



**BlueCross
BlueShield**
Association



THE ETHICAL AND TRANSPARENT USE OF DATA TO REDUCE HEALTH DISPARITIES

Introduction

Significant health inequities plague the U.S. health care system, mostly impacting historically marginalized racial (US Department of Health and Human Services Task Force on Black and Minority Health, 1985), ethnic, socioeconomic (Riley, 2012) and LGBTQ+ communities (Baptiste-Roberts et al., 2017). There are many causes for these inequities, including systemic racism that has enabled structural inequalities to become embedded in the health care system. For example, predominantly Black communities receive less public health investment than White communities (Bailey et al., 2021). When working to solve inequities, underlying issues of racism cannot be ignored. To improve health care across the nation, understanding what led to and what perpetuates such inequities is necessary for generating solutions.

To address health inequities that contribute to racial and other health disparities, getting the right data to understand where they exist is crucial. Without knowing where problems exist and the context in which those problems occur, effective solutions cannot be developed, and without knowing how proposed solutions impact health disparities, meaningful change cannot be measured. This paper will focus on the changes needed in data collection today across the health care ecosystem. Common industry standards for how data are collected and used are a critical step to address the significant disparities that exist.



Types of Inequities

Health inequities present as differential quality of care, differential access to care, and differential prevalence of disease.

Table 1. Main Types of Health Inequities

Categories of Health Inequities	Definition
Differential Quality of Care	Historically marginalized populations receive less quality care than the general population.
Differential Access to Care	Historically marginalized populations have less access to care than the general population.
Differential Prevalence of Disease	Higher prevalence of disease can be found amongst historically marginalized populations.

Differential Quality of Care: One troubling aspect of health inequities in the U.S. is that when a person sees a provider, a White individual may experience better quality of care than a Black individual. In 2002, a seminal study by the National Academy of Medicine found that Black individuals are less likely to receive kidney dialysis than White individuals when corrected for socioeconomic status (Nelson, 2002). Other studies have shown Black patients are less likely to receive common hospital procedures than White patients (McBean & Gornick, 1994).

Differential Access to Care: Historically marginalized communities have limited access to health care services. This can be attributed to various factors, which include lack of transportation to get to a provider, low density of providers in the community or poor access to high-speed internet for telehealth visits. Health plan coverage is another key factor, while work and family responsibilities can also make it difficult to see a provider when needed. Rural communities have a particular dearth of health care options (Weisgrau, 1995). Furthermore, LGBTQ+ individuals often face additional barriers to accessing care, including fear of whether they will be able to receive care and how they will be treated (Johnson et al., 2020).

Differential Prevalence of Diseases: Sickness and disease do not present equally across all people. For example:

- In the U.S., Black people have a higher prevalence of diabetes than White people. Research has shown that socioeconomic status may account, in part, for this difference (Bancks et al., 2017).
- Native Americans have the highest rate of diabetes in the U.S. (Centers for Disease Control and Prevention, 2017).
- Black males also have a higher rate of colon, stomach and prostate cancer compared to non-Hispanic White males (US Department of Health and Human Services Office of Minority Health, 2021).
- Black Americans have a higher mortality rate from cancer (American Cancer Society, 2022).
- A study looking at substance use disorder in Cuban Americans showed that individuals of Cuban descent who were born in the U.S. had higher rates of substance use disorder (Ibanez et al., 2017).

Data Needed to Address Racial and Other Health Disparities

To reduce racial, ethnic, sexual orientation and gender health disparities, each component of these disparities needs to be understood on its own terms. Root cause analysis of the problem followed by testing successful solutions in the real world is needed. To make that possible, a more robust and fine-tuned understanding of where disparities exist is required. While there is significant documentation of health disparities in the U.S., the data that are available are still lacking. Higher quality, disaggregated data that covers the population in more granularity and enables communities to monitor developments in disparities and the impact of interventions is needed.

