



HHS Public Access

Author manuscript

Per Med. Author manuscript; available in PMC 2019 September 13.

Published in final edited form as:

Per Med. 2018 September ; 15(5): 403–412. doi:10.2217/pme-2018-0037.

Enhancing diversity to reduce health information disparities and build an evidence base for genomic medicine

Lucia A. Hindorff¹, Vence L. Bonham², and Lucila Ohno-Machado³

¹Division of Genomic Medicine and Office of the Director, National Human Genome Research Institute, National Institutes of Health

²Division of Intramural Research, Social and Behavioral Research Branch and Office of the Director, National Human Genome Research Institute, National Institutes of Health

³UCSD Health Department of Biomedical Informatics, University of California, San Diego

Abstract

Advances in genomic medicine are arising from efforts to build a national learning health system (LHS) and large-scale precision medicine studies. However, the underlying evidence base lacks sufficient data from populations historically underrepresented in biomedical research. Although the literature on health and healthcare disparities is extensive, disparities in the availability and quality of health information about diverse and underrepresented populations are less well characterized. This Perspective describes scientific and ethical benefits to incorporating health information from diverse and underrepresented populations in the LHS, resulting in a more robust and generalizable LHS. Near-term recommendations for incorporating diversity into the evidence base for genomic medicine are proposed, even as the groundwork for national and international efforts is underway.

Keywords

diversity; learning healthcare system; evidence base race; ethnicity; LHS; knowledge base; health disparities; healthcare disparities; health information disparities

Introduction

The learning healthcare system (LHS) holds great promise for developing, studying, and refining scientific hypotheses that can lead to genomic medicine implementation. The LHS is defined as a healthcare system where data are continually collected and analyzed in an iterative way to drive health system improvement[1]. As this iterative process generates new knowledge, this knowledge can be used to improve clinical practice at the point of patient care. Thus, the LHS comprises the creation and application of knowledge for clinical decision-making within the system, and the evaluation of their outcomes. Knowledge used for patient care originates from a variety of sources, e.g., reports of scientific research that are available globally, public health data for the covered population, quality improvement (QI) efforts designed and implemented locally to improve an organization's efficiency and care outcomes, and specific clinical and social information about the patient. Collectively, this knowledge forms an evidence base from which clinicians and patients make decisions.

The LHS is often described as a collective enterprise proposed at national or international scale. In practice, at least in the United States, it comprises many individual healthcare systems working in parallel to achieve the goals of continuous improvement and innovation[2]. These systems can combine data to more rapidly detect clinically relevant signals related to patient outcomes, as well as share their experiences in implementing practical interventions. The cyclical nature of the learning healthcare system is well supported by the field of implementation science – the scientific study of methods to accelerate the uptake of research findings and other evidence-based practices to improve routine practice[3]. Implementation science can thus be regarded as an important scientific engine that drives the LHS, with the underlying evidence base serving as the fuel. Genomic medicine strives to understand the contribution of an individual’s genomic information, in the context of social and environmental factors, to his or her health. Rapidly evolving in scale and adoption, genomic medicine is ripe for using implementation science methods and the framework of the LHS[4]. Because the LHS is a product of the underlying evidence base, i.e., data from patients cared for in various health systems, it is crucial that the LHS and the underlying genomic medicine and implementation science endeavors represent the diversity of patients within the LHS. It is thus essential that different types of healthcare systems participate in the LHS to help generate generalizable knowledge now, even as the ultimate goal of having all institutions involved is not yet a reality.

What indicators are there that the LHS may not be appropriately reflecting diversity? Some observations indirectly point to the same general issues that affect biomedical and clinical research: large-scale biomedical research studies, including those embedded within healthcare systems, show that evidence accumulating from genome-wide association studies[5,6] and clinical trials[7] demonstrate an underrepresentation of racial and ethnic minorities and underserved individuals. Additionally, the disproportionate representation of European ancestry individuals in reference databases has been linked with reduced effectiveness of variant interpretation in individuals of non-European ancestry[8,9]. These deficiencies result in an evidence base that does not reflect populations in whom disease burden may be greatest[10]. It also results in the potential for misdiagnoses that can widen existing healthcare disparities[11]. By healthcare disparities, we refer to differences in the way that individuals access healthcare, provider biases, and the structural factors underlying healthcare systems that cause healthcare inequities[12]. Whenever clinical action is taken is based on information selectively generated within a limited group, there is a form of health information disparity. This is particularly relevant to genomic medicine, where the quality and quantity of information available for use in clinical decision-making differs by group and/or at a community level. The translational nature of genomic medicine results in the widening of health information disparities as resources that underrepresent key populations form the basis for implementation studies that will most directly improve the health and healthcare for those who are represented, and not necessarily for those who are not represented. This poses a fundamental scientific and ethical shortcoming. We suggest that, without specific attention to mitigating these information disparities by improving population diversity, efforts to implement a national-scale LHS, especially an LHS broadly representative of the U.S. population, will result in the same problems that have limited generalizability of genetic research studies. Conversely, building diversity into the LHS, and

into the design of implementation research and genomic medicine studies, will strengthen both the underlying data and the resulting conclusions. The goal of this paper is to describe opportunities to incorporate diversity in the LHS and to address health information disparities and improve the evidence base for genomic medicine, resulting in a more robust and generalizable LHS.

