

Evaluating Imputation to Reduce Race and Ethnicity Missingness in the Colorado All Payer Claims Database

Martha Meyer, PhD, MPH
Valarie Garrison, MES



Learn more about Colorado APCD data access, release applications, and supporting documentation at civhc.org



Research Objective

Evaluate whether integrating historical eligibility records and secondary Race/Ethnicity (R/E) fields improves completeness and accuracy of R/E data in the Colorado All Payer Claims Database (CO APCD).

Background

Limited evidence exists on how imputation affects R/E distributions in CO APCD datasets

R/E data in the CO APCD exhibits substantial missingness and inconsistency

Missingness threatens validity of:

- Disparity estimates
- Risk adjustment
- Equity monitoring

56% of records are missing R/E data, limiting completeness and equity-focused analysis

Study Design

Longitudinal member eligibility history (primary + secondary R/E fields)

Imputation using Analytics Group Benchmark approach

Comparison of:

- Imputed vs payer-reported R/E distributions

Analysis:

- Risk differences (imputed - payer reported)
- 95% confidence intervals
- Superiority testing (one-sided z-test)
- Non-inferiority testing (-3 percentage-point margin)

Population

The CO APCD includes data representing 5.7 million covered lives.

Included populations span:

- Commercial (large group, small group, individual markets)
- Medicare Advantage and Fee-for-Service
- Medicaid
- Non-ERISA self-insured claims and select ERISA submissions