Table 2. Room for Improvement in Health Equity Data

Components of Better Health Equity Data	Explanation
Quality Data	Data from gold-standard sources and collection methods
Granular Data	Data for communities that vary by region or culture (e.g., data for the U.S., for Illinois, for Chicago, for Hyde Park, for individuals)
Disaggregated Data	Detailed information about an ethnicity (e.g., Cuban, instead of Latino)
Data to Monitor Change	Continuous data collection instead of data collected at a point in time

Higher Quality Data:

Data are key to measuring, monitoring, and tracking changes in health disparities. There are several challenges that affect the quality of available data. The quality of available data is inadequate for the reasons described below:

- **Bias of smaller data sets and reliability:** Data on a sample of the population can be plagued with selection bias, impacting the reliability of data.
- **Contradicting data and validity:** In addition, there can be challenges with data validity. Contradictions can exist between different sources of data. For example, it may be difficult to resolve data about a person’s gender or race on insurance claims when they differ from enrollment data.
- **Self-reported data:** Self-identified data are considered by many to be the gold standard. However, when filling out a form in a doctor’s office, people may be hesitant to list their race or ethnicity for fear of receiving substandard care; the categories for race and ethnicity may also not be reflective of how an individual views themselves. There is evidence to suggest that Black and Hispanic Americans have a higher level of mistrust of the medical community than White Americans (Bazargan et al., 2021). Lack of trust can lead to challenges in getting complete data sets. Individuals from underrepresented populations may be concerned with how their data will be used and may not be confident that it will be adequately protected.

More Granular Data

More detailed information about regions and populations is needed to understand nuances that may not be present in national data sets. Certain documented health disparities may apply across the country. However,

there can be significant cultural and geographic differences that make it impossible to detect disparities that appear in different areas based on region. *Regional data* are needed to diagnose distinct, regional problems so proven solutions can be replicated and scaled effectively. In addition, more *individual level data* are needed. Outreach programs can be effective only if delivered to populations in need of such programs.

Disaggregated Data

Health care demographic data are often collected at an aggregate level that is less useful for health equity purposes. For example, the industry standards for race and ethnicity were defined in 1997 by the Office of Management and Budget (OMB) (Office of Management and Budget, 1997). These standards, widely used in health care, and required by law for certain health care data collection use cases, are useful for reporting, but do not enable the granular level insights required to be more effective at addressing health disparities. The OMB's definition of race is: White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander. The lack of specificity in racial and ethnic classifications does not enable the understanding of health care issues specific to subpopulations and impacts opportunities to address racial health disparities. The Office of Minority Health (OMH), the Affordable Care Act of 2010 (ACA) enrollment data capture and others have proposed more granular categories that map back to the OMB5 (US Department of Health and Human Services Office of Minority Health, 2021b). However, the OMH, ACA and other categories are not universally adopted.

Data to Monitor Change

Improved data collection also means collecting more data points over time. Health inequities are not static. It is crucial to collect data over time to monitor where new types of inequities have arisen and whether interventions have worked to reduce health disparities inequities. Getting better data is crucial to identify trends in health disparities (Bilheimer & Klein, 2010). In order to develop solutions tailored for a specific community, knowing who is impacted is key. More data are also needed to measure the quality and impact of interventions developed to target specific issues. If a community health center establishes a new program to improve the quality of care for people experiencing homelessness, it is critical to know if this program is successful or needs to be changed. This can only occur if data are collected at multiple points in time, before and after a change is implemented.

The Need for Unified Industry Standards Across the Health Care Ecosystem

Collecting and exchanging data with the right level of detail must be part of a multi-stakeholder effort. No single organization has enough data to tackle health inequities on its own. Health care leaders need to work together to address the challenges. Because such inequities are a universal challenge, it is imperative that the health care industry collaborates and coalesces around adoption of national data collection and exchange standards for health equity data sets. Without the use of consistent data standards, the ability to effectively address racial health and other disparities is compromised: inconsistent use of data standards creates problems for interoperability and makes it challenging to measure progress over time and identify emerging trends. Each group that uses its own way to exchange and represent information contributes to data that cannot be compared with every other group working to solve the same problem. As noted by the Office of the National Coordinator for Health IT (ONC), advancing the use of data standards and interoperability of health equity data are needed to improve the health and well-being of all individuals and communities. Beyond

adopting a common standard for health equity data, more work is needed across the ecosystem to solve this challenge (The Office of the National Coordinator for Health Information Technology, 2013).

