Deziree Jackson, Blackburn, Justin, and Hernandez, Elaine

Evaluating Race-Ethnicity Concordance and Discordance between Self-Reported and Administrative Enrollment

A brief in preparation for: Health Services Research

This study evaluates the concordance between self-identified and administratively recorded measures of race and ethnicity, with a focus on how classification discrepancies may affect health services research. As the U.S. population becomes increasingly racially and ethnically diverse, particularly among multi-racial and Latino populations, accurate and consistent measurement of race and ethnicity is essential for monitoring health disparities. By examining the extent and patterns of mismatches (e.g., non-matching classifications) across data sources, this study aims to assess the implications of these inconsistencies for data validity, population health monitoring, and the effective targeting of health services. We ask: To what extent measures of race and ethnicity from administrative data agree (or disagree) with self-reported race-ethnicity when considering (1) ethnicity alone, (2) race alone, (3), and race and ethnicity in combination?

To assess concordance, we link survey and administrative data to assess concordance between self-reported and administratively recorded race and ethnicity. Self-reported data will be collected from 78 in-depth interviews with Medicaid beneficiaries. Administrative data will be obtained from Indiana Medicaid enrollment records. We will conduct cross-tabulations and calculate agreement statistics using Cohen's kappa to quantify concordance across race and ethnicity categories, including multi-racial and Latino identifiers. To complement the quantitative analysis, we draw on semi-structured qualitative interviews to better understand how individuals interpret and self-identify their race and ethnicity across contexts. These interviews will explore meaning, salience, and lived experiences of choosing a racial and ethnic identity, particularly among multi-racial and Latino populations. The qualitative findings will provide contextual depth and insight into the causes and consequences of misclassification, offering recommendations for more accurate and inclusive data collection practices.