Measures of “Race” and the analysis of racial inequality in Brazil

Stanley R. Bailey\textsuperscript{a,*}, Mara Loveman\textsuperscript{b}, Jeronimo O. Muniz\textsuperscript{c}

\textsuperscript{a} Department of Sociology, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, United States
\textsuperscript{b} Department of Sociology, University of Wisconsin, Madison, 1180 Observatory Drive, 8128 Sewell Social Science Building, Madison, WI 53706-1393, United States
\textsuperscript{c} Department of Sociology at the Federal University of Minas Gerais, Av. Antônio Carlos, 6627, FAFICH, sala 4156, Belo Horizonte, MG 31270-901, Brazil

\textbf{A R T I C L E  I N F O}

Article history:
Received 16 September 2011
Revised 21 April 2012
Accepted 26 June 2012
Available online 6 July 2012

Keywords:
Race
Skin tone
Classification
Inequality
Brazil
Quantile regression

\textbf{A B S T R A C T}

Quantitative analyses of racial disparities typically rely on a single categorical measure to operationalize race. We demonstrate the value of an approach that compares results obtained using various measures of race. Using a national probability sample of the Brazilian population that captured race in six formats, we first show how the racial composition of Brazil can shift from majority white to majority black depending on the classification scheme. In addition, using quantile regression, we find that racial disparities are most severe at the upper end of the income distribution; that racial disparities in earnings are larger when race is defined by interviewers rather than self-identified; and that those classified as “black” suffer a greater wage penalty than those classified as “brown.” Our findings extend prior conclusions about racial inequality in Brazil. More generally, our analysis demonstrates that comparison of results across measures represents a neglected source of analytic leverage for advancing empirical knowledge and theoretical understanding of how race, as a multidimensional social construct, contributes to the production of social inequality.

© 2012 Elsevier Inc. All rights reserved.

\textbf{1. Introduction}

Social scientific understanding of racial inequality is shaped by the way race is measured and operationalized in quantitative research. This is true for at least three reasons: First, the choice of categories in a given data set conditions what we see in terms of that context’s racial composition. The decision to include any given set of categories is contingent, reflecting judgments about who should be counted and with what terminology (Harris, 2002; Nobles, 2000). Second, the categories we employ determine how we understand the distribution of resources along racial lines. Classification by skin color, for example, can reveal inter- and intra-racial category stratification (Hughes and Hertel, 1990; Keith and Herring, 1991; Gullickson, 2005; Goldsmith et al., 2006). Third, the particular method of classification we choose shapes how we understand racial inequality. For instance, interviewer classification versus self-classification can lead to different conclusions about the magnitude of income inequality between racial groups (Saperstein, 2006; Telles and Lim, 1998).

Although race is typically experienced as a stable and unitary individual trait, social constructionist theory posits that race is actually a contextual and multidimensional social construct. Racial identifications hinge on a constellation of factors, including self-perception, ascription by others, interactional cues, institutional contexts, and prevailing cultural understandings of consequential markers of human difference (Cornell and Hartmann, 1998; Jenkins, 1994; Harris and Sim, 2002; Roth, 2005; Brunnsma, 2005; Daynes and Lee, 2008). Since race is socially constructed, there is no a priori reason to assume that any given measure of race is more or less valid than others. Indeed, studies by Telles and Lim (1998), Saperstein (2006, 2008), and Campbell (2009) call attention to the fact that various measures that researchers have traditionally treated as alternative
proxies for “race” do not necessarily capture the same underlying “thing.” Far from being a liability, the fact that different measures of race may be used to operationalize different dimensions of race provides a valuable and underutilized source of analytic leverage for studies of racial inequality.

In this article, we present an analytical approach that exploits the comparison of results across different methods and formats of racial classification better understand patterns and dynamics of racial inequality. We demonstrate the substantive and theoretical value added by the use of multiple measures through an original analysis of racial inequality in wages in Brazil based on the 2002 Brazilian Social Survey (PESB). The PESB is a national probability sample that captures racial classification through six formats and methods. We first present descriptive statistics to show how the method used to collect and code racial data can dramatically affect starting assumptions about a population’s racial composition. We then demonstrate how the measurement of race can affect the analysis of racial inequality by using quantile regression methods to generate and compare levels of income inequality in Brazil based on different definitions of “race.”

Results show that differing classification schemes significantly alter the overall picture of the racial composition of the Brazilian population. According to some schemes, Brazil is a predominantly nonwhite country; in others, it becomes majority white. We also find that the magnitude of racial disparities in wages changes depending on how race is defined and according to location along the income distribution. Finally, comparisons of level of inequality across classification schemes provide clues to the underlying mechanisms fueling racial disparities. We observe, for example, that the darkest segment of the nonwhite population is especially vulnerable in Brazil, signaling the nuanced role of skin color in estimates of inequality, a factor that is obscured in schemes that treat all nonwhites as equally disadvantaged.

Our empirical findings build upon and extend recent scholarship on racial stratification in contemporary Brazil. Looking beyond the Brazilian context, our analysis demonstrates more generally how comparison of results across measures can advance understanding of descriptive trends and underlying dynamics that fuel racial inequality.

2. Race measures and inequality studies in Brazil

The vast majority of research on racial dynamics in Brazil focuses on the black-to-white continuum. Brazil’s large-scale social surveys typically use three racial or color terms to capture the range of identifications on this continuum: white (branco), brown (pardo, or “mixed”), and black (preto). Approximately 99% of the population self-classified in one of these three categories in the last three censuses (1991, 2000, and 2010).1

Researchers who analyze survey data to investigate racial inequality in Brazil use either a dichotomous (white versus nonwhite) or a ternary (white versus brown versus black) analytical framework. The choice between these approaches bears on ongoing debates about the nature and dynamics of racial stratification in Brazil. Two questions are central to these debates. First, is there a “mulatto escape hatch” in Brazil that enables “brown” Brazilians to experience more mobility than “blacks”? In other words, is the divide between “whites” versus “nonwhites” the only racial division of real substance, or is the social distinction between “browns” and “blacks” also sociologically consequential? In addition, what accounts for racial inequality in Brazil? Can racial inequality be explained through other more fundamental processes, such as social class inequalities, as opposed to racial discrimination (e.g., Wagley, 1952)?

