

Evaluating Race and Hispanic Origin Responses of Medicaid Participants Using Census Data*

Leticia Fernandez, Sonya Rastogi, Sharon Ennis and James Noon
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC, 20233

Abstract

Racial and ethnic disparities in health and health care are complex and continue to challenge researchers and policy makers. With the intention of improving the measurement and monitoring of disparities, certain provisions of the Patient Protection and Affordable Care Act (ACA) of 2010 will require states in the coming years to collect, report and analyze data on demographic characteristics of applicants and participants in Medicaid and other federally supported programs. By linking Medicaid records to 2010 Census, American Community Survey, and Census 2000, this new large-scale study examines and documents the extent to which pre-ACA Medicaid administrative records match self-reported race and Hispanic origin in Census data. Linked records allow comparisons between individuals with matching and non-matching race and Hispanic origin data across several demographic, socioeconomic and neighborhood characteristics, such as age, gender, language proficiency, education and Census tract characteristics. Identification of the groups most likely to have non-matching and missing race and Hispanic origin data in Medicaid relative to Census data can inform strategies to improve the quality of demographic data collected from the Medicaid population.

Key Words: Medicaid, Agreement of race and Hispanic origin responses, Linked administrative records and Census data.

1. Introduction

Although well documented, racial and ethnic disparities in health and health care are complex and continue to challenge researchers and policy makers (Smedley et al. 2003, Kawachi et al. 2005, Andrulis et al. 2010). In the past, as noted by the Institute of Medicine, the lack of a strong information infrastructure at the federal and state levels have hindered measurement and monitoring of health disparities across states by race, ethnicity and other characteristics (Institute of Medicine 2004). In the case of Medicaid administrative data in particular, states do not use standardized protocols to collect race or Hispanic origin data (McAlpine et al. 2007, Sebelius 2011).

Provisions of the Patient Protection and Affordable Care Act of 2010 (Section 4302, ACA) will require in the coming years that states collect, report and analyze data on race, Hispanic origin, gender, primary language, and disability status for applicants and participants in Medicaid and any other federally conducted or supported health care or public health program or survey (Sebelius 2011, Office of Minority Health 2013).

* *Disclaimer:* This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

In this paper, we link administrative records of Medicaid program participants enrolled in the period 2006 to 2008 to their race and Hispanic origin responses in decennial census and American Community Survey (ACS) data. With these linked data, we document the extent to which pre-ACA Medicaid administrative records match self-reported race and Hispanic origin in Census data. We also describe the socioeconomic, demographic and contextual characteristics associated with individuals whose race or Hispanic origin records in Medicaid (a) match Census data, (b) do not match Census data, or (c) have missing race or Hispanic origin data.

2. Factors that Impact Reporting of Race and Hispanic Origin Data

The Medicaid program is administered by states as a state-federal partnership. States submit Medicaid eligibility and claims data to the Centers for Medicare and Medicaid Services (CMS) through the Federal Medicaid Statistical Information System (MSIS). MSIS data includes demographic characteristics of Medicaid enrollees. However, whether and how these data are collected varies from state to state. Since there is no standard Medicaid application at the national level, race and Hispanic origin data are collected in each state using their own question format and a variety of modes, including paper and online applications, which may or may not include assistance from intake specialists (McAlpine et al. 2007, Sebelius 2011). Moreover, states may list questions about race and Hispanic origin as optional in the Medicaid application form. Clearly, these differences in the collection of race and Hispanic origin data make inter-state comparisons difficult.

We focus on two issues that arise in studies comparing data from different sources: discrepancies in race and Hispanic origin data in Medicaid compared to Census data and item non-response.

2.1 Non-matching Responses

Earlier studies comparing the agreement of race and Hispanic origin data between surveys and administrative records find discrepancies to be associated with race, Hispanic origin, age, gender, education and English proficiency. In particular, Hispanic, Asian and American Indian populations are less likely to have matching responses than individuals who report as non-Hispanic, White or Black (Arday et al. 2000, Kressin et al. 2003, Gomez et al. 2005, Zaslavsky et al. 2012). The exception is McAlpine et al. (2007) who find that American Indians are more likely to have a matching race response than Whites in their study using survey and Minnesota Medicaid data. In contrast, individuals who are younger, more educated, male, and those who speak English at home are associated with a lower likelihood of matching race and Hispanic origin responses between survey and administrative records compared to individuals who are older, less educated, female, and those who speak another language at home (Gomez et al. 2005, McAlpine et al. 2007).

