

Combining Electronic Health Record, Enrollment, and Imputed Race/Ethnicity for Commercial Members

Amanda Price¹, Winnie Chi², Marc Overhage³

¹Carelon Research, Norfolk, Virginia; ²Elevance Health, Wilmington, DE; ³Elevance Health, Indianapolis, IN

Background

- Health plans play an essential role in removing obstacles and reducing health inequities but are not required to collect data on race or ethnicity.
- To obtain health equity accreditation, health plans are required to report on race/ethnicity and HEDIS requires metric stratification by race/ethnicity.
- Unavailable or inaccurate race/ethnicity data creates roadblocks to measure and address health inequities and disparities within a population.¹
- Commercial health plans collect member race/ethnicity from electronic health records (EHRs) and enrollment sources, but data is only available for a small proportion of members² and could benefit from standardized data collection practices.^{3,4}
- Imputing race/ethnicity is an established practice that can provide race/ethnicity for large populations^{5,6} and be used for necessary race/ethnicity stratifications.
- Commercial health plans have been incentivized to maximize race/ethnicity data availability and should consider all available data sources of race/ethnicity to both maximize availability and maintain accuracy of member race/ethnicity.

Objectives

- This study validated individual and combined race/ethnicity data sources against a known sample to determine the best race/ethnicity data source or combination of sources for race/ethnicity
- This study's goal was to increase the availability and reliability of member race/ethnicity data in order to stratify health care metrics and identify health equity disparities.

Methods

Data Sources

- Race/ethnicity data (EHRs, enrollment sources, and imputed sources) from a large administrative claims database.
- Three vendors were used to impute race/ethnicity. Vendor 1 uses the Bayesian Improved Surname and Geocoding (BISG) algorithm, which combines geocoding with surname analyses.¹⁰ Vendor 2 uses consumer databases and Vendor 3 uses similar methodology to BISG.
- Self-reported race/ethnicity from human resources records from employer was used as gold standard for comparison.

Validation

- Primary validation compared agreement between member race/ethnicity for each source (electronic health records, enrollment files, and imputation algorithms) with the gold standard race/ethnicity.
- Secondary validation compared composite race/ethnicity agreement to the gold standard.
- Agreement was measured using Cohen's kappa statistic and true positive values.

Results

Figure 1. Prioritization Method Used to Create Standardized Composite Race/Ethnicity