Researchers broadly agree that contemporary Brazilian society is characterized by entrenched racial or color stratification, an empirical reality that contradicts Brazil’s carefully cultivated myth of racial democracy. Nonetheless, researchers disagree about whether “brown” Brazilians fare significantly better than “blacks” in the Brazilian racial order. Researchers also disagree about the underlying reasons for observed racial disparities in consequential social outcomes (Muniz, 2010). Our analyses using multiple race measures and classification formats provide evidence that bears on both of these debates. In addition, we address a complicating factor in the interpretation of any findings of racial inequality in Brazil: the possibility that racial classification is not independent of social status in Brazilian society. To situate the analyses and contributions of this study, we begin with a brief assessment of the state of the evidence on these questions in the literature.

2.1. A mulatto escape hatch?

There has been a proliferation of quantitative research on Brazilian racial inequality in the last two decades. Much of it builds on the work of two important scholars, Carlos Hasenbalg (1985) and Nelson do Valle Silva (1985). Analyzing survey data and the three census categories—black, brown, and white—they challenged a core assumption about the Brazilian context: that there existed a “mulatto escape hatch” permitting browns to occupy an intermediate social position between whites and blacks and mitigating the racial penalty associated with blackness (Degler, 1971). Hasenbalg and Silva challenged the idea of a “mulatto escape hatch” by documenting strikingly similar SES profiles of blacks and browns compared to whites. Hasenbalg (1985) concluded that in their ongoing struggle for status attainment “nonwhites” (an umbrella category designating browns and blacks) experienced a cycle of cumulative disadvantage relative to whites. The similar profiles of

---

1 Our analysis builds upon and extends an earlier investigation of how the picture of racial inequality in Brazil would change if the intermediate “brown” category were removed from the census (Loveman et al., 2012).

2 Brazil’s census form asks individuals to identify their “color or race.” Researchers disagree on whether these terms are interchangeable in the Brazilian context (e.g., Sheriff, 2001, pp. 29–58). We use “race” for the Brazilian terms raça and cor (color) for ease of readability.

3 The remaining two census terms are Asian descent (amarelo, or “yellow”) and indigenous (indígena).
blacks and browns, both Hasenbalg and Silva argued, justified treating them as a single nonwhite population in studies of racial inequality. Thus, they advanced a binary analytical framework (white/nonwhite) in place of the ternary approach (white/brown/black) for understanding racial dynamics in Brazil.

More recent scholarship is divided on whether the “mulatto escape hatch” question has been resolved. Studies of racial wage disparities have documented inequality using both binary and ternary analytical approaches. Silva (2000), Reis and Crespo (2005), and Lovell (2006), for example, adopt versions of binary classification. They compare whites to “nonwhites,” “negros,” or “Afro-Brazilians,” all categories formed by collapsing the brown and black census categories. By adopting a dichotomous analytical approach, these studies underscore the wage gap between whites and nonwhites, while downplaying or negating the relevance of distinctions within the nonwhite category. In contrast, Arcand and D’Hombres (2004) and Arias et al. (2004) retain the ternary format in their analyses and find important differences between blacks and browns in addition to differences between these two categories and whites. Telles and Lim (1998) also use a ternary format and claim to find evidence for a “mulatto escape hatch,” reporting that brown workers’ wages are close to the halfway point between those of whites and blacks. Randall (2009), however, argues that since the year 2000, differences in total income between blacks and browns are no longer evident.

Although there are some important exceptions, the use of the dichotomous format appears widely accepted among researchers who have followed the lead of Hasenbalg and Silva (on this tendency, see Telles, 2004, p. 146). The Brazilian state and black movements also generally embrace a binary racial lens (Telles, 2004).

2.2. Explaining racial disparities

Hasenbalg’s and Silva’s works also challenge the core assumption that racial inequality in Brazil can be explained fully by factors other than discrimination. Silva’s (1985, 2000) scholarship in particular uses a dichotomous analytical framework to investigate how much of the racial differentiation in earnings results from labor market discrimination. He employs wage decomposition methods (e.g., Oaxaca, 1973) that control for the effects on wages of a variety of compositional factors, such as human capital and labor market characteristics (e.g., education, age, region, occupation). Using 1960 census data, for example, Silva found that 82% of the wage difference between whites and nonwhites was due to compositional factors; he attributed the remaining 18% to racial discrimination.

The wage decomposition analyses that dominate studies of racial inequality in Brazil have produced some divergent findings. For example, Silva (2000) wrote that his work based on Brazil’s annual National Household Survey (PNAD) documented a reduction in the importance of discrimination over time, whereas Lovell (2006), in examining census data for the city of São Paulo, estimated that the gap in wages attributable to discrimination remained relatively stable from 1960 to 2000. Reis and Crespo (2005) analyzed PNAD data for 1987–2002 and reported a significant cohort effect in the earnings differential between white and black workers: they found smaller gaps among the young. A study using the 1996 PNAD data focused on regional differences and found wage disparities greater in the northeast compared to the southeast after other compositional factors were taken into account (Campante et al., 2004).

Other studies have addressed the “independent effect” of race on inequality using ternary classification. For example, using PNAD data from 1998, Arcand and D’Hombres (2004) found that although browns and blacks had similar overall earnings profiles compared to whites, the residual disparity unaccounted for by compositional factors was significantly greater for blacks than for browns. At the same time, they are critical of attributing all unexplained difference to discrimination; rather, they suggest that some of the residual disparity may be due to unmeasured differences in educational quality (public versus private) and family background.

