CARRA Working Paper Series

Working Paper *#2015-01*

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

Leticia E. Fernandez
U.S. Census Bureau

Sonya Rastogi
U.S. Census Bureau

Sharon R. Ennis
U.S. Census Bureau

James M. Noon
U.S. Census Bureau

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, D.C. 20233

Paper Issued: April 20, 2015

Evaluating Race and Hispanic Origin Responses of Medicaid Participants Using Census Data[1]

Leticia E. Fernandez, U.S. Census Bureau
Sonya Rastogi, U.S. Census Bureau
Sharon R. Ennis, U.S. Census Bureau
James M. Noon, U.S. Census Bureau

Abstract

Health and health care disparities associated with race or Hispanic origin are complex and continue to challenge researchers and policy makers. With the intention of improving the measurement and monitoring of these disparities, provisions of the Patient Protection and Affordable Care Act (ACA) of 2010 require states 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 variables. 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 Medicaid populations.

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

[1] An earlier version of this paper was presented at the Joint Statistical Meetings in August 2013 and is available in the 2013 JSM 2013 Proceedings, Social Statistics Section, Alexandria, VA: American Statistical Association.

I. Introduction

Although well documented, health and health care disparities associated with race or Hispanic origin 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 hindered measurement and monitoring of these disparities across states (Institute of Medicine 2004). In the case of Medicaid administrative data in particular, states lacked a standardized protocol to collect race or Hispanic origin data in systematic and comparable ways (McAlpine et al. 2007; Sebelius 2011).

Provisions of the Patient Protection and Affordable Care Act of 2010 (Section 4302, ACA) require 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).

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 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.

Our analysis contributes to a better understanding of the extent and factors associated with missing and inconsistent data in Medicaid when compared to Census/ACS data. By identifying the groups most at risk of missing and inconsistent race or Hispanic origin in Medicaid, our findings have the potential to inform strategies to improve the quality of data collection infrastructures, which may facilitate monitoring disparities in compliance with ACA provisions. Linked records also have the potential to supplement information needed to better assess racial and Hispanic origin disparities.

II. Prior Research

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 online and paper applications that 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 individual-level race and Hispanic origin between Medicaid administrative records and Census, and item non-response.

Non-matching Responses

Earlier studies comparing the agreement of race and Hispanic origin data between surveys and administrative records find discrepancies to vary by 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 comparing 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 2011).

Studies that have looked at the impact of changes in question format find that Hispanics are less likely to choose a race in addition to choosing the Hispanic category when they are asked a combined question that lists Hispanic as a racial option. In contrast, when race and Hispanic origin are asked in separate questions, Hispanics may choose a race in addition to reporting as Hispanic, but race item non-response tends to be higher, which suggests that some Hispanics view their ethnicity as their race (Campbell and Rogalin 2006; Compton et al. 2012).

The census race and Hispanic origin data used in this study come 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. Appendix 1 shows the different wordings of race and Hispanic origin questions in the data sources that are compared in this study.

Proxy Respondents – Although its impact has not been assessed, another important source of discrepancies is proxy responses, individuals reporting another person's race or Hispanic origin. For example, other household members may have completed the decennial census, and their answers may not match those of self-reporting individuals in Medicaid.

In the case of administrative records, even if instructions direct field staff and administrators to collect race and Hispanic origin self-responses, in some instances they may fill out this information on behalf of the respondent. In a study comparing self-reported race and Hispanic origin to cancer registry records, Gomez and Glaser (2006) point out 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 Hispanic origin of Medicaid applicants based on verbal reports or assist applicants to determine their race and Hispanic origin responses (Sebelius 2011).

Changes in how individuals identify racially/ethnically (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 and White or Black populations, respectively (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 response consistency.

Item Non-Response

There is not an extensive literature on the characteristics of those who do not respond to race and Hispanic origin questions. Some researchers 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, as mentioned above, Hispanics have higher non-response rates for race when they are asked a question on Hispanic origin separately from a race question rather than in a combined question that lists both race and Hispanic origin as options (Compton et al. 2012; Martin 1990). 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).

In this study we provide estimates of the relative contributions that missing and non-matching responses for race and Hispanic origin make to the mismatch between Medicaid administrative records and Census and survey data.

III. Data and Methods

The data used in this study come from linking at the individual level Medicaid administrative records for 2006-2008, Census 2000, 2010 Census, ACS 2001 through 2009 one-year data, and ACS 2006-2010 five-year data. Individuals are linked across these sources using a unique person identifier internal to the Census Bureau called the Protected Identification Key (PIK). To assign these identifiers, the Census Bureau uses the Person Identification Validation System (PVS), which employs probabilistic matching techniques. Personally identifiable information is then removed from the records to anonymize the data and preserve confidentiality so it can be used for statistical purposes and research (see Wagner and Layne 2014 for more information).

Medicaid Records - In the first step of the linking process in this study, PIKs were assigned to Medicaid records based on date of birth and Social Security Number. Overall, around 93 percent of the Medicaid records in each of the three years in this study received a PIK. This allowed us to unduplicate the records and combine them into a single file that kept only the most updated race and Hispanic origin data for each person. Appendix 2 shows that there are differences in the percent assigned a PIK by race, Hispanic origin, age group and state of residence. Observations with missing race information were less likely to receive a PIK than those with a race response. The PIK rate for Hispanics was lower (around 85 percent) than for non-Hispanics (96 percent). Individuals age 45 and older were more likely than younger ones to receive a PIK (98 percent and 91 to 92 percent, respectively), and most observations with missing age (date of birth) were not assigned a PIK. Rates also varied by state, ranging from 78.8 percent of the Medicaid observations from Montana in 2008 getting a PIK to 99.7 percent of the observations in West Virginia in 2006.

Census/ACS multiple years - In the next step, the resulting unduplicated Medicaid file that combines three years of program participants who received a PIK (67,051,122 individuals) was linked to Census/ACS data. In this step, we linked as many of the unduplicated Medicaid records as possible to a large combined Census/ACS file that contained the most recent unedited and unduplicated race and Hispanic origin responses from a household for individuals ever responding to Census 2000, the 2010 Census or the ACS 2001 through 2009. These are the race and Hispanic origin responses that will be compared to the Medicaid records.

We linked 90 percent of the PIKed Medicaid 2006-2008 records to a record in Census/ACS (60,312,876 individuals). Appendix 3 shows that the percent that was linked varied by race, Hispanic origin, age group and state of residence of Medicaid participants. White alone individuals were more likely to be linked to a Census/ACS record (92.4 percent) than any other racial group, especially compared to Asian, Native Hawaiian or Other Pacific Islanders (NHPI), and Two or More Races (around 86 percent). Hispanics in the Medicaid records were less likely than non-Hispanics to be linked (87 percent and 91 percent, respectively). Individuals aged 25 to 44 were more likely than their counterparts to be linked to Census/ACS (93 percent compared to 91 percent or lower for other age groups); and only about 60 percent of the individuals with

missing age (date of birth) in Medicaid records were found in a Census/ACS record. There were also variations by state, ranging from 85 percent of the PIKed Medicaid records from Hawaii to 95 percent of the PIKed records from Vermont linking to a Census/ACS record.

