Equity and Bias in Electronic Health Records Data

Andrew D. Boyd, MD,
Department of Biomedical and Health Information Sciences, University of Illinois Chicago, Chicago, IL

Rosa Gonzalez-Guarda, PhD,
Duke University School of Nursing, Durham NC

Katharine Lawrence, MD,
Department of Population Health, New York University Grossman School of Medicine, New York NY

Crystal L. Patil, PhD,
University of Illinois Chicago, College of Nursing, Chicago, IL

Miriam O. Ezenwa, PhD,
University of Florida College of Nursing, Gainesville, Florida

Emily C. O'Brien, PhD,
Department of Population Health Sciences, Duke University School of Medicine, Durham, NC

Hyung Paek, MD,
Yale University, New Haven, CT

Jordan M. Braciszewski, PhD,
Henry Ford Health, Detroit, MI

Oluwaseun Adeyemi, PhD,
New York University Grossman School of Medicine, Ronald O. Perelman Department of Emergency Medicine, New York, NY

Allison M Cuthel, MPH,
New York University Grossman School of Medicine, Ronald O. Perelman Department of Emergency Medicine, New York, NY

Juanita E. Darby, DC,
University of Illinois Chicago College of Nursing, Chicago, IL

Christina K. Zigler, PhD,
Duke University School of Medicine, Durham, NC

P. Michael Ho, MD,
Division of Cardiology, University of Colorado School of Medicine, Aurora, CO

Keturah R. Faurot, MPH,
Department of Physical Medicine and Rehabilitation, University of North Carolina School of Medicine, Chapel Hill, NC

Karen Staman, MS,
Duke University School of Medicine, Durham NC

Jonathan W. Leigh, MPH,
University of Illinois Chicago, College of Nursing, Chicago, IL

Dana L. Dailey, PhD,
St. Ambrose University, Davenport, IA and University of Iowa, Iowa City, IA

Andrea Cheville, MD,
Mayo Clinic Comprehensive Cancer Center, Rochester, MN

Guilherme Del Fiol, MD,
Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT

Mitchell R. Knisely, PhD,
Duke University School of Nursing

Keith Marsolo, PhD,
Department of Population Health Sciences, Duke University School of Medicine, Durham, NC

Rachel L. Richesson, PhD,
Department of Learning Health Sciences, University of Michigan Medical School

Judith M. Schlaeger, PhD
University of Illinois Chicago, College of Nursing, Chicago, IL

Abstract

Embedded pragmatic clinical trials (ePCTs) are conducted during routine clinical care and have the potential to increase knowledge about the effectiveness of interventions under real world conditions. However, many pragmatic trials rely on data from the electronic health record (EHR) data, which are subject to bias from incomplete data, poor data quality, lack of representation from people who are medically underserved, and implicit bias in EHR design. This commentary examines how the use of EHR data might exacerbate bias and potentially increase health inequities. We offer recommendations for how to increase generalizability of ePCT results and begin to mitigate bias to promote health equity.

Keywords

Health equity; patient-reported outcomes; social determinants of health; community engagement; health literacy

Introduction

By using data collected during clinical care, embedded pragmatic clinical trials (ePCTs) increase knowledge about the effectiveness of clinical interventions under real world

conditions. However, the electronic health record (EHR) data upon which many ePCTs are designed are subject to implicit bias in EHR design; bias from incomplete data and poor data quality; and overrepresentation of data from people with structural privilege.[1] These biases can limit the relevance and generalizability of results, and subsequently increase health inequities.

This commentary draws our collective experience to examine how the use of EHR data might exacerbate bias and potentially increase health inequities. We offer recommendations for how to increase generalizability of ePCT results to begin to mitigate bias and promote health equity.

Strategies to address bias in health research using the EHR

Research leveraging EHRs must be deliberately designed to identify and address bias to promote health equity. The Health Equity Lens framework, initially developed for public health professionals, outlines five health equity concepts for framing health disparities.[2] We use these concepts to explore sources of bias and provide recommendations.

1. Systemic, Social and Health Inequity Bias

Problem: The use of EHR-derived data requires careful attention to mitigate the unintended consequences of using data that mirrors US social and structural inequities. Moreover, insufficient attention has been paid to collecting data about the social determinants of health (SDoH).[3] Critically and more difficult to resolve, the available data from EHRs only reflects those who *access* healthcare. Those who are not represented in EHR datasets are a direct consequence of historical and ongoing forms of oppression causing ubiquitous health inequities that limit who can access care. Further, EHR data completeness and accuracy may reflect additional biases resulting from institutional policy, training practices, and implicit provider bias.[4] When patient-reported outcomes (PROs) are collected using patient-facing EHR modalities alone (i.e., patient portals), a portion of the population that does not use portals will also be excluded for various reasons (e.g., literacy and/or technology barriers).