Challenges in incorporating diversity

For this Perspective, we refer to diversity according to genetic-based ancestry, sex, and demographic factors (e.g. race, ethnicity, gender, marital status, sexual orientation, geographic location, country of origin, language, education, and social economic status – including the degree to which individuals are underserved by the healthcare system) that differ among groups of individuals and have been shown to affect their health and healthcare. For example, genetics as well as environmental and cultural factors have been shown to have consequences for individuals' health and healthcare[13], and both genetic and environmental factors impact an individual's epigenome[14]. Capturing measures of diversity in the LHS is crucial in ensuring that all groups of individuals are represented in the LHS. Capturing race and ethnicity components of diversity is required under Meaningful Use Stage 2, the national standard for electronic health records that must be met to qualify for federal incentives (<https://www.healthit.gov/providers-professionals/achieve-meaningful-use/core-measures-2/record-demographics>). This is clearly an important step forward[15] although the limited scope of required race and ethnicity measures has been criticized for not allowing sufficient granularity to self-report and distinguish among racial or ethnic groups with different health outcomes[16]. The lack of required environmental and social measures is another limitation[17]. With an eye toward incentivizing the inclusion of specific social and behavioral domains in electronic health records, the National Academy of Sciences Board of Population Health and Public Health Practice convened a committee in 2013 to recommend essential data elements for a robust LHS[18,19].

Additional effort is needed to increase representation of underrepresented populations in the organized health care system, including ethnic and racial minorities, recent immigrant populations and individuals of low socioeconomic status. For large healthcare systems with a relatively small proportion of diversity, it is imperative that data from diverse individuals be included in routine system queries, QI projects, as well as in observational and interventional studies. There is also an untapped opportunity to build LHS capacity in smaller, sometimes under-resourced healthcare systems whose patients might be underrepresented in research. These efforts are often focused on community health centers or health centers providing free care or treating primarily young, low-income, uninsured and minority groups[20–22]. Both approaches – maximizing use of existing health system infrastructure, as well as building new capacity in institutions such as federally qualified health centers (FQHCs) and minority-serving healthcare institutions – are needed. Diversity has been increasing for federally-funded research programs (e.g., in the StrongHeart study[23], the Study of Latinos[24], the All of Us Research Program[25], Population Architecture using Genomics and Epidemiology[26], Clinical Sequencing Evidence-generating Research[27] and PCORnet[28]), increasing representation of Native American, Hispanic/Latino, African, and Asian ancestry populations, as well as individuals with low

socioeconomic status or in medically underserved regions. However, parallel efforts for the LHS need to be developed.

For genomic medicine, there are several considerations related to the integration of information from diverse individuals into the LHS. Some, while not trivial to implement, are relatively straightforward. For example, technical efforts to integrate genomic information into the EHR[29,30] will benefit all individuals; however, the underlying underrepresentation needs to be addressed head on. Many genetic tests are not covered by certain insurance plans or have a high co-pay. Under- or non-insured patients will consequently have reduced access to these tests, and their results will not be captured in the EHR. Similarly, the logistical obstacles to genotyping or sequencing individuals more widely, as has been proposed for pre-emptive genotyping prior to prescription drug use[31,32], are also technically straightforward but present cost barriers. The public's beliefs and attitudes towards genetic data collection and sharing are evolving, and the resulting impact on genomic medicine research is itself an active area of research. There are technical and ethical challenges to preserving privacy and confidentiality in large scale data resources[33] and variable willingness to participate in research studies has been described in underrepresented populations[34,35]. A pilot study on patient preferences for sharing electronic health records for research showed that willingness to share those data within the healthcare system is much higher than with other institutions[36]. One source of tension is the need to increase inclusion of underrepresented populations with the rapid implementation of genomic information within LHS. As promising directions for integrating genomics into clinical care are identified, implementation will not be delayed because of challenges related to data generation or integration for a broad range of individuals. However, this implementation should be fair: care must be taken to ensure that recruitment or implementation are not compromised. Other unresolved questions address the permission and benefit of the LHS to use information about genetically-inferred ancestry when available, and whether and how discrepancies between genetic ancestry and self-reported race or ethnicity should be reconciled.