Adoption of universal industry standards are needed in three key areas: race, ethnicity, and language (REL), sexual orientation and gender identity (SOGI), and social determinants of health (SDoH). If successful, health plans, hospitals, providers, community health centers, national and local alliances, academic partners, policy makers and other stakeholders will be able to collaborate across the country in the collection and analysis of meaningful health equity data sets. When thinking about such standards, the fluid and evolving nature of SOGI data should be recognized. Commonly adopted categories for gender identity 10 years ago are no longer sufficient. One can expect that in 10 more years, additional categories are likely to arise. Therefore, data standards need to be agile.

In addition to technical standards, standards are needed to define how best to ask for health equity data. For example, when asking about gender identity, the Centers for Disease Control and Prevention (CDC) recommends first asking about a person’s sex at birth and then asking about current gender identity. How a question is framed can yield different answers. In an effort to standardize how data are collected and exchanged, it is equally important to uniformly ask the questions that generate such data.

Three Key Health Equity Data Sets, Three Separate Industry Standards

Each of the health equity data sets below is at a different stage in the journey to standardization, and each has unique considerations.

Table 3. Three Types of Health Equity Data Sets

Health Equity Data Sets
<ol style="list-style-type: none"> 1. Race, Ethnicity, and Language (REL) 2. Sexual Orientation and Gender Identity (SOGI) 3. Social Determinants of Health (SDoH)

REL: The most widely adopted standard for REL today is the OMB standard. Several approaches have been proposed to develop at more granular categories that map to the OMB standard, including both the CDC (Centers for Disease Control and Prevention, 2000) and the OMH. Such efforts can be reviewed and used as a starting point. However, several issues need to be solved in a cohesive manner (US Department of Health and Human Services Office of Minority Health, 2018), highlighting the need for next generation standards. For example, are there additional disaggregated categories that should be included? What categories will be most useful in addressing racial health disparities?

Because the OMB categories (to which the partially disaggregated OMH data set maps) are commonly used in the health care ecosystem (and are required by the Department of Health and Human Services (HHS) (Evaluation, 2011), creating categories that cannot be mapped to the OMB standard may be burdensome. Doing so would make it difficult to analyze OMB-derived data together with non-OMB derived data.

SOGI: The health equity data set that is the least developed is sexual orientation and gender identity. Currently, there is little standardization in the field. HL7 has embarked on a project named “Gender Harmony” to develop SOGI standards (HL7 International, 2022). This effort can become the starting point for SOGI standardization. In addition to the actual categories and definitions to be adopted, how people are asked for this information should also be standardized.

SDoH: Social determinants of health have received significant attention in recent years in academic research, community programs and the media. The way SDoH has been analyzed has evolved from the traditional aspects of access to housing, transportation, and food to include additional areas such as domestic violence. HL7, through the Gravity Project, has established a consensus-building community to expand SDoH core data for interoperability and accelerate standards-based information exchange (HL7 International). While individual organizations may wish to seek approval for additional domains to those in Gravity, the current Gravity Project’s domains can serve as a foundation.

Additional Considerations for Successful Data Collection

Imputation

Efforts to improve data collection and standards-based exchange will take time to implement. In the meantime, considerations on how available data can be used to promote health equity in the absence of data that is more robust is critical. One can consider using imputed data, which is the derivation of unknown data from known data. Many methods exist for data imputation. For example, the Bayesian Improved Surname and Geocoding (BISG) algorithm developed by RAND Corporation uses a person’s ZIP code and last name to impute their race (Fremont et al., 2016). Data imputation may not always be as reliable as self-identified data; however, it can serve as a useful tool until self-reported data are more widely available.

