Published in final edited form as:

JAMA. 2019 March 26; 321(12): 1217–1218. doi:10.1001/jama.2018.21947.

Race and Ethnicity Data in Research

John Heintzman, MD, MPH,

OCHIN Inc, Portland, Oregon

Miguel Marino, PhD

Department of Family Medicine, Oregon Health & Science University, Portland

To the Editor

Mr Bonham and colleagues¹ cautioned about using race and ethnicity in genomic research to make predictions about individuals, populations, and clinical outcomes. We agree with their call for consensus in the research uses of race and ethnicity and that “other types of data providing more nuanced insights should be collected in addition to race, ethnicity, and genetic ancestry, such as a person’s educational attainment, income, and geographic residence.”¹

However, we would argue that the authors should go further in their recommendations for more nuanced uses of race and ethnicity data. Calls for better collection and understanding of race and ethnicity data are not new. They have been documented as early as the 1980s in the Malone-Heckler Report, by the Institute of Medicine in *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care* in 2002, and, more recently, in the Affordable Care Act. Few authors, however, suggest concrete paths forward and, as a result, these recommendations span 40 years with relatively little change.

Even though national attention has drifted toward understanding race, ethnicity, and health with genomics, health services research methods for studying race and ethnicity have stayed stagnant in recent decades, largely using surveys developed in the mid-20th century, failing to properly use data innovations in health systems and, therefore, not uncovering the nuance for which these authors called.

A path toward better understanding of race and ethnicity and their association with health relies on the potential to increasingly employ and link novel data sets, which may have yet to be fully used for research in race and ethnicity. For example, national electronic health record data sets have provided insights on racial and ethnic disparities in preventive services² and have been used to study disparities in Medicaid coverage accuracy when linked to claims.³

In addition, there is burgeoning evidence that zip code is a better predictor of life expectancy than genetic code.⁴ The capacity now exists to link large patient data sets by patient address

to numerous community-level social determinants of health to understand the intersection between social determinants and race and ethnicity.⁵

To understand race and ethnicity and their role in population health, genomics data need to be used cautiously. In addition to calls for consensus and common ground, novel, diverse, and integrated data sets should be used to capture the biology, health, and experience of patients and populations from all backgrounds.

Acknowledgments

Conflict of Interest Disclosures: Dr Heintzman reported receiving grants from the National Institute of Minority Health and Health Disparities, the National Institute on Aging, the Agency for Healthcare Research and Quality (AHRQ), and the Patient-Centered Outcomes Research Institute; and being a contract investigator at OCHIN Inc, a nonprofit health information, AHRQ-sponsored, practice-based research network. No other disclosures were reported.

References

1. Bonham VL, Green ED, Pérez-Stable EJ. Examining how race, ethnicity, and ancestry data are used in biomedical research. *JAMA*. 2018;320(15):1533–1534. doi:10.1001/jama.2018.13609 [PubMed: 30264136]
2. Heintzman JD, Bailey SR, Muench J, Killerby M, Cowburn S, Marino M. Lack of lipid screening disparities in obese Latino adults at health centers. *Am J Prev Med*. 2017;52(6):805–809. doi:10.1016/j.amepre.2016.12.020 [PubMed: 28190691]
3. Marino M, Angier H, Valenzuela S, et al. Medicaid coverage accuracy in electronic health records. *Prev Med Rep*. 2018;11:297–304. doi:10.1016/j.pmedr.2018.07.009 [PubMed: 30116701]
4. Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, et al. Inequalities in life expectancy among US counties, 1980 to 2014: temporal trends and key drivers. *JAMA Intern Med*. 2017;177(7):1003–1011. doi:10.1001/jamainternmed.2017.0918 [PubMed: 28492829]
5. Bazemore AW, Cottrell EK, Gold R, et al. “Community vital signs”: incorporating geocoded social determinants into electronic records to promote patient and population health. *J Am Med Inform Assoc*. 2016;23(2):407–412. doi: 10.1093/jamia/ocv088 [PubMed: 26174867]