CARRA Working Paper Series

Working Paper *#2015-05*

# An outside view: What do observers say about others' races and Hispanic origins?

Sonya Rastogi Porter
U. S. Census Bureau

Carolyn A. Liebler
University of Minnesota

James M. Noon
U. S. Census Bureau

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

Paper Issued: August 26, 2015

# An outside view: What do observers say about others' races and Hispanic origins?

Sonya Rastogi Porter, United States Census Bureau
Carolyn A. Liebler, University of Minnesota
James M. Noon, United States Census Bureau

**Abstract**

Outsiders' views of a person's race or Hispanic origin can impact how she sees herself, how she reports her race and Hispanic origins, and her social and economic experiences. The way outsiders describe non-strangers in terms of their race and Hispanic origin may reveal popular assumptions about which race/Hispanic categories are salient for Americans, which kinds of people are seen as multiracial, and the types of cues people use when identifying another person's race. We study patterns of observer identification using a unique, large, linked data source with two measures of a person's race and Hispanic origin. One measure (from Census 2000 or the 2010 Census) was provided by a household respondent and the other (from the other census year) was provided by a census proxy reporter (e.g., a neighbor) who responded on behalf of a non-responsive household. We ask: *Does an outsider's report of a person's race and Hispanic origin match a household report?* We find that in about 90% of our 3.7 million (non-representative) cases, proxy reports of a person's race and Hispanic origin match responses given by the household in a different census year. Match rates are high for the largest groups: non-Hispanic whites, blacks, and Asians and for Hispanics, though proxies are not very able to replicate the race responses of Hispanics. Matches are much less common for people in smaller groups (American Indian/Alaska Native, Pacific Islander, Some Other Race, and multiracial). We also ask: *What predicts a matched response and what predicts a particular unmatched response?* We find evidence of the persistence of hypodescent for blacks and hyperdescent for American Indians. Biracial Asian-whites and Pacific Islander-whites are more often seen by others as non-Hispanic white than as people of color. Proxy reporters tend to identify children as multiple race and elders as single race, whether they are or not. The race/Hispanic composition of the tract is more powerfully predictive of a particular unmatched response than are tract-level measures of socioeconomic status; unmatched responses are often consistent with the race/Hispanic characteristics of the neighborhood.

## Introduction

Race is socially constructed through a variety of processes, some of which involve the perceptions of others. Outsider perceptions of a person's race and ethnicity[1] can affect how the person sees himself (Cooley 1902; Doyle and Kao 2007), how he is treated by others (Herman 2010), and may have mental health side effects (Campbell and Troyer 2007; Kramer et al. 2015). Most discrimination is probably based on outsiders' assumptions about a person's race rather than on the person's self-identification. By studying patterns in characteristics of those assumed to be in each race/Hispanic group, we illuminate an important aspect of the social processes involved in developing and sustaining race and Hispanic identities: perceptions by others.

By studying outsider views we learn about the types of information Americans use to decide what race and Hispanic origin someone else is, as well as which race/ethnic categories are in the forefront of the public mind. Also, race and ethnicity are multidimensional constructs; by studying two aspects of racial identification in relation to each other (responses by an outsider as compared to the race/ethnicity reported by a household member) we inform research on how observer identification is related to social and economic outcomes (Bailey et al. 2014; Saperstein 2011; Saperstein and Penner 2012). Finally, the relationship between a census proxy respondent's[2] views of race/ethnicity and household-reported[3] race/ethnicity has implications for

[1] We use the terms "ethnicity" and "ethnic" as synonyms for Hispanic origin, following the definitions employed in our data. The U.S. Census Bureau adheres to the Office of Management and Budget's 1997 race and ethnicity standards, which specify five major race groups: white, black or African American ("black" here), American Indian or Alaska Native ("American Indian" or "AIAN" here), Asian, and Native Hawaiian or Other Pacific Islander ("Pacific Islander" or "NHPI" here), and define two ethnic groups (Hispanic and non-Hispanic). The Census Bureau is also permitted to use the category Some Other Race ("SOR" here) for respondents who cannot be classified within the five major race groups.

[2] In a decennial census, every person and every address must be counted. If a household does not respond through the mail, an enumerator is sent to visit the address in person. If the enumerator tries to reach a resident unsuccessfully at least three times, or if the household refuses to participate, the enumerator can

Census data quality. We can learn the extent to which proxy reporters give the same responses as we would expect to receive from the person himself/herself.

We use a large, linked data set to address three research questions. First, do outsider reports of a person's race and ethnicity match the reports given by someone in the household? We use descriptive statistics to identify patterns of matched and unmatched responses across race and Hispanic origin groups. Second, what predicts a matched response? We use descriptive statistics and logistic regression analyses to identify correlates of a proxy's matched response versus an unmatched response. Third, what predicts a particular unmatched response? We use descriptive statistics and multinomial logistic regression models to identify patterns of particular unmatched responses (e.g., when does a proxy report single-race black for a black-white biracial respondent?).

Our data are uniquely suited to study these important questions. As described more thoroughly below, the data were created by carefully linking individuals' cases across the full count decennial censuses of 2000 and 2010. We focus on over three million cases which contain two measures of race and Hispanic origin for each person: one measure was provided by a census proxy reporter (in Census 2000 or the 2010 Census) and the other was provided by someone in the person's household (in 2010 or 2000).

---

find someone else to serve as a proxy (U.S. Census Bureau 2009). A proxy respondent must be at least 15 years old and have knowledge about the status of the household on census day. A proxy respondent is usually a neighbor, rental agent, building manager, or a person who recently moved into the address (U.S. Census Bureau 2009). In 2000, about 9 million people had records created using proxy responses and in 2010, proxy responses accounted for more than 13 million cases. This is 3 to 4% of the full census data.

[3] We do not know which person within the household provided responses for people in the home, though Census Bureau studies show that in most cases one person (often the householder or his/her spouse) answers the decennial census questions for the entire household (e.g., Sweet 1994).

The proxy respondents in this study were not strangers to the people for whom they were reporting; all proxy respondents in our data provided the first and last name of the person to the census enumerator. These remarkable data allow us to go beyond previous knowledge about how an outsider would describe a non-stranger's race. Other researchers have studied outsider race reports by enumerators (e.g., Campbell and Troyer 2007), by participants in small experiments (e.g., Mondloch et al. 2010; Peery and Bodenhausen 2008), and by co-resident parents (e.g., Xie and Goyette 1997). We also enhance knowledge by studying observer reports that were given for people of all race and ethnic groups, without limitation to a specific universe (e.g., women ages 15-44; Saperstein 2011). In addition, our research increases understanding of data provided by census proxy respondents. In public and restricted-use versions of census data, proxy responses are not distinguished from responses given by the household. Thus, our behind-the-scenes look at proxy response cases is especially useful for data users who need to know how proxy responses compare to those given by households.

*A note on terminology*

The race(s) and Hispanic origin reported for a person by someone in their household are likely to reflect that person's own identity at that time. Because we expect that it is often an accurate description, we write as if each person "is" the race(s) and Hispanic origin reported for/by them by the household respondent. However, racial identities are socially constructed and can be fluid, so household reports for the same person sometimes vary from one census to the next (Liebler et al. 2014a, 2014b). Thus, a proxy response may be different from the household response and yet be an accurate reflection of how the person would like to be identified in that year. There are probably some circumstances in which neither the proxy report nor the household report accurately describe a person's identity.

## Prior Research

Previous researchers studying outsiders' views of a person's race or Hispanic identity have been limited in several ways. First, usually due to data limitations, they have primarily focused only on larger race groups such as white and black. This has precluded extensive exploration of outsider identification patterns for other groups (e.g., American Indians or multiracials) who appear to have more complex racial and ethnic identities (Harris and Sim 2002; Liebler et al. 2014b). Second, prior studies have focused on specific universes such as adolescents (Campbell and Troyer 2007), multiracial people (Peery and Bodenhausen 2008), college-aged students (Herman 2010), or women of childbearing age (Saperstein 2011). Third, related studies conducted in laboratory settings are divorced from daily life and have very small sample sizes (Mondloch et al. 2010; Peery and Bodenhausen 2008). Nevertheless, we use these studies to frame our research because they ground our analyses and give some expectations for the content of our results.

*Does an outsider report of a person's race and Hispanic origin match the household report?*

Observer reports match self-reports 97% to 99% of the time for whites and blacks across studies with differing methodologies and covering a variety of specific sub-populations (Campbell and Troyer 2007; Herman 2010; Saperstein 2006, 2011). There have been mixed results for single-race Asians. Campbell and Troyer (2007) found a high match rate (96%) for Asians, but Saperstein (2011) found that observers correctly classified Asian and Pacific Islanders as "other" in only 76% of cases when asked to code people as white, black, or "other."

There is much less congruence for other groups (Campbell and Troyer 2007; Herman 2010; Lee 2008; Saperstein 2006, 2011). Observers correctly classified only 11% of American Indians as "other," as opposed to white or black (Saperstein 2011). Among adolescents and

young adults, Campbell and Troyer (2007) found relatively low agreement between observer and self-reported race for American Indians. Among college students evaluating yearbook pictures, Herman (2010) found relatively low congruence between self-reported identity and observer reports for American Indians, Asians, Middle Easterners, Hispanics, Pacific Islanders, and multiracials. A study of multiracial adults with black, Asian, or Hispanic heritage living in California suggests that Americans are accustomed to identifying people with black ancestries but are not as accustomed to identifying Asian and Latino ancestries (Lee 2008).

It is possible that observer race and ethnicity reports do not match self-reports for some people because the people have relatively complex and fluid identities and (perhaps) inconsistent self-reports across time or in different situations (Harris and Sim 2002; Liebler et al. 2014a). In a large study with linked census data, the groups with the highest levels of race response change across a decade were American Indians, Pacific Islanders, Some Other Race respondents, Hispanics (when answering the race question), and people who gave multiple-race responses (Liebler et al. 2014a, 2014b). The causal order of this identification process is unclear; perhaps one of the causes of individuals changing their census race/Hispanic responses over time is inconsistent messages from outsiders about a person's assumed race. Single-race non-Hispanic whites, blacks, and Asians gave relatively consistent race responses across the period in that study (Liebler et al. 2014a); people in these three groups may give easier-to-identify messages to outsiders about their preferred race/Hispanic response.