When using imputed data, there are several considerations: 1) whether the data are available from another source; 2) whether the imputation method is sound; and 3) whether the imputation method is validated with a subset of known data. Race is an attribute that is often imputed. However, it may not always be appropriate to impute race. This does not mean that race should never be included in modeling, but rather that special attention is needed for when and how to include it (Vyas et al., 2020).

Two components are needed to effectively use race in health care analyses:

1. Evaluation of the appropriateness of race to the specific question at hand.
2. Standardization across the industry of acceptable and unacceptable uses of race in health care analyses.

Another important consideration for the use of imputed data is the use case for such data. The more sensitive the application, the higher the confidence threshold must be. When using imputed data to guide a decision around an individual, one must be confident in the data. Even with a high confidence threshold, one must assess whether imputation is appropriate for the specific use case. For example, one must consider whether the use of imputed data could create lack of trust with a patient or a community. If imputed data are to be included in a model, traceability of its source must be ensured, and care taken that the imputed data are not propagated throughout a data ecosystem in the same way as self-reported data. When using imputed data,

consideration must be made as to how its use will impact the community about whom the imputation has been made. Imputation is widely used today (Jarrin et al., 2020), (Centers for Disease Control and Prevention, 2015) however not consistently across health care organizations (Xue, 2019). As an industry, there is a need for consistent standards for (1) how imputation is done, and (2) when imputation is appropriate.

Ethical and Transparent Use of Health Equity Data

When using health equity data sets, it can be helpful to focus on two principles:

1. “Do no harm”
2. “Do better”

The goal of using health equity data sets is principle #2, “Do better.” This means reducing health disparities and promoting health equity. However, collecting and using health equity data brings with it certain risks that must be avoided. The goal of health equity cannot be pursued at the expense of principle #1, “Do no harm.”

Bias

When including health equity data sets in health care analytics, there is the potential to use that data as a source of bias or unintended discrimination. Bias literature includes a concept called “fairness through unawareness” (Pechenizikiy & Gajane, 2018), which suggests that to avoid discrimination based on attributes such as race or ethnicity, one should exclude race and ethnicity from the data set. This is common practice within Human Resource departments, where hiring managers may not ask a person about their race. There have even been efforts to ban the use of race and ethnicity in medical research (Ioannidis et al., 2021) in line with the concept of “fairness through unawareness.”

Fairness through unawareness is not an option when working with health equity data sets. Discrimination and health inequities exist, and if the community does not study and develop solutions to address them, they will remain.

It is therefore important to include race and ethnicity, and other health equity data, in health care data collection. However, the risk of bias and discrimination must be mitigated. To do so, validated bias identification and mitigation strategies should be adopted. Even if not actively using race or ethnicity data, there is the risk for a model to lead to inequitable outcomes (Obermeyer et al., 2019). This underscores the need for a solid bias identification and mitigation strategy. This should include, but is not limited to, adopting methods to find bias in both data and models and to ensure that models are implemented in an equitable manner.

When bias is found in data or models, careful consideration should be taken as to how to mitigate the source of bias. This does not necessarily mean throwing out the data or model. More nuanced approaches such as rebalancing can be employed (e.g., if women are underrepresented in the data, give more weight to female data in the model).

Transparency and Trust

Self-reported data can be challenging to obtain. This may be in part due to lack of trust in the entities that are collecting the data. One must consider how the health care ecosystem can establish more trust with

individuals it cares for. Creating meaningful transparency that engenders trust can be a crucial step forward. Transparency engenders trust, and trust can make people more comfortable in sharing REL, SOGI and other data. This transparency–trust cycle can create a bond between a patient and the health care community. Where this bond exists, people may become more comfortable sharing their data. Where this bond is broken, or not yet formed, people will be wary of sharing their data. The interplay of trust and transparency in health care is increasingly recognized as an important piece of the health care equation (Esmaeilzadeh, 2019).