Similar questions relate to sources of data beyond the electronic health records. An important contributor to the knowledge base upon which clinical decisions are made is the biomedical literature. The large, randomized controlled trial is often regarded as the gold standard for evidence generation, with pragmatic studies conducted as part of QI projects sometimes considered a viable real-world alternative[37]. However, this literature has a similar bias to GWAS in that study participants are overwhelmingly of European descent[7]. Furthermore, clinical trials are typically implemented at substantial cost-per-trial, resulting in a tradeoff between minimizing bias with respect to study design. This results in a bias in the process of knowledge production that disfavors diversity[38,39]. Increasing diversity can and should be a primary focus of clinical trial design; in parallel, however, additional forms of evidence, such as observational studies or case reports, should be included in the knowledge bases underlying genomic medicine[40]. Referred to recently as a paradigm-shift from evidence-based medicine to medicine-based evidence, this approach relies on establishing longitudinal profiles of individual patients based on a constellation of biological, clinical, social and environmental factors[41]. Although not yet widely adopted, this approach may be particularly relevant to scientific questions for which randomized trials

including large numbers of diverse participants are not practicable. A recent recommendation from the National Academy of Science's *Evidence Framework for Genetic Testing* was to establish a decision repository to record, in a standardized way, previous decisions about genetic testing, and to capture criteria related to population, disease, outcome, approach, or test[42]. These standardized criteria could facilitate the inclusion of multiple lines of evidence from underrepresented populations, from case reports to population-stratified analyses of clinical trials.

How will these opportunities contribute to a more robust evidence base for genomic medicine?

Including populations from diverse backgrounds in the LHS and improving the ability of the LHS to integrate data that examines the diversity of the population will make it easier to study problems that will improve healthcare of all individuals. Below, we describe how incorporating genomic and socioeconomic diversity into the LHS will make it more robust.

First, generation and testing of novel hypotheses relevant to underrepresented ancestral and social demographic populations, especially broad subgroups in whom evidence is lacking, will be enabled. For example, pharmacogenomic studies enabled by EHRs[43] could be applied to multiethnic populations, facilitating identification of individuals susceptible to adverse or therapeutic drug reactions. If data suggest that evidence-based recommendations are not followed in some individuals, studies could be designed to evaluate alternate strategies for providers or patients to facilitate follow through with the recommendations. Routine system queries developed under QI efforts could be designed to look for individuals with notable benefit or harm, or for outliers. Subsequent opportunities to intervene directly could be facilitated, such as customizing clinical decision support rules, possibly leading to more effective treatments. Although truly population-specific variants are expected to be uncommon, a variant associated with a rare disorder and observed with much higher prevalence in certain ancestry groups could be more easily identified and studied for outcomes in defined subpopulations within the health system. Some variants of current unknown significance may prove harmless with the inclusion of currently underrepresented ancestry populations. Additionally, the interplay of genes and environment will be better understood, as a larger range of factors will be considered.

Next, evidence generated within a diversity-focused LHS is likely to be more generalizable to more settings. Using employment, education, income, geographic location, with self-reported race and ethnicity as examples, the consistent collection of these data within an LHS makes it feasible to integrate data from that health system with those of other health systems to minimize biases. This consistency benefits the collective LHS and the treatment of the patient at hand. This is particularly important for a health system with a homogenous population, since experiences with dissimilar patients may be very limited. To the extent that the LHS features continual recruitment of individuals, properly documenting diversity assures that underrepresented populations can be identified and disparities addressed.

One of the most ambitious aspects of integrating diversity into the LHS is the opportunity to recognize and reduce disparities in healthcare and health information. Healthcare disparities are by no means unique to genomics; indeed, closing the gap in access to primary care is

more likely to influence overall population health rather than access to specialty care. Nonetheless, genomic medicine is gaining a foothold in areas such as cancer genomics, undiagnosed diseases, and prenatal and neonatal genomics, and it is crucial that, wherever there is benefit to such approaches, those benefits be distributed equally among all groups. Health information disparities also need to be addressed (Table). For genomic medicine resources, the underlying evidence base is distorted in favor of findings coming from studies heavily biased toward European ancestry participants[5,6,8,44]. This bias can have clinical consequences such as misdiagnoses of individuals[11] and lower detection rates[9]. Gaps in allele frequency information needed to interpret clinically relevant variants also have consequences for the utility of genomic information among groups[8]. Additional factors contribute to the challenges. For example, information on ancestry, social and environmental data are not well captured in the EHR[45]. Although data are limited, variable quality and completeness of information in the EHR are also of concern. As another example, efforts to improve the quality of care by creating personal health records linked to the EHR were reported less successful in non-European patients[46].

Improvements in these areas will guard against the potential for genomic information to be misinterpreted or misapplied. Considering the above factors, a more diverse and robust LHS will generate and analyze data that address health in all populations while also identifying and solving challenges that are disproportionately present in subgroups. It will have the nimbleness to respond to unanticipated opportunities or scientific hypotheses.

What can be done now?

The call to diversify the LHS and its underlying evidence base fits well with calls to build capacity within the LHS[1,47] as well as within large-scale personalized medicine efforts[48]. The very nature of the LHS benefits from an iterative process upon which layers of information are continually being integrated. The time is right for researchers to incorporate diversity, whether based on genetics, race or ethnicity, or demographic factors, within their research in a deliberate and coordinated way. These efforts can drive standards-setting and data sharing in ways that are not incentivized by industry or for-profit healthcare.

