COMMENTARY: RACE AND ETHNICITY IN BIOMEDICAL RESEARCH – CLASSIFICATIONS, CHALLENGES, AND FUTURE DIRECTIONS

Dame Idossa, MD[1]; Narjust Duma[2]; Katerina Chekhovskiy, BS[3];
Ronald Go, MD[3]; Sikander Ailawadhi, MD[4]

The use of race and ethnicity in biomedical research has been a subject of debate for the past three decades. Initially the two major race categories were: White and Black, leaving other minorities uncounted or inappropriately misclassified. As the science of health disparities evolves, more sophisticated and detailed information has been added to large databases. Despite the addition of new racial classifications, including multiracial denominations, the quality of the data is limited to the data collection process and other social misconceptions. Although race is viewed as an imposed or ascribed status, ethnicity is an achieved status, making it a more challenging variable to include in biomedical research. Ambiguity between race and ethnicity often exists, ultimately affecting the value of both variables. To better understand specific health outcomes or disparities of groups, it is necessary to collect subgroup-specific data. Cultural perceptions and practices, health experiences, and susceptibility to disease vary greatly among broad racial-ethnic groups and requires the collection of nuanced data to understand. Here, we provide an overview of the classification of race and ethnicity in the United States over time, the existing challenges in using race and ethnicity in biomedical research and future research directions. *Ethn Dis.* 2018;28(4):561-564; doi:10.18865/ed.28.4.561.

Keywords: Biomedical Research; Race; Ethnicity; Health Disparities

[1] Mayo Clinic, Department of Internal Medicine; Rochester, MN
[2] Mayo Clinic, Department of Medical Oncology; Rochester, MN
[3] Mayo Clinic, School of Medicine, Rochester, MN
[4] Mayo Clinic, Department of Hematology-Oncology; Jacksonville, FL

INTRODUCTION

The use of race and ethnicity as a proxy for genetic variations and determinants of risk and outcomes is prevalent in biomedical research. Race is largely a sociopolitical construct and classification system based on the geographic origin of a person's ancestry.[1] Ethnicity is a related but broader construct encompassing culture, tradition, religion and shared heritage.[1] The significance and meaning of each construct varies, depending on global location and contingent upon multiple factors including economic, political, social, and cultural practices. Although the biological relevance of race in biomedical research is a topic of considerable debate, the role it plays in the lived experiences of people and its impact on the health and wellbeing of communities is well-documented.[2]

For the better part of its history, the United States used two major race categories: White and Black. Tremendous power and privileges came with being classified as White, while principles of classification such as the "one-drop rule" was used to assign all others into lower status ie, Black or colored.[3] As the country became more diverse and/or began recognizing more groups, race categories on the US census also multiplied. For instance, Asians were first acknowledged in 1860, when Chinese first appeared as a category. By the 1900s, there were several sub-categories within the Asian classification; however many groups were still left out and some, such as Hindus (a religious group, not a racial group) were added.[4]

Despite being the first occupants of the "new world," Native Americans were not included in the early US censuses as they were thought to live in separate nations. However, in 1860 the term Indian was added, but was meant to include only those who were considered assimilated (ie, those who settled in or near White communities). It wasn't until 1890 that the census attempted to count the entire Native American population.[5]

Furthermore, there was a lack of systematic effort to include Hispanics on the census until 1970. This population was predominantly Mexican and classified as White. In 1930, the term Mexican was added to the

---

Address correspondence to Narjust Duma, MD; Mayo Clinic, Department of Medical Oncology; 200 First Street SW; Rochester, MN 55905; 507.284.2511; duma.narjust@mayo.edu

census, but then eliminated in 1940 after successful lobbying from Mexican Americans, due to the power and privileges associated with "Whiteness."[6] The passage of the 1965 Immigration and Nationality Act increased the migration of people from Latin America and other non-European regions. This resulted in the influx of a larger Latino population to the United States; thus, the 1970 census asked about Mexican, Puerto Rican, Cuban, Central, South American, or other Spanish origin.[7] In 2000, the census added the category, Latino.

> *The category of White has been constant on the census; however, definitions of who can be considered White have changed significantly over time.*

The category of White has been constant on the census; however, definitions of who can be considered White have changed significantly over time. For example, more Hispanic-origin Americans identified as White in the 2010 census compared with the 2000 census. The 2000 census defined White people as "any people with origins in any of the original peoples of Europe, the Middle East, or North Africa."[8] Interestingly, it wasn't until 2000 that Americans have had the option of identifying with more than one race. The most recent decennial census, in 2010, had 63 possible race categories: six for single races and 57 for combined races, further highlighting the complexities involved classifying people into multiple sub-classifications.