Health care organizations must be transparent about the intended use of health equity data. For example, if a hospital asks questions about a patient’s socioeconomic status, such questions could be prefaced with an explanation of why they are being asked, what will be done with the information and how the information is safeguarded. Similarly, if a health plan asks its members about their gender identity, an explanation about why such information is requested, and how the information will be used to help guide the member’s care, is essential.

Access to Data

The source of health care data at its most fundamental level is the patient. Despite how far the data move from the original source, providing patients with access to their data is important, and often legally required depending on the entity that holds the data if the individual requests it. In addition to building trust as to how data are used, it is also important from a patient’s health perspective. Empowering patients with their own data can help them make more informed decisions about their own health care (Baldwin et al., 2017). Successfully executing this part of the health equation is an important step towards meaningful transparency.

Privacy and Security

When using health equity data sets, extra care is needed to ensure that this type of data is maintained with adequate privacy and security safeguards. Individually identifiable health equity data is considered protected health information (PHI) under the Health Insurance Portability and Accountability Act (HIPAA) of 1996. Organizations should ensure they comply with applicable federal and state laws and regulations regarding access, use and disclosure of such information, including but not limited to HIPAA and its implementing regulations.

Call to Action

The need to standardize and improve health equity data collection and usage are crucial steps towards reducing racial and other health disparities. Successful initiatives must take place at an ecosystem level, and collaboration across that ecosystem is necessary. A more uniform approach to data collection and usage will empower the health care ecosystem to collaborate in a way that fosters unity and equitable outcomes and makes meaningful progress in addressing health inequities.

Table 4. A Call to Action for Standardization Across the Health Care Ecosystem

Actions Needed to Standardize Health Equity Data
<ol style="list-style-type: none"><li data-bbox="134 310 1136 422">1. Develop and adopt data collection and exchange standards across REL and SOGI data sets while continuing to build standardization for SDoH through the Gravity Project.<li data-bbox="134 436 1136 506">2. Develop and adopt standards for sound imputation methods and use of health equity data.<li data-bbox="134 520 1136 590">3. Learn from patients how best to collect REL, SOGI and SDoH data and provide transparency to patients on how such data will be used.

The concepts and efforts described above are important steps forward in a collective goal to promote health equity and effectively address health disparities. The success of efforts to standardize data collection and use should be measured, in part, on how widely such standards are adopted. Members of the health care ecosystem should collaborate and strive to make these efforts as successful as possible, for communities, for individuals and for improving health equity.