Based on the research outlined above, we expect that proxy reports will almost always match household reports for non-Hispanic single-race white or black people, and will often match household reports for non-Hispanic single-race Asians. We expect the match rates for other non-Hispanic race groups –American Indians, Pacific Islanders, those who report multiple

races – to be considerably lower. We expect observer and household reports of Hispanic origin to match quite often, but we expect less congruence in the race responses for Hispanics.

*What predicts a matched response? What predicts a particular unmatched response?*

Proxy respondents could have a variety of relationships with a person in the non-responding household. They may know each other well and may have spoken about race and Hispanic identities. If a proxy does not have much information, she may draw on other clues and cues to develop answers. She might interpret the person's name, bumper stickers, decorations, or types of clothing. She could also make assumptions based on the type of home (e.g., rented apartment) or the type of neighborhood (e.g., known to have many of a particular race or with high income). In the sections below, we describe four sets of factors that may predict a matched response and/or help determine how a proxy answers in cases where it does not match. These factors are: (a) socially defined ideas about who "belongs" in which group, (b) observations or knowledge about the household, (c) local racial composition and urban status, and (d) local socioeconomic standing.

***Norms about claiming or belonging to a particular group:*** If the household respondent and the proxy respondent follow the same social norm about which people fit into which race/Hispanic categories then their responses will probably match.

The construction of whiteness appears deeply embedded in current norms such that both an individual and their observer would agree on whether the person should be reported as white. Whiteness has generally been defined, both legally and socially, as including *only* people with fully white heritage (Haney López 1996; Painter 2010).[4] For instance, the United States legal

[4] Some Europeans immigrant groups that were initially treated as non-white were able to permeate the boundary of whiteness in the United States (for an example, see Brodkin 1998).

system was instrumental in codifying who should be considered white (and non-white) through states' legal definitions of who is Black, immigration laws and quotas, anti-miscegenation laws, and segregation laws (Haney López 1996).

The historical social construction of race for blacks and American Indians may influence the way people decide what race to report. The "one drop rule" of hypodescent – the idea that someone is a group member if they have any ancestry from that group – has dominated the social construction of race for blacks (Davis 2001; Hollinger 2003). Hyperdescent – the idea that someone is not a "true" member of the group unless most/all of their ancestors are from that group (Snipp 2003) – affects who is seen as American Indian; part-white American Indians are often socially defined as white, effectively rendering their indigenous heritage invisible.

Historically, Asians have been defined and isolated through legal measures, for example: the Chinese Exclusion Act, limits on Japanese immigration, the case of United States v. Bhagat Singh Thind, and the Asiatic Barred Zone section of the Immigration Act of 1917. With the recent influx of Asian immigrants,[5] household respondents and proxies may utilize birthplace to help categorize Asians for the race question, which would result in high levels of agreement between proxy and household responses for Asians. However, high intermarriage rates between Asians and whites (Fredrickson 2003) may allow some Asians, particularly non-immigrants, to permeate the white boundary. Little is known about what norms (if any) affect how Pacific Islanders are seen.

Like Asians, Hispanics also have a long history in the United States. In the 1800s and 1900s, Mexican Americans were often defined as white but treated as non-white (Fredrickson

[5] In 2007, 67 percent of U.S. Asians were born abroad (Grieco 2010).

2003). Similar to Asians, the currently high proportion of Hispanics who are immigrants[6] would tend to increase the proxy-household response match rate if both respondents are using birthplace (or parent's birthplace) to decide how to report race and Hispanic origin.

In a study of household reports in linked data from the censuses of 2000 and 2010, there was cross-year consistency in whether a person was reported as Hispanic (vs. non-Hispanic), but race responses given by/for Hispanics were substantially less stable across the period (Liebler et al. 2014a). In many ways, "Hispanic" is currently constructed as a race, yet there was no "Hispanic" response category in the race question in these censuses. Many Hispanics have found the census race question challenging (Compton et al. 2012) and Hispanics do not seem to encounter clear nation-wide norms about how they should racially identify (Dowling 2014; Liebler and DeRousse-Wu 2012; Vargas 2015). Race responses by Hispanics and changes in these responses happen despite the fact that Hispanics with different race reports (e.g., white versus Some Other Race) have different residential and socioeconomic patterns (Logan 2003). Because of the complexity of race for Hispanics, we expect that proxy responses will be relatively unlikely to match household responses.

There are also age-related norms about race responses. Much of the societal conversation about multiracial people has centered on the idea that they are primarily a product of the recent increase in interracial unions (DaCosta 2007). This implies that multiracial people would be young and that few adults and elders would identify as multiracial. Proxy respondents might be open to the idea that a child is multiracial but not consider the possibility for an older person.

***Observations or knowledge about the household:*** A proxy respondent who has more information about the household will be better able to respond as the household would have**.** If

[6] Over 40 percent of Hispanics are immigrants and one third of Hispanics are second generation (Rumbaut, 2011).

the proxy reports a person's age fairly accurately, this may mean they know each other well. Families (particularly those with children) may have greater community involvement and more interaction with their neighbors as children play outside. Also, proxies can use information about one family member to categorize another, this is especially likely for children. Thus, we expect proxy responses to more accurately reflect household responses for children and people living in families.

A proxy may have different assumptions about a homeowner than a renter. For example, groups with stereotypes defining them as poor (Clawson and Trice 2000; Moses 2012) may be assumed to be renters, while groups with model minority stereotypes (Clawson and Trice 2000) might be assumed to be homeowners. We expect proxy responses to match household responses more often for homeowners than for renters, with mismatched responses following racial stereotypes.

***Local racial/ethnic composition:*** If the proxy respondent does not know the race and Hispanic origin of a person, he or she might look to outside clues based on the broader area – clues such as which race/ethnic groups are predominant in the area. Given patterns of residential segregation (Crowder et al. 2012; Iceland and Weinberg 2002), the proxy respondent may assume that the person has the same race and Hispanic origin as others in the local area. Race data from death certificates of American Indians shows this pattern – self-identified American Indians are much more likely to be identified as such on their death certificate if they pass away near other American Indians and/or on a reservation (Arias et al. 2008). Given that people are much better at identifying someone else's race if it is the same as their own (Mondloch et al. 2010) and residential racial segregation is particularly common for whites and blacks (Iceland

and Weinberg 2002), a white or black person is fairly likely to be the same race as their proxy, and this may increase the chances of a match.

In most states, people of color are concentrated in urban areas. Extrapolating from this, proxy respondents in urban areas might assume a person is a minority more often than do proxies in rural areas. Interviewers categorize inner city residents as black more often than they do other people (controlling for self-identified race; Saperstein and Penner 2012). When the local area contains a relatively high proportion of people from a particular minority group, we expect that a proxy's non-matching response will be that minority group. If the area is urban, we expect that a proxy's non-matching response will less often be non-Hispanic single-race white.

***Local socioeconomic standing:*** A proxy might rely on stereotypes about the socioeconomic status of people in each race/ethnic group when asked to describe the race and ethnicity of a person they do not know well. Socioeconomic status impacts both racial self-reports and observer assumptions about a person's race (Freeman et al. 2011; Penner and Saperstein 2008; Saperstein and Penner 2012). Proxies who do not know someone well might report them as non-Hispanic white if it is an area with relatively high income and education and report the person as Hispanic and/or non-white if people in the area have lower income and education, or if many in the area receive public assistance income.

## Data and Methods

### *Data*

***Record linkage:*** We use internal U.S. Census Bureau full-count data from Census 2000 and the 2010 Census and link individual's records across the two census years to form a longitudinal data set of individuals. To facilitate record linkage between the two files, the Census Bureau assigned each person a unique anonymized identifier called a Protected Identification

Key (PIK) via the Person Identification Validation System (PVS) (see Wagner and Layne 2014). The PIKs are assigned using probabilistic matching techniques (Fellegi and Sunter 1969) based on full name, gender, date of birth, and address. When personal information is missing or incomplete, or if a person is not in the reference files that are used to assign a PIK, then a PIK cannot be assigned.[7] We began with all individuals who (a) were assigned a PIK in both censuses, (b) were able to be linked across Census 2000 and the 2010 Census, (c) were in a household that directly responded to the census in 2000 or 2010 but not both, and (d) were in a household for which a proxy gave the response in 2010 or 2000 but not both. In all, the source data contain a total of 4,786,167 people who fit these criteria (2,034,426 proxy responses in 2000 and 2,751,741 proxy responses in 2010).

***Case selection:*** We applied several case selection criteria to maximize the meaningfulness of the cases under study (we excluded some cases for multiple reasons). People in group quarters are often not directly asked their race and Hispanic origin for the census,[8] so we took out those cases (112,610 cases excluded). Edited and imputed race or Hispanic origin information might not reflect the person's identity, so we dropped those cases (863,387 excluded). We wanted to be sure that the proxy was talking about the right person, so we insisted that the gender information was reported (108,260 excluded) and that the gender matched (66,350 excluded). There was a processing error in 2000 for those coded as Some Other Race multiracial, so we took those

[7] Of the 281 million people counted in Census 2000, 253 million (90%) received a PIK. Of the 308 million counted in the 2010 Census, 279 million (90%) received a PIK.

[8] Evidence suggests that race and Hispanic information for people in group quarters is often drawn from internal administrative records (Chun and Gan 2014).

people out of our analyses (63,386 excluded).[9] Once the case selection criteria were applied, our data include 1,624,384 cases where the proxy report was in 2000 and 2,054,508 cases with a proxy report in 2010.[10] We pooled the data for analysis for a total of 3,678,892 cases.

***Representativeness:*** Our data are not a representative sample of any particular population. In the Appendix, we provide an overview of the people in our data compared to the full set of people counted in each decennial census. Differences between our data and the full censuses are due to: uneven PIK rates (e.g., proxies know the full legal names of some types of people more than others), differences in match rates (e.g., different types of people are differentially likely to join or leave the population over the decade), our case selection criteria affect some groups more than others (e.g., some types of people are more likely to live in group quarters), and some types of people are overrepresented in non-responding households (see Walker et al. 2012). Our data include a relatively high proportion of non-Hispanics – particularly whites, blacks, and American Indians – and relatively low proportion of non-Hispanic Asians and Hispanics of any race. Our data also include a relatively high proportion of people in rural areas or the South. Despite their non-representativeness, these data are excellent for this project. The data cover the entire United States and include people of all race and Hispanic origin groups in such numbers that we can study American Indians, Pacific Islanders, Hispanics of different races, and people in four multiracial groups. Also, these data are restricted to cases where the

---

[9] U.S. Census Bureau (2007), Data Note 5, states: "In Census 2000, during the conversion process of making the race write-in entries on the enumerator-filled questionnaire consistent with those in the mailout/mailback questionnaire, a step was inadvertently omitted. This resulted in an overstatement by about 1 million people reporting more than one race (or about 15% of the Two or More Races population). This overstatement almost entirely affects race combinations involving Some Other Race [in combination] with the five race groups identified by the Office of Management and Budget …."