There may be several reasons an individual's race or Hispanic origin in Medicaid may not match her or his Census responses. Below we discuss some of these:

Question Format and Wording – Differences in question wording and design can have an impact on how people understand and respond to race and Hispanic origin questions. For instance, some states have Medicaid application forms that ask race and Hispanic origin as separate questions, while others ask the questions in a combined format. Moreover, some Medicaid forms explicitly indicate that the respondent may choose more than one

race, while others do not (McAlpine et al. 2007, Sebelius 2012).¹ Studies that have looked at the impact of changes in race question wording and response categories among Hispanics consistently find that they are less likely to report a race in question formats that list Hispanic origin as an option compared to when they are asked separate questions about Hispanic origin and race (Campbell and Rogalin 2006, Compton et al. 2012).

Proxy Respondents – Another important source of discrepancies is proxy responses, individuals reporting another person’s race. For example, other household members may have completed the decennial census, and their answers may not match those of self-reporting individuals in Medicaid. Also, for surveys and on administrative forms, even if instructions direct field staff and administrators to collect race and Hispanic origin self-responses, in some instances, field staff and administrators may fill out this information for the respondent. In a study comparing self-reported race and Hispanic origin to cancer registry records, Gomez and Glaser (2006) argued that females’ greater disagreement in their responses were likely to be a reflection of the assumptions by the staff regarding race and Hispanic origin based on the patients’ surnames, affecting cases in which married women had adopted their husband’s last name. In the case of Medicaid applications, in some states intake specialists record race and ethnicity of Medicaid applicants based on verbal reports or assist applicants to determine their race and ethnicity (Sebelius 2011).

Changes in how individuals identify racially (within a context or over time) – In recent years, a growing body of sociological research reports that individuals may change their self-reported race or Hispanic origin over time, with changes in social position or even in particular situations, for example to highlight a specific ancestry (Brown et al. 2006, Saperstein and Penner 2012). Various demographic, socioeconomic and contextual factors have been associated with a higher propensity for racial fluidity. Most consistently, studies report that Hispanics and individuals with non-Black, non-European ancestry tend to have higher racial fluidity than non-Hispanic, White and Black populations (Harris and Sim 2002, Brown et al. 2006, Saperstein and Penner 2012). While racial fluidity is likely to contribute to discrepancies in the comparisons of race and Hispanic origin between Medicaid and Census data, in this study it is not possible to separate racial fluidity from other factors affecting data consistency.

2.2 Item Non-Response

There is not extensive literature on the characteristics of those who do not respond to race and Hispanic origin questions. Some research suggests that Hispanics have higher race item non-response than non-Hispanics, and this may be partly explained by conceptual issues regarding race and Hispanic origin (Rodriguez 2000, Compton et al. 2012). In particular, some Hispanics do not report a race or have difficulty with answering a

¹ The census race and Hispanic origin response data used in this study are from slightly different question designs as well. Race and Hispanic origin questions for Census 2000 and 1-year ACS surveys from 2001 through 2007 were slightly different from 2010 Census questions and 1-year ACS surveys for 2008, 2009 and 2010. Starting in ACS 2008, an instruction was added stating that “For this survey, Hispanic origins are not races,” example groups were added for the Other Hispanic, Other Asian, and Other Pacific Islander checkbox response categories, and the term “origin” was added to the Hispanic origin question. Although these design changes did reduce Some Other Race responses in 2010, we do not expect them to have as large an impact on the overall results of this study relative to the different designs in Medicaid application forms across states.

question on race when they are asked a question on Hispanic origin separately (Compton et al. 2012). Previous studies comparing survey data to administrative records also find that White, younger, married, employed and more educated individuals are more likely to have missing race responses than their counterparts (Kressin et al. 2003, McAlpine et al. 2007).

3. Data and Methods

The data used in this study are Medicaid administrative records for 2006 to 2008 linked to race and Hispanic origin responses in Census 2000, the 2010 Census, and ACS 2001 through 2009 one-year data. For Census data, only records with unedited and non-imputed race and Hispanic origin were kept. In addition, responses from proxies, such as neighbors and building managers were not included in the Census data for this analysis. Linked Medicaid-Census race and Hispanic origin responses were then linked to ACS 2006-2010 five-year data to attach respondents' most recent demographic and socioeconomic characteristics.

In order to remove duplicate cases and link Medicaid program participants to their decennial census and ACS records, a Protected Identification Key (PIK) was assigned via the Person Identification Validation System (PVS), which employs probability record linkage techniques (see Wagner and Layne 2012 for more information). Personally Identifiable Information (PII) such as name, date of birth, and address were used to assign a PIK. The PII was then removed from the data file to anonymize the data and preserve confidentiality so it can be used for statistical purposes and research.