References

- American Cancer Society. (2022). *Cancer Facts & Figures for African American/Black People*. Retrieved 3/24/2022 from <https://www.cancer.org/research/cancer-facts-statistics/cancer-facts-figures-for-african-americans.html>
- American Public Health Association. (2021). *Racism is a Public Health Crisis*. Retrieved 3/24/2022 from <https://www.apha.org/topics-and-issues/health-equity/racism-and-health/racism-declarations>
- Bailey, Z. D., Feldman, J. M., & Bassett, M. T. (2021). How Structural Racism Works - Racist Policies as a Root Cause of U.S. Racial Health Inequities. *N Engl J Med*, 384(8), 768-773. <https://doi.org/10.1056/NEJMms2025396>
- Baldwin, J. L., Singh, H., Sittig, D. F., & Giardina, T. D. (2017). Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. *Healthc (Amst)*, 5(3), 81-85. <https://doi.org/10.1016/j.hjdsi.2016.08.004>
- Bancks, M. P., Kershaw, K., Carson, A. P., Gordon-Larsen, P., Schreiner, P. J., & Carnethon, M. R. (2017). Association of Modifiable Risk Factors in Young Adulthood With Racial Disparity in Incident Type 2 Diabetes During Middle Adulthood. *JAMA*, 318(24), 2457-2465. <https://doi.org/10.1001/jama.2017.19546>
- Baptiste-Roberts, K., Oranuba, E., Werts, N., & Edwards, L. V. (2017). Addressing Health Care Disparities Among Sexual Minorities. *Obstet Gynecol Clin North Am*, 44(1), 71-80. <https://doi.org/10.1016/j.ogc.2016.11.003>
- Bazargan, M., Cobb, S., & Assari, S. (2021). Discrimination and Medical Mistrust in a Racially and Ethnically Diverse Sample of California Adults. *Ann Fam Med*, 19(1), 4-15. <https://doi.org/10.1370/afm.2632>
- Bilheimer, L. T., & Klein, R. J. (2010). Data and measurement issues in the analysis of health disparities. *Health Serv Res*, 45(5 Pt 2), 1489-1507. <https://doi.org/10.1111/j.1475-6773.2010.01143.x>
- Centers for Disease Control and Prevention. (2000). Race and Ethnicity Code Set Version 1.0. Retrieved 3/24/2022, from <https://www.cdc.gov/phinf/resources/vocabulary/documents/cdc-race--ethnicity-background-and-purpose.pdf>
- Centers for Disease Control and Prevention. (2015). *Editing and imputation of race and Hispanic origin data in the NHIS*. Retrieved 3/24/2022 from https://www.cdc.gov/nchs/nhis/rhoi/rhoi_documentation.htm
- Centers for Disease Control and Prevention. (2017). *Native Americans with Diabetes: Better diabetes care can decrease kidney failure*. Retrieved 3/24/2022 from <https://www.cdc.gov/vitalsigns/aiand-diabetes/index.html>
- Centers for Disease Control and Prevention. (2021). *Racism and Health*. Retrieved 3/24/2022 from <https://www.cdc.gov/healthequity/racism-disparities/index.html>
- Centers for Disease Control and Prevention. (2022). *Collecting Sexual Orientation and Gender Identity Information Importance of the Collection and Use of These Data*. Retrieved 3/24/2022 from <https://www.cdc.gov/hiv/clinicians/transforming-health/health-care-providers/collecting-sexual-orientation.html>
- Esmailzadeh, P. (2019). The Impacts of the Perceived Transparency of Privacy Policies and Trust in Providers for Building Trust in Health Information Exchange: Empirical Study. *JMIR Med Inform*, 7(4), e14050. <https://doi.org/10.2196/14050>
- Evaluation, O. o. t. A. S. F. P. A. (2011). *HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status*. Retrieved 3/24/2022 from <https://aspe.hhs.gov/reports/hhs-implementation-guidance-data-collection-standards-race-ethnicity-sex-primary-language-disability-0>