[10] As mentioned above, proxy reports were more common in 2010 in 2000.

proxy knows the person's full name (and thus they are probably familiar with each other), which is a type of social relationship thought to be related to identity formation and maintenance (Cooley 1902).

*Methods*

We address the first research question – do observer reports match household reports? – using information about all people in our data. We present and discuss descriptive results (in Table 2 below) showing the overall numbers of matched responses. For the five non-Hispanic single-race groups in our study – whites, blacks, American Indians, Asians, and Pacific Islanders – we define a match based on the broad race group (e.g., a response of "Korean" and a response of "Vietnamese" would be coded as matching because they are both Asian groups). We also code matches broadly for the Hispanic single-race groups in our study – Hispanic white and Hispanic Some Other Race – (e.g., "Nicaraguan" and "Mexican" would be coded as matching because they are both Hispanic groups). We focus on four non-Hispanic multiple-race groups: black-white, American Indian-white, Asian-white, and Pacific Islander-white. In these cases we coded any non-Hispanic multiple-race response as a match because an exact match was very rare.[11]

In our second research question, we ask: In what cases do observers give the same general race and Hispanic origin responses that were given by the household? We address the question of what predicts a matched response by looking at ten subsamples based on the household report of race and ethnicity. We include all proxy responses, dividing them into those that match across the two census years and those that do not match. In Table 3 below, we show results for the four non-Hispanic biracial groups and the four non-Hispanic single-race groups. In

[11] Household responses for multiple race people are particularly inconsistent across time (Liebler et al. 2014a). If we required the proxy respondent to match exactly, we would be holding the proxy to a higher standard of reliability than is evident among household responses.

Table 4 below, we show two versions of "match" for the two Hispanic single-race groups – (a) we model predictors of a proxy identifying the person as Hispanic (regardless of their race response), and (b) we model predictors of a proxy reporting both their household-identified race and Hispanic origin.

In our third research question, we ask about particular types of non-matched responses. To understand reasons that the proxy's response does not match the household response, we construct a three-category dependent variable including a match and two common non-matches and restrict our subsamples to those in one of these categories. For example, the first subsample includes non-Hispanic black-white people and a proxy year response of either (a) non-Hispanic white, (b) non-Hispanic black, or (c) non-Hispanic multiple-race; the few non-Hispanic black-white people whose proxy gave a different race/Hispanic response are not included in the sample.

*Measures*

Above we described factors that might affect what race and Hispanic report the proxy gives, whether as a match or a specific type of non-match. These factors point to characteristics of the person, family, and household, as well as the race/ethnicity, urban status, and socioeconomic standing of people in the tract.[12] We measure all independent variables in the proxy year to capture the context in which the proxy is reporting; see Table 1. For the most part, this information was available to the proxy at the time; the proxy gave the information about the person's personal, family, and household characteristics, and was probably familiar with the

[12] Most census tracts are contiguous-area county subdivisions. The optimum size is 4,000 people, though tracts range from 1,200 to 8,000 people. There were over 73,000 census tracts in the U.S. in the 2010 Census. See https://www.census.gov/geo/reference/gtc/gtc_ct.html for details.

Table 1: Construction of independent variables using information from the proxy year

| | |
|---|---|
| *Knowledge about the Family* | |
| Age | = age of the person in four categories (0-9, 10-18, 19-64, and 65+) in the proxy year. |
| Proxy report of age within 2 years | = the age reported by the proxy is within two years of the household reported age, adjusted for time. |
| Age was imputed in either year | = the age of the person was imputed in the household year or the proxy year, or both. |
| Household structure | = household structure in four categories based on who lives in the home, their ages, and how they are related to the person listed first on the form. |
| Own, rent, or no rent paid | = proxy was asked whether the people in the household own or rent, or do not pay for housing. |
| *Local Racial Composition* | |
| Urban area | = the Census Bureau codes each tract as either urban or rural.^ |
| Race(s) and Hispanic origins of people in the tract | = for example, of all people in the tract in the full census data for the proxy year, the logged % who reported single-race, non-Hispanic white. Repeated for other groups. |
| *Local Socioeconomic Standing* | |
| % in tract using public assistance | = of all people in tract in census data for proxy year, % with any public assistance income.^^ |
| % adults in tract w/ HS educ+ | = of people ages 25+ in tract in census data for proxy year, % with a high school diploma/GED or more education.^^ |
| Avg. adult income in tract/$10,000 | = average income of all people ages 15+ in tract in census data for the proxy year, in units of $10,000.^^ |
| *Control Variables* | |
| Proxy report was in 2010 | = the household report was in 2000 and the proxy report was in 2010. |
| Female | = household report (and proxy report) for the person's sex was female. |

^ See https://www.census.gov/geo/reference/urban-rural.html for more information about what areas qualify as rural versus urban in 2000 and in 2010.

^^ We used Census 2000 long form data if the proxy report was in 2000. For 2010 proxy reports, we used 2006-2010 American Community Survey data.

local area. Thus, the independent variables are endogenous to the outcome, and we encourage the reader to see the models as representing statistical associations rather than telling a causal story.

**Results**

*Does an outsider report of a person's race and Hispanic origin match the household report?*

Overall, we found that the reports of a person's race and ethnicity by a knowledgeable outsider matched those given by the household in 90% of the cases; see Table 2 for more specific match rates. The largest groups had the highest match rates. Almost all of non-Hispanic single-race whites and blacks were reported as such by the proxy (98% and 94% respectively). The great majority (88%) of non-Hispanic single-race Asians' reports also matched across measures.

The proxies in our data are quite able to report whether a person was of Hispanic origin. Virtually all non-Hispanics (99%) were identified as non-Hispanic by the proxy, and 86% of Hispanics were reported as Hispanic by the proxy. Almost all Hispanics of Some Other Race in our data (93%) were reported as Hispanic by outsiders, while only 84% of Hispanic whites were reported as Hispanic. Thus, we find support for the idea that race responses of Hispanics distinguish different groups (for discussion of this idea, see Hitlin et al. 2007, Logan 2003, and Tafoya 2004). We note that the match rate for some groups (e.g., Asians and Hispanics) may be affected by the proxy's assumptions about the ethnicity of the person's surname or given name (Cotton et al. 2008; Lauderdale and Kestenbaum 2000; Word and Perkins 1996).

Proxies were less able to identify non-Hispanic single-race American Indians (62% matched) and Pacific Islanders (62%), and they were quite poor at identifying multiple-race people as multiracial (8% to 36%; see Table 2). It may be particularly difficult for a proxy to identify a multiracial person because the idea of acknowledging mixed heritage is relatively new.

Table 2: Case counts of proxy responses given for selected people, by race/Hispanic response in household year

| Response in household year | Total N | Proxy response matched: Race only N | Race only % | Hispanic only N | Hispanic only % | Race & Hispanic N | Race & Hispanic % |
|---|---|---|---|---|---|---|---|
| All people in data | 3,678,892 | 3,369,958 | *92%* | 3,602,723 | *98%* | 3,322,939 | *90%* |
| Non-Hispanic | 3,363,753 | 3,196,952 | *95%* | 3,332,604 | *99%* | 3,179,595 | *95%* |
| Hispanic | 315,139 | 173,006 | *55%* | 270,119 | *86%* | 143,344 | *45%* |

| Response in household year | Total N | Proxy response matched N | Proxy response matched % | Proxy response did not match. Proxy said… *NH white alone* | *NH black alone* | *NH multiple races* | *All other responses* |
|---|---|---|---|---|---|---|---|
| NH white alone* | 2,582,994 | 2,525,516 | *98%* | (matched) | 13,297 | 13,091 | 31,090 |
| NH black alone* | 564,949 | 529,591 | *94%* | 21,723 | (matched) | 7,041 | 6,594 |
| NH AIAN alone* | 45,228 | 27,834 | *62%* | 11,382 | 1,323 | 2,986 | 1,703 |
| NH Asian alone* | 98,989 | 87,174 | *88%* | 5,252 | 673 | 2,904 | 2,986 |
| NH NHPI alone* | 5,196 | 3,195 | *62%* | 623 | 122 | 697 | 559 |
| NH SOR alone* | 5,949 | 648 | *11%* | 2,084 | 1,408 | 822 | 987 |

| Response in household year | Total N | Proxy response matched N | Proxy response matched % | Proxy response did not match. Proxy said… *NH white alone* | *NH minority group alone* | *All other responses* |
|---|---|---|---|---|---|---|
| NH black & white^ | 14,973 | 5,396 | *36%* | 3,513 | Black: 5,313 | 751 |
| NH AIAN & white^ | 19,064 | 1,549 | *8%* | 14,896 | AIAN: 2,085 | 534 |
| NH Asian & white^ | 10,570 | 2,882 | *27%* | 4,813 | Asian: 2,189 | 686 |
| NH NHPI & white^ | 1,498 | 358 | *24%* | 724 | NHPI: 294 | 122 |

| Response in household year | Total N | Proxy response matched: Race & Hispanic N | Race & Hispanic % | Hispanic only N | Hispanic only % | Proxy response did not match. Proxy said… *Hispanic white alone* | *Hispanic SOR alone* | *All other responses* |
|---|---|---|---|---|---|---|---|---|
| Hisp. white alone | 166,552 | 79,166 | *48%* | 139,838 | *84%* | (matched) | 56,484 | 30,902 |
| Hisp. SOR alone | 121,304 | 62,320 | *51%* | 113,260 | *93%* | 46,952 | (matched) | 12,032 |

Data are from Census 2000 and the 2010 Census.

* Match = both household and proxy reported non-Hispanic and gave the same general single race group.

^ Match = both household and proxy reported non-Hispanic and the proxy reported that the person was of (any) Two or More Races.

NH = Non-Hispanic; AIAN = American Indian/Alaska Native; NHPI = Native Hawaiian and other Pacific Islander; SOR = Some Other Race.