All results presented in this analysis are unweighted. Since we focus on a unique population – the Medicaid population – and further limit the population to those who received a PIK and then matched to Census data, survey weights likely do not represent our universe. We also ran our regression analyses with weights. We generally find that observed patterns in the unweighted analyses are similar to the weighted analyses.²

There are about 1.5 million unique linked records in the sample, which is limited to individuals ages 25 and older. The race groups included in the analysis are White alone, Black alone, American Indian or Alaskan Native (AIAN) alone, Asian alone, and Native Hawaiian or Other Pacific Islander (NHPI) alone. Due to small sample sizes, Asian and NHPI were combined for the analysis. We excluded from the analysis individuals who reported their race in Census data as Some Other Race alone or Two or More Races because of no or negligible response matching across datasets. The Medicaid data does not include a category for Some Other Race as Census does, and not all states allow Medicaid applicants to report more than one race.

² The few exceptions are as follows: (a) In the model predicting non-matching rather than matching race responses the coefficient for never married becomes statistically significant, but does not change in magnitude. Also, the coefficient for individuals who speak English not well or not at all changes from 0.94*** to 1.06***, such that they are more likely to have non-matching race responses compared to English-only speakers. (b) In the model predicting missing rather than matching race responses, the coefficients for high school degree and some college become statistically significant, but magnitudes are similar. (c) In the model predicting missing rather than matching Hispanic origin responses, the coefficient for college or higher education changes from 1.08*** to 0.98*** such that they become somewhat less likely to be missing Hispanic origin data in Medicaid compared to those with no high school diploma.

We first discuss the large inter-state variability in the percentage of cases matching, non-matching and missing race and Hispanic origin responses compared to Census data. Next, we examine the percentage of cases in each racial and Hispanic origin group that had the same answer across Medicaid and Census data. Last, we present our finding from multinomial logistic regressions that examine the relationship between various demographic, socioeconomic and contextual variables and the outcome of comparing Medicaid and Census race and Hispanic origin records. We use two dependent variables in the analysis, one for race and one for Hispanic origin. Each dependent variable has three possible outcomes: (a) Same race (Hispanic origin) response in Medicaid and Census data; (b) different race (Hispanic origin) responses, and (c) missing race (Hispanic origin) data in Medicaid.

We include in our analysis as explanatory variables those that previous studies comparing survey and administrative data at a smaller scale have consistently identified as associated with discrepancies in race and Hispanic origin responses. These demographic and socioeconomic variables are age, gender, marital status, race, Hispanic origin, education and English proficiency. Except for marital status, these variables are also those identified in ACA provisions as important to collect because of their relevance to measures of healthcare disparities.

Inter-state differences in Medicaid data collection, as well as previous research on racial identification, suggest that state and local factors are likely to influence whether race and Hispanic origin responses will match in our comparisons of Medicaid and Census data (Eschbach and Gomez 1998; Herman 2004). Therefore, we include in the regressions as explanatory variables county population size, percent minority in the state, and percent Hispanic in the tract of residence of the respondent, as well as the median household income in the tract.

One important limitation in this research is that our analysis is limited by our ability to link Medicaid participants to decennial census records and ACS surveys. That is, Medicaid participants who did not get an assigned PIK because of insufficient information on the database are not in the sample. In addition, Medicaid participants with a PIK that did not match to decennial census and ACS records are not in the analysis. Therefore, our findings apply to the Medicaid subsample in the analysis and may not be generalizable to the Medicaid population in general.

4. Findings

Overall, we find that race responses matched for 83.6 percent of the linked Medicaid-Census individuals in the sample, 3.7 percent had non-matching races, and 12.7 percent had missing race responses in Medicaid. Hispanic origin responses matched in 88.9 percent of the observations in the linked sample, 2.7 percent did not match, and 8.4 percent had missing Hispanic origin data in Medicaid.

Figures 1 and 2 show the large variations in missing and matching rates for race and Hispanic origin data by state. As shown in Figure 1, the percentage of cases with missing race data ranged from 47.6 percent and 42.5 percent in two states (Colorado and Massachusetts) to less than one percent in seven states (Connecticut, Idaho, Montana, North Dakota, Oklahoma, South Dakota and Utah). Non-matching race ranged from 11.7

percent in California and 10.2 percent in New York to less than one percent in four states (Iowa, Minnesota, Vermont and Wisconsin).

Figure 2 shows that three states (Alaska, Louisiana and Maryland) had over 90 percent missing Hispanic origin response data for the linked sample. For these states, the percentage of cases with missing Hispanic origin data were similarly high for each of the three years included in the study, 2006-2008. Non-matching Hispanic origin ranged from 10.7 percent of the Medicaid-Census linked cases in New Mexico and 9.8 percent in New Jersey, to less than one percent in 14 states.