- Fremont, A. M., Weissman, J. S., Hoch, E., Elliott, M. N., & Rand, C. (2016). When race/ethnicity data are lacking : using advanced indirect estimation methods to measure disparities. *Research report RR-1162-COMMASS*. <https://www.rand.org/pubs/periodicals/health-quarterly/issues/v6/n1/16.html>
- HL7 International. *Gravity Project*. Retrieved March 24, 2022 from <https://www.hl7.org/gravity/>
- HL7 International. (2022). *The Gender Harmony Project*. Retrieved 3/24/2022 from <https://confluence.hl7.org/display/VOC/The+Gender+Harmony+Project>
- Ibanez, G. E., Agudo, M., Martin, S. S., O'Connell, D. J., Auf, R., & Sheehan, D. M. (2017). Offending Behavior, Drug Use, and Mental Health Among Foreign-Born versus U.S. Born Latino Criminal Justice Clients. *J Immigr Minor Health*, 19(3), 674-685. <https://doi.org/10.1007/s10903-016-0542-8>
- Institute of Medicine of the National Academies. (2003). *Unequal Treatment Confronting Racial and Ethnic Disparities in Health Care*. The National Academies Press.
- Ioannidis, J. P. A., Powe, N. R., & Yancy, C. (2021). Recalibrating the Use of Race in Medical Research. *JAMA*, 325(7), 623-624. <https://doi.org/10.1001/jama.2021.0003>
- Jarrin, O. F., Nyandege, A. N., Grafova, I. B., Dong, X., & Lin, H. (2020). Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-standard Self-reported Race Collected During Routine Home Health Care Visits. *Med Care*, 58(1), e1-e8. <https://doi.org/10.1097/MLR.0000000000001216>
- Johnson, A. H., Hill, I., Beach-Ferrara, J., Rogers, B. A., & Bradford, A. (2020). Common barriers to healthcare for transgender people in the U.S. Southeast. *Int J Transgend Health*, 21(1), 70-78. <https://doi.org/10.1080/15532739.2019.1700203>
- McBean, A. M., & Gornick, M. (1994). Differences by race in the rates of procedures performed in hospitals for Medicare beneficiaries. *Health Care Financ Rev*, 15(4), 77-90. <https://www.ncbi.nlm.nih.gov/pubmed/10172157>
- Nelson, A. (2002). Unequal treatment: confronting racial and ethnic disparities in health care. *J Natl Med Assoc*, 94(8), 666-668. <https://www.ncbi.nlm.nih.gov/pubmed/12152921>
- O'Reilly, K. B. (2020). *AMA: Racism is a threat to public health*. American Medical Association. Retrieved 3/24/2022 from <https://www.ama-assn.org/delivering-care/health-equity/ama-racism-threat-public-health>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. <https://doi.org/10.1126/science.aax2342>
- Office of Management and Budget. (1997). *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Retrieved 3/24/2022 from https://obamawhitehouse.archives.gov/omb/fedreg_1997standards
- Pechenizikiy, M., & Gajane, P. (2018). On Formalizing Fairness in Prediction with Machine Learning. https://www.fatml.org/media/documents/formalizing_fairness_in_prediction_with_ml.pdf
- Riley, W. J. (2012). Health disparities: gaps in access, quality and affordability of medical care. *Trans Am Clin Climatol Assoc*, 123, 167-172; discussion 172-164. <https://www.ncbi.nlm.nih.gov/pubmed/23303983>
- The National Committee for Quality Assurance and Grantmakers in Health. (2021). Federal Action Is Needed to Improve Race and Ethnicity Data in Health Programs. <https://www.gih.org/wp-content/uploads/2021/10/GIH-Commonwealth-Fund-federal-data-report-part-1.pdf>
- The Office of the National Coordinator for Health Information Technology. (2013). The Path to Interoperability. Retrieved 3/24/2022, from https://www.healthit.gov/sites/default/files/factsheets/onc_interoperabilityfactsheet.pdf

- US Department of Health and Human Services Office of Minority Health. (2018). *Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status*. Retrieved 3/24/2022 from <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=53>
- US Department of Health and Human Services Office of Minority Health. (2021a). *Cancer and African Americans*. Retrieved 3/24/2022 from <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=16>
- US Department of Health and Human Services Office of Minority Health. (2021b). *Explanation of Data Standards for Race, Ethnicity, Sex, Primary Language, and Disability*. Retrieved 3/24/2022 from <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=54>
- US Department of Health and Human Services Task Force on Black and Minority Health. (1985). *Report of the Secretary's Task Force on Black & Minority Health*. http://resource.nlm.nih.gov/8602912?_gl=1*1wq125a*_ga*MzU2NjY3Njg0LjE2NDc4OTA1NTk.*_ga_7147EPK006*MTY0ODE0NDk3OC4xLjEuMTY0ODE0NTA1Ni4w*_ga_P1FPTH9PL4*MTY0ODE0NDk3OC4xLjEuMTY0ODE0NTA1Ni4w
- Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*, 383(9), 874-882. <https://doi.org/10.1056/NEJMms2004740>
- Weisgrau, S. (1995). Issues in rural health: access, hospitals, and reform. *Health Care Financ Rev*, 17(1), 1-14. <https://www.ncbi.nlm.nih.gov/pubmed/10153465>
- Xue, Y., Harel, O., Aseltine, O. (2019). Comparison of Imputation Methods for Race and Ethnic Information in Administrative Health Data. 13th International conference on Sampling Theory and Applications,