Arias and colleagues (2004) directly addressed the effects of education quality and family background on earnings for whites, browns, and blacks. They found that the bulk of racial earnings inequality was indeed due to whites’ more favorable socioeconomic backgrounds and the fact that they attended better-quality schools. In addition, using quantile regression, they showed that racial wage gaps increased with average earnings. Lastly, Arias and his colleagues found some significant differences between the experiences of browns and blacks in terms of returns on education: whereas at the top of the wage scale browns and whites saw similar labor market rewards for their educational investments, at the bottom of the scale the rewards given to browns were similar to those of blacks.

2.3. Racial classification and social status

Synthesizing the evidence on racial inequality in Brazil is complicated by the possibility that racial classification is not independent of social status in Brazilian society. The idea that social mobility may lead to racial mobility is a popular cultural trope in Brazil, captured in colloquial phrases such as “money whitens.” Scholarly discussions of the mulatto escape hatch typically note that as nonwhite individuals experience social mobility, they may be treated as white (or whiter) (Ianni, 1960; Degler, 1971). In other words, “someone can be born black, yet become mulatto through an increase in social status” (Golash-Boza, 2010, p. 140). Researchers generally refer to this process as “whitening” (e.g., Telles and Lim, 1998).

---

4 Lovell (2006) bases the adoption of a dichotomous scheme on research detailing the ambiguous and unstable nature of the boundary between black and brown categories, in contrast to the boundary between white and nonwhite, which is “relatively unambiguous and remarkably stable” (Carvalho et al., 2004, p. 68).
Several important studies have found evidence of whitening processes. Wood (1991) and Carvalho et al. (2004) offered estimates of racial reclassification over time by comparing projected and observed racial composition across censuses. Finding deficits in the black category and gains in the brown category between 1950 and 1980, they argued that “a large proportion of individuals who declared themselves black in 1950 reclassified themselves as brown in 1980” (p. 331). Additional suggestive evidence of whitening comes from research on the factors that predict how individuals are racially classified. Telles and Lim (1998) examined interviewer classification versus self-identification and documented a disconnect between an individual’s perception of his or her racial classification and the way that a third party may classify that individual (see also Silva, 1996). Telles and Lim suggested that relative to self-identification, interviewers often “whiten those with higher status and darken those of lower status” (p. 473). Finally, a study by Schwartzman (2007) pointed to intergenerational whitening processes. Using 2005 PNAD data, Schwartzman found that higher educational levels raise the likelihood that a nonwhite parent will classify his or her child as white and lowers the likelihood that a white parent will classify his or her child as black or brown.

The picture that emerges from existing studies of race measures and racial inequality in earnings in Brazil is multifaceted, and at least three points merit special note. First, scholars disagree over the use of binary versus ternary classification schemes. It is thus generally difficult to compare findings across studies or to know whether studies with different conclusions actually speak to each other directly. One of our goals is to confront this problem head-on through systematic comparison of results obtained using different classification schemes in the same time period.

Second, there is some agreement in Brazilian inequality studies that the lion’s share of the gap between whites compared to browns and blacks is due to nonracial individual characteristics, such as education, and contextual differences, such as region. Nonetheless, researchers continue to find that important disparities persist, though their estimates of those disparities vary significantly, as do their suggestions regarding the dynamics driving them. Another of our goals, then, is to explore the sensitivity of estimates of inequality to varying measures of race, and to determine whether the difference in estimates across measures can generate insight into the underlying dynamics that may be fueling racial inequality.

Finally, there is broad agreement that an important caveat underlies all cross-sectional studies of racial inequality of income in Brazil: racial classification and socioeconomic status may be endogenous. To the extent that racial mobility is occurring, it likely contributes to overestimates of racial inequality in income because higher-income individuals tend to be reclassified from darker categories to lighter ones. The absence of longitudinal surveys in Brazil precludes direct investigation of the extent to which whitening occurs, and like prior studies of racial inequality in earnings in Brazil, our data and study design do not enable us to rule out the possibility of an interrelationship between racial classification and social status indicators. An advantage of an approach that compares results across formats and classification schemes within a single survey, however, is that it provides some built-in check of the sensitivity of findings to the way race is measured.

### 3. Data

Our analysis uses data from the Brazilian Social Survey (PESB), conducted in July and August 2002. PESB follows the model of the American General Social Survey (GSS). The data are based on a nationally representative sample covering the five regions of Brazil and all persons aged 18 and over. The data set consists of 2364 persons sampled across 102 municipalities. Because we are interested in the black–white continuum, we exclude 34 individuals who self-classified as of Asian or indigenous origin.

According to the 2000 census, the population of Brazil is 53% white, 39% brown, 6% black, 0.45% indigenous, and 0.4% Asian origin. The entire sample for the weighted PESB survey, following the same self-classificatory racial scheme as the census, is 46% white, 34% brown, 11% black, 5.7% indigenous, and 3.7% Asian origin. The Asian-origin and indigenous groups are proportionally larger than in the census due to oversampling in smaller regions where they are historically more concentrated.

The PESB captures racial classification in several novel formats. In addition to self-classification following the official census format, the survey includes an item that asks respondents who self-classified as brown in the census format to reclassify themselves as either white or black. In addition, the data include interviewer classification, ancestry (parental race), and self-classification in reference to perceived skin color. To capture perceived skin color, the data set employed a set of photographs of men with different skin tones as reference points, with respondents being asked to identify which photograph showed a skin color most like their own. We describe each classification scheme in the next section.

### 4. Methods

We begin by presenting descriptive statistics that show how the racial composition of the Brazilian population shifts when individuals are sorted according to six possible classification schemes. Each scheme captures a different dimension or definition of “race.” Table 1 summarizes the classification schemes used in our analyses.
Table 1
Brazilian racial classification schemes, categories, and dimensions.