2006-2010 ACS five-year data - For the last step, in order to bring into the analysis more recent and detailed individual-level characteristics (other than race and Hispanic origin, which were obtained from the previous linkage), we linked the Medicaid records to the ACS five-year file for 2006-2010. This allows us to examine the role of demographic, socioeconomic and contextual characteristics on the likelihood of matching race and Hispanic origin between Medicaid and Census records. The ACS is an ongoing survey that samples around 3.5 million household addresses annually. The five-year file we use in this study contains information for 22.6 million individuals. However, about 8 percent of the ACS observations in this file do not have a PIK that would allow linking to Medicaid records. Since previous research has reported that un-PIKed records differ from those assigned a PIK in their characteristics, we adjusted the ACS sample weights to account for observations left out of the MSIS-ACS due to not having a PIK (Bond et al. 2014; Rastogi and O'Hara 2012; Luque and Bhaskar 2014).

In all, about 5 percent of PIKed Medicaid records were linked to the PIKed 2006-2010 ACS file. The last column in Appendix 3 shows the percent linked by various characteristics in the Medicaid data. White alone, American Indian or Alaska Native (AIAN) alone, non-Hispanics, and individuals ages 65 and older were more likely to be linked to the five-year ACS file than their counterparts. These are unweighted linked percentages. In the Findings Section we discuss in more detail the weighted distribution of the linked observations.

Methods - In our analysis we first describe the consistency of race and Hispanic origin data between Medicaid and Census using the full merge between these files (60.3 million). In addition, we present inter-state variations in the extent of matching and missing race and Hispanic origin data. Next, using the sample linked to ACS for the years 2006-2010, we use multinomial logistic regressions to model the role of individual and contextual variables on matching responses. We have two dependent variables, 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; (b) different race (Hispanic origin) responses, and (c) missing race (Hispanic origin) response in Medicaid.

The explanatory variables included in the regression models have been associated with discrepancies in race and Hispanic origin responses in previous studies that compare survey and administrative data at a smaller scale: age, gender, marital status, race, Hispanic origin, nativity (U.S. or foreign born), education and English proficiency (Arday et al. 2000; Kressin et al. 2003; Gomez et al. 2005; McAlpine et al. 2007; Zaslavsky et al. 2012). Several of these variables have also been identified in ACA provisions as important to collect because of their relevance to measures of healthcare disparities (Sebelius 2011; Office of Minority Health 2013).

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.

The sample used for the regression analysis is further restricted to individuals ages 25 and older in order to include completed education, marital status and income variables. We also excluded individuals who reported their race as Some Other Race or Two or More Races because of no or negligible response matching in Medicaid records. Not all states allow Medicaid participants to report more than one race, such that only 0.3 percent of Medicaid participants overall were reported as Two or More Races in each year of Medicaid included in our analysis (see Table 1). Also, unlike Census, the Medicaid data does not include a category for Some Other Race. Prior to their exclusion, the Census/ACS race variable in the linked sample showed 4.4 percent of individuals reporting as Two or More Races and 8.2 percent reporting as Some Other Race (see Table 2). In addition, we excluded observations that have missing race and Hispanic origin responses in Census/ACS.

The sample with all its exclusions is about five percent of the Medicaid records for individuals age 25 and older (unweighted observations 1,451,169; weighted count 22,751,562). The race groups included in the analysis are White alone, Black alone, AIAN alone, Asian alone, and NHPI alone. Due to small sample sizes, Asian and NHPI were combined for the analysis. We present results from weighted regressions. We also ran our regressions without using weights and generally found that the patterns were similar to the weighted analyses.[2]

Limitations – Since not every Medicaid participant was assigned a PIK or could be linked to their census or ACS records, our findings should be taken with caution. Although we did reweight the ACS data to compensate for un-linkable observations, we did not correct for missing PIKs in the Medicaid data, which is likely to result in some unexamined biases in our results. In addition, Medicaid participants with a PIK that did not match to decennial census and ACS records are not in the analysis and are likely to differ from those in the analysis. Therefore, our findings apply to the Medicaid subsample in the analysis but may not be generalizable to the national Medicaid population.

[2] The few exceptions are as follows: (a) In the model predicting non-matching rather than matching race responses the coefficient for individuals who speak English very well or well changes direction from 1.11 ($p<.001$) in the weighted regression to 0.96 ($p<.01$) in the unweighted regression, so that they are *less* likely to have non-matching race responses compared to English only speakers. Also, the coefficient for personal income is statistically significant in the weighted regression but not in the unweighted regression. (b) In the model predicting missing rather than matching race responses, the coefficient for some college in the weighted regression is 0.95 and statistically significant ($p<.001$), while in the unweighted regression this coefficient is 0.99 and not statistically significant. (c) In the model predicting non-matching rather than matching Hispanic origin, the coefficient for population county size 100,000 to 500,000 is 0.98 and not significant in the weighted regression, but this coefficient becomes 0.97 and significant ($p<.05$) in the unweighted regression. (d) In the model predicting missing rather than matching Hispanic origin, the coefficient for college or higher education changes from 0.98 ($p=.21$) in the weighted regression to 1.06 ($p<.001$), so that in the unweighted regression individuals who complete at least college are more likely to have missing Hispanic origin data in Medicaid compared to less educated individuals. In the weighted model, their odds do not differ from those of individuals with no high school.

IV. Findings

We first discuss the distributions for race and Hispanic origin in the Medicaid records, and how these change as subsets of data are linked across Census/ACS data. Next we discuss the specific match by race and Hispanic origin, as well as inter-state variations in missing and matching Medicaid records as compared to decennial data.

The first three columns in Table 1 show the race and Hispanic origin distribution of Medicaid participants in each year in the study, 2006 through 2008; their distribution after the three years of data are combined and only individuals assigned a PIK are retained (fourth column in Table 1); the distribution in the Medicaid-Census linked sample (fifth column in Table 1); and the distribution in the sample linking Medicaid records to ACS data (last column in Table 1). These distributions are very similar for all ages as well as for individuals ages 25 and older, which may be a reflection of the high percent of Medicaid records assigned a PIK (93 percent) as well as the high percent that were linked to Census data (90 percent).

In the original files, which include records with and without PIKs, 44 percent are White alone, about 24 percent are Black alone, 1.5 percent are AIAN, 3 percent are Asian, 1 percent are NHPI, one-third of a percent are Two or More Races, and race is missing for 26 percent of the Medicaid records. In terms of Hispanic origin, around 70 percent are non-Hispanic, 23 percent are Hispanic, and 7 to 9 percent have missing Hispanic origin data.

Taking only the Medicaid records that were assigned a PIK, unduplicated and then combined into a single file show that particular groups are less likely to have been assigned a PIK, such that compared to the first three columns the combined sample of Medicaid records has a higher percent White alone (48 percent), a smaller percent with missing race data (23 percent), and fewer Hispanics (21 percent).

Ninety percent of the Medicaid records with a PIK were found in the combined Census/ACS data (Census 2000, Census 2010, and ACS for years 2001 through 2009), which resulted in further changes in the race and Hispanic origin characteristics of the Medicaid linked sample. The Medicaid-Census sample shows that the White alone population increased to 49 percent, those missing a race response declined to 22 percent, and Hispanics declined to 20 percent, but other groups retained a similar presence to the pre-Census linkage. Finally, only about 5 percent of the linked Medicaid-Census records were found in the reweighted 2006-2010 ACS file. However, the weighted distribution is very similar to the unduplicated Medicaid file that combines three years of data.

In Table 2 and Tables 3, we examine the match in race and Hispanic origin data between Medicaid and Census records in the linked file (60.3 million individuals). As mentioned above, this file contains 90 percent of the PIKed records in Medicaid for the years in the study and their corresponding response in Census/ACS data. Table 2 shows individuals' race reported in Medicaid along the columns, and the race response in Census along the rows. The diagonal shows the percent agreement for each race. A cursory look shows that Black alone and White alone have the highest consistency between the two data sources, while other racial groups have

lower rates of matching answers. Among White alone in Census, 78.1 percent have the same race in Medicaid, and 20.3 percent have missing race information.