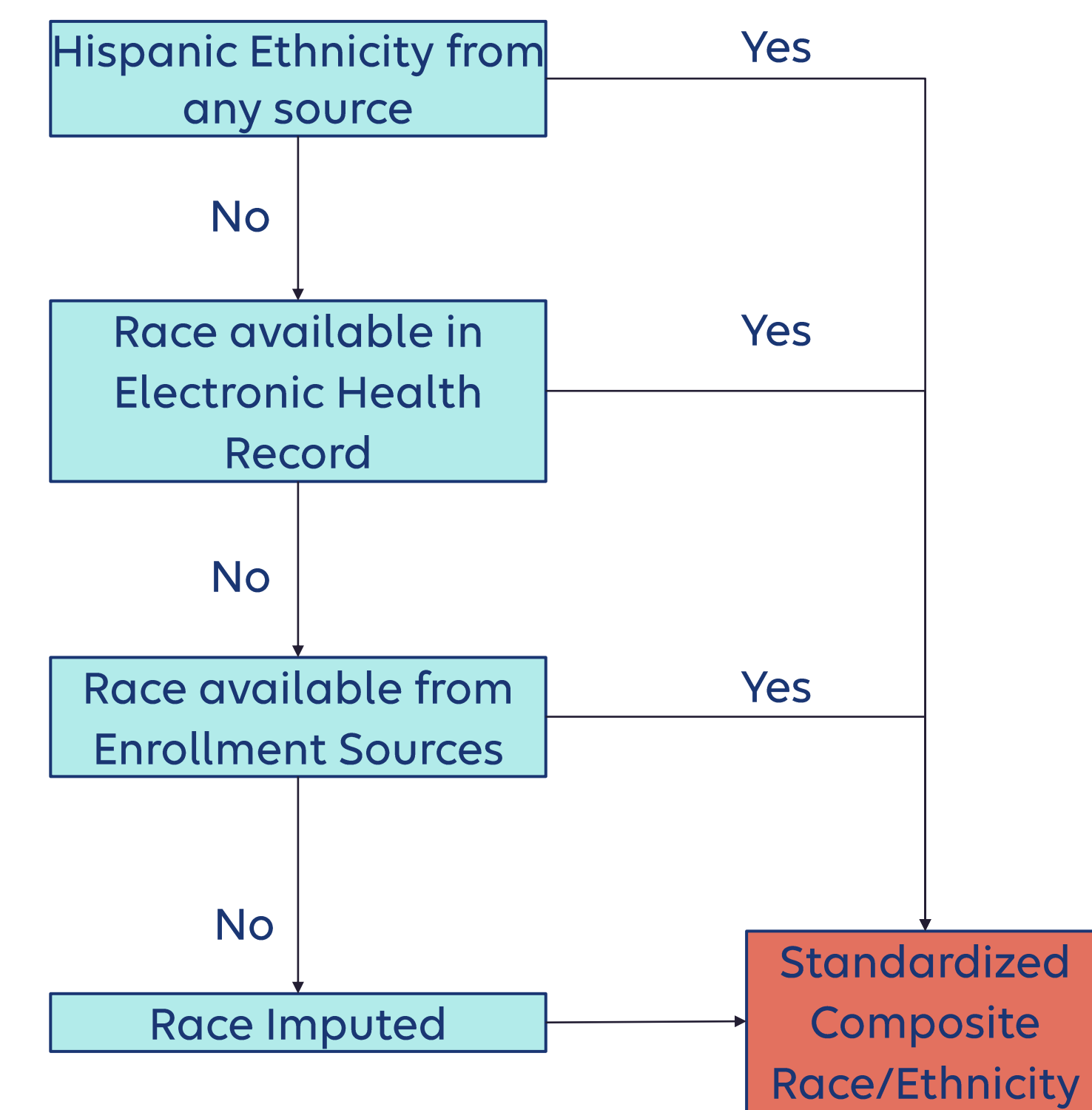


Table 1. Source of Member Race/Ethnicity for Standardized Composite

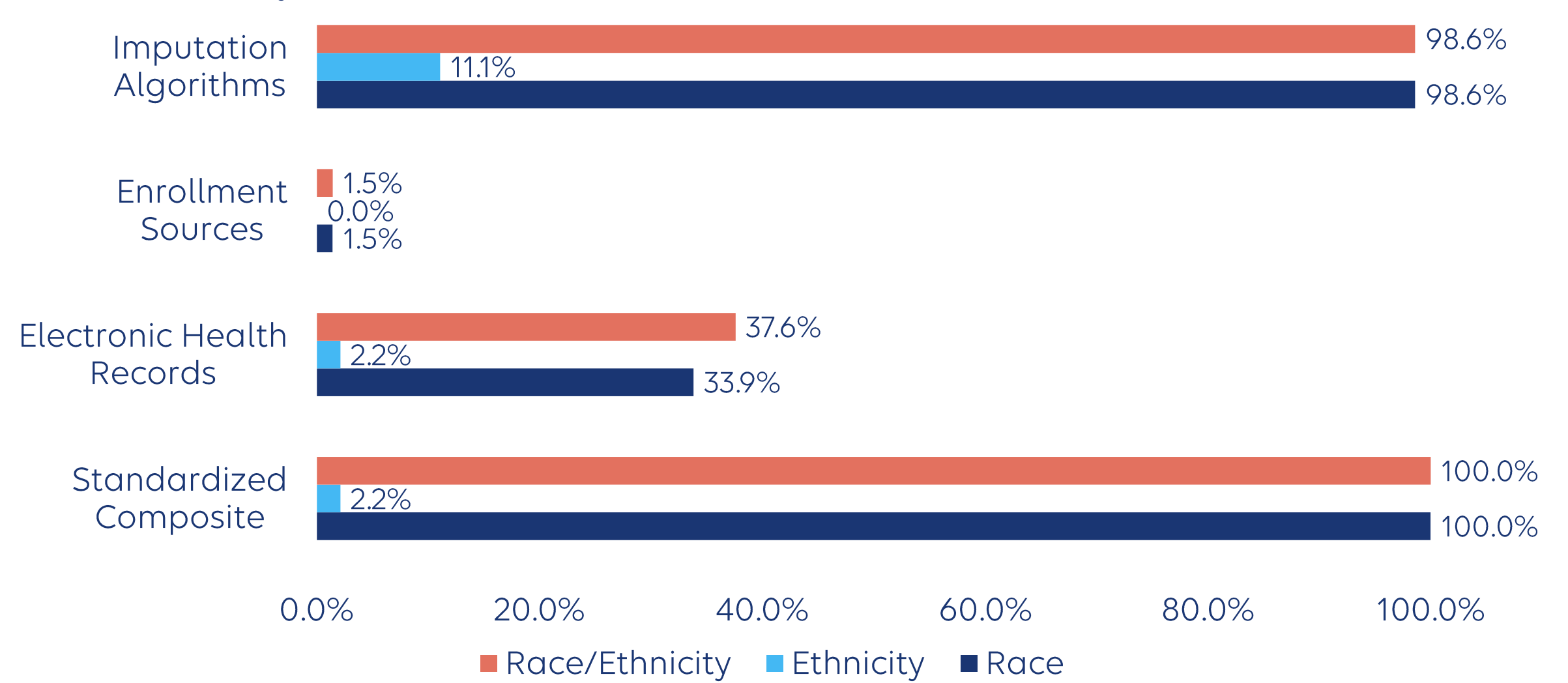
	Race/Ethnicity	
	N	%
Electronic Health Records	19,447	37.64
Enrollment Sources	479	0.93
Imputation Algorithm	31,741	61.43
Vendor 1	18,878	36.54
Vendor 2	2,818	5.45
Vendor 3	10,045	19.44
Total Members	51,667	

