Consistent Divisions or Methodological Decisions? Assessing the U.S. Racial Hierarchy Across Outcomes

Beka Guluma1 · Aliya Saperstein1

Accepted: 22 November 2021 / Published online: 27 January 2022
© Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Scholars have offered a range of perspectives on the twenty-first century racial landscape with little consensus about either the current state of the U.S. racial hierarchy or its future trajectory. We offer a more comprehensive assessment, using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to study racial stratification across a number of socioeconomic outcomes. We pay particular attention to the robustness of results across different categorization schemes that account for self-identification and interviewer classification, as well as racial fluidity. Although we observe that White and Asian Americans generally have the best socioeconomic outcomes, on average, while Black Americans and American Indians have the worst, we also find meaningful differences in patterns of stratification both across outcomes and depending on how race is operationalized. These differences in stratification are reflected in the estimated number of strata as well as the rank order of racial categories. Our results suggest that ongoing debates about the nature of the U.S. racial hierarchy can be partly explained by methodological decisions about which outcomes to study and how best to measure race.

Keywords Racial inequality · Racial conceptualization · Quantitative methods · Survey measurement

When Du Bois ([1903] 1994) offered his prediction on the future of race relations at the turn of the twentieth century, the American racial order was largely stratified along a Black/White divide. At the time, the United States was a generation removed from a civil war fought over Black peoples’ status in American society and in the midst of a transition to a new form of racial subjugation (Blackmon, 2009). One hundred years later, at the turn of the twenty-first century, scholars revisited Du Bois’s prediction in a more racially diverse context, positing a range of outcomes from shifting binary divides to triracial or more variegated stratification schemes. Two decades into that new century, whether or how the racial hierarchy has evolved in the U.S. remains an unsettled question.

This ongoing debate provides the backdrop for our attempt to offer more comprehensive empirical estimates of U.S. racial stratification. Specifically, we consider whether the persistence of multiple competing perspectives on the racial hierarchy could be the result of studies not only focusing on different outcomes but also relying on different measures, which could lead to differing conclusions. Using data from the National Longitudinal Study of Adolescent to Adult Health, we examine a range of socioeconomic outcomes and assess how our results compare to prominent predictions and perspectives offered by scholars of race and inequality. We build on previous studies that also aimed to assess whether observed disparities differ according to how race is operationalized (e.g., Howell & Emerson, 2017; Shiao, 2019) by including measures that account for racial fluidity and multidimensionality and by considering a broader range of socioeconomic outcomes.

As expected, based on previous research, we generally observe that White and Asian Americans have the best socioeconomic outcomes, on average, while Black Americans and American Indians have the worst. However, we also find meaningful differences in patterns of stratification both across outcomes and depending on how race is operationalized. Most notably, our results do not clearly support any single prediction about the degree of stratification nor the rank order of categories in the American racial hierarchy—whether it be a White/nonwhite binary, a Black/nonblack binary, or a triracial division. Rather, the variations we observe suggest that ongoing debates about the nature of the
U.S. racial hierarchy can be partly explained by methodological decisions about which outcomes to study and how best to account for a person’s racialized experiences.

Our conclusion implies that, as scholars of race and inequality, we will need to not only pursue broader studies that account for multiple outcomes (cf. Reskin, 2012), but also more carefully justify our decisions about the measures of race we use.¹ Unlike Du Bois’s time, we live in a data-rich world, and we should strive to continually revisit our theories of the “color line” in light of new research. Greater methodological rigor will not only improve quantitative research on racial inequality, but also how we understand inequality-producing processes and might seek to undo them. The analysis presented here is our contribution towards achieving these collective goals.

**Debating the Racial Hierarchy**

We begin by offering an overview of the most prominent perspectives on the nature of the U.S. racial hierarchy circa the turn of the twenty-first century.² As noted above, these encompass predictions of various types of binary divisions, several three-category schemes, and more continuous stratification based on skin color (see Table 1).