Individuals reporting as Black alone have the highest match with Medicaid records, at 91.0 percent. Only 5.0 percent of Black alone have missing race in Medicaid. Among those reporting as AIAN in Census, 58.0 percent had the same race in Medicaid, 20.8 percent were listed as White alone, and 15.5 percent had missing race. Asian alone and NHPI alone had matching rates of 63.6 percent and 51.3 percent, respectively. However, 15.2 percent of Asian in Census are listed as NHPI in Medicaid, and 18.8 percent of NHPI in Census are listed as Asian in Medicaid, which suggests the need for further research on whether these inconsistencies arise at the individual or at the administrative level. Among Asian, 6.8 percent are reported as White alone in Medicaid and 10.7 percent have missing race. Among NHPI, 13.0 percent are listed as White alone in Medicaid and another 13.4 percent have missing race.

Unlike Census, Medicaid does not have a Some Other Race category, and most reporting this race in Census, 74.7 percent, are missing a race response in Medicaid rather than coded as a different racial category. Twenty percent are listed as White alone, 3.6 percent as Black alone, and the rest are distributed among AIAN, Asian, NHPI and Two or More Races.

The lowest match comes from comparing Medicaid data for individuals who reported as Two or More Races in Census, 1.9 percent. Even though individuals may report as many races as they want according to the MSIS database instructions, the application forms for Medicaid may not indicate this or allow multiple racial responses. Most individuals who reported as Two or More Races in Census were classified in Medicaid records as White alone (42.2 percent), Black alone (21.3 percent) or Missing race (23.8 percent).

In terms of matching Hispanic origin records, Table 3 shows that 78.3 percent of the individuals who reported as Hispanic in Census were also Hispanic in Medicaid records. About 18 percent were recorded as non-Hispanic and 4 percent had missing Hispanic origin data. Non-Hispanics had a higher match at 89.7 percent, with 1.4 percent reported in Medicaid as Hispanic, and 8.9 percent missing Hispanic origin data.

Overall, the match for race is lower than for Hispanic origin. Sixty-seven percent of the records compared between Medicaid and Census have the same non-missing race response, 8.7 percent have a different (non-missing) race response, 20.2 percent have missing race in Medicaid but not in Census, 2.1 percent have missing race in both Medicaid and Census, and another 2.1 percent have missing race in Census but a non-missing race response in Medicaid. Including records in which both Medicaid and Census report missing race increases the match rate to 69 percent.

In contrast, 83.2 percent of Medicaid records report the same Hispanic origin as in Census, 5.0 percent do not match Hispanic origin, 7.4 percent have missing Hispanic origin in Medicaid but not in Census, 0.4 percent have missing data in both Medicaid and Census, and 4.0 percent have missing Hispanic origin in Census but not in Medicaid. Including records in which both Medicaid and Census have missing Hispanic origin increases the match rate to 84 percent.

Focusing only on the Medicaid records with missing race data for which there is a non-missing Census response (12.2 million or 21 percent in the linked sample), most are White alone (55 percent), followed by SOR (31 percent). Among the Medicaid records with missing Hispanic origin response for which there is a non-missing Census response (4.5 million or 7.4 percent of the linked sample), most are non-Hispanic (87 percent).

Next we turn to inter-state variations in non-matching and missing rates for race and Hispanic origin Medicaid data. As shown in Figure 1, the percentage of Medicaid cases with missing race data ranged from a high of around 55 percent (in New Mexico, Colorado, Texas and California) to less than one percent. The percentage of Medicaid records with non-matching race ranged from a high of 40 percent (in Hawaii) to around three percent in several states.

Figure 2 shows the inter-state variation in non-matching and missing Hispanic origin. Three states in the linked sample (Louisiana, Alaska and Maryland) had 90 percent or more of their cases with missing Hispanic origin. 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 around 18 percent (New Jersey and Idaho) to one percent or less in 10 states.

Next, Table 4 shows the results from the weighted multinomial logistic regressions modeling the relationship between individual and community factors on whether a person's race and Hispanic origin match between Census and Medicaid. As a reminder, for this part of the analysis we use the subset of Medicaid records that were linked to the 2006-2010 ACS data (about 5 percent of Medicaid records). 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 single-race minorities in the sample compared to White alone individuals. Black alone individuals are about three times more likely than White alone to have non-matching rather than matching race responses, although they are less likely to have missing race responses. In contrast, the odds of having non-matching race responses are around 34 and 14 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 non-matching Hispanic origin responses between Medicaid and 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.

Foreign-born individuals are more likely to have both non-matching and missing race responses than native born persons, but they are more likely to have the same Hispanic origin response in Medicaid and Census compared to their native born counterparts.

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 (Gomez et al. 2005; McAlpine et al. 2007). Older individuals are also more likely to have missing race and Hispanic origin data than those aged 25 to 44. Individuals with a college degree are also more likely to have missing race, but not more likely to have missing Hispanic origin than those with no high school. Men are more likely to have missing race data, and women are more likely to have missing Hispanic origin data.

Individuals who speak a language other than English 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.

In terms of contextual factors, in general 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.

V. 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 or missing race and Hispanic origin data in the linked Medicaid-Census sample varies substantially from state to state. Given that we are able to match 90 percent of the Medicaid cases with a PIK to Census/ACS data, this is not likely to be due only to PIK assignment. In some states, the race data of a quarter or more of their program participants in the linked file 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 relatively high percentage of matching race (67

percent) and Hispanic origin (83 percent) in the linked Medicaid-Census sample. Third, net of other demographic and socioeconomic factors, we find that racial minorities are more likely to have non-matching race and Hispanic origin data, and in some cases also missing race and Hispanic origin data. Hispanics and immigrants are more likely to have both non-matching and missing race data than non-Hispanic and native-born populations.

Policy changes in the coming years will expand eligibility for Medicaid health insurance coverage as well as increase requirements regarding data collection from program participants for better assessment and monitoring of program outcomes. The standards for data collection on race and Hispanic origin set forth by the Office of Management and Budget (OMB) include a minimum of five racial categories: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White (OMB 1997). The data on ethnicity has two categories: Hispanic or Latino and Not Hispanic or Latino. OMB encourages the collection of more detailed race and Hispanic origin data provided these could be aggregated back to one of the five minimum categories. The Department of Health and Human Services (HHS) has recently published final standards for data collection in accordance with Section 4302 of the Affordable Care Act (Office of Minority Health 2013) that include these OMB race and Hispanic origin categories in addition to sex, primary language and disability status. The law requires the use of these standards to the extent practicable when self-reported data are collected in national population health surveys sponsored by HHS. Use of standardized forms to collect demographic information is necessary to increase the completeness and accuracy of data that will facilitate much needed intra-state comparisons of disparities in health and health care. The Institute of Medicine (2009) also recommends the development and implementation of a standard form to collect race, ethnicity, granular ethnicity (ethnic origin or ancestry) and language as a way to identify, address and monitor disparities in all health and health care data systems.