It has grown in prominence recently as more and more organizations comply with the revised federal guidelines (Office of Management and Budget 1997). Proxies also have low success rates for matching the race reports for Hispanics (many Hispanics also have trouble identifying their own race on this question that has no "Hispanic" category; Compton et al. 2012). Fewer than half of Hispanic whites (48%) were reported as both Hispanic and white by their proxy and about half (51%) of Hispanics of Some Other Race were reported as Hispanic Some Other Race by their proxy.

In sum, proxy reports of race and Hispanic origin mirrored household reports for single-race non-Hispanic whites, blacks, and Asians. Proxies also successfully identified Hispanics, especially Hispanics of Some Other Race. In other groups, particularly the multiple-race groups, proxy reports matched household reports much less often. As discussed below, we see social assumptions about race when we look more closely at patterns in the non-matches.

*What predicts a matched response?*

In our second research question, we ask: In what circumstances does the proxy response match the household response? We address this question for ten race and Hispanic origin groups by using descriptive statistics and logistic regression models which predict whether or not a proxy response matched the household response (Tables 3 and 4). We note that the $r^2$ model fit[13] is especially good for the single-race American Indians ($r^2$ = 0.52; Table 3) but is reasonably good for all groups ($r^2$ = between 0.11 and 0.27 in Table 3).

***Norms about claiming or belonging to a particular group:*** The response match patterns we observe may relate to social norms about racial and ethnic identification. The high match

[13] We also present the AIC fit statistic which adjusts for very large sample sizes (Akaike 1976).

Table 3: Which characteristics predict a matched proxy response? Odds ratios from eight logistic regression models.

| | Household-year response was non-Hispanic and… | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Black & White | AIAN & White | Asian & White | NHPI & White | Black | AIAN | Asian | NHPI |
| *Observations of the household* | | | | | | | | |
| Age 0-9 | 2.36 *** | 1.80 *** | 1.99 *** | 2.42 *** | 0.78 *** | 0.76 *** | n.s. | n.s. |
| Age 10-18 | 2.00 *** | 1.62 *** | 1.82 *** | 1.77 ** | 0.80 *** | 0.90 * | 0.88 ** | 0.79 * |
| Age 65+ | 0.33 *** | n.s. | 0.32 *** | n.s. | 1.11 *** | 1.32 *** | 1.74 *** | n.s. |
| Proxy report of age within 2 years | 1.51 *** | 2.26 *** | 1.74 *** | 1.70 * | 1.13 *** | 1.24 *** | 1.26 *** | 1.32 * |
| Age was imputed in either year | 0.64 *** | 0.48 *** | 0.75 *** | n.s. | 0.62 *** | 0.69 *** | 0.66 *** | n.s. |
| Family HH with no minor child | 1.61 *** | 1.39 *** | n.s. | n.s. | 1.15 *** | 1.28 *** | 1.67 *** | 1.59 *** |
| Family HH with minor child | 1.75 *** | 1.43 *** | n.s. | n.s. | 1.55 *** | 1.29 *** | 1.55 *** | 1.77 *** |
| Non-family household | n.s. | n.s. | 0.83 * | n.s. | 0.94 ** | n.s. | 0.90 ** | 0.64 ** |
| Renter | 0.83 *** | 0.86 * | 0.59 *** | n.s. | n.s. | 1.23 *** | 0.66 *** | n.s. |
| No rent paid | n.s. | 1.46 ** | n.s. | n.s. | 1.12 ** | n.s. | n.s. | n.s. |
| *Local Racial Composition* | | | | | | | | |
| Urban area | n.s. | 1.18 ** | 1.23 ** | 1.91 *** | 1.04 * | 0.93 * | 0.91 * | n.s. |
| Log % in tract NH single-race white | | | | | 0.93 *** | 0.82 *** | 1.09 *** | n.s. |
| Log % in tract NH single-race black | 0.92 *** | | | | 1.82 *** | | | |
| Log % in tract NH single-race AIAN | | n.s. | | | | 2.46 *** | | |
| Log % in tract NH single-race Asian | | | n.s. | | | | 1.81 *** | |
| Log % in tract NH single-race NHPI | | | | 1.06 *** | | | | 2.00 *** |
| Log % in tract NH 2+ races | 1.53 *** | 1.66 *** | 1.82 *** | 1.72 *** | 0.68 *** | 0.42 *** | 0.67 *** | 0.34 *** |
| *Local Socioeconomic Standing* | | | | | | | | |
| % in tract using public assistance | 0.98 * | 1.04 * | n.s. | n.s. | 1.01 *** | 1.03 *** | 1.05 *** | n.s. |
| Avg. adult income in tract/$10,000 | n.s. | n.s. | n.s. | n.s. | 1.03 *** | n.s. | 0.98 *** | n.s. |
| % adults in tract w/ HS educ+ | 1.01 * | 1.02 *** | 1.01 * | n.s. | 1.00 *** | 1.02 *** | 0.99 *** | n.s. |
| *Control Variables* | | | | | | | | |
| Proxy report was in 2010 | 2.55 *** | 1.80 *** | 1.32 *** | n.s. | 1.06 *** | 1.10 ** | 0.95 * | 1.17 * |
| Female | n.s. | n.s. | n.s. | 0.75 * | 1.12 *** | 1.08 ** | 0.95 * | n.s. |
| Intercept | 0.41 ** | 0.04 *** | n.s. | n.s. | 7.45 *** | 0.09 *** | 16.54 *** | n.s. |
| N in model | 14,977 | 19,064 | 10,575 | 1,499 | 564,945 | 45,230 | 98,988 | 5,198 |
| *r-square* | 0.156 | 0.108 | 0.152 | 0.221 | 0.139 | 0.516 | 0.123 | 0.273 |
| AIC (model fit statistic) | 19,579 | 10,747 | 12,391 | 1,650 | 264,420 | 60,274 | 72,391 | 6,932 |

Data are from Census 2000 and the 2010 Census. All non-matched responses are included in these models. Comparison categories not listed are: Ages 19-64, living alone, home owner, and rural area.

*** p<=0.001; ** p<= 0.01; * p<=0.05; n.s. = the coeffient was not sigificant.

HH = Household; NH = Non-Hispanic; AIAN = American Indian/Alaska Native; NHPI = Native Hawaiian and other Pacific Islander; SOR = Some Other Race.

Table 4: Which characteristics predict a matched proxy Hispanic origin response? How is this different if the race response must also match? Odds ratios from four logistic regression models.

| | Predict match of Hispanic and exact race response | | Predict match of only Hispanic, regardless of proxy race response | |
|---|---|---|---|---|
| *Household-year response was …* | Hispanic White | Hispanic SOR | Hispanic White | Hispanic SOR |
| *Observations of the household* | | | | |
| Age 0-9 | 0.96 * | 0.94 * | 0.65 *** | n.s. |
| Age 10-18 | 0.93 *** | 0.96 * | 0.88 *** | n.s. |
| Age 65+ | 1.16 *** | n.s. | 1.29 *** | 1.21 *** |
| Proxy report of age within 2 years | 1.09 *** | 1.13 *** | 1.32 *** | 1.42 *** |
| Age was imputed in either year | 0.87 *** | 0.93 *** | 0.64 *** | 0.65 *** |
| Family HH with no minor child | 1.05 ** | 1.25 *** | 1.57 *** | 1.68 *** |
| Family HH with minor child | 1.11 *** | 1.16 *** | 1.73 *** | 1.94 *** |
| Non-family household | 0.93 ** | n.s. | 0.92 ** | 1.10 * |
| Renter | n.s. | 0.87 *** | 0.94 *** | 0.83 *** |
| No rent paid | 1.19 *** | n.s. | 1.40 *** | n.s. |
| *Local Racial Composition* | | | | |
| Urban area | 1.08 *** | 0.89 *** | 1.08 ** | 1.21 *** |
| Log % in tract NH single-race white | n.s. | 0.90 *** | 0.71 *** | 0.95 * |
| Log % in tract Hispanic single-race white | 3.06 *** | | 1.55 *** | |
| Log % in tract Hispanic single-race SOR | | 3.74 *** | | 1.29 *** |
| Log % in tract Hispanic non-white | 0.46 *** | | 1.17 *** | |
| Log % in tract Hispanic non-SOR | | 0.27 *** | | 1.31 *** |
| *Local Socioeconomic Standing* | | | | |
| % in tract using public assistance | n.s. | 0.99 *** | 0.98 *** | 0.97 *** |
| Avg. adult income in tract/$10,000 | 0.95 *** | 1.06 *** | n.s. | n.s. |
| % adults in tract w/ HS educ+ | 1.00 *** | 1.00 *** | 0.99 *** | 0.98 *** |
| *Control Variables* | | | | |
| Proxy report was in 2010 | 0.73 *** | 1.41 *** | 1.04 * | 1.14 *** |
| Female | 0.96 *** | 0.96 ** | 0.83 *** | 0.74 *** |
| Intercept | 1.26 *** | 0.71 *** | 37.76 *** | 143.47 *** |
| N in model | 166,550 | 121,304 | 166,550 | 121,304 |
| *r-square* | 0.207 | 0.189 | 0.253 | 0.139 |
| AIC (model fit statistic) | 230,484 | 168,073 | 146,668 | 59,197 |

Data are from Census 2000 and the 2010 Census. All non-matched responses are included in these models. Comparison categories not listed are: Ages 19-64, living alone, home owner, and rural area.

*** p<=0.001; ** p<= 0.01; * p<=0.05; n.s. = the coeffient was not sigificant.

HH = Household; NH = Non-Hispanic; SOR = Some Other Race.

rates for whites and blacks highlight the extent to which these socially constructed categories are clearly defined and consistently enforced through norms such as hypodescent for blacks. The relatively high Asian and Hispanic match rates may reflect the influence that recent immigration and/or distinctive names have on identification for these groups. Weaker or inconsistent norms may govern American Indians, Pacific Islanders, multiracial groups and also Hispanics when reporting a race. For instance, conflict about who counts as American Indian may result in mismatches between proxies and the household, as well as relatively high levels of racial fluidity (within-person race response change) in this group.

Throughout our results (Table 3; also in Tables 5 and 6 below), proxies are more likely to report a child as multiple-race than an adult, whether or not the person was actually reported as multiple-race in the household year; this pattern is clearer for younger children. In other words, proxies are more likely to provide a matched response for multiple-race children (compared to multiple-race adults) and they are more able to provide a matched report for single-race adults (compared to single-race children). Multiracial elders are rarely reported as multiple-race by proxies. These results give evidence of age-related societal ideas about what kinds of people identify as multiracial; proxies may be assuming that children of mixed heritage should be reported as multiple-race, and that multiple-race identities are a new phenomenon and thus there are unlikely to be multiracial elders. This may make it more difficult for elders who might otherwise identify as multiracial to maintain a multiracial identity (Cooley 1902; Tashiro 2011), potentially reducing the enumerated size of the multiple-race elder population.

***Observations or knowledge about the household***: In Tables 3 and 4, a match is positively predicted when the proxy knows the age of the person. As expected, a proxy is more likely to provide a matched response for a person living in a family (whether or not there is a co-

resident child) than for someone who lives alone. Homeownership (as reported by the proxy) is associated with higher odds of a match for Asians, multiple-race groups, and Hispanics of Some Other Race (Tables 3 and 4), whereas American Indians more often have matched responses if they are renters. Proxies may be relying on positive and negative stereotypes (Lee and Zhou 2015; Penner and Saperstein 2008), for example stereotyping Asians as homeowners but not American Indians.

***Local racial/ethnic composition***: We include several measures of the racial and ethnic composition of the tract.[14] We find that each of these is a powerful predictor of whether the proxy provides a correct response.[15] The likelihood that the proxy response will match is much higher if the person is living in an area with many others who report the same race/Hispanic group. As the share of the co-ethnic group in the tract increases, so do the odds of a match. We find this for the Pacific Islander-whites, all single-race groups, and the Hispanic groups. The one exception is notable. For non-Hispanic black-white people (Table 3), the odds of a match (defined as any non-Hispanic multiple-race response) decrease as the area's population of non-Hispanic single-race blacks increases.

Overall, the power of these measures of local racial/ethnic composition suggest that proxy respondents use local race and Hispanic composition as cues to understand the race(s) and Hispanic origins of people in that area. This implies that people living away from their

---

[14] For indigenous people, connection to the land is related to identity (and perhaps outsider identification) above and beyond the racial composition of the area (Kana'iaupuni and Liebler 2005; Liebler 2010). We do not use separate measures of living on a reservation and living in Hawaii because they are quite correlated with the percent non-Hispanic single-race American Indians in the tract ($r = 0.55$) and percent non-Hispanic single-race Pacific Islanders in the tract ($r = 0.69$).

[15] Without measures of tract race/ethnic composition, model fit is notably worse for all groups and dramatically worse for single-race blacks, Asians, and American Indians. Results are available on request.

race/Hispanic group, or people from very small groups who are not generally known to be living in the area, are relatively unlikely to be reported as members of that group when reported by the proxy. This pattern may bias segregation measures toward a higher index of segregation when proxy responses are used. Also, to the extent that their self-identified race is not supported by outsider's reflections (Khanna 2004), a person who does not live close to co-ethnics may have a more complex racial identity experience, potentially contributing to fluid identity and/or a feeling of being racially invisible.

***Local socioeconomic standing:*** We expected that socioeconomic status characteristics of people in the tract would be predictive of a match when the matches followed stereotypes. We find some support for this idea. In tracts with more people using public assistance, proxies are more likely to match responses of many of the minority groups. Tract-level education is not strongly related to the likelihood of a matched response.

*What predicts a particular unmatched response?*

In ten percent of our cases, the proxy's response does not match the household response (see Table 2). To understand factors predicting who receives which non-matched response, we discuss descriptive statistics (Table 2) and estimated ten multinomial logistic regression models (shown in Tables 5, 6, and 7). The measured characteristics we include in these multinomial logistic regression models give very good model fit. For example, values of $r^2$ range from 0.16 to 0.37 among the multiple-race groups shown in Table 5. The specific non-matching response given for single-race American Indians is quite predictable ($r^2$ = 0.50 in Table 6) with other $r^2$ values ranging from 0.14 to 0.37 for single-race groups in Table 6. The non-matched response given for Hispanics of Some Other Race and Hispanic whites are also quite able to be predicted with our measures ($r^2$ = 0.44 and 0.34, respectively; Table 7).

Table 5: Multinomial logistic regression models predicting a specific non-matching proxy response, for four non-Hispanic multiple-race samples: Relative risk ratios as compared to a match

| | Response in household year was non-Hispanic and … | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Black & White | | AIAN & White | | Asian & White | | NHPI & White | |
| *Proxy said non-Hispanic …* | *Black* | *White* | *AIAN* | *White* | *Asian* | *White* | *NHPI* | *White* |
| *Observations of the household* | | | | | | | | |
| Age 0-9 | 0.35 *** | 0.53 *** | 0.62 *** | 0.53 *** | 0.47 *** | 0.55 *** | 0.31 *** | 0.47 ** |
| Age 10-18 | 0.46 *** | 0.57 *** | n.s. | 0.54 *** | 0.54 *** | 0.56 *** | n.s. | 0.40 *** |
| Age 65+ | 3.09 *** | 2.98 *** | n.s. | n.s. | 3.20 *** | 3.36 *** | n.s. | n.s. |
| Proxy report of age within 2 years | 0.67 *** | 0.64 *** | 0.71 * | 0.42 *** | 0.64 *** | 0.56 *** | n.s. | 0.54 * |
| Age was imputed in either year | 1.49 *** | 1.76 *** | 1.41 * | 2.23 *** | n.s. | 1.51 *** | n.s. | n.s. |
| Family HH with no minor child | 0.63 *** | 0.59 *** | n.s. | 0.68 *** | n.s. | n.s. | n.s. | n.s. |
| Family HH with minor child | 0.67 *** | 0.45 *** | n.s. | 0.66 *** | n.s. | 0.81 ** | n.s. | 0.59 * |
| Non-family household | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. |
| Renter | 1.30 *** | 1.13 * | n.s. | 1.16 * | 1.49 *** | 1.72 *** | n.s. | n.s. |
| No rent paid | n.s. | n.s. | 0.56 *** | 0.72 ** | n.s. | n.s. | n.s. | n.s. |
| *Local Racial Composition* | | | | | | | | |
| Urban area | n.s. | n.s. | 0.81 ** | 0.77 *** | n.s. | 0.82 * | 0.52 * | 0.49 *** |
| Log % this NH single-race group in tract | 1.33 *** | 0.90 *** | 1.93 *** | 0.91 *** | 1.34 *** | 0.94 ** | 1.48 *** | 0.92 *** |
| Log % in tract NH 2+ races | 0.63 *** | 0.69 *** | 0.47 *** | 0.58 *** | 0.46 *** | 0.51 *** | 0.40 *** | 0.45 *** |
| *Local Socioeconomic Standing* | | | | | | | | |
| % in tract using public assistance | 1.03 * | n.s. | 0.96 * | 0.94 *** | n.s. | n.s. | n.s. | n.s. |
| Avg. adult income in tract/$10,000 | n.s. | 1.06 ** | n.s. | n.s. | 0.94 ** | n.s. | n.s. | n.s. |
| % adults in tract w/ HS educ+ | n.s. | 0.99 * | 0.99 * | 0.99 ** | 0.99 ** | n.s. | n.s. | n.s. |
| *Control Variables* | | | | | | | | |
| Proxy report was in 2010 | 0.40 *** | 0.37 *** | 0.63 *** | 0.54 *** | 0.75 *** | 0.76 *** | n.s. | n.s. |
| Female | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | 1.36 * |
| Intercept | n.s. | n.s. | 6.16 *** | 11.66 *** | n.s. | 0.29 *** | n.s. | n.s. |
| N in model | 14,222 | | 18,530 | | 9,884 | | 1,376 | |
| *r-square* | 0.194 | | 0.261 | | 0.164 | | 0.374 | |
| AIC (model fit statistic) | 30,750 | | 23,306 | | 20,634 | | 2,805 | |

Data are from Census 2000 and the 2010 Census. Only people with the proxy race/Hispanic responses shown are included in each model. Comparison categories not listed are: Ages 19-64, living alone, home owner, and rural area.

*** p<=0.001; ** p<= 0.01; * p<=0.05; n.s. = the coeffient was not sigificant.

HH = Household; NH = Non-Hispanic; AIAN = American Indian/Alaska Native; NHPI = Native Hawaiian and other Pacific Islander.

Table 6: Multinomial logistic regression models predicting a specific non-matching proxy response, for four non-Hispanic single-race samples: Relative risk ratios as compared to a match

| | Response in household year was non-Hispanic and… | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Black | | AIAN | | Asian | | NHPI | |
| *Proxy said non-Hispanic …* | *White* | *2+ races* | *White* | *2+ races* | *White* | *2+ races* | *White* | *2+ races* |
| *Observations of the household* | | | | | | | | |
| Age 0-9 | 0.92 * | 2.89 *** | 1.14 * | 1.88 *** | n.s. | 1.64 *** | n.s. | n.s. |
| Age 10-18 | 0.74 *** | 2.03 *** | 0.86 ** | 1.41 *** | n.s. | 1.62 *** | n.s. | 1.66 *** |
| Age 65+ | 0.89 *** | n.s. | 0.78 *** | 0.67 *** | 0.62 *** | 0.55 *** | n.s. | n.s. |
| Proxy report of age within 2 years | 0.80 *** | 1.19 *** | 0.71 *** | 1.52 *** | 0.74 *** | 0.87 * | 0.58 *** | n.s. |
| Age was imputed in either year | 2.14 *** | 0.66 *** | 1.66 *** | n.s. | 2.13 *** | 0.76 *** | 1.42 * | 0.61 ** |
| Family HH with no minor child | 0.73 *** | 1.34 *** | 0.72 *** | 0.88 * | 0.53 *** | 0.77 *** | 0.46 *** | n.s. |
| Family HH with minor child | 0.43 *** | 1.44 *** | 0.67 *** | n.s. | 0.50 *** | n.s. | 0.38 *** | n.s. |
| Non-family household | n.s. | 1.13 * | n.s. | n.s. | 1.17 *** | 1.21 * | 1.46 * | 2.00 *** |
| Renter | 1.06 ** | 0.76 *** | 0.82 *** | 0.64 *** | 1.65 *** | n.s. | n.s. | 0.64 *** |
| No rent paid | 0.83 *** | n.s. | 0.78 ** | n.s. | 0.77 * | 1.42 ** | n.s. | n.s. |
| *Local Racial Composition* | | | | | | | | |
| Urban area | 1.05 * | 0.81 *** | n.s. | 1.40 *** | n.s. | 1.22 ** | n.s. | n.s. |
| Log % non-Hispanic white in tract | 1.43 *** | n.s. | 2.91 *** | 1.39 *** | 1.37 *** | n.s. | 1.63 *** | n.s. |
| Log % this NH single-race group in tract | 0.57 *** | 0.55 *** | 0.40 *** | 0.49 *** | 0.56 *** | 0.56 *** | 0.53 *** | 0.53 *** |
| Log % in tract NH 2+ races | 1.27 *** | 2.12 *** | 2.40 *** | 3.33 *** | 1.21 *** | 1.99 *** | 1.92 *** | 5.25 *** |
| *Local Socioeconomic Standing* | | | | | | | | |
| % in tract using public assistance | 0.96 *** | 1.01 * | 0.96 *** | n.s. | 0.95 *** | 0.98 * | n.s. | n.s. |
| Avg. adult income in tract/$10,000 | 0.97 *** | 0.91 *** | n.s. | 1.07 * | 1.03 ** | 1.03 * | n.s. | n.s. |
| % adults in tract w/ HS educ+ | 1.00 ** | 1.01 *** | 0.97 *** | 0.98 *** | 1.01 *** | n.s. | n.s. | n.s. |
| *Control Variables* | | | | | | | | |
| Proxy report was in 2010 | 0.70 *** | 2.23 *** | 0.88 *** | 1.47 *** | n.s. | 1.42 *** | 0.73 ** | n.s. |
| Female | 0.87 *** | n.s. | 0.87 *** | n.s. | 1.18 *** | n.s. | n.s. | n.s. |
| Intercept | 0.08 *** | 0.03 *** | 33.05 *** | 3.38 *** | 0.01 *** | 0.05 *** | n.s. | n.s. |
| N in model | 558,351 | | 42,201 | | 95,329 | | 4,515 | |
| *r-square* | 0.181 | | 0.503 | | 0.139 | | 0.370 | |
| AIC (model fit statistic) | 258,654 | | 68,820 | | 66,323 | | 7,286 | |