Most assessments of the structure and future direction of the racial hierarchy continue to adhere to a binary distinction; they simply disagree about whether the divide has become one between White and “nonwhite” individuals, or between Black and “nonblack” individuals. For example, Warren and Twine (1997) argued that because whiteness is rooted in anti-blackness, the U.S. racial hierarchy will continue to be characterized by a strong divide between Black and White Americans. Although other nonblack racial groups may “become” White, even with expanded boundaries, whiteness will still be constructed in opposition to blackness. In contrast, Lee and Bean (2007) argued that a distinct Black/nonblack color line was emerging that positions White, Hispanic, and Asian Americans on one side and Black Americans on the other. Gans (1999) also predicted that the U.S. racial hierarchy was moving towards a Black/nonblack binary with a “residual” group of unclassified individuals. Numerous empirical studies have offered support for such binary distinctions, whether framed as a Black/White divide (e.g., Bratter & Gorman, 2011; Charles & Hurst, 2006; Couch & Fairlie, 2010; King & Johnson, 2016; Kreisman & Rangel, 2015; Rugh et al., 2015; Sharp & Hall, 2014) or as a Black/nonblack divide (e.g., Emeka, 2009; Fischer & Hout, 2006).

Alternative perspectives claim that the binary hierarchy of the past has been replaced by more complex “tri-partite” or “triracial” divisions. The most prominent such prediction has been articulated by Bonilla-Silva (2002, 2004). In his proposed three-category scheme, White Americans and others who have assimilated into whiteness comprise the top stratum of “Whites.” A middle stratum of “honorary Whites” includes light-skinned Hispanic Americans and several Asian-origin groups such as Korean and Japanese Americans. The bottom stratum, the “collective Black,” includes Americans of African descent and marginalized Americans with Hispanic and Asian origins as well as reservation-bound American Indians.³ Bonilla-Silva (2004) predicted each stratum will become further stratified internally by skin color, leading to what he described as a Latin-American-like “pigmentocracy.”

More recently, Gans (2012) revised his earlier binary depiction to a tri-partite hierarchy with White, light-skinned Hispanic Americans, and Asian Americans in the top stratum; darker-skinned Hispanic Americans and Asian

---

¹ Scholars have also pushed for the development of measures that capture the effects of structural racism in addition to, or instead of, more traditional individual-level approaches (e.g., Homan et al., 2021; Sewell, 2021). Incorporating measures of structural racism is unfortunately beyond the scope of our data because we do not have detailed information on place of residence with which to append them but is a crucial addition to the methodological toolkit of inequality scholars that is further supported by our findings about the sensitivity of results to different measurement strategies.

² Recent Google Scholar citation counts offer one illustration of the prominence of these perspectives: Bonilla-Silva (2004, 961 citations), Gans (1999, 357 citations), Lee and Bean (2007, 346 citations), Warren and Twine (1997, 317 citations).

³ The names of the three categories in Bonilla-Silva’s proposed hierarchy are consistent across the various published versions, though the specific composition of each group does differ somewhat. For example, Bonilla-Silva (2002) classifies Filipinos as collective Black but Bonilla-Silva (2004) classifies Filipinos as honorary White. (See also Bonilla-Silva and Dietrich, 2008).
Americans in the middle stratum; and Black and Afro-Hispanic Americans in the bottom stratum. Other scholars have argued for simpler trichotomous schemes, with clear divisions separating White, Hispanic, and Black Americans (e.g., Feliciano, 2016; Stokes-Brown, 2012). Still others have allowed for more continuous patterns of stratification by including measures of skin color (e.g., Branigan et al., 2013; Burch, 2015; Garcia et al., 2015; Monk Jr., 2015; White, 2015).