The relatively high percentage of observations with matching race and Hispanic origin responses in Medicaid and Census data was not distributed evenly among the race groups in the analysis. Table 1 shows the extent of agreement in race (Part I) and Hispanic origin (Part II) between Medicaid records and Census data. The Black alone population has the highest match at 90.6 percent, followed by the White alone population at 83.8 percent. For those who reported White alone in Census data, only small percentages of responses were classified into other race groups, but a large share (14.9 percent) had missing race data in Medicaid. In contrast, only 5.3 percent of those who reported as Black alone in Census data were missing in Medicaid.

The matching race response rates were lower for the rest of the groups. Sixty-two percent of the AIAN alone population in Census data matched to Medicaid AIAN alone responses, 22.0 percent were classified as White alone, and 11.8 percent had missing race data in Medicaid. The finding that AIAN alone may be classified as White alone is consistent with research suggesting that single race AIAN sometimes change their race responses to White alone or White/AIAN (Doyle and Kao 2007).

Among individuals who reported as Asian alone in Census data, the match rate in Medicaid data was 60.4 percent, with 19.6 percent classified as NHPI alone, 6.6 percent classified as White alone, and 11.0 percent missing race data in Medicaid. Similarly, 53.5 percent of those who reported as NHPI alone in Census data were classified as NHPI alone in Medicaid, 19.3 percent were classified as Asian alone, 14.4 percent were classified as White alone, and 10.3 percent had missing race responses in Medicaid.

Part II in Table 1 shows the match for Hispanic origin responses in the linked Census and Medicaid sample. Non-Hispanics have a higher match at 90.5 percent compared to Hispanics at 72.9 percent. For Hispanics in Census data, 21.9 percent are classified as non-Hispanic in Medicaid. In contrast, less than one percent of non-Hispanics in Census were classified as Hispanic in Medicaid, but a higher percentage have missing Hispanic origin data (8.7 percent compared to 5.3 percent missing among Hispanics).

In Table 2 we show the results from the multinomial logistic regressions modeling the relationship between individual and community factors on whether a person's race and Hispanic origin responses match between Census and Medicaid. The coefficients in the race and Hispanic origin models show that controlling for demographic, socioeconomic and contextual factors, Medicaid race and Hispanic origin records are less likely to match Census responses for the single race racial minorities in the sample compared to White alone individuals. Black alone individuals are about three times more likely to have non-matching rather than matching race responses than White alone, although they are less likely to have missing race responses than White alone. In contrast, the odds of having non-matching race responses are around 33 and 21 times larger for AIAN alone and for

the combined Asian alone/NHPI alone group, respectively, than for White alone. In addition, the odds of missing race information in Medicaid are larger for AIAN than for White alone.

Racial minorities are also significantly more likely to have different Hispanic origin responses in Medicaid than in Census compared to White alone, and except for Asian/NHPI alone, they are more likely than White alone to have missing Hispanic origin data in Medicaid. Similarly, Hispanics in the sample are more likely than non-Hispanics to have both non-matching and missing race responses in Medicaid. They are also more likely to have non-matching Hispanic origin responses (i.e., to be classified as non-Hispanic), although they are less likely to have missing Hispanic origin data in Medicaid than non-Hispanics.

Individuals who speak another language at home, regardless of proficiency, are more likely to have missing race and missing Hispanic origin data in Medicaid than those who speak only English at home. However, we find differences in matching by English proficiency. Those who speak English not well or not at all are more likely to have matching race responses as well as matching Hispanic origin responses than English-only speakers. Those who speak English well or very well, are less likely than English-only speakers to have matching race responses but more likely to have matching Hispanic origin responses.

Consistent with findings from earlier research, individuals who are older, male, and more educated are more likely to have non-matching race and Hispanic origin responses in Medicaid compared to Census than younger, females and less educated individuals. Older individuals and individuals with a college degree are also more likely to have missing race and Hispanic origin data. Men are more likely to have missing race data, and women are more likely to have missing Hispanic origin data.

In terms of contextual factors, individuals living in counties with more than 500,000 people and in more affluent neighborhoods are more likely to have non-matching and missing race and Hispanic origin responses than those in counties with smaller populations or in neighborhoods with lower median household incomes. Similarly, individuals living in states where minorities make up 35 percent or more of the population are more likely to have non-matching race responses and non-matching or missing Hispanic origin responses, but less likely to have missing race responses than those in states with lower representation of minorities.

Individuals living in tracts with a large Hispanic presence are more likely to have different race responses in their Medicaid and Census records, and more likely to have missing race data in Medicaid, than those living in tracts where the concentration of Hispanics is lower. At the same time, living in Hispanic neighborhoods is associated with greater agreement of Hispanic origin responses in Medicaid and Census data, and a lower likelihood of missing Hispanic origin data in Medicaid.