Data are from Census 2000 and the 2010 Census. Only people with the proxy race/Hispanic responses shown are included in each model. Comparison categories not listed are: Ages 19-64, living alone, home owner, and rural area.

*** p<=0.001; ** p<= 0.01; * p<=0.05; n.s. = the coeffient was not sigificant.

HH = Household; NH = Non-Hispanic; AIAN = American Indian/Alaska Native; NHPI = Native Hawaiian and other Pacific Islander.

Table 7: Multinomial logistic regression models predicting a specific non-matching proxy response, for two Hispanic single-race samples: Relative risk ratios as compared to a match

| | Response in household year was … | | | |
|---|---|---|---|---|
| | Hispanic white | | Hispanic of Some Other Race | |
| *Proxy said …* | *Non-Hispanic white* | *Hispanic, non-white* | *Non-Hispanic white* | *Hispanic, non-SOR* |
| *Observations of the household* | | | | |
| Age 0-9 | 1.49 *** | 0.92 *** | 1.62 *** | 1.11 *** |
| Age 10-18 | 1.16 *** | 1.04 * | 1.24 *** | 1.11 *** |
| Age 65+ | 0.76 *** | 0.92 *** | 1.13 *** | 1.48 *** |
| Proxy report of age within 2 years | 0.76 *** | n.s. | 0.77 *** | n.s. |
| Age was imputed in either year | 1.56 *** | n.s. | 1.55 *** | n.s. |
| Family HH with no minor child | 0.66 *** | 1.13 *** | 0.47 *** | 0.71 *** |
| Family HH with minor child | 0.60 *** | 1.09 *** | 0.41 *** | 0.69 *** |
| Non-family household | 1.14 *** | n.s. | n.s. | n.s. |
| Renter | n.s. | 0.95 *** | n.s. | 0.97 * |
| No rent paid | 0.72 *** | 0.91 ** | 0.78 *** | n.s. |
| *Local Racial Composition* | | | | |
| Urban area | 0.91 *** | n.s. | 1.06 * | 1.19 *** |
| Log % non-Hispanic white in tract | 1.41 *** | 0.97 ** | 1.66 *** | 1.13 *** |
| Log % Hispanic white in tract | 0.47 *** | 0.29 *** | | |
| Log % Hispanic SOR in tract | | | 0.17 *** | 0.16 *** |
| Log % Hispanic non-white in tract | 1.16 *** | 3.84 *** | | |
| Log % Hispanic non-SOR in tract | | | 3.56 *** | 6.92 *** |
| *Local Socioeconomic Standing* | | | | |
| % in tract using public assistance | 1.02 *** | 0.98 *** | n.s. | 0.97 *** |
| Avg. adult income in tract/$10,000 | 1.01 * | 1.03 *** | 0.97 *** | 0.97 *** |
| % adults in tract w/ HS educ+ | 1.01 *** | 1.00 ** | 1.01 *** | 1.00 *** |
| *Control Variables* | | | | |
| Proxy report was in 2010 | 1.10 *** | 1.45 *** | 0.40 *** | 0.38 *** |
| Female | 1.19 *** | n.s. | 1.23 *** | 1.04 *** |
| Intercept | 0.06 *** | 0.84 ** | 0.16 *** | 2.38 *** |
| N in model | 164,793 | | 170,629 | |
| *r-square* | 0.337 | | 0.435 | |
| AIC (model fit statistic) | 331,542.8 | | 340,919.2 | |

Data are from Census 2000 and the 2010 Census. Only people with the proxy race/Hispanic responses shown are included in each model. Comparison categories not listed are: Ages 19-64, living alone, home owner, and rural area.

*** p<=0.001; ** p<= 0.01; * p<=0.05; n.s. = the coeffient was not sigificant.

HH = Household; SOR = Some Other Race.

***Norms about claiming or belonging to a particular group:*** Non-Hispanic single-race white is the most common non-matching race/ethnicity reported for all single-race minority groups in our data (see Table 2). Consistent with the idea of hyperdescent (Snipp 2003), the vast majority of American Indian-whites (78%) were reported as single-race white. Asian-whites and Pacific Islander-whites are also often misreported reported as non-Hispanic white. There are several potential reasons for proxies to tend toward a non-Hispanic white response. First, social rules (e.g., hyperdescent) may impact assumptions about who should be considered white. Second, centuries of interracial unions have created some people who may seem phenotypically white but who identify with a different group. Third, a person might not have discussed race and Hispanic origin with their proxy. Because non-Hispanic whites are often not described racially (Bonilla-Silva 2012), the lack of discussion might lead the proxy to assume the person is white. Fourth, some may be living in predominantly white areas and be assumed to be like their neighbors. Finally, record linkage is more successful for people with higher socioeconomic standing (Bond et al. 2014) – the same people who are more often assumed to be white (Saperstein and Penner 2012; Vargas 2015).

One group does not follow this pattern. When not reported as biracial, black-white people were more commonly reported as single-race black than as single-race white. This might reflect long-standing norms that define blacks by rules of hypodescent (i.e., part-blacks are seen as just black; Davis 2001).

***Observations or knowledge about the household:*** In many cases, a proxy whose age report was (relatively) accurate was more likely to give a matched race and Hispanic origin response (Tables 3 through 7). The exceptions are interesting. For example, when the proxy's age report was accurate, the proxy tended to report single-race blacks and American Indians as

multiple-race (Table 6). In cases where age was imputed in either year, the proxy tended to give a non-Hispanic white response for a biracial American Indian-white, black-white, Asian-white (Table 5) or any of the single-race minorities (Tables 6 and 7). If the need for age imputation is caused by unfamiliarity with the person, this is consistent with the idea that people in the United States assume a person is non-Hispanic white if their race is not mentioned (Bonilla-Silva 2012).

In most models, our results are consistent with the idea that people in family households are more well-known to proxy respondents; proxies more often give a matched response in these cases. However, single-race blacks in families tend to be reported as multiple-race (Table 6) and Hispanic whites in families tend to be reported as Hispanic non-white (Table 7).

***Local racial/ethnic composition:*** Mismatching proxy responses closely follow the tract's race and Hispanic composition. For example, a non-matched response for a biracial white-minority person in our data is extremely likely to be their single-race minority group if that group is relatively prevalent in the tract, and their odds of being misreported as single-race white are low in those places (Table 5). These variables measuring the tract's racial/ethnic composition are powerfully predictive of matched and unmatched responses across all groups in our models (Tables 5, 6, and 7).

Taking other factors into account, we do not see evidence that people in urban areas are assumed to be minorities. Instead, there is a tendency for proxies to report urban people as multiracial. Proxy reporters tend to give matched responses for urban-dwelling multiracial people and assign multiracial status to single-race American Indians and Asians who are living in urban areas (Tables 5 and 6).

***Local socioeconomic standing:*** There is some evidence that stereotypes are informing proxy responses. In tracts where public assistance income is more common, proxies tend to

either match the (non-white) household report or give a non-white unmatched response (Tables 5 and 6). As the average income in the tract rises, black-whites are more often reported as white, single-race Asians are more often reported as white or multiracial, and American Indians are more often reported as multiracial. On the other hand, blacks and Hispanics of Some Other Race in areas with higher mean income more often have a matched response (Tables 5 and 7); in these higher-income areas, these minorities are clearly identified by their proxies as people of color.

People in tracts with more highly educated residents have a slight tendency to give multiple-race responses for multiple-race people (Table 5). Perhaps experiences in higher education raise awareness of multiracial people and nuances of racial identity, and even those proxies without such education have learned from their neighbors.

Although there is some evidence that socioeconomic characteristics of the area translate into particular non-matched responses, the relative risk ratios are small and many coefficients are not significant. These measures of local socioeconomic standing have less predictive power than the measures of the race/ethnic composition of the tract; in supplementary analyses (available on request), model fit is much worse in models with local socioeconomic standing measures but not measures of local race/ethnic composition. For example, the $r^2$ values for models in Table 6 are half as large when the measures of tract race/ethnic composition are removed. Future research on the idea that "money whitens" (people are more often assumed to be white if they have more resources) should include measures of the area's race/ethnic composition.

**Discussion and Conclusion**

In our investigation of outsiders' reports of a person's race and ethnicity, we asked: do non-strangers' reports of a person's race and ethnicity match the reports given by someone in the household? The general answer to this question is "yes." In 90% of the 3.7 million cases in our

study, the proxy respondent reported the same general race and Hispanic origin that was reported by the household in a different census year. It can be said that proxy respondents who know the person's first and last name are a good source for information about the person's race and ethnicity.

We found that there are a few categories that proxy respondents use effectively: they can identify people who are non-Hispanic white, black, or Asian, and they can tell whether a person is Hispanic. This set of larger groups (white, black, Asian, and Hispanic) are (generally) consistently identified by others and (usually) self-identify consistently across time and circumstance (del Pinal and Schmidley 2005; Doyle and Kao 2007; Liebler et al. 2014a). Consistency in a person's self-reports may be related to consistency in how outsiders interpret their race and ethnicity.

Proxy reporters do much worse at matching the race responses of American Indians, Pacific Islanders, multiracial people, and Hispanics. People in these same groups were found to be less consistent in their race/ethnicity responses when two household year race/Hispanic reports (from 2000 and 2010) were compared (Liebler et al. 2014a, 2014b). Answering the race question for a Hispanic person may be challenging because people often view Hispanic as a race (Compton et al. 2012; Hitlin et al. 2007), yet Hispanic was not a race response option on these two censuses.

Proxies may have less day-to-day interaction with people from small race groups (American Indians, Pacific Islanders, and multiracials). Some people in these groups report feeling "invisible" (Higgins 2004) and our results suggest that this is more than a feeling – they are unlikely to be racially identified correctly by an outsider, even one who knows their full name. Our research and prior work has shown that people in these groups have extensive

experience with outsider misidentification and personal changes in identification and identity. For many of them, identity processes and decisions about how to report their race(s) on a form are especially complex and multifaceted. Because their experiences do not seem to be parallel to those of people in the larger, consistently-identified groups, these smaller groups should be considered explicitly in research aiming to understand dynamic and intertwined experiences with identity, identification, and life chances. Survey researchers should consider oversampling people in these groups to achieve sufficient sample sizes to expand researchers' ability to follow this important avenue of research.

Using multivariate models, we addressed ideas about factors that influence proxy respondents to provide one type of mismatched response (e.g., white) versus another (e.g., multiple-race). When reported as single race, multiracial black-whites are most often reported as black (consistent with the idea of hypodescent), while multiracial American Indian-whites were more often reported as white (consistent with the idea of hyperdescent). Misreported Asian-whites and Pacific Islander-whites are more often reported as non-Hispanic white than as people of color. Those perceived to be white may be experiencing some of the advantages conveyed by whiteness and might fit the "honorary white" category described by Bonilla-Silva (2004).

We identified a variety of factors predictive of a match between the proxy response and the response given by the household. Our evidence suggests that proxy respondents rely on outside cues to decide how to report a non-stranger's race/ethnicity – particularly age, household structure, and the racial composition of the area. Proxies tend to identify children and urban residents as multiracial and identify elders and rural residents as single race, whether they are or not. If the tract's population held relatively many people from a race/ethnic group, the proxy was likely to report that group for the person, whether they are or not.

Although proxy respondents in our data match household responses 90% of the time, our data are non-representative and probably overestimate the extent of matches. The Census Bureau has other methods for handling non-responses, including hotdeck imputation in which a nearby household's characteristics are allocated to people in a non-responding household. Both proxy respondents and hotdeck imputation rely on information from the local area when providing race and Hispanic origin information for others. A likely unintended effect of both of these strategies is that they slightly reduce the diversity within census tracts in the data (relative to reality) by mis-categorizing people who are racially or ethnically different from their neighbors. Both proxy responses and hotdeck imputation are affected by residential segregation; data quality may be better for highly segregated areas than in more integrated areas.

An alternate strategy for gathering information about non-responding households is to use household-provided information gathered in a prior census data or from administrative records (Bhaskar et al. 2014; Rastogi et al. 2014) when available. This strategy was used in the 2010 Census with prior census data (Rothhaas et al. 2012) and is being explored for more extensive use in the 2020 Census (Vitrano and Chapin 2012). Although using race and Hispanic origin responses from prior census data or administrative records does not allow for person-level response fluidity, it has the benefit of relying more closely on self-identification rather than assumptions of outsiders. Using a person's (or household member's) response from another data source could avoid potential biases introduced by patterns/biases in how proxies tend to respond. These biases could artificially increase measures of residential segregation, decrease the average age of multiracials, increase the rural/urban divide for multiracials, and reduce the enumerated size of the less-often-matched race/ethnic groups. Each of these could have significant implications if a high proportion of responses are gathered by proxy.

**References**


Akaike, Hirotugu. 1976. “An Information Criterion (AIC).” *The Mathematical Scientist* 14(153): 5-9.

Arias, Elizabeth, William S. Schauman, Karl Eschbach, Paul Sorlie, and Eric Backlund. 2008. “The Validity of Race and Hispanic Origin Reporting on Death Certificates in the United States.” *Vital and Health Statistics* 2(148).

Bailey, Stanley R., Aliya Saperstein, and Andrew M. Penner. 2014. “Race, Color, and Income Inequality across the Americas.” *Demographic Research* 31: 735-756

Bhaskar, Renuka, Adela Luque, Sonya Rastogi, and James Noon. 2014. “Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data.” Center for Administrative Records Research and Applications Working Paper 2014-14. Washington, DC: U.S. Census Bureau.

Bond, Brittany, J. David Brown, Adela Luque, and Amy O’Hara. 2014. “The Nature of Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey.” Center for Administrative Records Research and Applications Working Paper 2014-08. Washington, DC: U.S. Census Bureau.

Bonilla-Silva, Eduardo. 2004. “From Bi-Racial to Tri-Racial: Towards a New System of Racial Stratification in the USA.” *Ethnic and Racial Studies* 27(6): 931-950.

Bonilla-Silva, Eduardo. 2012. “The Invisible Weight of Whiteness: The Racial Grammar of Everyday Life in Contemporary America.” *Ethnic and Racial Studies* 35(2): 173-194.

Brodkin, Karen. 1998. *How Jews Became White Folks & What That Says About Race in America.* New Brunswick, NJ: Rutgers University Press.

Campbell, Mary and Lisa Troyer. 2007. “The Implications of Racial Misclassifcations by Observers.” *American Sociological Review* 72: 750-765.

Chun, Asaph Young and Jessica Gan. 2014. “Analysis of the Source of Group Quarters Enumeration Data in the 2010 Census.” Joint Statistical Meetings Proceedings, Survey Research Methods Session. Alexandria, VA: American Statistical Association.

Clawson, Rosalee A. and Rakuya Trice. 2000. “Poverty as We Know It: Media Portrayals of the Poor.” *The Public Opinion Quarterly* 64(1): 53-64.

Compton, Elizabeth, Michael Bentley, Sharon Ennis, and Sonya Rastogi. 2012. “2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment.” 2010 Census Program for Evaluations and Experiments, 2010 Census Planning Memoranda Series No. 211. Washington, DC: U.S. Census Bureau.

Cooley, Charles Horton. 1902. *Human Nature and the Social Order*. NY: Scribner’s Sons.

Cotton, John L., Bonnie S. O'Neill, and Andrea Griffin. 2008. "The 'Name Game': Affective and Hiring Reactions to First Names." *Journal of Managerial Psychology* 23(1):18-39.

Crowder, Kyle, Jeremy Pais, and Scott J. South. 2012. "Neighborhood Diversity, Metropolitan Constraints, and Household Migration." *American Sociological Review* 77(3): 325-353.

DaCosta, Kimberly McClain. 2007. *Making Multiracials: State, Family, and Market in the Redrawing of the Color Line.* Stanford: Stanford University Press.

Davis, James F. 2001. *Who is Black? One Nation's Definition.* 10th Anniversary Edition. Pennsylvania State University Press, University Park, PA.

del Pinal, Jorge and Dianne Schmidley. 2005. "Matched Race and Hispanic Origin Responses from Census 2000 and Current Population Survey February to May 2000." Population Division Working Paper No. 79. Washington, DC: U.S. Census Bureau

Dowling, Julie A. 2014. *Mexican Americans and the question of race*. Austin, TX: University of Texas Press.

Doyle, Jamie M. and Grace Kao. 2007. "Are Racial Identities of Multiracials Stable? Changing Self-Identification among Single and Multiple Race Individuals." *Social Psychology Quarterly* 70(4):405-423.

Fellegi, Ivan P., and Sunter, Alan B. 1969. A theory for record linkage. *Journal of the American Statistical Association, 64,* 1183-1210.

Fredrickson, George M. 2003. "The historical construction of race and citizenship in the United States." Identities, Conflict and Cohesion Programme Paper 1. United Nations Research Institute for Social Development.

Freeman, Jonathan B., Andrew M. Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady. 2011. "Looking the Part: Social Status Cues Shape Race Perception." *PLoS ONE* 6(9).

Grieco, Elizabeth M. 2010. "Race and Hispanic Origin of the Foreign-Born Population in the United States: 2007." American Community Survey Reports ACS-11. Washington, DC: U.S. Census Bureau.

Haney López, Ian. 1996. *White by Law: The Legal Construction of Race.* New York: New York University Press.

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

Herman, Melissa R. 2010. "Do You See What I Am?: How Observers' Backgrounds Affect Their Perceptions of Multiracial Faces." *Social Psychology Quarterly* 73(1): 58-78.

Higgins, Nicholas P. 2004. *Understanding the Chiapas Rebellion: Modernist Visions and the Invisible Indian.* Austin: University of Texas Press.

Hitlin, Steven, J. Scott Brown, and Glen H. Elder. 2007. "Measuring Latinos: Racial vs. Ethnic Classification and Self-Understandings." *Social Forces* 86(2): 587-611.

Hollinger, David A. 2003. "Amalgamation and Hypodescent: The Question of Ethnoracial Mixture in the History of the United States." *The American Historical Review* 108(5):1363-1390.

Iceland, John and Daniel H. Weinberg. 2002. *Racial and Ethnic Residential Segregation in the United States 1980-2000.* Washington, DC: U.S. Census Bureau.

Kana'iaupuni, Shawn and Carolyn A. Liebler. 2005. "Pondering Poi Dog: The Importance of Place to the Racial Identification of Mixed-race Native Hawaiians." *Ethnic and Racial Studies* 28:687–721.

Khanna, Nikki. 2004. "The Role of Reflected Appraisals in Racial Identity: The Case of Multiracial Asians." *Social Psychology Quarterly* 67(2):115-131.

Kramer, Rory, Ruth Burke, and Camille Z. Charles. 2015. "When change doesn't matter: Racial identity (in)consistency and adolescent well-being." *Sociology of Race and Ethnicity, 1,* 270-286.