Recommendation: Data sources such as PROs and Z-codes (included in the International Classification of Diseases-10) can be used to collect the demographic and SDoH variables needed to understand outcomes and, ultimately, improve clinical practice. While there is no consensus about best practices for equity-based data collection and which SDoH measures should be minimally included, we suggest that ePCT teams should strive to collect and report standardized SDoH measures. The HL7 Gravity[5,6] is one initiative aiming to identify and harmonize SDoH data so these are interoperable for electronic health information exchange. The increased national and global attention to health equity is driving not only standards but also incentives and tools to support SDOH data collection. To reduce bias in patient reported data often collected through patient portals to EHR systems, health systems and researchers will need to invest in the design of portals and engagement features, such as text messaging, and conduct specific research efforts to better understand the clinical effectiveness of these optimized EHR features in improving patients' effective use of EHRs and engagement in their health and health care.

2. Representation and Diversity

Problem: Much of current medical evidence was generated from clinical trials with predominately white participants which does not ensure conclusions drawn are safe and effective for all populations.[7] From these trials and knowledge, algorithms are built into EHR clinical decision support (CDS) tools to suggest risk factors, diagnoses, treatments and supportive services, with potentially the same omission of areas of study. Since the range of patient populations are not proportionally represented, the underlying logic of these algorithms and CDS tools limit applicability.[8]

Recommendation: Given the identified limitations of EHR data sets, greater transparency is needed regarding sources, input, and missing data and modelling choices underlying clinical decision support tools.[9] When planning ePCTs, sponsors and investigators should actively seek out and engage with a variety of settings serving diverse populations; efforts that support participatory research design should be prioritized. To address data collection barriers among people who have been historically marginalized and underrepresented, some investigators have enabled interventions using bidirectional text messaging that collect PRO measures and facilitate engagement with underrepresented populations who have high rates of cell phone ownership. To reduce bias that may arise from translated PRO or patient-facing measures that are used without cross-cultural validation, we recommend investing in the testing and psychometric validation of instruments used among different populations prior to use.[10,11]

3. Community Engagement

Problem: Community engaged approaches to EHR research are underutilized.

Recommendation: More than 25 years of evidence supports following the principles of community-based participatory health research.[12,13][14] In ePCTs, patients and communities ultimately affected by the health condition of concern should inform the research questions, variables and instrument selection, implementation, and interpretation of clinical research to ensure the research is relevant. Human-centered design[15,16] is one strategy that incorporates diverse stakeholders in the design and development of health technology interventions. Increasingly, these approaches focus on understanding and engaging with patients,[17] and incorporate equity-centered or emancipatory lenses that place equity more centrally in the process.[18–20]

4. Intersectionality

Problem: EHR-based research rarely captures variables that allow for intersectionality analyses.

Recommendation: Intersectionality[21] conceptualizes how political and economic power and oppression are linked and create systems of discrimination or disadvantage that are experienced by individuals based on identities (e.g., race, gender, sexual orientation, disability, immigration status, housing, education and income). Intersectionality is a lens that can be used to understand the differential effects of interventions tested through ePCTs. However, more refined data collection is needed to capture the identities and SDoH variables

that are not typically documented in the EHR. For example, offering identities write-in options can allow the social categories that are important to patient's experiences to emerge. Such nuances will enable analyses of individual and combined (additive or multiplicative) associations. This level of clarity and granularity can help prevent inappropriate data aggregation and increase transparency regarding decisions on how variables are produced and used in analyses. Although studies may not be powered to control for every variable, allowing for more refined social categories will help ensure more people will benefit from the interventions being tested which promotes health equity.

5. Literacy and Health Literacy

Problem: Health information collected in PROs is often written above the NIH-recommended 5th-grade level, or are developed without the input of patient end-users. Misunderstanding of the PROs due to reading grade level or lack of community knowledge could lead to incorrected data or unvalidated data collected by patients impacting clinical decisions.

Recommendation: The reading level of PROs should be formally evaluated, with potential cognitive testing to ensure suitability for the population of interest. As mentioned above, community partners should be involved in the review and validation of PRO content. As literacy and health literacy has a material impact on how patients interpret and respond to PRO tools, efforts should be made to appropriately capture SDoH of respondents; this includes the "digital" domains of literacy (e.g., digital health literacy, digital competence, digital agency) that may influence PRO data collection and interpretation.

Conclusion

EHR-based data collection within PCTs is increasing, leaving research vulnerable to biases in the design, collection, and use of electronic health data, and potentially propagating inequities in health and the healthcare system. Complex multilevel (national, state, and local) strategies and support from stakeholders are needed to address bias stemming from the use of EHR data for research and healthcare delivery. The embedded, ubiquitous, and often unknown biases in EHR data (due to variations in care delivery, experience, data capture or data quality, and lack of diverse representation) can limit the relevance and generalizability of results from pragmatic trials, and subsequently increase health inequities.

