools.

Over the past 5 years, the LCP has organised 38 international events in 16 countries including Thailand, Brazil, Colombia, Uganda, Ghana, and Kenya. These include data hackathons, or datathons, and machine learning workshops.[9] Our ecosystem incorporates cross-disciplinary expertise representing intersections between data science, healthcare, and the communities they serve. Their input is complemented by pedagogical insights from teachers and education experts; the result is everyone learning with and from each other to produce insights that could not have been achieved by any one of these specialty areas working in isolation. These initiatives take time and have yet to produce concrete examples of success stories in LMICs.

Big data is proliferating in diverse forms within the healthcare field, not only because of the adoption of EHRs, but also because of the growing use of wireless technologies for ambulatory monitoring.[10] The availability of bigger and better data to build digital health tools including algorithms requires a new breed of healthcare providers as well as additional non-clinical members of the team. Investments in transforming the culture towards a shared objective of team learning will ultimately translate into data systems that scale. Advanced hardware and software technologies will not generate dividends for population health unless the organisational culture fosters these kinds of teams which can collaborate and innovate to leverage these technologies.

Contributors

All authors contributed equally to the conceptualization, writing, review, and editing of this commentary. They all approved the final version of this work.

Declaration of interests

The authors declare no conflicts of interest.

References

1 Salager-Meyer F. Scientific publishing in developing countries: Challenges for the future. J English Acad Purpos 2008;7(2):121–32.
2 Weiss S, Peters M, Agus M. Children's Surviving Sepsis Campaign International Guidelines for the Management of Septic Shock and Sepsis-Associated Organ Dysfunction in Children Taskforce: Perspective of the surviving sepsis campaign on the management of pediatric sepsis in the era of coronavirus disease 2019. Pediatr Crit Care Med 2020;21(11):e1031–7.
3 Czerniewicz L. It's time to redraw the world's very unequal knowledge map. The Conversation 2015;24:2020.
4 Mandl KD, Kohane IS. A 21st-century health IT system-creating a real-world information economy. N Engl J Med 2017;376(20):1905–7.
5 Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low-and middle-income countries. Bull World Health Organ 2015;93:203–8.
6 Governing Council: Asia eHealth Information Network; Available from: https://www.asiaehealthinformationnetwork.org/about_aehin/governing-council/.
7 Cosgriff CV, Charpignon M, Moukheiber D, Lough ME, Gichoya J, Stone DJ, et al. Village mentoring and hive learning: The MIT Critical Data experience. iScience 2021;24(6).
8 Johnson AE, Stone DJ, Celi LA, Pollard TJ. The MIMIC Code Repository: enabling reproducibility in critical care research. J Am Med Inform Assoc 2018;25(1):32–9.
9 Aboab J, Celi LA, Charlton P, Feng M, Ghassemi M, Marshall DC, et al. A "datathon" model to support cross-disciplinary collaboration. Sci Transl Med 2016;8(333):333.ps8-ps8.
10 Mantena S, Celi LA, Keshavjee S, Beratarrechea A. Improving community healthcare screenings with smartphone-based AI technologies. The Lancet Digital Health 2021;3(5):e280–2.