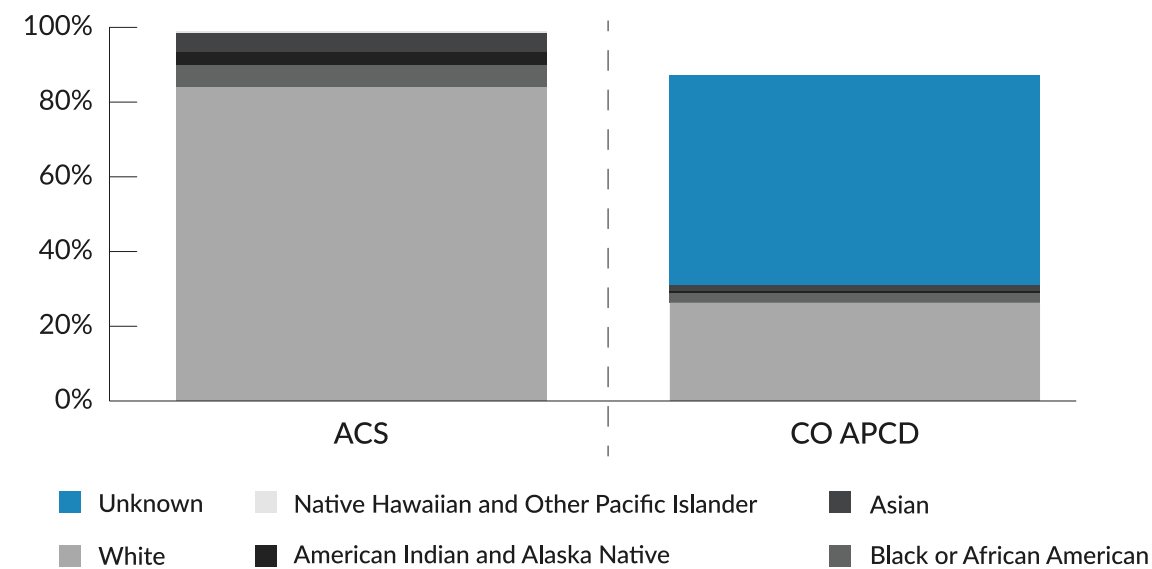
Key Findings

Missing Race/Ethnicity Data in Payer Submissions

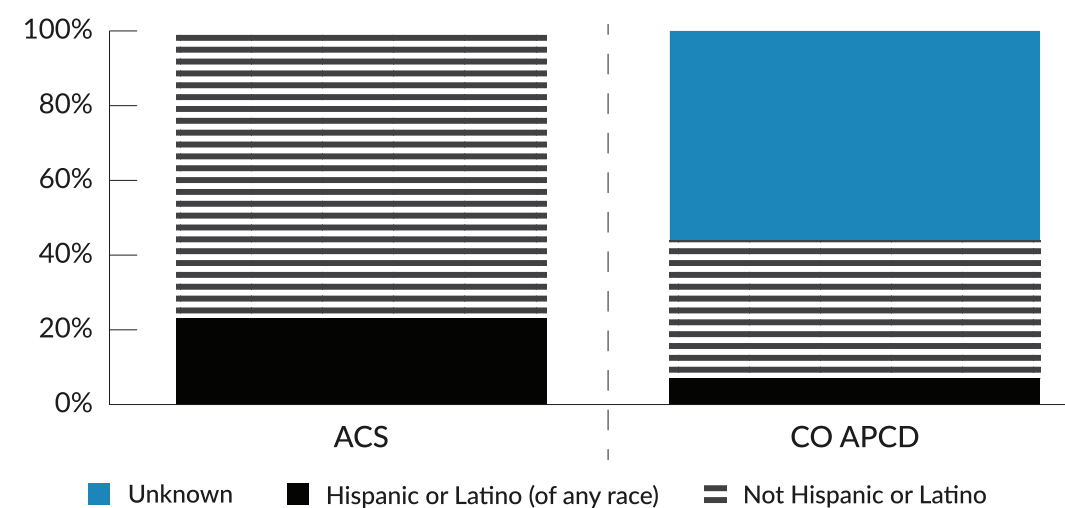
Comparison to American Community Survey (ACS) benchmarks highlights substantial missingness in CO APCD race and ethnicity data.

56% of race/ethnicity data is missing in payer submissions

CO APCD RACE DATA | Contains a High Proportion of "Unknown" Values Compared to ACS



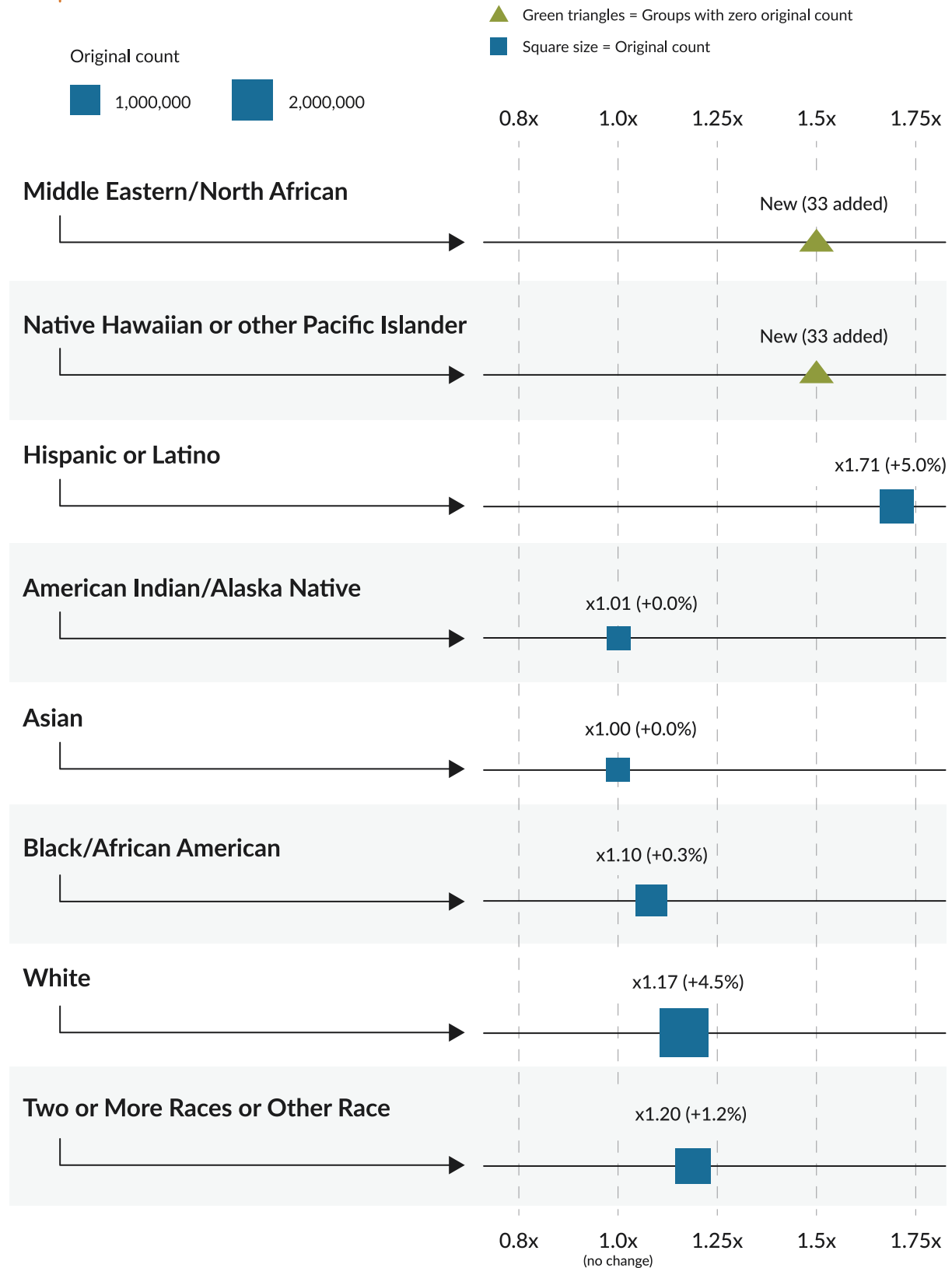
CO APCD ETHNICITY DATA | Shows Substantial Missingness Relative to ACS Benchmarks