<table>
<thead>
<tr>
<th>Classification scheme</th>
<th>Categories</th>
<th>Dimension or perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census</td>
<td>White, brown, black</td>
<td>Self-classified race</td>
</tr>
<tr>
<td>Ascribed census</td>
<td>White, brown, black</td>
<td>Interviewer ascribed race</td>
</tr>
<tr>
<td>Post-hoc binary</td>
<td>White, nonwhite</td>
<td>Researcher’s gaze</td>
</tr>
<tr>
<td>Descent rule</td>
<td>White, black</td>
<td>Ancestry via parental race</td>
</tr>
<tr>
<td>Forced racial binary</td>
<td>White, black</td>
<td>Black movement perspective</td>
</tr>
<tr>
<td>Photo comparison</td>
<td>White, black</td>
<td>Appearance</td>
</tr>
</tbody>
</table>

1. Census: **Self-classification as black, brown, or white.** The Brazilian Census Bureau (IBGE) has employed these three terms for much of the last two centuries. This scheme captures self-identification with options restricted to the official census categories.

2. Ascribed census: **Interviewer classification as black, brown, or white.** Several scholars argue that beyond one’s own sense of racial or color identity, what matters most in a racially hierarchical society is classification by a third party. Telles and Lim (1998), as well as Saperstein (2006, 2008), suggest that interviewer classification can be seen as a proxy for perceptions of potential discriminators.

3. Post-hoc binary: **Collapsing black and brown from the census scheme into a unified nonwhite category.** This classification scheme reflects a statistical reallocation of self-classified race into the binary categories often used by researchers (e.g., Silva, 1985; Hasenbalg, 1985).

4. Descent rule: **Classification according to parental race (white if both parents are white, or black if at least one parent is black or brown).** Although scholars generally remark that Brazilian racial or color classification follows phenotype rather than descent, with this scheme we simulate a scenario in which a US-style one-drop rule prevailed in Brazil. Compared to other classification schemes, the descent rule scheme generates leverage to assess empirically the relative weight of ancestry versus phenotype in Brazilians’ racial calculus.

5. Forced racial binary: **Self-classification in a binary format as white or black.** This survey item asks respondents who self-classified as brown in the census scheme to reclassify themselves as either white (branco) or black (preto). The Brazilian black movement prefers the term negro over preto to describe the sum of browns and blacks. Telles (2004), however, argues that the terms are interchangeable in colloquial use, a view supported by results of a 2003 survey (Rosemberg, 2004). Hence, we treat the forced binary format in the 2002 PESB survey (branco or preto) as analogous to binary schemes that employ the categories branco and negro. By 2010 it appears that the momentum is towards greater differentiation of preto and negro; nonetheless we believe the binary scheme reflects dichotomous racial self-classification in 2002.

6. Photo comparison: **Classification by photographs as white or black.** This scheme is based on two survey items. One asks respondents to examine photos of eight young Brazilian adults with various skin tones (see Online Supplement Section A). They are then asked to divide them into two groupings, whites versus blacks (pretos). Two questions later, respondents are asked to examine the photos again and to indicate which pictured individual has a color or race most similar to the respondent’s own. Combined responses to these two items enabled the sorting of the population into a photo-referenced racial classification.

Next we investigate how levels of income vary across racial categories in each classification scheme. We graph the ratio of white to nonwhite income for each classification scheme by deciles of the income distribution.

Finally, we examine the relationships between racial categories and income levels, controlling for standard compositional variables. To investigate these relationships, we use income as the dependent variable and age, age squared, sex, education, marital status, region, and large metropolitan areas as independent variables to estimate models for each of our six classification schemes.\(^5\) We measure income as the natural logarithm of individual income. Valid income values were smoothed and 5–17%, respectively, of Brazil’s total wage inequality (Ferreira et al., 2006). Research on gender discrimination suggests that this variable accounts for about 5% of the wage gap, with women earning about 70% of the value paid to men with the same productive characteristics (Silva, 2003, p. 455; Barros and Mendonça, 1995, p. 37). Evidence also shows that the wage gap is larger among married workers than among single workers (Blau and Kahn, 1996).

\(^5\) Studies suggest that education is decisive in Brazil, accounting for between one-third and one-half of total inequality (Barros and Mendonça, 1995; Ramos and Vieira, 2000). In addition, regional differences as well as the urban versus rural divide (for which our metropolitan area acts as a proxy) account for 7–13% and 5–17%, respectively, of Brazil’s total wage inequality (Ferreira et al., 2006). Research on gender discrimination suggests that this variable accounts for about 5% of the wage gap, with women earning about 70% of the value paid to men with the same productive characteristics (Silva, 2003, p. 455; Barros and Mendonça, 1995, p. 37). Evidence also shows that the wage gap is larger among married workers than among single workers (Blau and Kahn, 1996).
mean” (Arias et al., 2004, p. 356); hence, it provides “a more complete picture of the relationships between variables missed by other regression methods” (Cade and Noon, 2003, p. 412).

Quantile regression was first developed by econometricians Koenker and Basset (1978) to estimate rates of change when factors of interest vary in their effect on the response variable. Since then researchers in ecology (Cade and Noon, 2003), meteorology (Sousa et al., 2009), economics (Machado and Mata, 2005; Koenker and Hallock, 2001; Arias et al., 2004), and sociology (Budig and Hodges, 2010) have applied this method to several areas of study, including determinants of wage or trends in income inequality.

A linear form of the model is expressed as

$$Q_Y(\tau | X) = \beta_0(\tau) + \beta_1(\tau)X_1 + \cdots + \beta_n(\tau)X_n$$

where \(\tau \in [0, 1]\) indicates the quantile and its corresponding parameter \((\tau)\); \(\beta_0\) is the intercept of the quantile-specific regression; \(X_1, \ldots, X_n\) represent a set of individual characteristics, including racial classification. The interpretation of the parameters \(\beta_1, \ldots, \beta_n\) is the same as in other linear models: they represent the change in the response variable resulting from a change in explanatory variables after controlling for the effects of other covariates in the model. The crucial difference is that the effects are now defined for some specific quantile. For example, for \(\tau = 0.5\), \(Q_Y(0.5|X)\) is the 50th percentile (or median) of the distribution of \(Y\) conditional on the values of \(X\) (Cade and Noon, 2003, p. 414).

