Introduction

Demographic characteristics of Marketplace enrollees can inform health equity policy and research. However, more than 30% of HealthCare.gov enrollees do not report race and ethnicity. This cross-sectional study uses imputation to describe HealthCare.gov enrollment data and assess patterns of missing enrollee race and ethnicity.

Methods

This study was reviewed and approved by RAND Corporation's institutional review board. Informed consent is not applicable to this study. This study followed the STROBE reporting guideline. We used open enrollment data from 2015 through 2021 for states using the HealthCare.gov platform from the Centers for Medicare and Medicaid Services' Multidimensional Information and Data Analytics System (MIDAS). We categorized enrollees into 7 mutually exclusive, self-reported groups: American Indian or Alaska Native; Asian American, Native Hawaiian, and Pacific Islander; Black; Hispanic (regardless of race); multiracial; White; and missing. These categories were based on information available in MIDAS and were consistent with open enrollment reports.

For enrollees missing data in one year but not in another, we used the closest available plan year to fill in the missing year(s). We then used the modified bayesian improved first name surname geocoding (mBIFSG) method to generate probabilities for the 6 nonmissing categories for remaining...
This approach uses 6 million surnames from 2010 Census data and 4250 first names from 2.5 million mortgage applications in 2007 and 2010 to assign race and ethnicity probabilities and links an enrollee’s address to Census block group racial and ethnic composition. Finally, we calibrated the posterior probabilities to the observed distribution of self-reported race and ethnicity.

We evaluated results using C statistics, combined self-reported and imputed data, and assessed trends in Marketplace enrollment by race and ethnicity. Estimates used 6 imputed probabilities summing to 1 for each enrollee rather than assigning race or ethnicity to enrollees.

Analyses were conducted using SAS, version 9.4 (SAS Institute).

Results

From 2015 to 2021, 24.4 million unique enrollees selected 61.4 million plans. Race and ethnicity was missing in 31.5% of plans (22.2% after using other plan-year data; <0.01% after imputation). Imputation had an overall C statistic of 0.94, ranging from 0.62 to 0.97 across categories. Discrimination was very high (C statistic >0.90) for Asian American, Native Hawaiian, and Pacific Islander; Black; Hispanic; and White but lower for American Indian or Alaska Native (C statistic = 0.62) and multiracial (C statistic = 0.67).

Annual percent missing ranged from 26.3% to 36.1% (Figure); imputation results indicated nonreporters were disproportionately Black or Hispanic. Estimated distribution of enrollment by category, before and after imputation (Figure, A vs B), for 2021 changed from 6.3% to 9.5% for Asian American, Native Hawaiian, and Pacific Islander; 6.2% to 11.3% for Black; 12.5% to 22.6% for Hispanic; and 38.7% to 54.0% for White.

Discussion

Imputation of race and ethnicity in HealthCare.gov data using the mBIFSG imputation method had high overall predictive accuracy except for American Indian or Alaska Native and multiracial categories. Imputed results indicate that enrollees with missing data were disproportionately Black and Hispanic and that the Marketplace may serve a more diverse population than suggested by self-reported data.

Imputation of race and ethnicity may improve researchers’ and policy makers’ ability to examine disparities and design better outreach and enrollment strategies. Imputation of race and ethnicity is common in Medicare data; here, we demonstrate its potential use in Marketplace data. Nonetheless, policy efforts should aim to improve self-reported race and ethnicity data collection.

Limitations include the imputation assumption that nonreporters are otherwise similar to self-reporters with the same names and Census block group, the use of older Census data owing to unavailability of 2020 data during the analysis, and the potential that mortgage application data may misrepresent racial and ethnic groups because of longstanding structural inequities in housing. A Census first-name list and additional Marketplace data could further improve imputation. Future research may use imputation to understand consumer experiences in the Marketplace and improve outreach effectiveness to address racial and ethnic inequity.
Corresponding Author: D. Keith Branham, DrPH, MPH, US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Ave SW, Washington, DC 20201 (Keith.Branham@hhs.gov).

Author Affiliations: US Department of Health and Human Services (HHS), Washington, DC (Branham, Finegold, Chen, Sommers); Office of the Assistant Secretary for Planning and Evaluation (ASPE), Washington, DC (Branham, Finegold, Chen, Sommers); Harvard Graduate School of Arts and Sciences, Cambridge, Massachusetts (Chen); Harvard Business School, Boston, Massachusetts (Chen); RAND Corporation, Santa Monica, California (Sorbero, Euller, Elliott).

Author Contributions: Drs Branham and Sorbero had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Branham, Sorbero, Elliott, Sommers.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Branham, Finegold, Chen, Euller.

Critical revision of the manuscript for important intellectual content: Branham, Chen, Sorbero, Elliott, Sommers.

Statistical analysis: Branham, Sorbero, Euller, Elliott.

Obtained funding: Finegold.

Administrative, technical, or material support: Branham, Finegold.

Supervision: Branham, Finegold, Sorbero, Sommers.

Conflict of Interest Disclosures: Ms Chen reported receiving grants from the National Institute on Aging (award no. T32AG51108), grants from the Harvard Business School Doctoral Fellowship, grants from the Robert Wood Johnson Foundation, and grants from the National Institute of General Medical Sciences (award no. T32GM007753) outside the submitted work, and the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Dr Sorbero reported having an ASPE contract that focuses on using claims-based algorithms to identify frailty among Medicare beneficiaries outside the submitted work. Dr Sommers reported holding a Health Research & Educational Trust editor’s position at Health Services Research and receiving grants from Baylor Scott & White, honorarium from the Urban Institute for grant review (2019), honorarium from AcademyHealth for grant review, honorarium from American Economics Journal for journal review, consulting fees from the Illinois Department of Healthcare and Family Services, grants from the Commonwealth Fund, and grants from the Robert Wood Johnson Foundation outside the submitted work. No other disclosures were reported.

Funding/Support: This work was supported by a contract with the ASPE to RAND Corporation, supporting Drs Sorbero and Elliott and Mr Euller.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Meeting Presentation: This study was presented at the AcademyHealth Annual Research Meeting; Washington, DC; June 6, 2022.

REFERENCES
1. Fremont A, Weissman JS, Hoch E, Elliott MN. When race/ethnicity data are lacking: using advanced indirect estimation methods to measure disparities. RAND Corporation. Published June 20, 2016. Accessed February 14, 2022. https://www.rand.org/pubs/research_reports/RR1162.html
2. Elliott MN, Morrison PA, Fremont A, McCaffrey DF, Pantoja P, Lurie N. Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. Health Serv Outcomes Res Methodol. 2009; 9(2):69. doi:10.1007/s10742-009-0047-1
3. Tzioumis K. Demographic aspects of first names. Sci Data. 2018;5:180025. doi:10.1038/sdata.2018.25
4. Hand DJ, Till RJ. A simple generalization of the area under the ROC curve for multiple class classification. Mach Learn. 2001;45(2):171-186. doi:10.1023/A:1010920819831
5. McCaffrey DF, Elliott MN. Power of tests for a dichotomous independent variable measured with error. Health Serv Res. 2008;43(3):1085-1101. doi:10.1111/j.1475-6773.2007.00810.x
6. Haas A, Elliott MN, Dembosky JW, et al. Imputation of race/ethnicity to enable measurement of HEDIS performance by race/ethnicity. Health Serv Res. 2019;54(1):13-23. doi:10.1111/1475-6773.13099