Much work in curation and integration of genomic medicine data is already being done through resources such as ClinGen (<https://www.clinicalgenome.org/>), ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), Exome Aggregation Consortium (ExAC, <http://exac.broadinstitute.org/>), genome Aggregation Database (gNOMAD, <http://gnomad.broadinstitute.org/>), and Global Alliance for Genomics and Health (GA4GH, <https://www.ga4gh.org/>). These resources have improved the integration of diverse populations in genomic studies by: serving as a repository for data from large numbers of research participants; efficiently integrating population-level data with data needed to interpret clinical variants; making this information readily available at the point of use by researchers, patients and providers; and convening researchers from around the globe to facilitate responsible data sharing. We suggest that additional efforts in the following key areas are needed.

Collect better data on diversity markers.—We recommend greater inclusion of data from individuals underrepresented in biomedical research in genomic medicine studies and clinical trials that generate evidence used for clinical decision-making, and special attention be given to the correct documentation of race, ethnicity, gender identity and social determinants of health for all patients. Additionally, recommendations for genetic testing should be extended to all patients who could benefit from it, and providers should familiarize themselves with proper authorization processes by different insurance plans. The first and most difficult step is to bring in more diverse participants into the healthcare systems that collectively comprise the LHS. This step is not unique to research and is more complex for healthcare institutions since they typically cover specific geographic areas and care for underserved populations is not compensated at the same level as that of private insurance. Additionally, inclusion of underrepresented populations requires addressing access to care, including cultural and linguistic barriers, related to seeking healthcare, and communicating with providers. Piecemeal access to healthcare through urgent care or public health clinics for low- income and uninsured populations contrasts with access for resourced and fully insured populations, creating a socioeconomic chasm. Until uninsured populations are part of the health system, data about their health will be elusive, as will an understanding of how healthcare disparities are worsening this chasm. Given the highly fragmented healthcare system in the US, even those who are fully insured may have their data scattered across many healthcare providers. EHR linkage (i.e., integrating records for a given patient across multiple institutions) is typically done in a probabilistic manner to account for spelling errors and other factors[49,50]. Additionally, a systematic approach to increasing the utilization and value of case reports and observational studies could help further develop the evidence base used for clinical decision-making, whether as part of an evidence-based medicine or a medicine-based evidence paradigm. Also, as genomic data are often generated and interpreted by clinical laboratories, these should consider whether to standardize practices to include ancestry in analyzing results.

Collect social and environmental data.—We also recommend that standards for capturing diversity be integrated into health systems. Data efforts related to diversity should align well with current data sharing norms, such as FAIR (findable, accessible, interoperable, reusable[51]) and evolving standards (GA4GH, <https://www.ga4gh.org>). The use of social and environmental data in electronic health records[52] and community level data[53] will enhance the evidence base for understanding health differences within and between populations and ability to identify high risk populations[54]. The current limited availability of geographic, social and behavioral data in genomic research adversely impacts the ability to address social and behavioral determinants of health and appropriately incorporate genomic medicine in diverse communities.

Make the data available and in a privacy-protecting manner.—Data generation and dissemination is no longer the exclusive provenance of large research studies. Whether large- or small-scale efforts, responsible dissemination of aggregate data from underrepresented populations via common resources such as gNOMAD or ClinGen should continue, to aid in variant interpretation. Funding agencies and journals could clarify expectations for diversity that are used in making funding and publication decisions,

respectively, as there are currently no explicit guidelines. However, it should be kept in mind that privacy threats differ and could be higher for certain groups. For all populations, appropriate levels of data protection should be implemented, commensurate with the need for patient privacy[55].

Acknowledge the heterogeneity in priorities and challenges faced by the LHS across different settings.—It would be remiss not to acknowledge that the LHS will be driven by the values in which this system is designed and implemented. Particularly in the United States, the LHS will comprise a patchwork assortment of for-profit, non-profit, and government enterprises, each with its own stakeholders to which it is accountable. The costs and benefits of implementation may be borne alternately by shareholders, healthcare institutions, insurance providers, or taxpayers. Although data are collected to care for patients, analyses on data from multiple institutions may be done under the auspices of biomedical research in one setting, quality improvement in another, profit-yielding innovation in a third, and evidence development in yet another, requiring different types of agreements. This heterogeneity poses a challenge to leadership and coordination of the overall effort. Comparing the U.S. to other countries with single payer national healthcare systems, the uncertainty arising when, for example, government budgets or stock prices fluctuate, leaves the U.S particularly susceptible to a lack of continuity. Agreement to a common set of standards, such as those under development by GA4GH[56], can promote consistency and progress, however incremental, in periods of uncertainty. For example, data standards and standards for protection of patient/participant privacy provide common bounds around how data can be collected and used. These bounds will also be influenced by the development of regulatory frameworks for exchanging and processing data, as is now implemented in the European Union General Data Protection Regulation (GDPR)[57]. Standards for measuring clinical utility or improving healthcare utilization will also promote a common purpose even though the motivations for collecting the data may vary. The activation of stakeholders around the development and implementation of these standards will be key to ensuring the continuity of effort around the LHS.

