

PERSPECTIVE OPEN



Harnessing big data for health equity through a comprehensive public database and data collection framework

Cameron Sabet¹, Alessandro Hammond^{1,2,3}, Nim Ravid², Michelle Sun Tong⁴ and Fatima Cody Stanford^{1,5}✉

The United States Department of Health and Human Services (HHS) pledged \$90 million to help reduce health disparities with data-driven solutions. The funds are being distributed to 1400 community health centers, serving over 30 million Americans. Given these developments, our piece examines the reasons behind the delayed adoption of big data for healthcare equity, recent efforts embracing big data tools, and methods to maximize potential without overburdening physicians. We additionally propose a public database for anonymized patient data, introducing diverse metrics and equitable data collection strategies, providing valuable insights for policymakers and health systems to better serve communities.

npj Digital Medicine (2023)6:91; <https://doi.org/10.1038/s41746-023-00844-5>

MAIN

In April 2022, Health and Human Services (HHS) pledged 90 million USD to the American Rescue Plan Uniform Data System Patient-Level Submission (ARP-UDS+) funding award in support of new data-driven efforts for Health Resources and Services Administration (HRSA) Health Center Programs to identify and reduce health disparities¹. This sum is planned to be distributed to 1400 community health centers (CHCs) serving over 30 million Americans¹. Specifically, this pledge will allow health centers to expand analytics and reporting capabilities to enhance healthcare services while supporting patient-level UDS+ data submissions to collect more precise data on health disparities². We herein propose a national public database featuring voluntarily self-disclosed and deidentified patient data to increase the granularity of health disparities metrics, and propose how this project can incorporate diverse metrics and equitable data collection procedures.

Big data has already been proposed to promote health equity at the federal level. For instance, in 2021, the Agency for Healthcare Research and Quality (AHRQ) released the National Healthcare Quality and Disparities Report (NHQR)³. The NHQR is a comprehensive review of American healthcare comparing health outcomes across both state and national levels⁴. This report revealed that, although communities of color have enjoyed increased healthcare access since 2000, racial inequity persisted in 2021 due to lack of focused interventions addressing disparities³. While CHCs disproportionately serve marginalized populations, their lack of granular data collection is likely leading to underestimations of health disparities in areas requiring the greatest resource support⁴.

Big data has impacted patient care for decades by helping health insurance companies incentivize preventative care among patients and physicians, ultimately decreasing the use of costly acute care and improving care equity⁴. For instance, the Community Health Needs Assessment (CHNA) in the Patient Protection and Affordable Care Act (ACA) requires CHCs to investigate the biggest health-related challenges in their communities and outline solutions⁵. However, CHNAs often lack subjective data like medical trust and are only collected periodically, producing an unnecessary lag time between when

systemic health issues arise and when healthcare providers can identify and respond to them.

Given these challenges with current healthcare metrics, the Centers for Medicare and Medicaid Services (CMS) should develop a national database like the CDC's Vaccine Adverse Event Reporting System (VAERS), which collects self-reported issues with vaccine reactions, to regularly collect self-reported data on symptoms, diseases, adverse reactions, medical trust, and perceptions of health equity through metrics like transportation time or food insecurity⁴. This platform can also be used to monitor medications while empowering CHCs to pinpoint and respond to systemic issues more quickly. Although the database may be skewed towards higher-income individuals with more time to self-report, the resource can help compare assess healthcare quality without overburdening physicians, ultimately improving equity interventions.

Although the data would be self-reported, the requested data should include many new metrics. For instance, the National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS), the basis for many equity measurement systems, HEDIS was inaugurated in 1991 and continued for 31 years without race or ethnicity included in its nearly one hundred performance measures⁶. These metrics were only introduced to eight of these measurements, and that only in 2022⁶. HEDIS also does not stratify metrics by other determinants of health like food insecurity or employment status⁶. Metrics for further development could also include language access, transportation barriers, and health literacy.

Healthcare organizations can better understand language barriers by offering patient surveys after each visit. Depending on the extent of the issue, Community Health Centers (CHCs) can then adopt interpreter services and track its effectiveness through surveys. To address transportation barriers, patients can opt to disclose transportation service use and reasons for missed appointments. CHCs can then partner with transportation providers like Uber WAV (Wheelchair Accessible Vehicles) and Uber Health to improve patient accessibility and cover for deficiencies in available wheelchair-accessible vehicles⁷. CHCs can also assess the readability of patient education materials and track the effectiveness of health literacy interventions like

¹Georgetown University School of Medicine, Washington, DC, USA. ²Harvard University, Cambridge, MA, USA. ³Division of Hematology/Oncology and Department of Pediatric Oncology, Boston Children's Hospital, Boston, MA, USA. ⁴University of California San Francisco School of Medicine, San Francisco, CA, USA. ⁵Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. ✉email: fstanford@mgh.harvard.edu

counseling, visual aids, staff training on health literacy communication, and simplified language in educational pamphlets through validated health literacy screening tools like REALM-SF (Rapid Estimate of Adult Literacy in Medicine - Short Form)⁸. REALM-SF examines abilities to read aloud medical terminology and is available free for download on the National Center for Education Statistics (NCES) website⁹.