We estimate regression coefficients for each quantile of the income distribution using each classification scheme and controlling for the regional segmentation of the labor market and other individual characteristics. By assessing the results side-by-side, we determine whether and how differing measures of race affect estimates of race-based inequality in income in Brazil.

5. Findings

5.1. Population distribution by classification scheme

Fig. 1 shows the distribution of persons by racial category in each of the six different schemes. Different classification schemes clearly generate very different pictures of Brazil’s racial composition. The two ternary formats look most alike and show that browns make up a significant share of Brazil’s population. In contrast, the “post hoc binary” scheme, which takes the picture generated by the census format and then eliminates the distinction between browns and blacks, portrays a population nearly evenly divided between whites and nonwhites. The other three binary classification schemes present varied pictures of Brazil’s racial composition. Two of these, “forced racial binary” and “photo comparison,” suggest that whites comprise upwards of 70% of the population. In contrast, judged by a “descent rule” based on parental race (similar to the rule of hypodescent), the nonwhite segment of the population forms a significant majority (at about 60%). In all of the dichotomous classification schemes, by definition, Brazil appears as a country divided between whites and nonwhites with no separate social space occupied by an intermediate or mixed-race population.

Fig. 1. Racial composition of Brazil according to six classification schemes.
These differing pictures of Brazil’s racial composition demonstrate significant levels of movement between categories and classification inconsistency. In Table 2, we present a cross-tabulation of the two ternary schemes: self-classification versus interviewer (ascribed) classification in the census format. The highest percentage of agreement, or overlap, between these two schemes occurs with whites, at 42.8%. Consistently classified browns and blacks are 25.9% and 6.5% of the total sample, respectively. The overall agreement between self- and interviewer classification was 75%, which coincides with the 79% agreement that Telles and Lim (1998) found using a national sample from 1995.6 Table 3 presents a matrix of levels of agreement between the four binary classification schemes for whites (below the main diagonal) and nonwhites (above the main diagonal). Of those classified as black in the descent rule scheme, 27.2% are also classified as black in the photo comparison scheme. For whites this percentage is equal to 38.3%. The sum of these two figures shows that the lowest level of agreement is between the descent rule and the photo comparison classification schemes (65.5%). This inconsistency supports the view that classification based on appearance operates differently than classification based on notions of ancestry.7

In sum, there is indeed ambiguity and inconsistency in racial classification in Brazil, but it is not boundless (Muniz, 2012). Additionally, it is clear that the racial composition that we “see” hinges on the racial classification scheme we employ. Brazil’s population oscillates between 40.7% and 70.4% white, between 0% and 40.1% brown, and between 10.8% and 59.3% black/nonwhite, depending on means of classification.

### 5.2. Racial inequality by classification scheme: bivariate results

To explore how the categories used to sort the population affect our understanding of income disparities in Brazil, we show a set of graphs that present ratios of white to black incomes across deciles of the wage distribution. Fig. 2 presents those ratios according to “post hoc binary,” “forced racial binary,” “descent rule,” and “photo comparison” classification schemes (those reflecting binary divides). Exploring this bivariate perspective is important because many researchers, the state, and social movement actors rely heavily on descriptive statistics for much of their discussion of income inequality in Brazil.8 Results show that income inequality along racial lines in Brazil is pronounced and varies across classification schemes. For example, in the “descent rule” scheme, nonwhites located at the fourth decile of the nonwhite income distribution earn on average about 50% of what whites at that decile earn. Compare that to the “photo comparison” scheme, where blacks at the fourth decile earn on average about 63% of what whites earn.

Fig. 3 presents ratios of the wage distribution according to the “census” and the “ascribed census” formats (the two ternary schemes). Results reveal an important distinction obscured in the binary schemes: brown-to-white and black-to-white ratios differ. For example, blacks located at the sixth decile of the black income distribution in the “ascribed census” scheme earn on average about 35% of what whites at the same decile earn. Compare that to browns in the same classification scheme at the sixth decile: they earn on average about 52% of what whites at that decile of the income distribution earn.

---

**Table 2**

Self-classification by ascribed classification.

<table>
<thead>
<tr>
<th>Self-identified census format</th>
<th>Ascribed census format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
</tr>
<tr>
<td>White</td>
<td>42.8</td>
</tr>
<tr>
<td>Brown</td>
<td>6.1</td>
</tr>
<tr>
<td>Black</td>
<td>0.3</td>
</tr>
<tr>
<td>Total</td>
<td>49.1</td>
</tr>
</tbody>
</table>

N = 1916.

**Table 3**

Percentage agreement between racial schemes.

<table>
<thead>
<tr>
<th>Classification Scheme</th>
<th>Post-hoc binary</th>
<th>Forced racial binary</th>
<th>Descent rule</th>
<th>Photo comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.2</td>
<td>31.6</td>
<td>43.5</td>
<td>27.3</td>
<td></td>
</tr>
<tr>
<td>36.3</td>
<td>29.8</td>
<td>38.8</td>
<td>27.2</td>
<td></td>
</tr>
<tr>
<td>49.8</td>
<td>61.4</td>
<td>38.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 1916.

*Note: Agreement on the black category is above the diagonal and agreement on the white category is below the main diagonal of the matrix.*

---

6. Comparison with the US context reveals the relatively high level of classification inconsistency in Brazil. Using GSS data from 1996 and 2000, Saperstein (2006) found in the US that 95% of “black,” “whites,” and “others” were consistently classified.

7. The fact that many Brazilians prefer to self-identify using non-census terms, such as moreno (brownish) or negro (black) may contribute to the comparatively high level of classificatory inconsistency (Silva, 1996; Bailey and Telles, 2006).