5. Discussion and Recommendations

In this paper we have linked and compared race and Hispanic origin data of Medicaid participants nationwide with their responses in decennial censuses and the ACS. We have three main findings from this research: First, the percentage of individuals with matching and missing race and Hispanic origin data in the linked Medicaid-Census

sample varies substantially from state to state. This could be, in part, the result of differential PIK assignment. Still, for some states close to half of the race data in the linked sample are missing, and three states in our sample have over 90 percent of the Hispanic origin data missing. Second, despite differences in how states collected race and Hispanic origin data from Medicaid program participants, and the optional nature of reporting race and Hispanic origin on Medicaid application forms in some states, there is a surprisingly high percentage of matching race (84 percent) and Hispanic origin (89 percent) in the overall linked Medicaid-Census sample. Third, when we focus on the 16 percent of individuals in the linked sample with non-matching or missing race data, and the 11 percent with non-matching or missing Hispanic origin data in Medicaid, we find that these are disproportionately racial minorities, Hispanics, foreign-born, males and older individuals.

Policy changes in the coming years are expected to expand eligibility for Medicaid health insurance coverage as well as to increase requirements regarding data collection from program participants. Standardized forms to collect demographic information are likely to increase the completeness and accuracy of Medicaid data as well as the ability to conduct much needed inter-state analysis of disparities in health and health care.

Clearly, data documenting disparities among subpopulations is critical for policy makers to identify and target groups that may be in need of additional resources. However, some of the issues that we find regarding data quality do not have an obvious explanation and may remain undetected even with standardized data collection across states unless their causes are better understood.

Among Hispanics, we have argued that differences in question wording or racial category options may explain their higher rates of non-matching and missing race data in Medicaid compared to non-Hispanics. For example, Medicaid does not have the Some Other Race option, which 36.7 percent of Hispanics reported as their race in the 2010 Census (Humes et al. 2011). As a result, Hispanic program participants may choose a different race that will not match Census data or may not report a race. Future standardization of questions on race should consider strategies to reduce race item non-response in this population.

The higher rates of non-matching and missing race data in Medicaid among immigrants do not seem to be associated with English proficiency. Those who do not speak English well are more likely (not less) to have matching race responses than their more English proficient counterparts and English-only speakers. Research on racial fluidity suggests that part of the explanation may be that immigrants may change their racial identification as they come to identify with particular native groups (see for example Golash-Boza and Darity 2008). It is also possible that proxy responses and question wording influence their racial responses. Immigrants' higher likelihood of missing responses relative to natives may be partly due to having less familiarity with the U.S. system of racial categories. Further research would help identify the role that these and other issues may play in race and Hispanic origin item non-response among immigrants.

Proxy responses may also be associated with the higher rates of non-matching and missing race and Hispanic origin Medicaid data among males and older adults. For example, women may be more likely to provide information for Medicaid records about the race and Hispanic origin of other household members, and these responses may not match the other individuals' self-reports in Census forms.

It is challenging to explain non-matching and missing race and Hispanic origin data in Medicaid among racial minorities and more educated respondents. Would these groups be more likely to change their responses and/or skip these questions altogether? We propose that there is need to understand better the reasons for these patterns in non-matching and missing data in order to develop strategies to collect better quality data that will encompass the diverse Medicaid population.

While missing data is readily obvious, non-matching racial and ethnic identification is likely to go undetected unless records are linked to surveys, such as this study has done. Non-matching racial and Hispanic origin responses are a concern because they may result in misleading racial and ethnic disparities in health and health outcomes. For example, in this study, about 22 percent of individuals reporting as Hispanic in Census were classified as non-Hispanic in Medicaid. To the extent that Hispanics experience lower access to health care, computations of disparities would underestimate the differential between non-Hispanic and Hispanic populations.