Clearly, improving the ability to document and monitor disparities among subpopulations is critical for policy makers in order to target groups that may be in need of additional resources or interventions. However, there are additional issues that if not examined further may hinder data quality and may go undetected even with the implementation of standardized data collection forms across states. Specifically, we suspect that differences in question wording or racial category options may contribute to Hispanics higher rates of non-matching and missing race data in Medicaid compared to the rates among non-Hispanics. Since Medicaid does not have the Some Other Race option, which 36.7 percent of Hispanics reported as their race in the 2010 Census, Hispanic in the Medicaid program may choose a different race that will not match their Census response or may not report a race at all. Future standardization of questions on race should consider adding the Some Other race category, as well as using other strategies to reduce race item non-response in this population (Humes et al. 2011; Institute of Medicine 2009).

In addition, the higher rates of non-matching and missing race data in Medicaid among the foreign born remain even when controlling for English proficiency. Research on racial fluidity suggests that part of the explanation may be that foreign-born individuals 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 responses from others in the household and question wording influence their racial responses. Immigrants' higher likelihood of missing race

responses relative to natives may be partly due to having less familiarity with the U.S. system of racial categories. Further research would also help to identify the role that these and other issues play in race item non-response among immigrants. Similarly, it is challenging to explain the higher rates of non-matching and missing race and Hispanic origin data in Medicaid among racial minorities and more educated respondents compared to White alone and individuals with lower levels of education. We propose that there is need to understand these patterns in order to develop strategies to collect better quality data that will encompasses the diverse Medicaid population.

Finally, while the extent of missing data can be directly measured, non-matching race and Hispanic origin may go undetected unless records are linked to alternative data sources such as surveys. Thus, it would be useful to conduct with some regularity surveys for consistency checks. Non-matching racial and Hispanic origin responses are a concern because they may result in misleading measures of racial and Hispanic origin disparities in health and health outcomes. For example, in this study, about 18 percent of individuals reporting as Hispanic in Census were classified as non-Hispanic in Medicaid, and another 4 percent were listed as missing Hispanic origin data. To the extent that Hispanics experience lower access to health care, computations of disparities would underestimate, and may even mask, the differential between non-Hispanic and Hispanic populations.

Acknowledgments

We would like to thank Renuka Bhaskar, Kathleen Call, John Iceland, Carolyn Liebler, Amy O'Hara, and Jessie Kemmick Pintor for their review and thoughtful comments on earlier versions of this paper.

References

American Community Survey Research Notes (ACSRN). 2008. Changes to the American Community Survey Between 2007 and 2008 and Their Potential Effect on Estimates of Hispanic Origin Type, Nativity, Race, and Language. Available from https://www.census.gov/population/hispanic/files/acs08researchnote.pdf

Andrulis, Dennis P., Nadia J. Siddiqui, Jonathan P. Purtle, and Lisa Duchon. 2010. "Patient Protection and Affordable Care Act of 2010: Advancing Health Equity for Racially and Ethnically Diverse Populations." Washington D.C.: Joint Center for Political and Economic Studies.

Arday, Susan L., David R. Arday, Stephanie Monroe and Jianyi Zhang. 2000. HCFA's Racial an Ethnic Data: Current Accuracy and Recent Improvements. *Health Care Financing Review*, 21(4): 107-116.

Bond, Brittany, J. David Brown, Adela Luque and Amy O'Hara. 2014. "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey." *CARRA Working Paper Series* #2014-08. U.S. Census Bureau. Available from https://www.census.gov/srd/carra/The_Nature_of_the_Bias_When_Studying_Only_Linkable_Person_Records.pdf

Brown, J. Scott, Steven Hitlin and Glen H. Elder, Jr. 2006. The Greater Complexity of Lived Race: An Extension of Harris and Sim. *Social Science Quarterly*, 87(2): 411-431.

Campbell, Mary E. and Christabel L. Rogalin. 2006. Categorical Imperatives: The Interaction of Latino and Racial Identification. *Social Science Quarterly*, 87(5): 1030-1052.

Centers for Medicare and Medicaid Services (CMS). 2010. Medicaid and CHIP Statistical Information System (MSIS) File Specifications and Data Dictionary. Release 3.1 (MSISD7-2010.docx). Center for Medicaid, CHIP and Survey and Certifications, and Office of Information Services. Available from https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MSIS/downloads/msisdd2010.pdf

Chestnut, John. 2008. "Effects of Using a Grid versus a Sequential Form on the ACS Basic Demographic Data." Final Report. U.S. Census Bureau, Washington, DC.

Compton, Elizabeth, Michael Bentley, Sharon Ennis and Sonya Rastogi. 2012. 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment. 2010 Census Planning Memoranda Series #211, U.S. Census Bureau, Washington, DC.

Eschbach, Karl and Christina Gomez. 1998. Choosing Hispanic Identity: Ethnic Identity Switching among Respondents to High School and Beyond. *Social Science Quarterly*, 79(1): 74-90.

Golash-Boza, Tanya and William 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, Scarlett L. and Sally L. Glaser. 2006. Misclassification of Race/Ethnicity in a Population-Based Cancer Registry (United States). *Cancer Causes & Control*, 17(6): 771-781.

Gomez, Scarlett L., Jennifer L. Kelsey, Sally L. Glaser, Marion M. Lee and Stephen Sidney. 2005. Inconsistencies between Self-Reported Ethnicity and Ethnicity Recorded in a Health Maintenance Organization. *Ann Epidemiol,* 15(1): 71-79.

Harris, David R. and Jeremiah Joseph Sim. 2002. Who is Multiracial? Assessing the Complexity of Lived Race. *American Sociological Review*, 67(4): 614-627.

Herman, Melissa. 2004. Forced to Choose: Determinants of Racial Identification in Multiracial Adolescents. *Child Development*, 75(3): 730-748.

Humes, Karen R., Nicholas A. Jones and Roberto 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. 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: The National Academies Press. [Accessed March 31, 2014]. Available from http://www.iom.edu/Reports/2009/RaceEthnicityData.aspx

Institute of Medicine (IOM). 2004. *Eliminating Health Disparities: Measurement and Data Needs.* Washington, DC: The National Academies Press.

Kawachi, Ichiro, Norman Daniels and Dean E. Robinson. 2005. Health Disparities By Race And Class: Why Both Matter. *Health Affairs*, 24(2): 343-352.

Kressin, Nancy R., Bei-Hung Chang, Ann Hendricks and Lewis E. Kazis. 2003. Agreement between Administrative Data and Patients' Self-Reports of Race/Ethnicity. *American Journal of Public Health*, 93(10): 1734-1739.

Luque, Adela and Renuka Bhaskar. 2014. 2010 American Community Survey Match Study. *CARRA Working Paper Series #2014-03.* Center for Administrative Records Research & Applications U.S. Census Bureau, Washington, D.C. Available from http://www.census.gov/srd/carra/2010_American_Community_Survey_Match_Study.pdf

Martin, Elizabeth. 2002. "Questionnaire Effects on Reporting of Race and Hispanic Origin: Results of a Replication of the 1990 Mail Short Form in Census 2000." U.S. Census Bureau, Washington, DC. Available from https://www.census.gov/cac/race_ethnic_advisory_committees/docs/2006_Nov_questionnaire_effects.pdf

Martin, Elizabeth, Theresa Demaio, and Pamela Campanelli. 1990. "Context Effects for Census Measures of Race and Hispanic Origin." *The Public Opinion Quarterly* 54(4): 551-566.

McAlpine, Donna D., Timothy J. Beebe, Michael Davern, and Kathleen 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 Management and Budget (OMB). 1997. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Federal Register Notice 62:58781-58790. October 30, 1997. [Accessed March 31, 2014]. Available from http://www.whitehouse.gov/omb/fedreg_1997standards/

Office of Minority Health (OMH). 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

Rastogi, Sonya and Amy O'Hara. 2012. 2010 Census Match Study. *2010 Census Planning Memoranda Series No. 247.* U.S. Census Bureau, Washington, DC. Available from https://www.census.gov/2010census/pdf/2010_Census_Match_Study_Report.pdf

Rodriguez, Clara E. 2000. *Changing Race: Latinos, the Census, and the History of Ethnicity in the United States*. New York: New York University Press.