Importantly, both Figs. 2 and 3 also show that differences between whites and nonwhites increase as individuals move up the wage distribution. In the ternary schemes, for example, "census" blacks at the second decile earn almost 60% of what whites earn at the same decile, whereas "census" blacks at the eighth decile earn only 30% of whites' income. This wage distribution perspective generates important information that is lost in analyses focusing on mean incomes. For example, Telles (2004, p. 115) reports results of a bivariate analysis that found blacks earn on average 40% of what whites earn, missing the effect of location along the wage distribution.

5.3. Racial inequality by classification scheme: regression results

Our quantile regression analysis examines racial disparities in wages under different classification schemes taking several key control variables into account (age, age squared, sex, schooling, marital status, region, and metropolitan area). We present our results in graph form in Figs. 4 and 5. The graphed values represent the wages of nonwhites (as grouped in the different schemes) as a fraction of whites' wages by deciles of the wage distribution. The predicted wages of each racial category are based on the race-specific intercepts from the different schemes when age and age-squared are kept at their means and all other control variables are set to 0.

The results for both binary and ternary formats confirm the previous bivariate findings of persistent income differences between whites and nonwhites/blacks regardless of classification scheme. However, the racial wage gaps are reduced and flattened considerably by the inclusion of our control variables. For example, Fig. 2 shows that nonwhites in the "descent rule" scheme located at the fourth decile of the nonwhite income distribution earn on average about 50% of what whites at the same decile of the white income distribution earn. In contrast, regression results in Fig. 4 show that nonwhites in the "descent rule" format located at the fourth decile of the nonwhite income distribution earn on average about 84% of what whites earn.

---

Fig. 2. Ratios of blacks’ to whites’ hourly wage by decile: four binary schemes.

Fig. 3. Ratios of nonwhites’ to whites’ hourly wage by decile: two ternary schemes.

---

9 See Supplement Table S1 for means of hourly wage and independent variables by racial classification scheme.
whites at the same decile of the white income distribution earn once we control for other variables associated with earnings.\textsuperscript{10} Fig. 4 also reveals some variation between binary schemes after other factors are controlled for. For example in the “photo comparison” scheme, blacks in the seventh decile earn about 88\% of what whites earn at the same decile of the white wage distribution. Contrast that to the “forced racial binary” scheme at the same decile, where blacks make only 76\% of white wages.\textsuperscript{11}

In addition, the gaps between whites and nonwhites vary by decile of the wage distribution. Quantile regression results generally reveal greater wage inequality at the upper deciles of the income distribution. For example, according to the “forced racial binary” scheme (Fig. 4), blacks located at the first decile of the income distribution earn on average 87\% of what whites earn at the same decile. Compare that gap to the ninth decile in the same classification scheme, where blacks earn only 67\% of white earnings. Ordinary Least Squares (OLS) results, available in the online supplement,\textsuperscript{12} reveal a mean value of 79\%, which is well below blacks at the first decile but well above blacks at the ninth decile.

Turning to the relative wage gap by decile in the ternary “census” and “ascribed census” schemes (Fig. 5), results show that gaps between whites and browns and between whites and blacks are generally greatest at the upper decile, paralleling the pattern seen in the binary formats.\textsuperscript{13} In addition, Fig. 5 suggests some important differences between brown and black populations that are lost when estimating inequality using a binary classification scheme. For example, the relative gaps between browns and whites and between blacks and whites according to the “ascribed census” scheme at the ninth decile are 72\% and 54\%, respectively.

Our findings reveal that the method of classification affects the picture of inequality. Fig. 5 shows that interviewer classification compared to self-classification increases the level of inequality suffered by blacks relative to whites, but not so for

\textsuperscript{10} See Supplement Table S4, which lists the graphed points for the binary schemes.

\textsuperscript{11} Paired t-tests on the differences between mean inequality comparing the possible pairings of the binary schemes reveal the following statistically significant differences: “post hoc binary” versus “descent rule,” “post hoc binary” versus “photo comparison,” and “descent rule” versus “photo comparison.”

\textsuperscript{12} See Supplement Table S4 for OLS predictions for the binary schemes.

\textsuperscript{13} An exception is the larger white-black gap at the lowest end of the wage distribution for “ascribed census” blacks. The data set sample weight that adjusts for nonracial characteristics produces this anomaly. See Supplement Table S5, which lists the graphed points for ternary schemes.
browns. For example, in the “census” scheme blacks and browns at the eighth decile earn on average 81% and 85%, respectively, of what whites earn at that decile. According to the “ascribed census” scheme, browns at the eighth decile earn about 86% of whites’ income. In contrast, ascribed blacks at the eighth decile earn on average only 70% of what ascribed whites earn.14 These patterns captured through interviewer classification would be lost using only self-classification and only regression methods that report mean differences. These patterns would also be obscured, of course, by use of a binary classification scheme.15

6. Discussion

Our analysis of race-based wage inequality in Brazil under alternative racial classification schemes demonstrates how the way race is measured influences descriptive and analytic conclusions about the nature and extent of racial inequality. In this discussion, we highlight how the use and comparison of multiple measures deepens our understanding of patterns and sources of racial inequality in Brazil.

The choice of classification scheme and method clearly affects how we perceive Brazil’s racial composition: as majority white or majority nonwhite, as a substantially mixed population or one composed of two discrete racial groups. Which picture of Brazil is the more accurate? We suggest this question may be wrong-headed; rather, each classification scheme captures a distinct dimension of the multidimensional social construct of race. As is true elsewhere, “race” in Brazil is partly a matter of self-identification and partly a matter of identification by others. Perceptions of skin color and ideas about ancestry factor into, but are not entirely coterminous with, self-identification or other-identification of race. Thus, the degree of consistency across measures will vary across time and place, and no single measure of “race” can be presumed, a priori, to be a proxy for others. By scrutinizing the divergences in the various pictures of Brazil’s racial composition captured using different classification schemes and methods, we gain information about how the constituent dimensions of race are related to one another.