DISPARITIES IN HEALTH

Racial disparities in health and health care are widespread and well documented.[9,10] The initial report highlighting health disparities in the United States was released by the US Department of Health and Human Services in 1983.[11] This report suggested that while the overall health of the nation was improving significantly, persistent disparities existed in communities of color. In response to these findings, the US government formed a task force to explore the disparities further. In 1985, "Report of the Secretary's Task Force on Black and Minority Health" was released to raise awareness of the ongoing disparities.[12] In the early 1990s, large epidemiologic studies, such as the Harvard Medical Practice Study,[9] suggested that there was practice variability and substandard care experienced by people of color and those from low socio-economic status.[9]

The Indian Health Service (IHS) was established in 1995 to improve the health of the Native American population in the United States. Disparities in infectious diseases (eg tuberculosis), lower life expectancy, and high infant mortality are only a few of the issues affecting the Native American population. Since its establishment, the IHS has improved health conditions dramatically; but, unfortunately, Native Americans continue to experience some of the worst health conditions in the United States.[13]

On the other hand, research suggests that recent Latino immigrants experience the least disparity in all-cause mortality. Although the rates of poverty, lack of insurance, unstable income and other social determinants of health are unfavorable for Latino immigrants, all-cause mortality is surprisingly lower than that of the general population and US-born Latinos.[14] There are multiple theories to explain this paradox including: selective arrival of younger and healthier immigrants; "salmon bias" of return migration whereby the less healthy immigrants return to their country of origin; and misclassification of surnames and ethnicity/nationality of origin on death certificates, resulting in systematic omission (estimated at 15%–20%) of Latino deaths.[15] In addition, there is significant heterogeneity in outcomes by disease process, so assessing overall all-cause mortality may not be appropriate, as it would miss capturing this variability. For example, when comparing the Latino population with the general population, there are significant disparities with poor outcomes in chronic diseases such as type 2 diabetes, infectious diseases (eg, human immunodeficiency virus [HIV]); cirrhosis; psychiatric conditions; and cancers (eg, stomach, liver, and cervical).[16]

CHALLENGES

Researchers view race and ethnicity as significant factors in research and often include them in studies; however, terms are rarely defined and differences in findings between groups are often unexplained. In 2000, in an effort to expand health disparities research, congress called for the creation of the National Center on Minority Health and Health Disparities (NCMHD) within the National Institute of Health (NIH). Minority groups were defined based on previous legislation and included: American Indians; Asian Americans; Native Hawaiians/Pacific Islanders; Blacks; and Hispanics.[17] Though the primary purpose of NCMHD was to support research efforts focusing on disparities experienced by populations of color in the United States, the definition of minorities provided less-than-optimal guidance on how to employ these terms in research. These large-group classifications combine groups of people who may have significant differences in health status and do not leave room to differentiate between sub-groups. For example, the term Hispanic includes individuals of Mexican, Puerto Rican, Cuban, Central and South American and other Spanish-speaking origin. Furthermore, there are significant genetic, behavioral, and environmental changes that can occur between first generation and the subsequent generation of immigrants. As a result, large-group classifications provide inaccurate insight into differences and mask variations within groups.

Although race is viewed as an imposed or ascribed status, ethnicity is an achieved status,[18] making it a more challenging variable to include. Ethnicity is a product of self-identification and may not always fit the standard classifications currently employed in research. In addition, self-identified ethnic classifications do not necessarily match assigned ethnic classification. For example, a study comparing population-based cancer registry from the Surveillance, Epidemiology, and End Results (SEER) database with self-reported data from 13,538 patients found that although the data agreed on race, agreement was only moderate on Hispanic ethnicity, and poor on immigrant status. SEER data was found to underclassify minority groups, resulting in inconsistent cancer statistics.[1,18]

While it is helpful and necessary to collect racial and ethnic identity information for tracking disparities, a person's racial and ethnic identity is not necessarily a full accounting of how they could be defined by ancestry. Moreover, history shows us that self-identification is fluid and changes over time. As the United States becomes more diverse, individuals with multiethnic backgrounds may find it challenging to identify with one specific ethnicity. Furthermore, groups that may not fit within the current category or those who have been systematically misclassified fall through the cracks. For example, when using the current classification, Pakistanis and Indians are classified as Asians. The lack of nuanced and disaggregated data collection masks the unique health challenges that exist in these communities and extent of health disparities at large. To better understand these challenges, it is important to systemize and standardize the collection of the data across of the consortium of health disparities research, including comprehensive ethnic identification and clearly defined racial and ethnic terms, as well as obtaining ethnic data directly from study participants or patients.

FUTURE DIRECTIONS

To better understand specific health outcomes or disparities of groups, it is necessary to collect subgroup-specific data. Cultural perceptions and practices, health experiences, and susceptibility to disease vary greatly among broad racial-ethnic groups and requires the collection of nuanced data to understand. Some databases such as the National Cancer Database have recently added sub-classification of Hispanics based on geographic region: Central American, South American, and Caribbean. Similar changes to identify other subgroups within the current classification could greatly increase the accuracy of our data.[18]

A comprehensive and systematic collection of racial and ethnic data is beneficial to biomedical research in multiple ways. Epigenetic studies have shown significant genetic variation across racial subpopulations. Relationships between susceptibility to disease and the frequency and expression of certain alleles have been established. We can study the unique nature of disease and the differences in tolerability and response to treatment in each racial/ethnic group, if we have well-defined groups.

