Payne Philip Richard Orrin (Orcid ID: 0000-0002-9532-2998)

Learning from Data: A Recurring Feature on the Science and Practice of Data-Driven Learning Health Systems

Embi PJ, Payne PR, Friedman CP

Peter J. Embi, MD, MS, FACP, FACMI, FIAHSI
Professor and Chair, Department of Biomedical Informatics
Professor of Medicine
Senior Vice-President for Research and Innovation
Vanderbilt University Medical Center
peter.embi@vumc.org
Philip R.O. Institute for Informatics, Washington

Philip R.O. Payne, PhD
Institute for Informatics
Washington University in St. Louis
prpayne@wustl.edu

Charles P. Friedman, PhD
Josiah Macy Jr. Professor of Medical Education
Chair, Department of Learning Health Sciences, Medical School
Professor of Information and Public Health
University of Michigan
cpfried@umich.edu

Leveraging increasingly available and rich phenotypic data sources has long been recognized as essential to the success of Learning Health Systems. Indeed, learning from the data that are generated at the point-of-care--as well as through bio-molecular diagnostics or the measurement of socio-demographic, behavioral, and environmental factors--is key to the afferent or "discovery" arm of a virtuous learning cycle.[1,2] The field of biomedical informatics, and particularly sub-fields such

as clinical and translational research informatics, has developed a significant focus on data science methods necessary to systematically and reproducibly "learn from data."[3,4]

While the promise of leveraging this wealth of available data sources is great, there are also many well-recognized challenges to doing so. The primary challenge stems from the fact that data collected through routine healthcare practice often has limited utility for secondary uses such as research and quality improvement activities. These limitations have manifested themselves in a number of recent scientific studies that have had to be retracted or adjusted due to incorrect assumptions and conclusions.[5,6] One thing is clear: The complexities of leveraging and learning from data require careful attention to several key principles in order to avoid mistaken conclusions and potential harm. The importance of these principles is the motivation for the "Learning from Data" feature of *Learning Health Systems*, which launches in this issue.

As described by Bastarache and colleagues in this first article in the Learning from Data series, real-world data such as that derived from EHRs contains vast amounts of information that can be used to generate knowledge and evidence through practice. [7] Such patient-derived data include physical measurements, diagnoses, interventions, exposures and outcomes. However, as the authors describe, each category of data requires careful consideration with regard to data quality, level of detail, provenance, as well as organizational and contextual factors that can cause unwanted variability in data across and between sites. Understanding and addressing such issues requires advances in

As we look ahead toward building learning health systems, we must seek to overcome the myriad challenges described in order to realize the promise of leveraging real-world data through advanced informatics and data science approaches. For instance, while the potential for AI-derived solutions is increasingly evident, there are emerging examples of largely unintentional negative effects of

algorithms developed and deployed based on incorrect assumptions or lack of awareness regarding underlying biases in the data used to develop them.[8] It is for this reason that there are growing calls for approaches such as "algorithmovigilance", a process for ongoing monitoring of algorithms and their applications.[9]

Therefore, given the promise of systematically generating evidence and enabling learning health systems by leveraging the vast and growing data assets across the health domain, there is a need for a focus on developing and disseminating the best practices needed to make beneficial use of these data. Starting with this issue, the "Learning from Data" feature will become a regular addition to the Journal. Dr. Embi will serve as the feature editor. The articles published in this feature will create a collection of manuscripts focused on the science and practice of applying informatics and data-science methods to enable the creation, operations, and sustainability of Learning Health systems.

Conflicts of Interest

Dr. Friedman is Chair of the Department of Learning Health Sciences at the University of Michigan Medical School. The authors assert they have no conflicts of interest.

References

1.	Friedman CP, Allee NJ, Delaney BC, et al. The science of Learning Health Systems: Foundations for a new journal. *Learn Health Syst.* 2017;1(1):e10020.
2.	Friedman C, Rubin J, Brown J, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc.* 2015;22(1):43-50.
3.	Embi PJ, Payne PR. Advancing methodologies in Clinical Research Informatics (CRI): foundational work for a maturing field. *J Biomed Inform.* 2014;52:1-3.
4.	Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. *J Am Med Inform Assoc.* 2009;16(3):316-327.

5. Madhavan S, Bastarache L, Brown JS, et al. Use of electronic health records to support a public health response to the COVID-19 pandemic in the United States: a perspective from 15 academic medical centers. *J Am Med Inform Assoc.* 2021;28(2):393-401.

6. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care.* 2013;51(8 Suppl 3):S30-37.

7. Bastarache L, Brown JS, Cimino JJ, et al. Developing real-world evidence from real-world data: Transforming raw data into analytical datasets. *Learning Health Systems.*n/a(n/a):e10293.

8. Maddox TM, Rumsfeld JS, Payne PRO. Questions for Artificial Intelligence in Health Care. *JAMA.* 2019;321(1):31-32.

9. Embi PJ. Algorithmovigilance-Advancing Methods to Analyze and Monitor Artificial Intelligence-Driven Health Care for Effectiveness and Equity. *JAMA Netw Open.* 2021;4(4):e214622.