Thus, for example, comparison of the population’s racial composition when categorization follows a descent rule versus self-identification suggests that Brazilians, as surveyed in 2002, do not rely heavily on notions of ancestry in deciding their individual racial identity. Comparative studies of race in Brazil and the United States have long argued that ancestry weighs more heavily in the calculus of race in the latter than in the former (e.g., Nogueira, 1985); our finding provides an empirical referent for assessing the extent to which parental racial classification determines individuals’ self-identification. Comparison across measures that capture different dimensions of race allows for empirical investigation of the cultural rules that govern racial classification in Brazil; replication of such comparative analyses using future surveys would provide evidence of how tacit rules of racial classification change over time.

Comparison across classification schemes can also illuminate implicit preferences and beliefs about race that likely inform racial categorization and identification in social surveys. For instance, comparing the “forced binary” to the “post hoc binary,” we see that when self-identified “brown” Brazilians are constrained to a choice between white and black, many opt for the former (Bailey, 2008). That preference suggests the continuing coveted nature of whiteness in Brazil. This comparison also suggests that popular notions of whiteness and blackness differ significantly from those of the state which, in some policy contexts, straightforwardly considers census browns to be nonwhites (Telles, 2004; Bailey, 2008; SEPPIR, 2010). Thus, when “race” is analyzed as a multidimensional social construct, simple descriptive differences in population composition across measures can be a source of new insight into the particular way race is socially constructed in that context (Muniz, 2012).

Our analysis also shows the sensitivity of racial disparities in income to the way race is measured. This sensitivity is brought into relief by our analyses of nonwhite to white income using quantile regression. Building on Arias and colleagues (2004), a first core finding of our quantile regression analysis is that OLS regression models overestimate the amount of income inequality between white and nonwhite populations in roughly the lower two-thirds of the income distribution, while underestimating inequality at the upper end of the distribution.16 Moving a step beyond Arias and colleagues, the fact that we find higher levels of income inequality in the upper end of the wage distribution in all six of our racial classification schemes—which shuffle the population quite significantly—demonstrates the robustness of this finding. Also in line with Arias and colleagues, we show that at lower levels of the income distribution, race can lose its independent significance for affecting wages altogether.17

Hence, we find that the salience of race as a basis for social closure that affects individual earnings is uneven. Moreover, we locate the position in the income distribution at which that closure becomes especially onerous for nonwhites: in the top

---

14 Paired t-tests on the differences between mean inequality comparing the possible pairings of blacks and browns in relation to whites show the following differences to be statistically significant: “census” browns versus “ascribed census” blacks, “census” blacks versus “ascribed census” blacks, and “ascribed census” browns versus “ascribed census” blacks.

15 See Supplement Table S2 for an example of the specific regression coefficients by quantile and their interpretation for the “forced racial binary.” In addition, see Fig. S1, where we plot the intercept and the coefficients for the “forced racial binary” representing the association of hourly wages to race, geographic location, and the completion of postsecondary education.

16 See Supplement Tables S4 and S5, which show OLS predictions along with the graphed quantile prediction for each scheme and across the wage distribution.

17 See Supplement Table S2 for an example of the absence of a race effect at the first decile once we controlled for regional segmentation of the labor market and important individual characteristics.
10–20% of the income structure. Whereas Telles (2004, pp. 220–222) suggests a “glass ceiling” effect in Brazil, in which the direct and indirect effects of race restrict the entrance of nonwhites into the middle class, our analyses suggest a more elite barrier of white social closure. Notably, our results echo the findings of Grodsky and Pager (2001) that US blacks face greater wage inequality in higher positions of the occupational earnings hierarchy.

The parallel between our findings and those reported by Grodsky and Pager (2001) suggests two important broader implications of our quantile regression analysis. First, efforts to generalize about the nature of racial inequality based on mean differences may obscure important findings fueling inequality. Second, adoption of a distributional perspective in combination with multiple measures of race in analyses of racial inequality could facilitate more rigorous cross-national comparative studies of racial stratification.\(^{18}\)

A second core finding of our regression analysis is that the white versus black wage gap is greater when race is measured by interviewer rather than self-classification. Assuming interviewer classification serves as a proxy for race “in the eye of the observer,” this finding suggests that direct discrimination by employers may contribute to racial wage disparities (Telles and Lim, 1998; Saperstein, 2006). Our finding of greater wage inequality with interviewer classification corroborates the earlier findings by Telles and Lim (1998) based on a 1995 survey.\(^{19}\) Our analysis extends beyond Telles and Lim, however, in the use of quantile regression to identify specific locations in the income distribution that may represent the loci of racial disadvantage.

Our results also depart from Telles and Lim’s findings in another critical respect. We find that interviewer classification compared to self-classification increases the level of mean income inequality suffered by blacks compared to whites, but not so for browns.\(^{20}\) This finding suggests that blackness per se—as opposed to gradations of non-whiteness—bears a distinct negative cost in Brazil. The specific disadvantages faced by individuals whom others see not as “brown” or “nonwhite” but as black are obscured by classification schemes that group all nonwhites under a single umbrella category. A similar dynamic has been identified in research on intra-racial category stratification in the United States (e.g., Keith and Herring, 1991), suggesting the need for more systematic attention in comparative race scholarship to how different dimensions of “race,” such as skin color versus categorical identification, may contribute differently to observed disparities in socially consequential outcomes.

Finally, taken together, our findings from comparing racial disparities in wages across classification schemes and levels of income provide compelling evidence for the value of retaining a ternary classification scheme when analyzing racial income inequality in Brazil. Juxtaposing our results based on ternary versus binary classification schemes suggests that the latter can obscure some important differences between black and brown populations. Silva’s (1985) and Hasenbalg’s (1985) seminal work endorsed the practice of collapsing browns and blacks into a single nonwhite category for inequality studies, and many researchers followed their lead. Nonetheless, the ternary measure continues to be available in virtually all data sets in Brazil; our analysis cautions researchers against ignoring it. More broadly, our results support the value of retaining “intermediate” categories in research on racial inequality whenever possible. At a minimum, the results of analyses that collapse nonwhites together, or that “reallocate” multi-race responses to single-race categories in the United States, should be compared to results obtained when retaining the original categorical distinctions in the data. Such comparisons yield valuable insights that affect interpretation of results (Harris, 2002; Snipp, 2003; Campbell, 2007).