Collectively, these efforts to account for changes in the U.S. racial hierarchy have generated considerable scholarly attention and provoked plenty of debate. For example, King (2019) argues that scholars have strayed from Du Bois’s original interpretation of the color line in part by over-valuing individual actions (such as intermarriage) and under-valuing the role of structural factors (such as immigration policy) in shaping the racial hierarchy. Bonilla-Silva’s claims about the “Latin Americanization” of race in the U.S. and the coming triracial hierarchy have also prompted critical comments (e.g., Sue, 2009) and inspired several symposia, including a special issue of *Race and Society* in 2002. This sustained scholarly discussion surrounding the racial hierarchy demonstrates both the importance of the topic to scholars of race and inequality and the difficulties to date in adjudicating between competing perspectives.

From Theory to Methodology

To provide a new perspective on these important issues, we focus on a relatively underexplored aspect of research on the U.S. racial hierarchy: the empirical methodology. Before turning to a discussion of our own methods, we organize approaches from previous research around three key methodological decisions: 1) how many and which outcomes to study, 2) how many and which racial categories to compare, and 3) whether a person’s race is accounted for using a single measure at a single point in time. We pay particular attention to recent work by Howell and Emerson (2017), which provides a foundation for our own analysis.

Variation Across Outcomes

In many studies, variation in the outcome of interest coincides with variation in how race is operationalized, which complicates the task of synthesizing research on U.S. racial inequality. For example, studies of housing often focus on Black/White differences (Charles & Hurst, 2006; Freeman, 2005; Rugh et al., 2015; Sharp & Hall, 2014) while many studies that attend to skin color examine educational outcomes (Branigan et al., 2013; Hannon et al., 2013; Monk Jr., 2014). Howell and Emerson (2017) offer a rare example that enables disentangling the consequences of these separate methodological decisions.

Findings from Howell and Emerson’s study suggest that the selection of outcomes could explain differing conclusions about the racial hierarchy, though the authors do not explicitly state this as a conclusion of their study. Instead, Howell and Emerson set out to compare several common racial operationalizations, or ways of coding responses to questions about race and origins in surveys, to determine which operationalization explains the most statistical variation in common outcomes of interest in studies of inequality. To determine the “best-fitting” categorization scheme, they used data from the Portraits of American Life Study (PALS) to compare results from various racial operationalizations on three outcomes: educational attainment, household income, and self-rated health. Howell and Emerson concluded that the categorization scheme they call *pentagon race* (which includes five categories distinguishing American Indian, Asian, Black, Hispanic, and White respondents) was the best-fitting operationalization across socioeconomic outcomes, with *combined race* (a more detailed scheme that also separated out Pacific Island-origin and multiple race respondents) representing an equivalent alternative depending on the outcome of interest.

| Table 2 Comparing racial hierarchy by outcome, from Howell and Emerson (2017) |
|---------------------------------|-------------------|-----------------|
|                                  | Years of education | Household income | Self-rated health |
| Pentagon                        | Asian             | Asian           | White            |
| White                           | Hispanic          | American Indian | Hispanic         |
| Black                           | Hispanic          | American Indian | Hispanic         |
| Hispanic                        | Black             | American Indian | Hispanic         |
| American Indian                 |                  |                 |                 |
| Combined                        | Asian             | Other           | White            |
| White                           | Pacific Islander  | White           | Pacific Islander |
| Black                           | Multiple races    | Hispanic        | Hispanic         |
| Hispanic                        | Black             | Multiple races  | Asian            |
| American Indian                 |                  |                 |                  |

N=2579. Categories are ordered according to regression estimates, as reported in Howell and Emerson (2017). Hierarchical strata, demarcated with blank lines, are depicted based on the presence of statistically significant differences from the White reference group.
Notably, however, the racial hierarchy implied by their results was not consistent across all three outcomes. We summarize Howell and Emerson’s findings in Table 2 to illustrate these differences. For example, when modeled using pentagon race, they found that variation in educational attainment exhibits a tri-partite hierarchy with Asian respondents at the top, White respondents in the middle, and Black, Hispanic, and American Indian respondents at the bottom of the hierarchy. However, the implied hierarchy in household income appears to be binary with Asian and White respondents at the top, and Hispanic, American Indian, and Black respondents at the bottom. A different binary hierarchy follows from their self-rated health results with White respondents alone at the top followed by Black, Asian, Hispanic, and American Indian respondents grouped together at the bottom. Moreover, only the White/nonwhite binary for self-rated health is consistent with the prominent perspectives noted above (Table 1). These observations motivate our first research question: Is the observed racial hierarchy in the U.S. consistent when we examine it across multiple socioeconomic outcomes?