Saperstein, Aliya and Andrew M. Penner. 2012. Racial Fluidity and Inequality in the United States. *American Journal of Sociology*, 118(3): 676-727.

Sebelius, Kathleen. 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, Brian D., Adrienne Y. Stith and Alan R. Nelson. (eds). 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: Institute of Medicine.

Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software." *CARRA Working Paper Series* #2014-01. U.S. Census Bureau,

Washington, DC. Available from https://www.census.gov/srd/carra/CARRA_PVS_Record_Linkage.pdf

Zaslavsky, Alan M., John Z. Ayanian and Lawrence B. Zaborski. 2012. The Validity of Race and Ethnicity in Enrollment Data for Medicare Beneficiaries. *Health Services Research,* 47(3): 1300-1321.

Table 1. Race and Hispanic Origin Distribution in Medicaid, Linked Medicaid-Census and Linked Medicaid-Census-ACS Datasets, MSIS 2006-2008

Race and Hispanic Origin as Reported in Medicaid (MSIS)	2006 MSIS	2007 MSIS	2008 MSIS	PIK'ed Unduplicated MSIS 2006-8	PIK'ed Unduplicated MSIS 2006-8 linked to Census	PIK'ed Unduplicated MSIS 2006-8 linked to Census and ACS 2006-10	
Percent with a PIK	92.8%	92.7%	93.3%	100.0%	100.0%	100.0%	
	All Ages						
Percent PIK'ed MSIS linked to Census / ACS	n/a	n/a	n/a	n/a	90.0%	5.3%	
Race							
White alone	44.3%	44.0%	44.4%	47.7%	49.0%	47.6%	
Black alone	24.4%	23.7%	23.3%	23.5%	23.1%	23.1%	
AIAN alone	1.4%	1.5%	1.4%	1.5%	1.5%	1.4%	
Asian alone	2.7%	2.7%	2.8%	3.0%	2.9%	3.0%	
NHPI alone	0.9%	1.0%	0.9%	1.0%	1.0%	1.0%	
Two or more races	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	
Missing	26.0%	26.7%	26.8%	22.9%	22.2%	23.6%	
Hispanic origin							
Non-Hispanic	70.7%	70.1%	68.5%	71.7%	72.5%	70.8%	
Hispanic	22.3%	22.9%	22.7%	20.5%	19.8%	21.5%	
Missing	7.0%	7.0%	8.8%	7.7%	7.8%	7.7%	
N	62,814,724	63,251,758	65,763,995	**67,051,122**	60,312,876	3,556,229	
Weighted N						61,670,269	
	Ages 25 and older, no additional restrictions						**With additional restrictions***
Percent PIK'ed MSIS linked to Census / ACS	n/a	n/a	n/a	n/a	91.9%	5.5%	5.0%
Race							
White alone	49.8%	49.6%	50.1%	53.2%	54.1%	51.7%	54.9%
Black alone	22.5%	21.7%	21.4%	22.0%	21.6%	21.5%	23.0%
AIAN alone	1.2%	1.3%	1.2%	1.3%	1.3%	1.2%	1.2%
Asian alone	3.4%	3.6%	3.8%	3.8%	3.5%	3.7%	4.0%
NHPI alone	1.3%	1.4%	1.3%	1.4%	1.3%	1.4%	1.4%
Two or more races	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	
Missing	21.4%	22.0%	21.9%	18.0%	17.8%	20.1%	15.5%
Hispanic origin							
Non-Hispanic	75.5%	75.1%	73.2%	77.1%	77.3%	74.5%	80.4%
Hispanic	17.7%	18.2%	17.6%	14.7%	14.5%	17.3%	11.1%
Missing	6.9%	6.8%	9.1%	8.2%	8.2%	8.2%	8.5%
N	26,051,241	25,287,943	26,005,178	**28,788,197**	26,469,007	1,597,118	1,451,169
Weighted N						25,916,567	22,751,562

*The sample used for analysis excludes individuals reporting as Two or More Races in Census or Medicaid, SOR in Census, or have missing race/Hispanic origin responses in Census.

Source: MSIS 2006-2008, Census 2000, 2010 Census, 2001 through 2009 ACS 1-year data, and 2006-2010 ACS 5-year data.

Table 2. Comparison of Race Responses in Linked Medicaid-Census Records, All Ages

Number and percent matched	Medicaid Race Records								
Census Race Response	**White alone**	**Black alone**	**AIAN alone**	**Asian alone**	**NHPI alone**	**Two or More Races**	**Missing**	**Total**	**% Census Race**
White alone	**25,841,286**	282,639	86,841	90,736	48,466	38,770	6,714,806	33,103,544	54.9%
(%)	**78.1%**	0.9%	0.3%	0.3%	0.1%	0.1%	20.3%	100.0%	
Black alone	468,066	**12,435,493**	15,611	25,442	12,032	36,638	678,477	13,671,759	22.7%
(%)	3.4%	**91.0%**	0.1%	0.2%	0.1%	0.3%	5.0%	100.0%	
AIAN alone	218,462	41,684	**610,255**	5,660	2,242	10,444	162,739	1,051,486	1.7%
(%)	20.8%	4.0%	**58.0%**	0.5%	0.2%	1.0%	15.5%	100.0%	
Asian alone	150,719	34,228	21,603	**1,416,464**	337,984	27,751	238,164	2,226,913	3.7%
(%)	6.8%	1.5%	1.0%	**63.6%**	15.2%	1.2%	10.7%	100.0%	
NHPI alone	17,323	3,619	809	25,103	**68,581**	439	17,867	133,741	0.2%
(%)	13.0%	2.7%	0.6%	18.8%	**51.3%**	0.3%	13.4%	100.0%	
Some Other Race alone	993,194	177,189	29,329	21,615	23,886	14,465	3,711,992	4,971,670	8.2%
(%)	20.0%	3.6%	0.6%	0.4%	0.5%	0.3%	74.7%	100.0%	
Two or More Races	1,116,547	564,702	95,922	103,298	85,559	**49,829**	630,791	2,646,648	4.4%
(%)	42.2%	21.3%	3.6%	3.9%	3.2%	**1.9%**	23.8%	100.0%	
Missing	766,163	374,765	24,464	49,311	23,405	8,122	**1,260,885**	2,507,115	4.2%
(%)	30.6%	14.9%	1.0%	2.0%	0.9%	0.3%	**50.3%**	100.0%	
Total	29,571,760	13,914,319	884,834	1,737,629	602,155	186,458	13,415,721	**60,312,876**	100.0%
(%)	49.0%	23.1%	1.5%	2.9%	1.0%	0.3%	22.2%	100.0%	