In summary, the LHS should be designed with opportunities to conduct iterative analyses that will beget good scientific research questions and quality improvement projects[1]. Implementation scientists and genomic medicine researchers should consider diversity of the participants in designing appropriately powered studies and seek to drive health care improvement and efficiency across all populations. Some have commented that the lines between clinical care and research, or between quality improvement projects and research, are blurring[58]. Others have posited that the opportunity for advances in personalized medicine lie not in the traditional evidence-based medicine paradigm, but that of medicine-based evidence[41]. While the governance and goals of each area is distinct, regardless of the approach, a diversity-enabled LHS benefits all areas. Indicators of a healthy and fully-functioning LHS should be identified that will permit evidence to be generated within the LHS while preserving autonomy and protecting the privacy of patients, providers, and institutions.

The goal of building a diverse, robust evidence base for genomic medicine acknowledges that the current state of knowledge is incomplete, and that new efforts are needed even as

existing ones continue to be strengthened. These changes are unlikely to occur overnight, given the iterative nature of the LHS and the multi-staged approach to implementing certified health information technology via meaningful use, which has not yet specifically addressed genomic information. However, this measured pace may serve a practical benefit: diversity in the LHS and in designing implementation science studies can be regularly evaluated and improved.

Future perspective

As the costs of genome sequencing continue to decrease over the next decade, genomic sequence information will continue to be generated through research studies, in clinical care, and through the direct to consumer route. Increased inclusion of underrepresented populations in genomic medicine research will be enhanced by consortia focused on studying underrepresented populations, such as the Clinical Sequencing Evidence-Generating Research (CSER; <https://cser-consortium.org/>) and the Implementing Genomics in Practice (IGNITE II; <https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-17-009.html>) programs. Efforts to recruit underrepresented individuals as part of large-scale efforts such as the Million Veteran Program[60] and All of Us (<https://allofus.nih.gov/>) will also improve inclusion of diverse populations. The All of Us program, in particular, has taken a participant-centric approach to recruitment, having established privacy and trust principles, funded communications and engagement awardees, and strategically launched recruitment with the support of multiple communities around the U.S. These efforts to improve diversity, coupled with the curation and data integration resources mentioned in this Perspective, will strengthen the underlying evidence base for genomic medicine. The LHS will need to continue to evolve in other, non-genomic ways to address the technical and ethical challenges of integrating longitudinal data on large numbers of individuals across disparate healthcare systems. Identifying use cases for which individuals will query and access the data are critical, as patients, providers, researchers and policymakers have different preferences and requirements. Despite these improvements, it is possible, and perhaps even likely, that health information disparities will continue to worsen. Until access to healthcare is available to all, the benefits of a diversity-enabled LHS will be incomplete, and the benefit of collecting, analyzing, and using data from diverse populations to improve health and healthcare will be unfilled. Although we have focused primarily on recommendations for researchers in this Perspective, a companion call to reduce health information disparities to improve the LHS is also needed for policymakers and healthcare systems stakeholders.

Financial disclosure/Acknowledgements

LO-M is funded by grants or contracts from the NIH (R01HG008802, OT2OD026552, R01GM118609, R01HL136835), VA (I01HX000982), and PCORI (CDRN-1306-04819). She serves on a scientific advisory panel for Pfizer. VLB is supported in part by the Intramural Research Program of the National Human Genome Research Institute. The authors would like to thank Teri Manolio, MD, PhD, for insightful feedback.

References

* of interest:

** of considerable interest:

1. Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America., (2012).
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*, 22(1), 43–50 (2015). [PubMed: 25342177]
3. Fisher ES, Shortell SM, Savitz LA. Implementation Science: A Potential Catalyst for Delivery System Reform. *JAMA*, 315(4), 339–340 (2016). [PubMed: 26813203]
4. Chambers DA, Feero WG, Khoury MJ. Convergence of Implementation Science, Precision Medicine, and the Learning Health Care System: A New Model for Biomedical Research. *JAMA*, 315(18), 1941–1942 (2016). [PubMed: 27163980] ** Describes a framework through which implementation science, precision medicine, and the LHS are inter-related.
5. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*, 538(7624), 161–164 (2016). [PubMed: 27734877]
6. Morales J, Welter D, Bowler EH et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol*, 19(1), 21 (2018). [PubMed: 29448949]
7. Roundtable on the Promotion of Health Equity and the Elimination of Health Disparities; National Academies of Sciences, Engineering, and Medicine. Strategies for Ensuring Diversity, Inclusion, and Meaningful Participation in Clinical Trials: Proceedings of a Workshop. (2016).
8. Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol*, 17(1), 157 (2016). [PubMed: 27418169]
9. Landry LG, Rehm HL. Association of Racial/Ethnic Categories With the Ability of Genetic Tests to Detect a Cause of Cardiomyopathy. *JAMA Cardiol*, (2018).
10. Landry LG, Ali N, Williams DR, Rehm HL, Bonham VL. Lack of Diversity In Genomic Databases is A Barrier to Translating Precision Medicine Research Into Practice. *Health Affairs*, 37(5), In press (2018).
11. Manrai AK, Funke BH, Rehm HL et al. Genetic Misdiagnoses and the Potential for Health Disparities. *N Engl J Med*, 375(7), 655–665 (2016). [PubMed: 27532831] * Describes an example where a lack of diversity in underlying evidence had clinical implications
12. Williams JS, Walker RJ, Egede LE. Achieving Equity in an Evolving Healthcare System: Opportunities and Challenges. *Am J Med Sci*, 351(1), 33–43 (2016). [PubMed: 26802756] ** A thorough review including the state of health disparities in several disease areas, challenges and opportunities in addressing them.
13. Adler NE, Glymour MM, Fielding J. Addressing Social Determinants of Health and Health Inequalities. *JAMA*, 316(16), 1641–1642 (2016). [PubMed: 27669456]
14. Zoghbi HY, Beaudet AL. Epigenetics and Human Disease. *Cold Spring Harb Perspect Biol*, 8(2), a019497 (2016). [PubMed: 26834142]
15. Zhang X, Perez-Stable EJ, Bourne PE et al. Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century. *Ethn Dis*, 27(2), 95–106 (2017). [PubMed: 28439179]
16. Douglas MD, Dawes DE, Holden KB, Mack D. Missed policy opportunities to advance health equity by recording demographic data in electronic health records. *Am J Public Health*, 105 Suppl 3, S380–388 (2015).
17. Blizinsky KD, Bonham VL. Leveraging the Learning Health Care Model to Improve Equity in the Age of Genomic Medicine. *Learn Health Syst*, 2(1) (2018).
18. Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records, Institute of Medicine. Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1. (2014).
19. Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records, Institute of Medicine. Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2. (2015).
20. Bernstein JA, Friedman C, Jacobson P, Rubin JC. Ensuring public health's future in a national-scale learning health system. *Am J Prev Med*, 48(4), 480–487 (2015). [PubMed: 25700654]

21. Association of American Medical Colleges Community Partnership for Patient Access to Supportive Services (ComPPASS). (Ed.^(Eds)
22. Nath JB, Costigan S, Hsia RY. Changes in Demographics of Patients Seen at Federally Qualified Health Centers, 2005–2014. *JAMA Intern Med*, 176(5), 712–714 (2016). [PubMed: 27064681]
23. Lee ET, Welty TK, Fabsitz R et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*, 132(6), 1141–1155 (1990). [PubMed: 2260546]
24. Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*, 20(8), 629–641 (2010). [PubMed: 20609343]
25. National Institutes of Health National Institutes of Health: All of Us Research Program. (Ed.^(Eds) (2018)
26. Wojick GG M; Nishimura KK; Tao R; Haessler J; Gignoux CR; Highland HM; Patel YM; Sorokin EP; Avery CL; Belbin GM; Bien SA; Cheng I; Hodonsky CJ; Huckins LM; Jeff J; Justice AE; Kocarnik JM; Lim U; Lin BM; Lu Y; Nelson SC; Park SL; Preuss MH; Richard MA; Schurmann C; Setiawan VW; Vahi K; Vishnu A; Verbanck M; Walker R; Young KL; Zubair N; Ambite JL; Boerwinkle E; Bottinger E; Bustamante CD; Caberto C; Conomos MP; Deelman E; Do R; Doheny K; Fernandez-Rhodes L; Fornage M; Heiss G; Hindorff LA; Jackson RD; James R; Laurie CC; Li Y; Lin D; Nadkarni G; Pooler LC; Reiner AP; Romm J; Sabati C; Sheng X; Stahl E; Stram DO; Thornton TA; Wassel CL; Wilkens LR; Yoneyama S; Buyske S; Haiman K; Kooperberg C; Le Marchand L; Loos R; J.F Matisse T.C North K.E Peters U Kenny E.E Carlson C.S. Genetic Diversity Turns a New PAGE in Our Understanding of Complex Traits. Preprint available at <https://www.biorxiv.org/content/early/2017/09/15/188094>, (2017).
27. National Institutes of Health. Clinical Sequencing Evidence-Generating Research (CSER2) - Clinical Sites with Enhanced Diversity (U01). (Ed.^(Eds) (2016)
28. Frank L, Basch E, Selby JV, Patient-Centered Outcomes Research I. The PCORI perspective on patient-centered outcomes research. *JAMA*, 312(15), 1513–1514 (2014). [PubMed: 25167382]
29. Chute CG, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. *Genet Med*, 15(10), 802–809 (2013). [PubMed: 24008998]
30. Shirts BH, Salama JS, Aronson SJ et al. CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. *J Am Med Inform Assoc*, 22(6), 1231–1242 (2015). [PubMed: 26142422]
31. Schildcrout JS, Denny JC, Bowton E et al. Optimizing drug outcomes through pharmacogenetics: a case for preemptive genotyping. *Clin Pharmacol Ther*, 92(2), 235–242 (2012). [PubMed: 22739144]
32. Dunnenberger HM, Crews KR, Hoffman JM et al. Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annu Rev Pharmacol Toxicol*, 55, 89–106 (2015). [PubMed: 25292429]
33. Wang S, Jiang X, Singh S et al. Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States. *Ann N Y Acad Sci*, 1387(1), 73–83 (2017). [PubMed: 27681358]
34. Sanderson SC, Brothers KB, Mercaldo ND et al. Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US. *Am J Hum Genet*, 100(3), 414–427 (2017). [PubMed: 28190457]
35. Nodora JN, Komenaka IK, Bouton ME et al. Biospecimen Sharing Among Hispanic Women in a Safety-Net Clinic: Implications for the Precision Medicine Initiative. *J Natl Cancer Inst*, 109(2) (2017).
36. Kim H, Bell E, Kim J et al. iCONCUR: informed consent for clinical data and bio-sample use for research. *J Am Med Inform Assoc*, 24(2), 380–387 (2017). [PubMed: 27589942]
37. Weinfurt KP, Hernandez AF, Coronado GD et al. Pragmatic clinical trials embedded in healthcare systems: generalizable lessons from the NIH Collaboratory. *BMC Med Res Methodol*, 17(1), 144 (2017). [PubMed: 28923013]

38. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. *N Engl J Med*, 374(22), 2175–2181 (2016). [PubMed: 27248626]
39. Stronks K, Wieringa NF, Hardon A. Confronting diversity in the production of clinical evidence goes beyond merely including under-represented groups in clinical trials. *Trials*, 14, 177 (2013). [PubMed: 23768231] * A helpful perspective that shows the consequences of lack of diversity on production of clinical evidence and suggests opportunities for improvement.
40. Petticrew M 'More research needed': plugging gaps in the evidence base on health inequalities. *Eur J Public Health*, 17(5), 411–413 (2007). [PubMed: 17962338]
41. Horwitz RI, Hayes-Conroy A, Caricchio R, Singer BH. From Evidence Based Medicine to Medicine Based Evidence. *Am J Med*, 130(11), 1246–1250 (2017). [PubMed: 28711551]
42. Committee on the Evidence Base for Genetic Testing, National Academies of Sciences, Engineering, and Medicine. *An Evidence Framework for Genetic Testing*. (2017).
43. Denny JC, Van Driest SL, Wei WQ, Roden DM. The Influence of Big (Clinical) Data and Genomics on Precision Medicine and Drug Development. *Clin Pharmacol Ther*, 103(3), 409–418 (2018). [PubMed: 29171014]
44. Chen MS, Lara PN, Dang JH, Paterniti DA, Kelly K. Twenty years post-NIH Revitalization Act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual: renewing the case for enhancing minority participation in cancer clinical trials. *Cancer*, 120 Suppl 7, 1091–1096 (2014).
45. Beck AF, Sandel MT, Ryan PH, Kahn RS. Mapping Neighborhood Health Geomarkers To Clinical Care Decisions To Promote Equity In Child Health. *Health Aff (Millwood)*, 36(6), 999–1005 (2017). [PubMed: 28583957]
46. Garrido T, Kanter M, Meng D et al. Race/ethnicity, personal health record access, and quality of care. *Am J Manag Care*, 21(2), e103–113 (2015). [PubMed: 25880485]
47. Lu CY, Williams MS, Ginsburg GS, Toh S, Brown JS, Khoury MJ. A proposed approach to accelerate evidence generation for genomic-based technologies in the context of a learning health system. *Genet Med*, 20(4), 390–396 (2018). [PubMed: 28796238]
48. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*, 372(9), 793–795 (2015). [PubMed: 25635347]
49. Chen F, Jiang X, Wang S et al. Perfectly Secure and Efficient Two-Party Electronic-Health-Record Linkage. *IEEE Internet Comput*, 22(2), 32–41 (2018). [PubMed: 29867290]
50. Kum HC, Krishnamurthy A, Machanavajjhala A, Reiter MK, Ahalt S. Privacy preserving interactive record linkage (PIRL). *J Am Med Inform Assoc*, 21(2), 212–220 (2014). [PubMed: 24201028]
51. Wilkinson MD, Dumontier M, Aalbersberg IJ et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3, 160018 (2016). [PubMed: 26978244]
52. Hollister BM, Restrepo NA, Farber-Eger E, Crawford DC, Aldrich MC, Non A. Development and Performance of Text-Mining Algorithms to Extract Socioeconomic Status from De-Identified Electronic Health Records. *Pac Symp Biocomput*, 22, 230–241 (2017). [PubMed: 27896978]
53. Bush WS, Crawford DC, Briggs F, Freedman D, Sloan C. Integrating Community-Level Data Resources for Precision Medicine Research. *Pac Symp Biocomput*, 23, 618–622 (2018). [PubMed: 29218920]
54. Oreskovic NM, Maniates J, Weilburg J, Choy G. Optimizing the Use of Electronic Health Records to Identify High-Risk Psychosocial Determinants of Health. *JMIR Med Inform*, 5(3), e25 (2017). [PubMed: 28807893]
55. Dyke SO, Dove ES, Knoppers BM. Sharing health-related data: a privacy test? *NPJ Genom Med*, 1(1), 160241–160246 (2016). [PubMed: 27990299]
56. Regulatory, Ethics Working Group, Global Alliance for Genomics Health, Sugano S. International code of conduct for genomic and health-related data sharing. *Hugo J*, 8(1), 1 (2014). [PubMed: 27090249]
57. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *Eur J Hum Genet*, 26(2), 149–156 (2018). [PubMed: 29187736]