**Variation in Racial “Operationalization”**

As is the case with variation across outcomes, variation in operationalization across studies can lead to perspectives about racial inequality that are difficult to reconcile. Although all federally funded data collection is required to include a minimum number of racial categories (Office of Management & Budget, 1997), not all the categories collected end up appearing in published analyses. Whether the result of sample size constraints or researcher discretion, studies vary widely in whether and how they include Hispanic and multiracial Americans in analyses. It is also common to find American Indians, Asian and Pacific Islander Americans either combined in a residual category or excluded from analysis entirely. These “lumping” and “splitting” decisions could have important implications for understanding stratification. For example, it is possible that studies are more likely to find support for binary hierarchies when they compare outcomes for just two or three categories, compared to studies that account for four or more racial categories (see e.g., Akee et al., 2019; Bailey et al., 2014; and Krivo & Kaufman, 2004). However, few studies are specifically designed to compare results for the same outcome across different categorization schemes. We again focus attention on Howell and Emerson’s (2017) approach but note that studies by Light and Nandi (2007) and Shiao (2019) were motivated by similar aims.

In their study, Howell and Emerson compared five racial operationalizations, which they selected based on examples from previous research and dub: census race, combined race, pentagon race, triracial, and skin color. Since the turn of the twenty-first century, the U.S. census and many national surveys have ascertained racial and ethnic information about respondents by asking two questions: the first about whether the person has “Hispanic, Latino, or Spanish” origins, and the second about which race(s) best describe the person. Their census race scheme includes only responses to the “race” question and categorizes respondents mutually exclusively as: monoracial White, monoracial Black, monoracial Asian, monoracial Pacific Islander, monoracial American Indian, or multiracial. Pentagon race uses information from both questions to identify the five groups corresponding to Hollinger’s (1995) characterization of the U.S. ethnорacial “pentagon”: American Indian, Asian, Black, Hispanic, and White. Combined race also draws on responses to both questions and retains a separate multiracial category. Triracial represents their attempt to operationalize Bonilla-Silva’s (2004) hierarchy of “White,” “honorary White,” and “collective Black” using survey data, and their skin color measure is a continuous ranking of respondents’ skin color from “very light” to “very dark” according to classification by the PALS interviewer.

Although Howell and Emerson concluded that pentagon race was the overall best-fitting operationalization for the outcomes they examined, they did find that different categorization schemes provided better fits to the observed data for some outcomes compared to others. It is also possible for different operationalizations to yield different characterizations of the racial hierarchy on the same outcome. This could be because they identify different levels of stratification (i.e., a tri-partite vs. a binary hierarchy), a different ranking of categories, or both. Their comparison between pentagon race and combined race suggested little substantive difference between the two coding schemes, with both indicating a tri-partite hierarchy for educational attainment and similar binary hierarchies for household income and self-reported health (see Table 2). However, these hierarchy depictions are based solely on statistically significant differences with the White reference group. Additional information from post-estimation tests of pair-wise differences between estimates for the various racial categories included in each model might have yielded more nuanced differences in stratification between the two racial categorization schemes.

Thus, we aim to replicate and extend Howell and Emerson’s efforts using a different dataset. Like them, we will assess the best-fitting operationalization across a series of

---

1 Howell and Emerson (2017) run standard regressions and did not test all pair-wise differences in the race estimates (e.g., estimate for Black respondents compared to the estimate for Asian respondents) which would allow for more precision in determining observed stratification for each outcome-specific hierarchy. So, our depiction in Table 2 relies solely on the reported coefficients and their statistical significance compared to White respondents, as the reference group.
socioeconomic outcomes, but we also will consider potential substantive variation according to the different racial categorization schemes. In doing so, we will answer our second research question: Does the observed racial hierarchy also differ by how “race” is operationalized?