Imputed vs Payer-Reported Race/Ethnicity Estimates

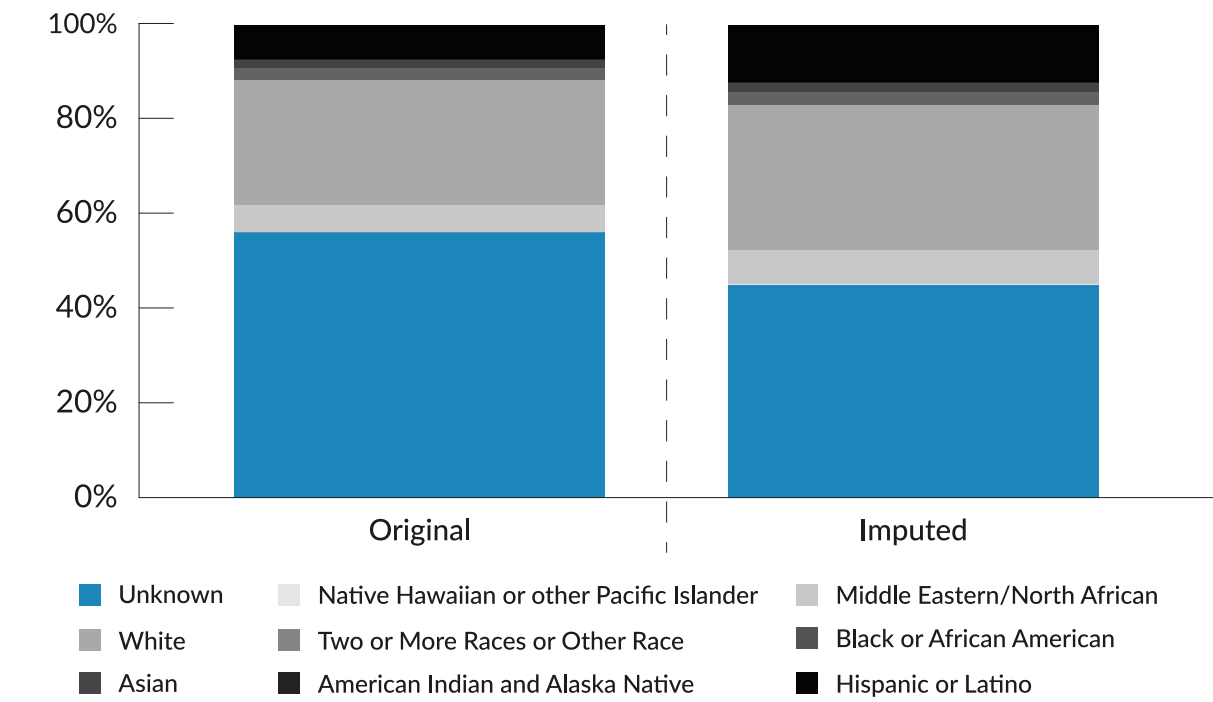
Differences between imputed and payer-reported race/ethnicity distributions were evaluated using risk differences and non-inferiority testing.

Imputed estimates show no meaningful distortion in subgroup distributions, with differences remaining within acceptable statistical bound.



Changes in Race/Ethnicity Distribution Following Imputation

Imputation reduced unknown race/ethnicity by 11 percentage points.



Implications for Policy or Practice

Improved completeness supports more accurate:

- Disparity measurement
- Risk adjustment and stratification
- Program evaluation
- Resource allocation
- Population health monitoring across diverse communities

Imputation Conclusion

Reduces missingness while preserving subgroup distributions

Enhances analytic validity of CO APCD data for population health and equity analysis

Supports more consistent and interpretable race/ethnicity reporting

*Submitting race and ethnicity data became a requirement for payers in late 2019 when a Data Submission Guide update mandated submission of one race and one ethnicity field. CIVHC modeled the data requirements on the classification guidelines from the U.S. Office of Management and Budget (OMB), which include an indicator for Latino/Non-Latino and five race fields: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, or White. CIVHC also enables payers to select race as "other" or "unknown."