To ensure that the system captures data from marginalized populations, we propose CHCs invest in technology infrastructure and staff training to prepare for more comprehensive data collection by, for instance, leveraging telehealth services to provide care management while streamlining data collection through remote patient monitoring and automated post-checkup surveys. Then, outside partnerships with research organizations can help CHCs analyze this data and devise tailored solutions through partnerships with relevant nonprofits like the National Partnership for Women and Families, which has developed best practices for patient engagement in healthcare, and Leapfrog, which helps analyze data to promote healthcare safety and quality^{10,11}.

Furthermore, to promote interoperability with electronic health records, we suggest following the technical standards developed by organizations such as the Office of the National Coordinator for Health Information Technology (ONC) and the Health Level Seven International (HL7) organization. These standards ensure that data can be exchanged seamlessly between different healthcare systems while safeguarding self-reported data. We also recommend consulting with the ONC's Trusted Exchange Framework and Common Agreement (TEFCA) to help establish policies for securely exchanging health information across organizations.

However, marginalized groups must be involved in the design of this database. They could provide input on how to create a user-friendly interface with icons instead of words on buttons to accommodate those with low literacy or limited technological proficiency. The database should also collect data on the widest possible range of demographic factors; provide technical support through online tutorials or help desks staffed by technical support personnel; establish robust privacy policies like strict data access controls, encryption protocols, and other technical safeguards; and implement data quality controls, like automated data validation checks; and seamlessly deliver the data to researchers and policymakers through data dashboards.

As big data revolutionizes the world, we must not forget its potential impact on improving health outcomes and disparities. By investing funds into health centers and hospitals for patient-driven reporting and data collection on race and ethnicity and developing minimum standards for health equity metrics collection and reporting across states, public discourse and media then draw on subsequent patient data analyses to bring greater attention to regional and national healthcare inequities, leading lawmakers to prioritize health equity initiatives and healthcare systems to have a clearer picture of the communities they serve.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Received: 10 November 2022; Accepted: 12 May 2023;

Published online: 20 May 2023

REFERENCES

1. Editors, H. HHS Announces \$90 Million to Support New Data-Driven Approaches for Health Centers to Identify and Reduce Health Disparities. *News* 2023.

<https://www.hhs.gov/about/news/2022/04/21/hhs-announces-90-million-support-new-data-driven-approaches-health-centers-identify-reduce-health-disparities.html> (2023).

2. Editors, H. FY 2022 American Rescue Plan Act Uniform Data System Supplemental Funding for Health Centers (ARP-UDS+): Award Recipient Presentation. in *Health Resources & Services Administration (HRSA) Health Center Program* (ed. H.R.S.A.H.-B.o.P.H.C. (BPHC)) (2022).
3. Masnik, T. *NCQA Updates & Releases New Quality Measures for HEDIS® 2023 with a Focus on Health Equity*. *News* 2022 (2023).
4. Dang, D. A. & Mendon, D. S. The value of big data in clinical decision making. *Int. J. Comput. Sci. Inf. Technol.* **6**, 3830–3835 (2015).
5. Artiga, S. & Hinton, E. Beyond health care: the role of social determinants in promoting health and health equity. *Health* **20**, 1–13 (2019).
6. Proposed Changes to Existing Measures for HEDIS®1 MY 2022: Introduction of Race and Ethnicity Stratification Into Select HEDIS Measures. in *Draft Document for HEDIS Public Comment—Obsolete After March 11, 2021* (ed. Editors, H.) (National Committee for Quality Assurance, 2021).
7. Health, U. <https://www.uberhealth.com/> (2023).
8. Arozullah, A.M. et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. *Med. Care* **45**, 1026–1033 (2007).
9. Editors, A. Personal Health Literacy Measurement Tools. <https://www.ahrq.gov/health-literacy/research/tools/index.html> (2023).
10. Bartosz, K. National partnership for women & families. *Coll. Res. Libraries N.* **64**, 342–342 (2003).
11. Tai, T. W. C. et al. An examination of Leapfrog safety measures and Magnet designation. *J. Healthc. Risk Manag.* (2023).

ACKNOWLEDGEMENTS

This work was funded by NIH P30-DK040561 (FCS) and U24 DK132733.

AUTHOR CONTRIBUTIONS

C.S. and A.H. were responsible for writing the piece and editing it, N.R. and M.S.T. were responsible for editing and providing input for ideas, and F.C.S. was responsible for editing, providing guidance on ideas, and overseeing the project.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-023-00844-5>.

Correspondence and requests for materials should be addressed to Fatima Cody Stanford.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023