Racial Fluidity and Multiple Measures of Race

The final motivation for our study comes from recent research that calls into question static, single-measure approaches to accounting for race in studies of inequality. This body of work has considered both the multidimensionality of race and its potential fluidity over time or across contexts. We briefly summarize key takeaways from each of these approaches, in turn.

A recent review by Roth (2016) highlights the importance of acknowledging conceptual and empirical differences between different measures of race. Roth identifies multiple “dimensions” of race, including racial identity, racial self-classification, observed race, reflected race, phenotype, skin color, and racial ancestry. Not all studies (or surveys) include all these dimensions, but studies that do make use of multiple measures across these dimensions often find that each provides different estimates for the degree of inequality (Saperstein et al., 2016). These differences may also point to different mechanisms that reproduce racial inequality varying from contemporary discrimination based on physical appearance to accumulated advantage or disadvantage depending on ancestry (e.g., Penner & Saperstein, 2015). Collectively, this research highlights the importance of using multiple measures of race to triangulate findings.

Studies also show that, despite the tendency to treat race as a static identity or classification, racial fluidity in the U.S. is not uncommon (Davenport, 2020; Liebler et al., 2017; Nagel, 1995). The presence of such fluidity is sociologically intriguing in and of itself but takes on additional importance given evidence suggesting that fluidity is likely related to inequality. In a series of studies across several datasets, Saperstein and colleagues find that changes in both racial self-identification and interviewer classification coincide with changes in various socioeconomic indicators (Saperstein & Gullickson, 2013; Saperstein & Penner, 2012, 2016; Saperstein et al., 2014). They find that respondents who experienced negative outcomes such as unemployment or contact with the criminal justice system were more likely to self-identify and be classified by an interviewer as Black. At the same time, respondents who experienced positive outcomes such as college completion or marriage were more likely to self-identify and be classified by an interviewer as White. The significance of accounting for racial fluidity in studies of inequality is echoed by similar research in other countries (Simonovits & Kézdi, 2016) and for other outcomes (Agadjanian & Lacy, 2021).

A few studies also have considered whether people with fluid or inconsistent racial categorizations have significantly different socioeconomic outcomes from people with consistent or stable categorization. For example, Kramer et al. (2015) examined mental health and educational outcomes for adolescents who (in)consistently identified their race. They find that respondents who inconsistently identified their race do not have worse mental health or educational outcomes than respondents who consistently identified their race. In contrast, Shiao (2019) found that accounting for inconsistent racial classification helps to explain more variation in interracial relationship history, educational attainment, and self-rated health than accounting for consistent racial classification alone. Research on “racial contestation” also suggests substantively meaningful stratification exists with people who are seen as White by others generally having better outcomes even when they do not self-identify as White (see Vargas & Kingsbury, 2016 for a review).

Taken together, this research indicates that studies of racial inequality that rely solely on one measure of race collected at one point in time may miss important patterns and associations that have implications for life outcomes. This brings us to our third and final research question: Does accounting for both multiple measures of race and racial fluidity among our racial “operationalizations” offer different conclusions about the U.S. racial hierarchy?

Overall, among recent quantitative studies that aimed to understand the racial hierarchy in the United States, we observed wide-ranging methodological variation in the socioeconomic outcomes examined, the number and type of racial categories included, and whether a person’s race was allowed to vary over time or across different observers. Beyond these three dimensions of methodological variation, conclusions about the nature and extent of the racial hierarchy may also depend on whether racial divisions are examined in combination with other axes of difference. For example, Kim (2015) observed that the racial hierarchy shifts from White/nonwhite when studying U.S.-born Americans to Black/nonblack when studying immigrants (see also Rugh, 2020). The extent of this methodological variation, along with its potential effect on substantive conclusions, is central to our general takeaway from reviewing the literature: without consistency across studies, it is perhaps unsurprising the field has yet to support definitive claims about the racial hierarchy.