Lauderdale, Diane S. and Bert Kestenbaum. 2000. "Asian American Ethnic Identification by Surname." *Population Research and Policy Review* 19:283-300.

Lee, Jennifer. 2008. "A Post-Racial America? Multiracial Identification and the Color Line in the 21st Century." *Nanzan Review of American Studies* 30: 13-31.

Lee, Jenifer and Min Zhou. 2015. *The Asian American Achievement Paradox*. NY: Russell Sage.

Liebler, Carolyn A. 2010. "Homelands and Indigenous Identities in a Multiracial Era." *Social Science Research* 39: 596-609.

Liebler, Carolyn A. and Marie DeRousse-Wu. 2012. "Intergenerational Transmission of Race: Permeable Boundaries between 1970 and 2010." Center for Economic Studies Working Paper 12-23. Washington, DC: U.S. Census Bureau.

Liebler, Carolyn A., Sonya Rastogi, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis 2014a. "America's Churning Races: Race and Ethnic Response Changes between Census 2000 and the 2010 Census." Center for Administrative Records Research and Applications Working Paper 2014-09. Washington, DC: U.S. Census Bureau.

Liebler, Carolyn A., Renuka Bhaskar, and Sonya Rastogi. 2014b. "Dynamics of Race: Joining, Leaving, and Staying in the American Indian/Alaska Native Race Category between 2000 and 2010." Center for Administrative Records Research and Applications Working Paper 2014-10. Washington, DC: U.S. Census Bureau.

Logan, John. 2003. “How Race Counts for Hispanic Americans.” Lewis Mumford Center for Comparative Urban and Regional Research. University at Albany, State University of New York.

Mondloch, Catherine, Natalie Elms, Daphne Maurer, Gillian Rhodes, William Hayward, James Tanaka, and Guomei Zhou. 2010. “Processes Underlying the Cross-race Effect: An Investigation of Holistic, Featural, and Relational Processing of Own-race versus Other-race Faces.” *Perception* 39: 1065-1085.

Moses, Joy. 2012. “Moving Away from Racial Stereotypes in Poverty Policy.” Washington, DC: Center for American Progress.

Office of Management and Budget. 1997. “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.” *Federal Register*, October 30, 1997. http://www.whitehouse.gov/omb/fedreg_1997standards

Painter, Nell Irvin. *The History of White People.* 2010. New York: W.W. Norton & Co.

Peery, Destiny and Galen V. Bodenhausen. 2008. “Hypodescent in Reflexive Categorization of Racially Ambiguous Faces.” *Psychological Science* 19(10): 973-977.

Penner, Andrew M. and Aliya Saperstein. 2008. “How Social Status Shapes Race.” *Proceedings of the National Academy of Sciences of the United States of America* 105(50): 19628-19630.

Rastogi, Sonya, Leticia Fernandez, James Noon, Ellen Zapata, and Renuka Bhaskar. 2014. “Exploring Administrative Records for Race and Hispanic Origin Item Non-Response.” Center for Administrative Records Research and Applications Working Paper 2014-06. Washington, DC: U.S. Census Bureau.

Rothhaas, Cynthia, Frederic Lestina, and Joan M. Hill. 2012. “2010 Decennial Census: Item Nonresponse and Imputation Assessment Report.” 2010 Census Program for Evaluations and Experiments, 2010 Census Planning Memoranda Series #173. Washington, DC: U.S. Census Bureau.

Rumbaut, Ruben. 2011. “Pigments of our imagination: The racialization of the Hispanic-Latino category.” Migration Policy Institute.

Saperstein, Aliya. 2006. “Double-Checking the Race Box: Examining Inconsistency between Survey Measures of Observed and Self-Reported Race.” *Social Forces* 85(1): 57-74.

Saperstein, Aliya. 2011. “Capturing Complexity in the United States: Which Aspects of Race Matter and When?” *Ethnic and Racial Studies* 35(8): 1484-1502.

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

Snipp, C. Matthew. 2003. “Racial Measurement in the American Census: Past Practices and Implications for the Future.” *Annual Review of Sociology* 29:563-588.

Sweet, Elizabeth. 1994. “Who Fills Out the Census Questionnaire Study.” 1990 DSSD REX Memorandum Series #PP-9. Washington, DC: U.S. Census Bureau.

Tafoya, Sonya. 2004. “Shades of Belonging: Latinos and Racial Identity.” Washington, DC: Pew Hispanic Center.

Tashiro, Cathy J. 2011. *Standing on Both Feet: Voices of Older Mixed-Race Americans*. Boulder: Paradigm.

U.S. Census Bureau. 2007. “Summary File 1: 2000 Census of Population and Housing Technical Documentation.” Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2009. “2010 Census Nonresponse Followup (NRFU) Enumerator Manual.” Washington, DC: U.S. Census Bureau.

Vargas, Nicholas. 2015. “Latina/o Whitening?: Which Latinas/os Self-Classify as White and Report Being Perceived as White by Other Americans?” *Du Bois Review* 12(1): 119-136.

Vitrano, Frank and Maryann Chapin. 2012. “Possible 2020 Census Designs and the Use of Administrative Records: What is the Impact on Cost and Quality?” Federal Committee on Statistical Methodology Research Conference Proceedings.

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.” Center for Administrative Records Research and Applications Working Paper 2014-01. Washington, DC: U.S. Census Bureau.

Walker, Shelley, Susanna Winder, Geoff Jackson, and Sarah Heimel. 2012. “2010 Census Nonresponse Followup Operations Assessment.” 2010 Census Program for Evaluations and Experiments, 2010 Census Planning Memoranda Series No. 190. Washington, DC: U.S. Census Bureau.

Word, David L. and R. Colby Perkins Jr. 1996. “Building a Spanish Surname List for the 1990’s— A New Approach to an Old Problem.” Population Division Technical Working Paper No. 13. Washington, DC: U.S. Census Bureau.

Xie, Yu and Kimberly Goyette. 1997. “The Racial Identification of Biracial Children with One Asian Parent: Evidence from the 1990 Census.” *Social Forces* 76(2): 547-570.

**Acknowledgements**

We thank Haley Pilgrim for her diligent assistance with this research. At various stages of this research, we have benefited from thoughtful comments from many people: Amy O'Hara, Renuka Bhaskar and Fabián Romero of the U.S. Census Bureau and Caren Arbeit, Julia Rivera Drew, Catherine Fitch, Liying Luo, and Susan Mason of the University of Minnesota. The Minnesota Population Center (University of Minnesota) provided important support for the second author through programs made possible by an NIH Center Grant (R24HD041023).

Appendix: Comparison of our analysis sample to full censuses of 2000 and 2010

| % of all people in data with each characteristic | Census 2000 100% data | Census 2010 100% data | Two parts of analysis sample: Part 1 In 2000 from HH | Part 1 In 2010 from Proxy | Part 2 In 2000 from Proxy | Part 2 In 2010 from HH |
|---|---|---|---|---|---|---|
| *Race/Hispanic origin* | | | | | | |
| Non-Hispanic white | 69.1% | 63.7% | 71.8% | 72.7% | 69.8% | 68.2% |
| Non-Hispanic black | 12.1% | 12.2% | 14.4% | 14.2% | 16.8% | 16.6% |
| Non-Hispanic AIAN | 0.7% | 0.7% | 1.2% | 1.0% | 1.0% | 1.2% |
| Non-Hispanic Asian | 3.6% | 4.7% | 2.6% | 2.5% | 2.7% | 2.8% |
| Non-Hispanic NHPI | 0.1% | 0.2% | 0.1% | 0.1% | 0.2% | 0.2% |
| Non-Hispanic white-black | 0.2% | 0.6% | 0.3% | 0.4% | 0.3% | 0.5% |
| Non-Hispanic white-AIAN | 0.3% | 0.5% | 0.5% | 0.4% | 0.3% | 0.6% |
| Non-Hispanic white-Asian | 0.3% | 0.5% | 0.3% | 0.2% | 0.2% | 0.3% |
| Non-Hispanic white-NHPI | 0.0% | 0.1% | 0.0% | 0.0% | 0.0% | 0.1% |
| Hispanic white | 6.0% | 8.7% | 4.2% | 3.9% | 4.3% | 5.0% |
| Hispanic Some Other Race | 5.3% | 6.0% | 3.6% | 3.6% | 3.7% | 2.9% |
| Another race/Hispanic comb. | 2.1% | 2.2% | 1.0% | 1.0% | 0.7% | 1.7% |
| Hispanic (any race) | 12.5% | 16.3% | 8.3% | 8.1% | 8.3% | 8.9% |
| Non-Hispanic (any race) | 87.5% | 83.7% | 91.7% | 91.9% | 91.7% | 91.1% |
| *Age and sex* | | | | | | |
| Age 0-9* | 14.1% | 13.1% | 11.6% | 0.9% | 14.6% | 0.1% |
| Age 10-18 | 13.0% | 10.9% | 19.4% | 9.5% | 9.1% | 13.2% |
| Age 19-64 | 60.4% | 62.9% | 59.7% | 72.7% | 71.0% | 77.5% |
| Age 65 or older | 12.4% | 13.0% | 9.3% | 16.9% | 5.4% | 9.2% |
| Male | 49.1% | 49.2% | 47.7% | 47.7% | 50.4% | 50.4% |
| Female | 50.9% | 50.8% | 52.3% | 52.3% | 49.6% | 49.6% |
| *Location* | | | | | | |
| Urban | 79.0% | 80.7% | 75.0% | 73.3% | 77.4% | 71.6% |
| Rural | 21.0% | 19.3% | 25.0% | 26.7% | 22.6% | 28.4% |
| Northeast | 19.0% | 17.9% | 19.7% | 19.2% | 17.3% | 16.6% |
| Midwest | 22.9% | 21.7% | 23.8% | 23.3% | 23.0% | 22.3% |
| South | 35.6% | 37.1% | 37.4% | 38.5% | 39.7% | 41.2% |
| West | 22.5% | 23.3% | 19.1% | 19.0% | 20.0% | 19.9% |

Data sources: Census 2000 and the 2010 Census 100% data are from SF1 and PL files (census.gov). Analysis data are from Census 2000 and the 2010 Census; see text for details.

AIAN = American Indian/Alaska Native; NHPI = Native Hawaiian and other Pacific Islander.

* A few children were enumerated in 2000 but were reported as less than 10 years old in 2010.