7. Conclusion

We began this article with the observation that social scientific understanding of racial inequality is shaped by the way race is measured and operationalized in quantitative research. To conclude, we point to three specific ways in which the adoption of an analytic approach that compares results across multiple measures can advance sociological understanding of racial stratification.

First, comparison across measures provides a means to investigate empirically how race is socially defined in a given context: whether, for example, boundaries are “bright” or “blurry” (Alba, 2005), whether self-classification diverges from interviewer-ascribed race, or what dimension supports the definition of race (ancestry versus appearance). Rather than rely on broad-brushstroke assumptions that Brazilian race dynamics are ambiguous and fluid whereas US racial boundaries are clear and rigid, for example, the interrogation of multiple measures treats the social definitions of race at a given time as an empirical question. The advantage of an approach that can discern among distinct dimensions of race is especially evident for research on shifting racial dynamics in the contemporary United States. This point is driven home by the surge of recent research on classification of multiracial and Hispanic populations in the United States. Scholars have investigated identification choices under alternative survey question formats; the effects of interview context on self-classification; discrepancies between interviewer and self-identification; how interracial couples classify their children; and the socioeconomic correlates of classificatory decisions (Kao, 1999; Harris and Sim, 2002; Rockquemore and Arend, 2002; Roth, 2005, 2010; Brunnsma, 2005; Campbell and Troyer, 2007; Hitlin et al., 2007; Campbell, 2007, 2009). Focused on two rapidly growing segments of the American population, these studies anticipate the moment when direct investigation of racial ambiguity and

\(^{18}\) Regarding cross-national comparison, the ability to capture the multidimensionality of race through multiple schemes could facilitate isolating a common dimension for comparison across contexts where survey self-classification preferences may be structured by different dimensions, as is the case for Brazil and the United States (i.e., appearance versus ancestry, respectively).

\(^{19}\) Corroboration of Telles and Lim’s influential study is important due to the limitations of the 1995 Datafolha data set used in their analysis. Most importantly, it included only an indirect measure of income based on five broad intervals from which the authors assigned midpoint values to respondents.

\(^{20}\) See OLS results presented in Supplement Table S5.
classificatory inconsistency will become indispensable components of rigorous research on racial stratification in the United States (Snipp, 2003; Saperstein, 2006; Liebler and Halpern-Manners, 2008; Penner and Saperstein, 2008).

Second, comparison using multiple measures is critical to assess whether, or the extent to which, findings on racial inequality are tied to the way race data is collected or coded. Results derived using any particular approach should be recognized as providing a partial view. The racial dynamics under scrutiny might well look different if investigated using a measure that captured a different dimension of race or a coding scheme that imposed a different delineation of racial boundaries. For example, our results suggest that interrogating the skin color aspect of racial dynamics through a ternary lens is important for studies of income inequality. Several studies that employ skin color in the US context also show that alternative racial coding reveals nuanced dynamics (e.g., Goldsmith et al., 2006). A skin color measure was also recently applied in Mexico (Villarreal, 2010), a context that often does not easily dialogue with the racial stratification literature due to Mexico’s characterization as a mestizo nation. Nonetheless, the analysis made possible by the skin color scheme reveals the distinct disadvantage associated with darker skin tones in the Mexican context. The key point is that the specific measure and the classification scheme used to analyze racial inequality matter and ought to be taken into account when interpreting results.

Third, precisely because different measures of race may tap into different dimensions of the sociological concept of race, comparison of findings using alternate classification schemes can yield valuable insight into the underlying dynamics that contribute to the production and reproduction of racial inequality. In Brazil, the distinct disadvantage of the darkest skin color population as identified through ascribed race when compared to self-identification suggests that direct discrimination in the labor market deserves continued scrutiny. In the US context, if a racial gap in educational achievement outcomes is larger when race is defined by survey interviewers than it is when race is self-defined, this could steer analytical attention toward theories focused on the role of ascription and discrimination in educational outcomes, as opposed to theories focused on self-esteem or identity (Bruch and Loveman, 2011). Comparison of findings across measures cannot in itself resolve the question of underlying causal mechanisms, but it can definitely point researchers toward one set of mechanisms and away from others.

In sum, comparison across measures represents a promising and currently underutilized source of analytic leverage in research on racial inequality. Several major data sets used to investigate racial disparities in social outcomes in the United States contain multiple measures of race. When working with census, GSS, or Add Health data, for example, comparison of results across question formats can help analysts to better understand the multidimensionality of race and gain insights into possible dynamics fueling disparities. Moreover, when a data set has only one measure, analysts should recognize that it provides but a partial view. Furthermore, when analysts make decisions to code race data one way rather than another, comparison and reporting of whether and how such decisions may affect results should be routine.

Going forward, social scientific research on racial inequality has much to gain from experimenting with strategies to incorporate key tenets of social constructionist theories of race into quantitative analyses of racial disparities. Efforts to better understand how the method and categories used to measure race affect description, estimation, and explanation of the racial divides in a given population represent an important step in this direction.

Acknowledgments

Earlier versions of this article were presented at the 2009 meetings of the American Sociological Association and of the Associação Nacional de Pós-graduação e Pesquisa em Ciências Sociais (ANPOCS) in Brazil. We thank participants in the sessions for their comments. The authors wish to thank Laura Randall, Sarah Bruch, Andrew Penner, Andreas Wimmer, and Tom DiPrete for comments on earlier drafts.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ssresearch.2012.06.006.

References