Disclosures

References

1. Siber-Sanderowitz S, Glasgow A, Chouake T, Beckford E, Nim A, Ozdoba A. Developing a Structural Intervention for Outpatient Mental Health Care: Mapping Vulnerability and Privilege. APT [Internet] 2022 [cited 2023 Apr 19];75(3):134–40. Available from: http://psychotherapy.psychiatryonline.org/doi/10.1176/appi.psychotherapy.20200057

2. CDC. CDC's Health Equity Guiding Principles for Inclusive Communication [Internet]. Centers for Disease Control and Prevention. 2021 [cited 2022 Jun 30];Available from: https://www.cdc.gov/healthcommunication/Health_Equity_Lens.html

3. Lyles CR, Wachter RM, Sarkar U. Focusing on Digital Health Equity. JAMA [Internet] 2021 [cited 2022 Jun 30];326(18):1795. Available from: https://jamanetwork.com/journals/jama/fullarticle/2785583 [PubMed: 34677577]

4. Adam H, Yang MY, Cato K, et al. Write It Like You See It: Detectable Differences in Clinical Notes By Race Lead To Differential Model Recommendations [Internet]. 2022 [cited 2022 May 17];Available from: http://arxiv.org/abs/2205.03931

5. Health Level Seven International. HL7 Data Type Definition Standards. Available at: http://www.hl7.org/implement/standards/product_section.cfm?section=2&ref=nav.

6. HL7 International. HL7 FHIR Realese 4 [Internet]. Available from: https://www.hl7.org/fhir/

7. Knepper TC, McLeod HL. When will clinical trials finally reflect diversity? Nature [Internet] 2018 [cited 2022 Aug 23];557(7704):157–9. Available from: http://www.nature.com/articles/d41586-018-05049-5 [PubMed: 29743700]

8. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med [Internet] 2020 [cited 2022 May 27];383(9):874–82. Available from: http://www.nejm.org/doi/10.1056/NEJMms2004740 [PubMed: 32853499]

9. Nong P, Williamson A, Anthony D, Platt J, Kardia S. Discrimination, trust, and withholding information from providers: Implications for missing data and inequity. SSM Popul Health 2022;18:101092. [PubMed: 35479582]

10. Linguistic Validation Manual for Patient Reported Outcomes(PRO) Instruments – ScienceOpen [Internet]. [cited 2023 Apr 19];Available from: https://www.scienceopen.com/document?vid=3af0e787-95c1-4b5c-8ac4-350e672054cc

11. Fowler FJ. Improving survey questions: design and evaluation. Thousand Oaks: Sage Publications; 1995.

12. Bentley AR, Callier S, Rotimi CN. Diversity and inclusion in genomic research: why the uneven progress? J Community Genet [Internet] 2017 [cited 2022 May 17];8(4):255–66. Available from: http://link.springer.com/10.1007/s12687-017-0316-6 [PubMed: 28770442]

13. Israel BA, editor. Methods in community-based participatory research for health. 1st ed. San Francisco, CA: Jossey-Bass; 2005.

14. Wallerstein N, editor. Community-based participatory research for health: advancing social and health equity. Third edition. Hoboken, NJ: Jossey-Bass & Pfeiffer Imprints, Wiley; 2017.

15. Fischer M, Safaeinili N, Haverfield MC, Brown-Johnson CG, Zionts D, Zulman DM. Approach to Human-Centered, Evidence-Driven Adaptive Design (AHEAD) for Health Care Interventions: a Proposed Framework. J Gen Intern Med 2021;36(4):1041–8. [PubMed: 33537952]

16. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A Framework and Toolkit of Strategies for the Development of More Effective Digital Interventions to Change Health Behavior. J Med Internet Res 2016;18(12):e317. [PubMed: 27986647]

17. Flood M, Ennis M, Ludlow A, et al. Research methods from human-centered design: Potential applications in pharmacy and health services research. Research in Social and Administrative Pharmacy [Internet] 2021 [cited 2023 Apr 5];17(12):2036–43. Available from: https://linkinghub.elsevier.com/retrieve/pii/S155174112100231X

18. Coulter RWS, Siconolfi DE, Egan JE, Chugani CD. Advancing LGBTQ Health Equity via Human-Centered Design. Psychiatr Serv 2020;71(2):109. [PubMed: 32008475]

19. Stiles-Shields C, Cummings C, Montague E, Plevinsky JM, Psihogios AM, Williams KDA. A Call to Action: Using and Extending Human-Centered Design Methodologies to Improve Mental and Behavioral Health Equity. Front Digit Health 2022;4:848052. [PubMed: 35547091]

20. Holeman I, Kane D. Human-centered design for global health equity. Inf Technol Dev 2019;26(3):477–505. [PubMed: 32982007]

21. Crenshaw K. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics.