

Accuracy of Bayesian Improved First Name Surname Geocoding (BIFSG) for Race and Ethnicity Imputation in a Kidney Care Management Program to Assess Racial Disparities

Liana Bruce, PhD; Christopher Krasniak, PhD; Cliff Eddings, MS; Bassem Mikhael, MD, MBA; Joe Kimura, MD, MPH; Brandon Phan, MBA



Background



Improving health equity and closing health outcome disparities requires data on patient race and ethnicity. Self-reported race and ethnicity data are ideal, but these data have low response rates.

The National Committee for Quality Assurance Health Equity Accreditation¹ requires that programs with < 80% response rates impute missing values to accurately measure health disparities² based on race and ethnicity.

Objectives



Assess accuracy and data coverage gained from imputation over observed self-reported data, by comparing race and ethnicity imputed with the BIFSG algorithm against self-reported data in a renal care management population.

Determine self-reported race coverage over time in our population to assess progress and remaining needs.

Methods



Patients may self-report their race and ethnicity at initial assessment, which is collapsed into Office of Management and Budget standard categories.

For the 84% of patients who declined to self-report, **we used residence, first names, and surnames to impute their race/ethnicity based on RAND's indirect estimation method**^{3,4,5} using a Bayesian combination of expected race from their names and census geocode identifier.

The final race value was generated by using 16% of self-report, 57% BIFSG-imputed, 9% BISG-imputed, and 11% geolocation-imputed values.

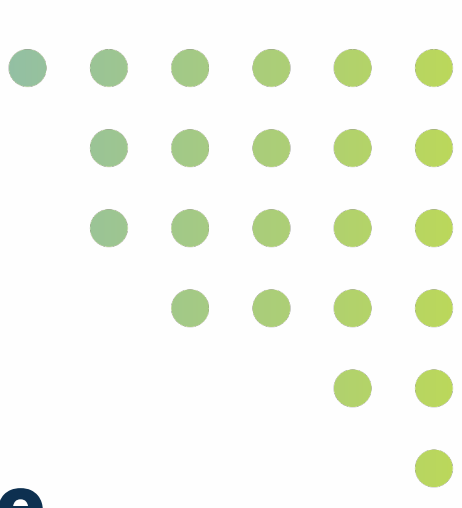
Conclusion

With 99% accuracy, we can impute missing race to increase our data coverage to 93% of patients.

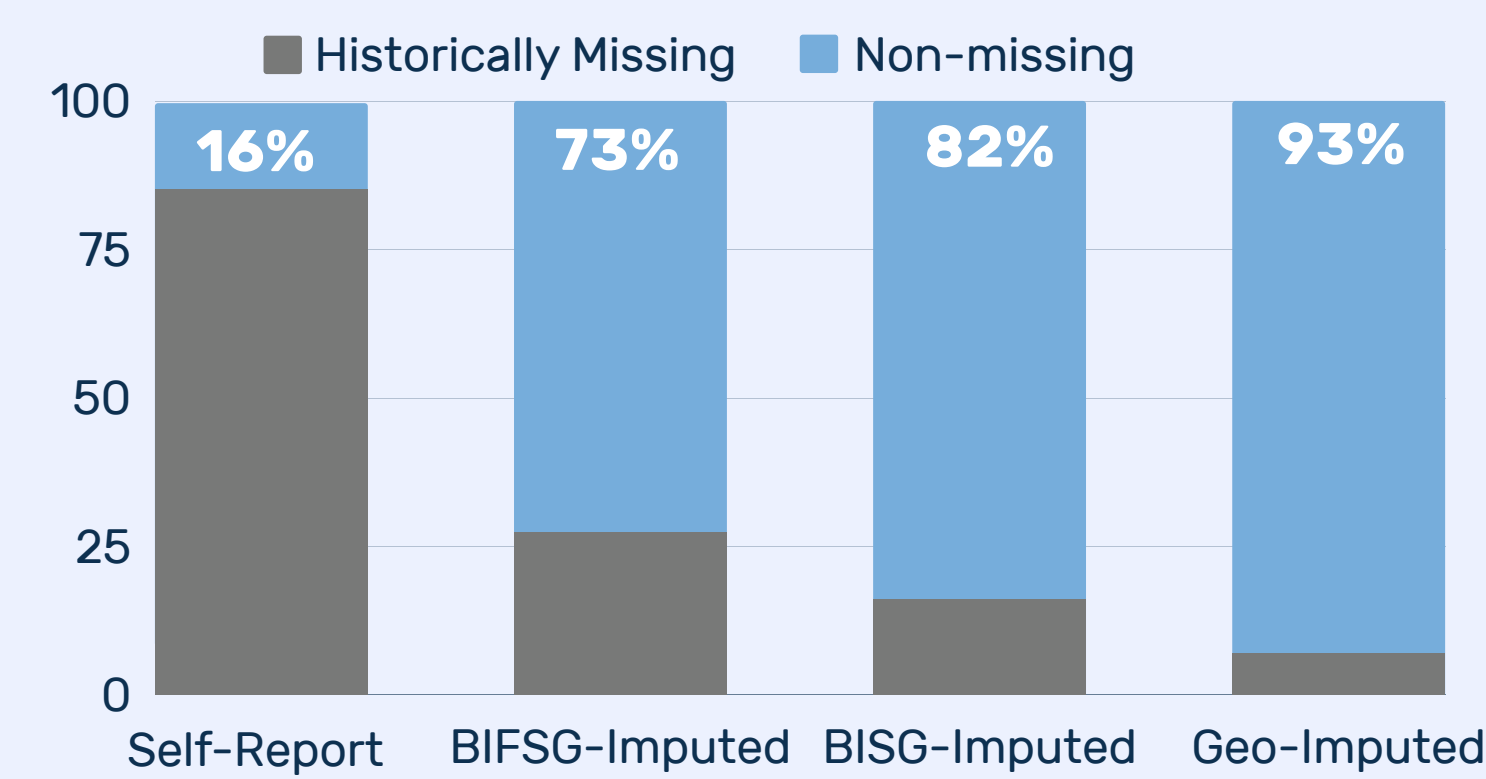
Somatus data collection efforts have historically increased our self-report rates from 6% to 29%. In parallel with these advances, imputation is an accurate and effective tool to increase coverage.

While our findings are consistent with national estimates⁶ and previous research⁷, this study is the first to impute missing race and ethnicity using BIFSG for a renal care population.

This imputation will allow us to more accurately assess² and close health disparities across race and ethnicity in our renal care management population.



Results



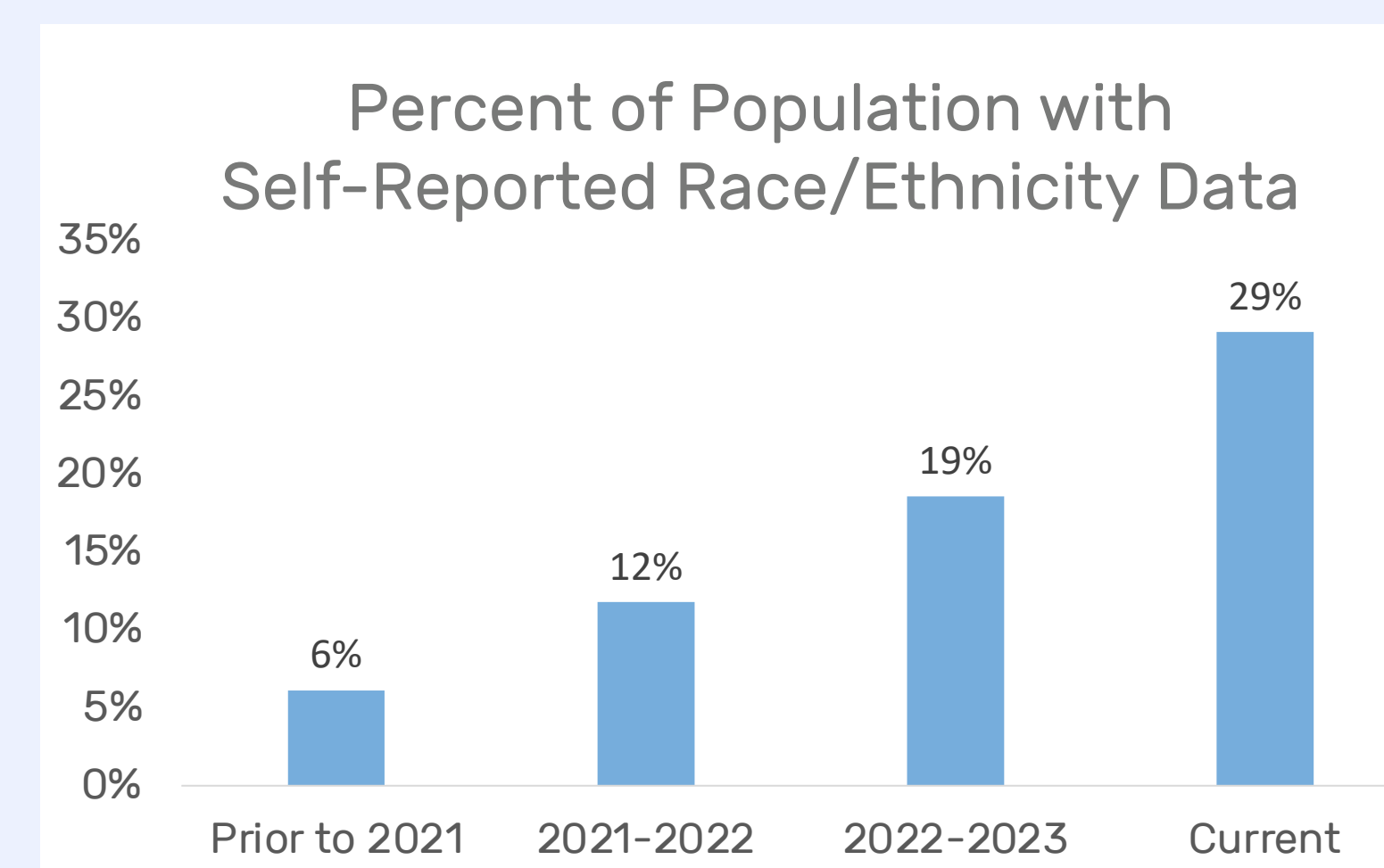
Historically only 16% of patients self-report their race.

Using a series of imputation rounds we increase our data coverage from 16% to 93%.

Metric	Calculation	Overall	White	Black	Hispanic	Asian or Pacific Islander	American Indian or Alaskan Native
Precision (PPV)	$(TP)/((TP+FP))$	80.53%	75.10%	89.14%	92.23%	69.20%	51.35%
Recall (Sensitivity)	$(TP)/((TP+FN))$	80.48%	93.96%	64.66%	82.79%	60.49%	6.55%
Specificity	$(TN)/((TN+FP))$	99.31%	88.71%	99.45%	99.90%	99.98%	99.99%
Accuracy	$(TP + TN) / (TP+TN+FP+FN)$	98.67%	90.11%	97.17%	99.67%	99.94%	99.91%

TP: True Positives, FP: False Positives, FN: False Negatives, TN: True Negatives

Imputation of race/ethnicity is 99% accurate when compared to self-report data.



Growth in self-reported race over time.

In parallel to increasing self-reported data capture over time, imputation is an effective tool to increase coverage to assess health disparities.



1. <https://www.ncqa.org/programs/health-equity-accreditation/>
 2. Yee, K., et al. (2022). "Implications of missingness in self-reported data for estimating racial and ethnic disparities in Medicaid quality measures." Health Serv Res 57(6): 1370-1378.
 3. Sorbero, M. E., et al. (2022). "Imputation of Race and Ethnicity in Health Insurance Marketplace Enrollment Data, 2015-2022 Open Enrollment Periods." Rand Health Q 10(1): 4.
 4. Dembosky, J. W., et al. (2019). "Indirect Estimation of Race/Ethnicity for Survey Respondents Who Do Not Report Race/Ethnicity." Med Care 57(5): e28-e33.
 5. Imai, K., et al. (2022). "Addressing census data problems in race imputation via fully Bayesian Improved Surname Geocoding and name supplements." Sci Adv 8(49): ead9824.
 6. United States Renal Data System. 2022 USRDS Annual Data Report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2022.
 7. Adjaye-Gbewonyo, D., et al. (2014). "Using the Bayesian Improved Surname Geocoding Method (BISG) to create a working classification of race and ethnicity in a diverse managed care population: a validation study." Health Serv Res 49(1): 268-283.

Category and Subcategory Augmented Intelligence, Digital Health, and Data Science
Keywords Health Disparities, Race Estimation
Funding Sources/Commercial Support (if applicable) Funded by Somatus Inc.

Authors
 Liana Bruce, PhD, Christopher Krasniak, PhD, Cliff Eddings, MS, Bassem Mikhael, MD, MBA, Joe Kimura, MD, MPH, Brandon Phan, MBA.