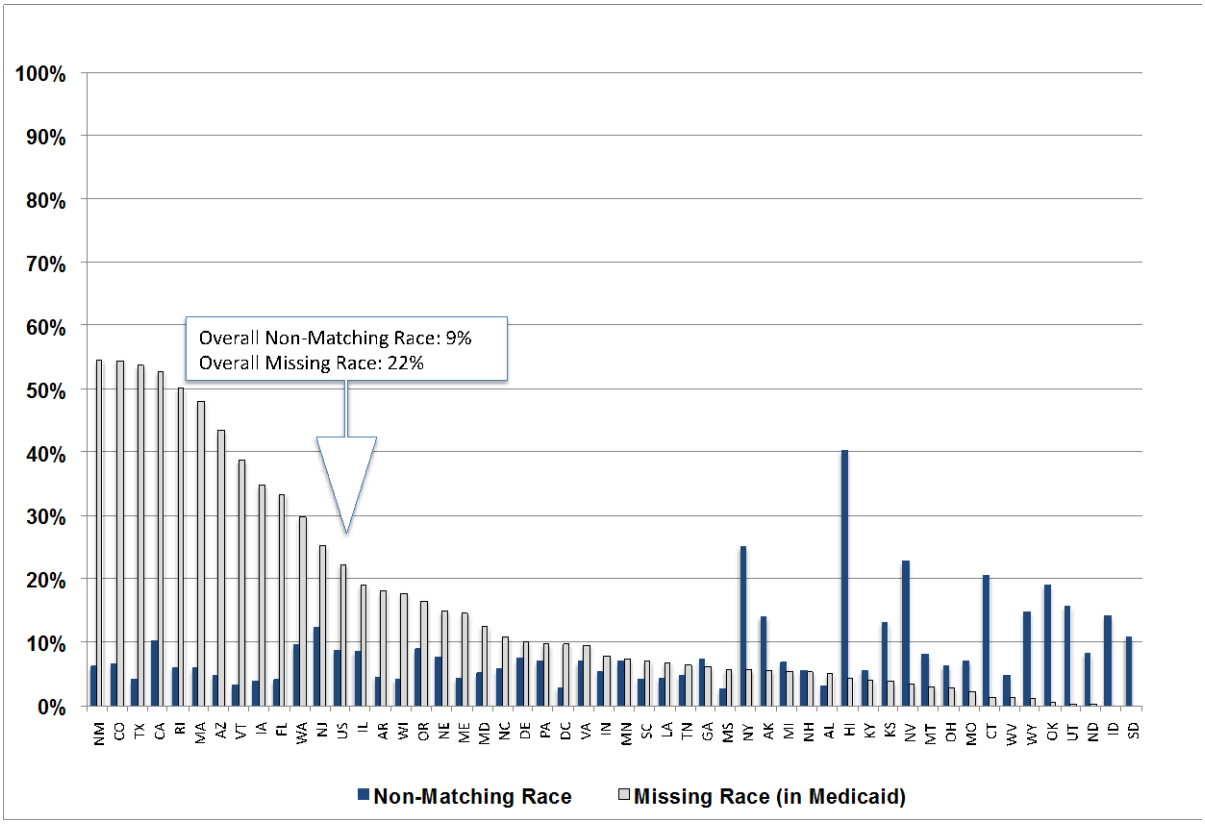
Source: MSIS 2006-2008, Census 2000, 2010 Census and 2001 through 2009 ACS 1-year data

Table 3. Comparison of Hispanic Origin Responses in Linked Medicaid-Census Records, All Ages

Number and percent matched	Medicaid Hispanic Origin Records				
Census Hispanic Origin Response	**Hispanic**	**Not Hispanic**	**Missing**	**Total**	**% Census Hispanic Origin**
Hispanic	**10,746,577**	2,418,326	561,335	13,726,238	22.8%
(%)	**78.3%**	17.6%	4.1%	100.0%	
Not Hispanic	616,813	**39,458,026**	3,896,620	43,971,459	72.9%
(%)	1.4%	**89.7%**	8.9%	100.0%	
Missing	554,080	1,833,913	**227,186**	2,615,179	4.3%
(%)	21.2%	70.1%	**8.7%**	100.0%	
Total	11,917,470	43,710,265	4,685,141	**60,312,876**	100.0%
(%)	19.8%	72.5%	7.8%	100.0%	

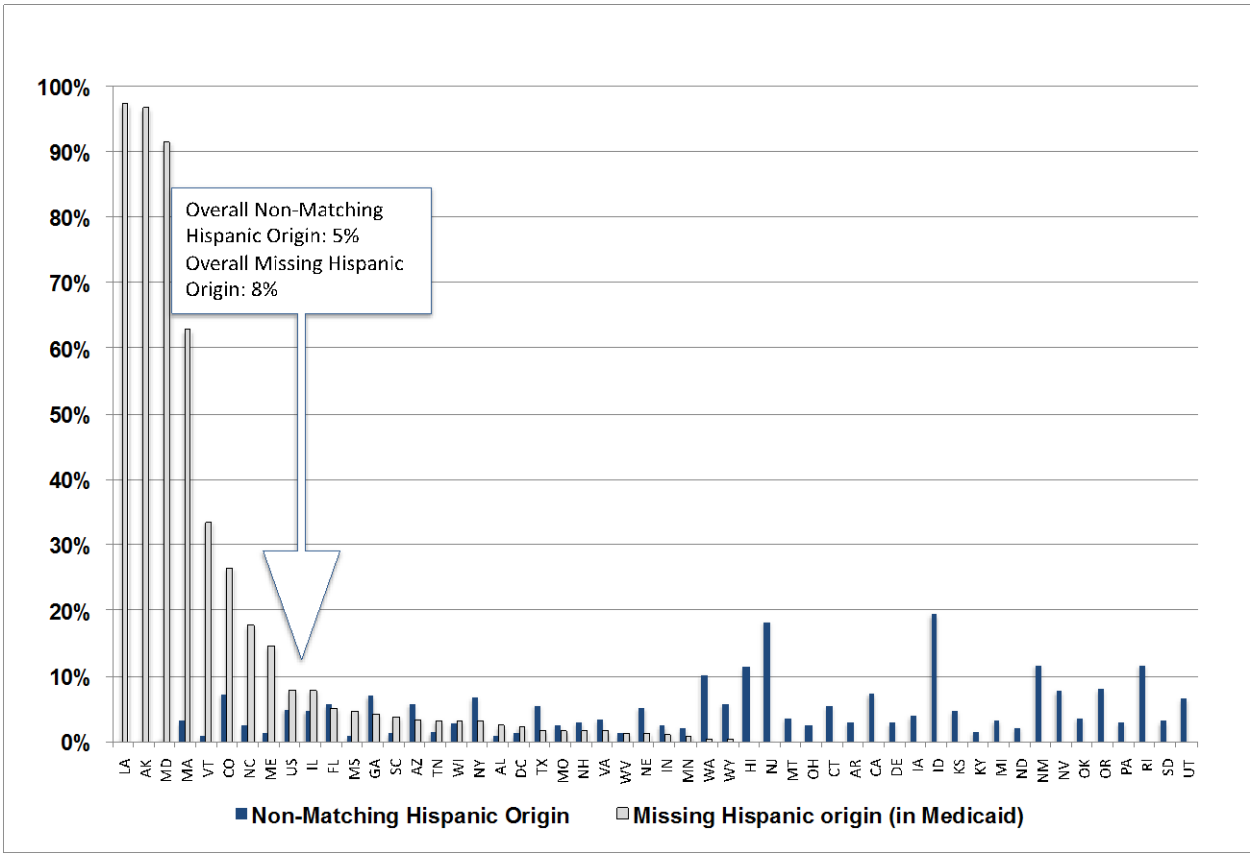
Source: MSIS 2006-2008, Census 2000, 2010 Census and 2001 through 2009 ACS 1-year data

Figure 1. Percent of Linked Medicaid-Census Records (N= 60,312,876) with Missing or Non-Matching Race Compared to Census, by State*



* Only observations with a PIK in MSIS were linked to Census/ACS data; cells with small counts have been suppressed.
Source: Authors' computations, MSIS 2006-2008, Census 2000, 2010 Census, ACS 2001 through 2009.