Acknowledgments

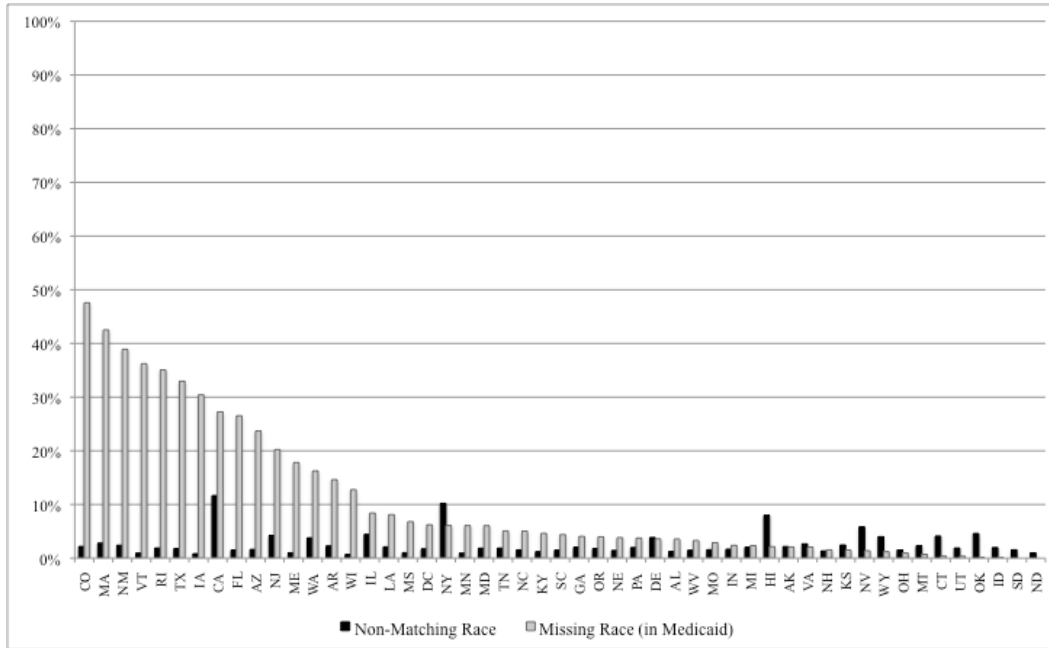
We would like to thank Amy O'Hara, John Iceland and Carolyn Liebler for their review and thoughtful comments in earlier versions of this paper.

References

- Andrulis, D.P., N.J. Siddiqui, J. Purtle and L. Duchon. 2010. *Patient Protection and Affordable Care Act of 2010: Advancing Health Equity for Racially and Ethnically Diverse Populations*. Washington, DC: Joint Center for Political and Economic Studies.
- Arday, S.L., D.R. Arday, S. Monroe and J. Zhang. 2000. HCFA's Racial and Ethnic Data: Current Accuracy and Recent Improvements. *Health Care Financing Review*, 21(4): 107-116.
- Brown, J.S., S. Hitlin and G.H. Elder, Jr. 2006. The Greater Complexity of Lived Race: An Extension of Harris and Sim. *Social Science Quarterly*, 87(2): 411-431.
- Campbell, M.E. and C.L. Rogalin. 2006. Categorical Imperatives: The Interaction of Latino and Racial Identification. *Social Science Quarterly*, 87(5): 1030-1052.
- Compton, E., M. Bentley, S. Rastogi and S. Ennis. 2012. 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment. DSSD 2010 CPEX Memorandum Series #B-05-R, 2010 Census Planning Memoranda Series #211, U.S. Census Bureau.
- Doyle, J.M. and G. Kao. 2007. Are Racial Identities of Multiracials Stable? Changing Self-Identification among Single and Multiple Race Individuals. *Social Psychology Quarterly*, 70(4): 405-423.
- Eschbach, K. and C. Gomez. 1998. Choosing Hispanic Identity: Ethnic Identity Switching among Respondents to High School and Beyond. *Social Science Quarterly*, 79(1): 74-90.
- Golash-Boza, T. and W. Darity Jr. 2008. Latino Racial Choices: The Effects of Skin Colour and Discrimination on Latinos' and Latinas' Racial Self-Identifications. *Ethnic and Racial Studies*, 31(5): 899-934.
- Gomez, S.L. and S.L. Glaser. 2006. Misclassification of Race/Ethnicity in a Population-Based Cancer Registry (United States). *Cancer Causes & Control*, 17(6): 771-781.
- Gomez, S.L., J.L. Kelsey, S.L. Glaser, M.M. Lee and S. Sidney. 2005. Inconsistencies between Self-Reported Ethnicity and Ethnicity Recorded in a Health Maintenance Organization. *Ann Epidemiol*, 15(1): 71-79.
- Harris, D.R. and J.J. Sim. 2002. Who is Multiracial? Assessing the Complexity of Lived Race. *American Sociological Review*, 67(4): 614-627.
- Herman, M. 2004. Forced to Choose: Determinants of Racial Identification in Multiracial Adolescents. *Child Development*, 75(3): 730-748.