58. Lee SS, Kelley M, Cho MK et al. Adrift in the Gray Zone: IRB Perspectives on Research in the Learning Health System. *AJOB Empir Bioeth*, 7(2), 125–134 (2016). [PubMed: 27917391]
59. Weber GM, Adams WG, Bernstam EV et al. Biases introduced by filtering electronic health records for patients with “complete data”. *J Am Med Inform Assoc*, 24(6), 1134–1141 (2017). [PubMed: 29016972]
60. Gaziano JM, Concato J, Brophy M et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*, 70, 214–223 (2016). [PubMed: 26441289]

Executive summary

Introduction

- The LHS is a model for implementing genomic medicine and relies on a robust evidence base from which to test scientific hypotheses and direct improvements to patient care.
- The current evidence base is lacking in diversity of characteristics such as race, ethnicity, geographic location, and socioeconomic status that differ among individuals and are known to impact their health and healthcare.
- There is a notable health information disparity by which the quality and quantity of information available for use in clinical decision-making differs by group. These observations are supported by current studies documenting the clinical impact of a lack of diversity.
- The goal of this paper is to describe opportunities to incorporate diversity in the LHS and to address health information disparities and improve the evidence base for genomic medicine, resulting in a more robust and generalizable LHS.

Challenges in incorporating diversity

- Although Meaningful Use requirements will help ensure that race and ethnicity are captured consistently in the EHR, there are barriers to inclusion of underrepresented individuals of diverse backgrounds in health systems. These include technical, social and scientific challenges.
- The evidence base should expand to take on forms of evidence that go beyond the gold standard of RCTs, which often underrepresent the populations who may stand to gain or lose the most from the outcome.

How will these opportunities contribute to a more robust evidence base for genomic medicine?

- Hypothesis testing and generation in a diverse LHS could identify groups who are likely to benefit from or be harmed by targeted interventions or opportunities to intervene.
- A diverse LHS is likely to yield findings that are generalizable to more settings.
- The long-term goal of integrating diversity is to recognize and reduce healthcare and health information disparities.

What can be done now?

- More and better data from a greater diversity of individuals are needed. Diverse participants need to be integrated into healthcare systems.
- Standards for collecting social and environmental data should be developed and adopted.

- Data should be made available in a privacy-protecting manner.
- Heterogeneity in priorities and challenges faced by the LHS across different settings, as well as in the United States compared to internationally, should be considered as standards for exchanging data and protecting patient privacy are implemented.
- The iterative nature of the LHS yields results that will beget additional scientific research questions.

Table.

Sources of health information disparities in genomic medicine.

Example	Clinical consequences	References
Individuals from non-EA populations are less likely to participate in genomic research studies	<ul style="list-style-type: none"> Approaches or algorithms requiring synthesis of existing knowledge for disease diagnosis, prediction or management (e.g., genetic risk scores) will perform better in EA compared to non-EA populations 	<ul style="list-style-type: none"> Popejoy[5]; Morales[6]
Individuals from non-EA populations are less likely to participate in clinical trials	<ul style="list-style-type: none"> Evidence base on which clinical impact (eg, safety, efficacy, utility) is assessed is skewed Results from trials have unknown real-world impact on individuals from populations who are not represented 	<ul style="list-style-type: none"> Strategies for Ensuring Diversity, Inclusion and Meaningful Participation in Clinical Trials[7]; Chen[44]
Allele frequency data used for variant interpretation are less likely to be publicly available for non-EA populations	<ul style="list-style-type: none"> Misdiagnosis of patients: false positive results Greater VUS (Variants of Unknown Significance) rates 	<ul style="list-style-type: none"> Manrai[11]; Petrovski[8]; Landry[9]
Race, ethnicity and other social identity population-level data are not well integrated into the electronic health record	<ul style="list-style-type: none"> Challenges in identifying health and healthcare disparities Inefficient distribution of healthcare resources and potential missed opportunities for redirecting them 	<ul style="list-style-type: none"> Beck[45], Weber[59]
Non-EA patients are less likely to register for personal health records, enabling better access to their health information and supporting health-related materials	<ul style="list-style-type: none"> Variable quality and quantity of information in the electronic health record (EHR) Clinical decision making based on variable completeness of information from EHR Reduced patient awareness and engagement 	<ul style="list-style-type: none"> Garrido[46]

EA = European Ancestry