Table 2. Standardized Composite Race and Ethnicity Agreement, and True Positive Percentages by Source, Sex, and Age Groups

	Kappa Statistic	Percentage of True Positives			
		Asian N (%)	Black or African American N (%)	Hispanic or Latino N (%)	White N (%)
Standardized Composite	0.8232	3,074 (94.6)	10,180 (93.1%)	4,395 (84.89%)	25,910 (88.51%)
Source of Race/Ethnicity					
Electronic Health Records	0.9467	1,205 (93.7)	4,991 (98.4)	1,047 (77.6)	10,651 (99.1)
Enrollment Sources	0.8893	15 (83.3)	389 (97.7)	54 (65.9)	196 (98.0)
Imputation Algorithm	0.7347	2,898 (79.1)	8,871 (72.2)	4,463 (82.7)	24,004 (90.7)
Vendor 1	0.8228	1,333 (76.2)	3,016 (73.6)	2,824 (83.2)	18,582 (98.3)
Vendor 2	0.7795	105 (30.0)	1,602 (89.2)	2,872 (87.2)	3,401 (87.6)
Vendor 3	0.7099	2,954 (80.0)	8,195 (67.2)	4,263 (79.9)	23,643 (91.1)
Sex					
Female	0.8156	1,486 (91.5)	8,794 (94.0)	3,194 (83.8)	18,849 (87.8)
Male	0.8350	1,559 (97.7)	1,195 (86.2)	1,155 (88.2)	6,637 (90.1)

- Self-reported race/ethnicity was available for 51,667 commercial health plan members who self-reported their race or ethnicity as Asian, Black or African American, Hispanic or Latino, or White. 74% of members were female and between the ages of 40 and 64.
- Imputation methods provided race/ethnicity for over 98% of the members, while EHR sources was available for about 38%, and enrollment for less than 2%.
- Imputation methods were used to create standardized composite race/ethnicity, for 61.43% of members.
- Agreement between standardized composite member race/ethnicity compared to gold-standard data was very high (k=0.8232). Asian race had the highest TPV (94.64%), and Hispanic or Latino ethnicity had the lowest (84.89%).
- EHR and enrollment data sources had very high agreement with the gold standard (k=0.9467 and k=0.8893, respectively).
- Vendor 1 had very high agreement (0.8228), while vendors 2 and 3 had substantial agreement, with vendor 3 having the lowest agreement of all sources at 0.7099.
- Race/ethnicity agreement was highest for males (k=0.8350) and the 30-39 age group (k=0.8245)

Figure 2. Race and Ethnicity Data Completeness and Availability for Standardized Composite and by Source



Conclusions

- In order for health plans to identify health equity they need available and reliable race/ethnicity to stratify necessary performance metrics.
- Based on validation, imputation methods can provide race and ethnicity for large populations and are most reliable and useful when combined with available EHR and enrollment data.
- Combining EHR and enrollment data with imputation methods increases the availability of race and ethnicity without decreasing the accuracy which allows health plans to stratify performance metrics designed to address and improve health equity.

References

- Polubriaginof FCG, Ryan P, Salmasian H, et al. Challenges with quality of race and ethnicity data in observational databases. J Am Med Inform Assoc. 2019;26(8-9):730-736.
- Ng JH, Ye F, Ward LM, Haffer SC, Scholle SH. Data On Race, Ethnicity, And Language Largely Incomplete For Managed Care Plan Members. Health Aff (Millwood). 2017;36(3):548-552.
- Institute of Medicine (US) Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington (DC)2009.
- Nead KT, Candice L, Hinkston, and Mackenzie R. Wehner. Cautions When Using Race and Ethnicity in Administrative Claims Data Sets. JAMA Health Forum. 2022;3(7).
- Elliott MN, Fremont A, Morrison PA, Pantoja P, Lurie N. A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity. Health Serv Res. 2008;43(5 Pt 1):1722-1736.
- Derosé SF, Contreras R, Coleman KJ, Koebnick C, Jacobsen SJ. Race and ethnicity data quality and imputation using U.S. Census data in an integrated health system: the Kaiser Permanente Southern California experience. Med Care Res Rev. 2013;70(3):330-345.