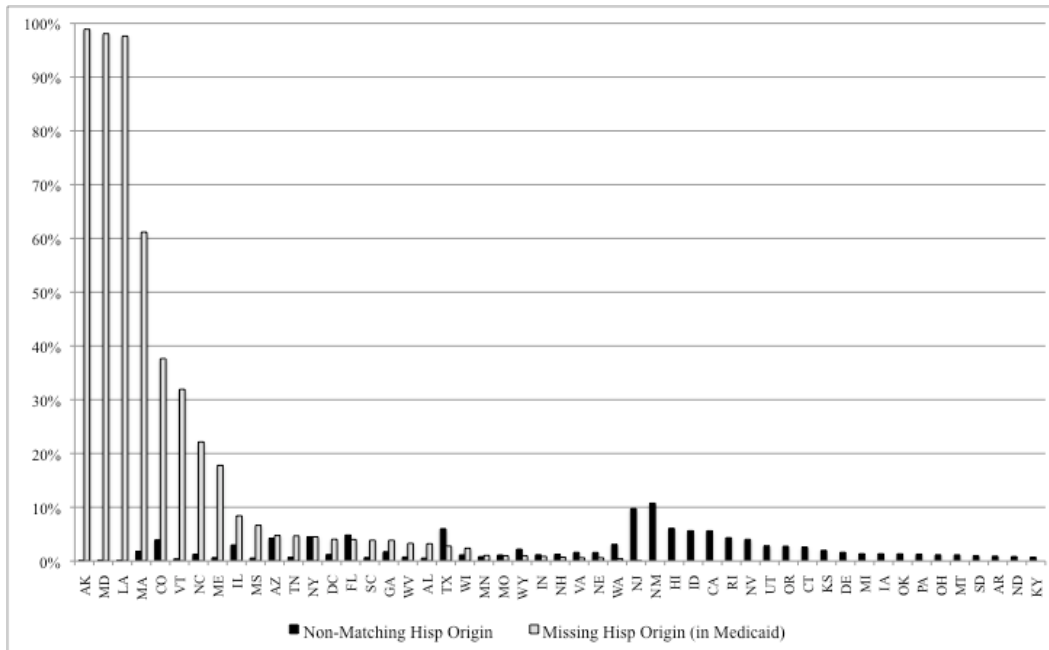
- Humes, K.R., N.A. Jones and R.R. Ramirez. 2011. Overview of Race and Hispanic Origin: 2010. U.S. Department of Commerce, U.S. Census Bureau: 2010 Census Briefs. [Accessed March 20, 2013]. Available from <http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>.
- Institute of Medicine. 2004. *Eliminating Health Disparities: Measurement and Data Needs*. Washington, DC: National Academies Press.
- Kawachi, I., N. Daniels and D.E. Robinson. 2005. Health Disparities By Race And Class: Why Both Matter. *Health Affairs*, 24(2): 343-352.
- Kressin, N.R., B.H. Chang, A. Hendricks and L.E. Kazis. 2003. Agreement between Administrative Data and Patients' Self-Reports of Race/Ethnicity. *American Journal of Public Health*, 93(10): 1734-1739.
- McAlpine, D.D., T.J. Beebe, M. Davern, and K. T. Call. 2007. Agreement between Self-Reported and Administrative Race and Ethnicity Data among Medicaid Enrollees in Minnesota. *Health Services Research*, 42(6): 2373-2388.
- Office of Minority Health. 2013. Final Data Collection Standards for Race, Ethnicity, Primary Language, Sex, and Disability Status Required by Section 4302 of the Affordable Care Act. [Accessed September 18, 2013]. Available from <http://minorityhealth.hhs.gov/templates/browse.aspx?lvl=2&lvlid=208>
- Rodriguez, C. 2000. *Changing Race: Latinos, the Census, and the History of Ethnicity in the United States*. New York: New York University Press.
- Saperstein, A. and A.M. Penner. 2012. Racial Fluidity and Inequality in the United States. *American Journal of Sociology*, 118(3): 676-727.
- Sebelius, K. 2011. Approaches for Identifying, Collecting, and Evaluating Data on Health Care Disparities in Medicaid and CHIP. U.S. Department of Health and Human Services. Report to Congress. [Accessed March 20, 2013]. Available from <http://www.healthcare.gov/news/reports/disparities09292011a.pdf>
- Smedley, B., A. Stith and A. Nelson. (eds). 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: Institute of Medicine.
- Wagner, D. and M. Layne. 2012. The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software. *Center for Administrative Records Research and Applications Internal Paper, U.S. Census Bureau*.
- Zaslavsky, A.M., J.Z. Ayanian, and L.B. Zaborski. 2012. The Validity of Race and Ethnicity in Enrollment Data for Medicare Beneficiaries. *Health Services Research*, 47(3): 1300-1321.

Figure 1. Percent of Medicaid Participants in the Sample with Missing or Non-Matching Race Data by State*



* Cells with small counts have been suppressed.
 Source: Authors' computations, MSIS 2006-2008, Census 2000, 2010 Census, ACS 2001-2009.

Figure 2. Percent of Medicaid Participants in the Sample with Missing or Non-Matching Hispanic Origin Data by State*



* Cells with small counts have been suppressed.

Source: Authors' computations, MSIS 2006-2008, Census 2000, 2010 Census, ACS 2001-2009.