Figure 2. Percent of Linked Medicaid-Census Records (N= 60,312,876) with Missing or Non-Matching Hispanic Origin Compared to Census, by State*



* Only observations with a PIK in MSIS were linked to Census/ACS data; cells with small counts have been suppressed.
Source: Authors' computations, MSIS 2006-2008, Census 2000, 2010 Census, ACS 2001 through 2009.

Table 4. Weighted Regressions Predicting the Outcome from Comparing Medicaid and Census Race and Hispanic Origin Responses, Odds Ratios (OR)

Variables in the Analysis	Model 1: Race		Model 2: Hispanic Origin	
	Non-matching	Missing	Non-matching	Missing
	(Matching race is the reference)		(Matching Hispanic origin is the reference)	
	OR	OR	OR	OR
Demographic Characteristics				
Race (as reported in Census)				
White alone (omitted)				
Black alone	2.88***	0.55***	1.45***	1.57***
American Indian or Alaska Native alone	34.3***	1.81***	1.86***	1.11***
Asian alone / Native Hawaiian or Pacific Islander alone	13.56***	0.73***	3.36***	0.73***
Ethnicity				
Not Hispanic or Latino (omitted)				
Hispanic or Latino	6.27***	18.54***	51.45***	0.96*
Place of birth				
Born in the U.S. (omitted)				
Foreign born	1.92***	1.68***	0.77***	0.77***
Age				
25-44 (omitted)				
45-64	1.19***	1.12***	1.14***	1.28***
65 and older	1.50***	1.43***	1.48***	1.48***
Gender				
Male (omitted)				
Female	0.88***	0.97***	0.95***	1.03***
Marital Status				
Married (omitted)				
Never married	1.09***	0.90***	1.29***	1.22***
Separated, divorced or widow(er)	1.06***	0.87***	1.15***	1.06***
Socio-Economic Characteristics and Language Proficiency				
Language Proficiency				
Speaks English only (omitted)				
English very well or well, other language at home	1.11***	1.52***	0.64***	1.46***
English not well or not at all, other language at home	0.75***	1.30***	0.50***	1.46***
Education				
No high school (omitted)				
High school	1.03*	0.94***	1.04***	0.89***
Some college	1.05**	0.95***	1.16***	0.81***
College degree or higher	1.28***	1.24***	1.27***	0.98
Personal income (log)	1.01***	1.01***	1.00	1.01***
Contextual Factors				
County population				
Over 500,000 pop (omitted)				
100,000 to 500,000 pop	0.73***	0.95***	0.98	1.18***
Less than 100,000 pop	0.62***	0.79***	0.85***	0.84***
Percent minority in the state				
Less than 35 percent (omitted)				
35 percent or higher	1.27***	0.68***	1.14***	1.53***
Percent population Hispanic in tract				
Less than 10 percent (omitted)				
10 to less than 25 percent	1.35***	1.21***	1.19***	0.67***
25 percent or higher	1.79***	2.12***	0.78***	0.39***
Median household income in tract (log)	1.37***	1.64***	1.33***	1.46***
Unweighted N	1,451,169			
Weighted N	22,751,562			

*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

Appendix 1: Differences in format and wording of Hispanic origin and race questions in Medicaid, Census and American Community Surveys

- **States reporting of Hispanic origin and race to the national Medicaid database**

Although states do not have a nationally standardized Medicaid application form, they are required to submit Hispanic origin and race data for their Medicaid eligible populations to the Centers for Medicare and Medicaid Services (CMS) using the Medicaid Statistical Information System (MSIS). The ethnicity codes and the five race binary variables below have been used by Medicaid since 2005.[3] Each of these variables is considered independent of the other so that individuals could report any combination of races and Hispanic origin (McAlpine et al. 2007, CMS 2010).

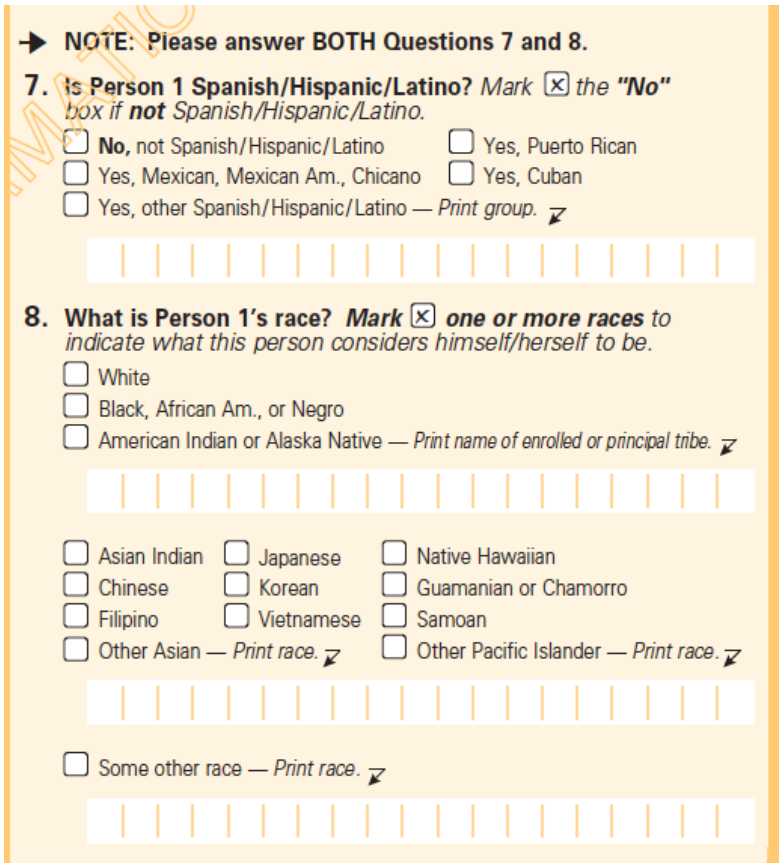
MSIS ETHNICITY CODE: 0 Not Hispanic or Latino
1 Hispanic or Latino
9 Ethnicity Unknown

MSIS RACE CODES:

RACE-CODE-1: 0 Non-White or Race Unknown
1 White
RACE-CODE-2: 0 Non-Black or African American or Race Unknown
1 Black or African American
RACE-CODE-3: 0 Non-American Indian or Alaska Native or Race Unknown
1 American Indian or Alaska Native RACE-CODE-4:
RACE-CODE-4: 0 Non-Asian or Race Unknown
1 Asian
RACE-CODE-5: 0 Non-Native Hawaiian or Other Pacific Islander or Race Unknown
1 Native Hawaiian or Other Pacific Islander

[3] Prior to 2005, states provided race and Hispanic origin information in one single variable with nine categories: 1=White, 2=Black or African American, 3=American Indian or Alaska Native, 4=Asian, 5=Hispanic or Latino (no race information available), 6=Native Hawaiian or Other Pacific Islander, 7=Hispanic or Latino and one or more races, 8=More than one race (Hispanic or Latino not indicated), and 9=Unknown (McAlpine et al. 2007).

- **Census 2000 and ACS 2001-2007 questions about Hispanic origin and race**



NOTE: Please answer BOTH Questions 7 and 8.