Studying health status and health outcomes based on race and ethnicity brings attention to inequalities in health care and helps identify barriers to receiving adequate care. It is also important in understanding group-associated risks (such as environmental risk factors) for specific diseases and creation of targeted interventions to reduce them. Improved racial and ethnic classification brings attention to previously understudied populations and promotes inclusion of groups historically overlooked in research. Studying the profound role race and ethnicity plays in the lived experiences of individuals is an opportunity to understand the unique health challenges of different groups and make research more inclusive and applicable to our diverse populations. It also helps identify areas where funding and other resources need to be allocated. The ongoing interest and attention that disparities in health and health care are receiving is encouraging and while we work out the details, the overall direction is optimistic.

Studying the profound role race and ethnicity plays in the lived experiences of individuals is an opportunity to understand the unique health challenges of different groups and make research more inclusive and applicable to our diverse populations.

Conflict of Interest

No conflicts of interest to report.

Author Contributions

Research concept and design: Idossa, Duma, Chekhovskiy, Go; Acquisition of data: Chekhovskiy; Data analysis and interpretation: Chekhovskiy, Go, Ailawadhi; Manuscript draft: Idossa, Duma, Chekhovskiy, Go, Ailawadhi; Statistical expertise: Chekhovskiy; Administrative: Idossa, Duma, Go; Supervision: Duma, Go, Ailawadhi

References

1. Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med*. 2003;348(12):1170-1175. https://doi.org/10.1056/NEJMsb025007 PMID:12646676
2. Adler NE, Rehkopf DH. U.S. disparities in health: descriptions, causes, and mechanisms. *Annu Rev Public Health*. 2008;29(1):235-252. https://doi.org/10.1146/annurev.publhealth.29.020907.090852 PMID:18031225
3. Davis FJ. *Who Is Black? One Nation's Definition.* State College, PA: Pennsylvania State University Press; 2010.
4. Mezey N. Erasure and Recognition: The Census, Race and the National Imagination. 2003;196. Last access August 20, 2018 from https://scholarship.law.georgetown.edu/facpub/196.
5. Meyer ML. *American Indian Blood Quantum Requirements: Blood is Thicker than Family. Over the Edge: Remapping the American West. California, USA.* Berkeley: University of California Press; 1999.
6. Hochschild JL, Powell BM. Racial reorganization and the United States Census 1850-1930: Mulattoes, half-breeds, mixed parentage, Hindoos, and the Mexican race. *Stud Am Polit Dev*. 2008;22(01):59-96. https://doi.org/10.1017/S0898588X08000047
7. Tienda M, Mitchell F. Hispanics and the future of America. In: *National Research Council (US) Panel on Hispanics in the United States*; Tienda M, Mitchell F, editors. Washington (DC): National Academies Press (US); 2006. 1. Last accessed August 20, 2018 from https://www.ncbi.nlm.nih.gov/books/NBK19892/
8. Grieco EM. *The White Population, 2000.* Washington, DC: US Dept. of Commerce, Economics and Statistics Administration, US Census Bureau; 2001.
9. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991;324(6):370-376. https://doi.org/10.1056/NEJM199102073240604 PMID:1987460
10. Ayanian JZ, Landon BE, Newhouse JP, Zaslavsky AM. Racial and ethnic disparities among enrollees in Medicare Advantage plans. *N Engl J Med*. 2014;371(24):2288-2297. https://doi.org/10.1056/NEJMsa1407273 PMID:25494268
11. Health, United States, 1983. *Ann Intern Med*. 1985;103(1):168-168. https://doi.org/10.7326/0003-4819-103-1-168_1
12. Heckler M. Report of the Secretary's Task Force on Black & Minority Health. Washington, DC: US Department of Health and Human Services. 1985. Last accessed August 20, 2018 from https://minorityhealth.hhs.gov/assets/pdf/checked/1/ANDERSON.pdf
13. Jones DS. The persistence of American Indian health disparities. *Am J Public Health*. 2006;96(12):2122-2134. https://doi.org/10.2105/AJPH.2004.054262 PMID:17077399
14. Markides KS, Coreil J. The health of Hispanics in the southwestern United States: an epidemiologic paradox. *Public Health Rep*. 1986;101(3):253-265. PMID:3086917
15. Rosenberg HM, Maurer JD, Sorlie PD, et al. Quality of death rates by race and Hispanic origin: a summary of current research, 1999. *Vital Health Stat 2*. 1999;(128):1-13. PMID:10611854
16. Vega WA, Rodriguez MA, Gruskin E. Health disparities in the Latino population. *Epidemiol Rev*. 2009;31(1):99-112. https://doi.org/10.1093/epirev/mxp008 PMID:19713270
17. Byrd WM, Clayton LA. Racial and ethnic disparities in healthcare: a background and history. In: Institute of Medicine, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: The National Academies Press. 2003:pp 455-527. Last accessed August 20, 2018 from https://www.nap.edu/read/10260/chapter/20
18. Lee C. "Race" and "ethnicity" in biomedical research: how do scientists construct and explain differences in health? *Soc Sci Med*. 2009;68(6):1183-1190. https://doi.org/10.1016/j.socscimed.2008.12.036 PMID:19185964