Table 1. Comparison of Race and Hispanic Origin Responses in Medicaid and Census, Ages 25 and Older*							
Number and percent matched	I. Medicaid Race Records						
Census Race Response	White alone	Black alone	AIAN alone	Asian alone	NHPI alone	Missing	Total
White alone	890,853	8,233	2,155	2,305	1,277	157,825	1,062,648
(%)	83.8%	0.8%	0.2%	0.2%	0.1%	14.9%	100.0%
Black alone	10,881	265,818	360	587	309	15,408	293,363
(%)	3.7%	90.6%	0.1%	0.2%	0.1%	5.3%	100.0%
AIAN alone	6,063	1,009	17,114	100	52	3,253	27,591
(%)	22.0%	3.7%	62.0%	0.4%	0.2%	11.8%	100.0%
Asian alone	4,480	1,039	589	40,994	13,297	7,442	67,841
(%)	6.6%	1.5%	0.9%	60.4%	19.6%	11.0%	100.0%
NHPI alone	379	51	14	505	1,404	270	2,623
(%)	14.4%	1.9%	0.5%	19.3%	53.5%	10.3%	100.0%
Total	912,656	276,150	20,232	44,491	16,339	184,198	1,454,066
(%)	62.8%	19.0%	1.4%	3.1%	1.1%	12.7%	100.0%
Number and percent matched	II. Medicaid Hispanic Origin Records						
Census Hispanic Origin Response	Hispanic		Not Hispanic		Missing		Total
Hispanic	98,433		29,536		7,095		135,064
(%)	72.9%		21.9%		5.3%		100.0%
Not Hispanic	10,180		1,194,260		114,562		1,319,002
(%)	0.8%		90.5%		8.7%		100.0%
Total	108,613		1,223,796		121,657		1,454,066
(%)	7.5%		84.2%		8.4%		100.0%
* Excludes Some Other Race, Two or More Races, and Census observations with missing race or Hispanic origin responses.							
Source: MSIS 2006-2008, Census 2000, 2010 Census, and 2001 through 2009 ACS 1-year data.							

TABLE 2. Multinomial Logistic Regressions Predicting the Outcome in Medicaid-Census Race and Hispanic Origin Comparisons, Odds Ratios (OR)				
Variables in the Analysis	Model 1: Race		Model 2: Hispanic Origin	
	Non-matching race	Missing race	Non-matching Hispanic origin	Missing Hispanic origin
	(Matching race is the reference)		(Matching Hispanic origin is the reference)	
	OR	OR	OR	OR
Demographic Characteristics				
Age				
25-44 (omitted)				
45-64 years old	1.20***	1.20***	1.13***	1.28***
65 years and older	1.52***	1.51***	1.41***	1.48***
Gender				
Male (omitted)				
Female	0.87***	0.96***	0.97**	1.03***
Marital Status				
Married (omitted)				
Never married	1.02	0.89***	1.33***	1.21***
Separated, divorced or widow(er)	1.03*	0.83***	1.19***	1.03***
Race (as reported in Census 2010)				
White alone (omitted)				
Black alone	3.01***	0.59***	1.46***	1.50***
American Indian or Alaska Native alone	33.16***	1.41***	1.96***	1.56***
Asian alone or Native Hawaiian and Other Pacific Islander alone	20.63***	0.91***	4.04***	0.65***
Ethnicity				
Not Hispanic or Latino (omitted)				
Hispanic or Latino	6.11***	17.01***	79.45***	0.92***
Socio-Economic Characteristics & Language Proficiency				
Education				
No high school (omitted)				
High school diploma	1.04***	0.99	1.07***	0.91***
Some college	1.08***	1.00	1.20***	0.84***
College degree or higher	1.38***	1.42***	1.34***	1.08***
Personal income (log)	1.00**	1.01***	1.00	1.01***
Language Proficiency				
Speaks English only (omitted)				
English very well or well, other language at home	1.24***	1.82***	0.53***	1.34***
English not well or not at all, other language at home	0.94***	1.85***	0.39***	1.26***
Contextual Factors				
County population				
Over 500,000 pop (omitted)				
100,000 to 500,000 pop	0.68***	0.86***	0.97*	1.10***
Less than 100,000 pop	0.53***	0.72***	0.83***	0.80***
Percent minority in the state				
Less than 35 percent (omitted)				
35 percent or higher	1.26***	0.66***	1.21***	1.88***
Percent population Hispanic in tract				
Less than 10 percent (omitted)				
10 to less than 25 percent	1.38***	1.19***	1.11***	0.66***
25 percent or higher	1.78***	2.04***	0.67***	0.37***
Median household income in tract (log)	1.19***	1.42***	1.11***	1.18***
N	1,454,066		1,454,066	
*p<.05, **p<.01, ***p<.001				
Source: MSIS 2006-2008, Census 2000, 2010 Census, 2001 through 2009 ACS 1-year data, and 2006-2010 ACS 5-year data.				