7. Is Person 1 Spanish/Hispanic/Latino? *Mark* ☒ *the* ***"No"*** *box if* ***not*** *Spanish/Hispanic/Latino.*

- ☐ **No,** not Spanish/Hispanic/Latino
- ☐ Yes, Puerto Rican
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Cuban
- ☐ Yes, other Spanish/Hispanic/Latino — *Print group.*

8. What is Person 1's race? ***Mark*** ☒ ***one or more races*** *to indicate what this person considers himself/herself to be.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Japanese
- ☐ Native Hawaiian
- ☐ Chinese
- ☐ Korean
- ☐ Guamanian or Chamorro
- ☐ Filipino
- ☐ Vietnamese
- ☐ Samoan
- ☐ Other Asian — *Print race.*
- ☐ Other Pacific Islander — *Print race.*
- ☐ Some other race — *Print race.*

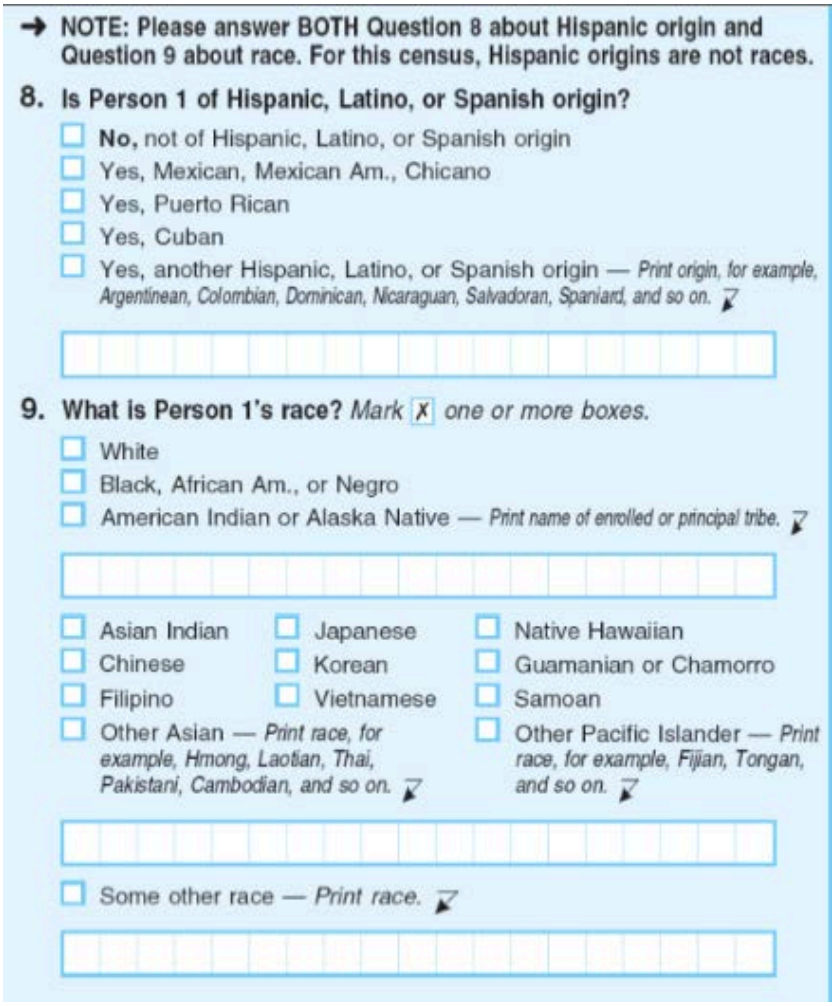
The race and Hispanic origin data from Census 2000 and ACS 2001-2007 were collected using the question wording shown above.

- **2010 Census and ACS 2008-2010 questions about Hispanic origin and race:**

Starting in ACS 2008 and used in the 2010 Census, the note that precedes the Hispanic origin and race questions added the sentence: "For this survey, Hispanic origins are not races." Also, the term "origin" was added to the Hispanic origin question, and the instruction, "Mark (X) the **"No"** box if **not** Spanish/Hispanic/Latino" was removed. There were also example groups added in ACS 2008 for the Other Hispanic in the Hispanic origin question, and for Other Asian and Other Pacific Islander in the race question. Finally, differences in layout were also introduced. The ACS 2007 used a grid format compared to the sequence format used in the ACS 2008 (Chestnut 2008, ACSRN 2008).

Previous studies have found that changes in wording and listing examples influence Hispanic origin reporting (Martin 2002). Consistent with earlier reports, a comparison of Hispanic origin and race questions in ACS 2007 with ACS 2008 (which would be similar to comparing Census

2000 and 2010 Census questions) found an increase in the percent of Hispanics reporting a specific origin (rather than a general response, i.e., Hispanic). There were also other changes resulting from the revisions to the ACS questionnaire between 2007 and 2008, such as an increase in the proportion of Hispanic who reported as White alone, with a corresponding decline in the proportion reporting as Some Other Race (ACSRN 2008).



→ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark* ✗ *one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

- ☐ Some other race — *Print race.*

Appendix 2: MSIS 2006-2008 PIK rates by selected characteristics

Percent of MSIS 2006-2008 Assigned a PIK by Selected Characteristics			
Characteristics	**2006**	**2007**	**2008**
Total observations	62,814,724	63,251,758	65,763,995
Percent assigned a PIK	92.8	92.7	93.3
Race			
White alone	96.3	96.1	96.5
Black alone	96.1	96.0	96.3
AIAN alone	95.8	96.2	96.0
Asian alone	94.6	94.3	94.4
NHPI alone	94.6	95.3	95.7
Two or more races	92.0	93.3	93.5
Missing	83.5	83.8	85.1
Hispanic Origin			
Hispanic	84.5	85.1	85.9
Non-Hispanic	95.5	95.5	96.0
Missing	92.3	90.6	91.1
Age group			
0-24 years old	91.8	91.8	92.5
25-44 years old	90.8	90.8	91.4
45-64 years old	97.8	98.0	98.2
65 and older	98.0	98.1	98.2
Missing	0.1	0.0	0.0
State			
Percent PIK'ed ranged from:			
Min	81.4 (California)	81.9 (California)	78.8 (Montana)
Max	99.7 (West Virginia)	99.5 (Vermont)	99.5 (Vermont)
Source: MSIS 2006-2008			

Appendix 3: Link rate of unduplicated MSIS 2006-2008 with a PIK to Census/ACS data

Percent of PIK'ed MSIS 2006-2008 Linked to Census and ACS by Selected Characteristics		
Characteristics	**Census/ACS[1]**	**2006-2010 ACS Data 5-Yr File**
Number	60,312,876	3,556,229
Percent	90.0%	5.3%
Race		
White alone	92.4	6.2%
Black alone	88.2	4.4%
AIAN alone	88.6	6.1%
Asian alone	86.3	4.5%
NHPI alone	86.4	4.6%
Two or more races	86.7	5.1%
Missing	87.5	4.4%
Hispanic Origin		
Hispanic	86.6	4.1%
Non-Hispanic	90.9	5.7%
Missing	90.3	5.3%
Age group		
0-24 years old	88.5	4.9%
25-44 years old	93.3	5.3%
45-64 years old	91.4	5.9%
65 and older	90.1	6.6%
Missing	59.6	1.8%
State		
Percent linked ranged from:		
Min	84.8 (Hawaii)	4.0 (Florida)
Max	94.5 (Vermont)	9.9 (Vermont)

[1] Linked MSIS 2006-2008 to the most recent "as reported, no proxy" record found combining Census 2000, 2010 Census and 2001 through 2009 ACS 1-year files.

Source: MSIS 2006-2008, Census 2000, 2010 Census, 2001 through 2009 ACS 1-year data, and 2006-2010 ACS 5-year data.