Data and Methods

The data used in this study are from the National Longitudinal Study of Adolescent to Adult Health (Add Health) (Harris, 2009). Add Health is a survey of both health and
socioeconomic outcomes that began in 1994 when the respondents were in grades 7–12. Our analysis draws on data from Wave I (1994–1995), Wave III (2001–2002), and Wave IV (2008). The key feature of Add Health that makes it especially useful for this study is the multiple measurements of race—both self-identified and interviewer classified—across multiple waves of data.

**Static Racial Operationalizations**

Our measures of static racial self-identification are drawn from questions in Wave III. Respondents were first asked whether they have Hispanic or Latino origins, and those who answered “yes” were asked about their specific national origin or ancestry (e.g., Mexican, Cuban, etc.). Respondents then were asked for their racial identification using the categories White, Black, Asian or Pacific Islander, and American Indian. Respondents could select multiple race categories, but a subsequent question asked respondents who did so to also select their “best” single race. We use these measures to replicate Howell and Emerson’s (2017) coding for the census, combined, and pentagon race categorization schemes. Frequencies are fairly stable across these three coding schemes, though cell counts for “White” and “American Indian” fluctuate most depending on whether Hispanic is included as a separate category (see Table 3).

For example, when Hispanic origin responses effectively override race responses (as in combined race and pentagon race) there are roughly 200 fewer respondents who self-identify as American Indian and roughly 1000 fewer respondents who self-identify as White.

We also combine information from self-identification and other survey responses from Wave III, such as skin color and urban/rural residence, to operationalize Bonilla-Silva’s (2004) triracial hierarchy. We aimed to match Bonilla-Silva’s formulation as closely as possible, given the available data. The coding of triracial in the analysis below offered the best fit across our outcomes of the several options we tested (see Online Appendix A for more details).

We include a static measure of interviewer-classified race and a stand-alone measure of skin color in our comparisons as well. We use Wave III data for our static interviewer-classified race measure to match the Wave III measures of self-identification described above. Add Health interviewers were asked to assign each interviewee to only one race from among the following categories: White, Black, Asian or Pacific Islander, and American Indian. The exclusion of a “Hispanic or Latino” category and the inability of interviewers to select multiple races does not allow for a direct comparison between self-identified and interviewer-classified race. Nevertheless, the inclusion of interviewer classification allows us to consider an additional operationalization of race beyond what is available in most surveys, including PALS.

Skin color also was recorded by the interviewer at the end of the Wave III survey using a five-point nominal scale (white, light brown, medium brown, dark brown, and black). Like Howell and Emerson (2017), we tested skin color coded in three ways—as a continuous variable, a 5-point categorical variable, and a 3-point categorical variable (see Online Appendix A for details). Also like Howell and Emerson, our best-fitting operationalization of skin color was the continuous measure, which we present in the analysis below. To create the continuous variable, the five skin color categories are recoded on a scale from 0 to 1. This recoding of skin color from categorical to continuous means that the model result for a given outcome represents the change in the outcome from white to black skin color.

---

**Table 3** Cell frequencies across static racial operationalizations

| Racial category | Combined race | Census race | Pentagon race | Interviewer-classified race |
|----------------|--------------|-------------|--------------|----------------------------|
| American Indian | 67           | 276         | 88           | 160                        |
| Asian          | 536          | 592         | 574          | 636                        |
| Black          | 1999         | 2076        | 2103         | 2178                       |
| White          | 5731         | 6855        | 5929         | 7195                       |
| Multiracial    | 361          | 370         |              |                            |
| Hispanic       | 1475         |             | 1475         |                            |

*Source* Add Health Wave III

Total *N* = 10,169. The combined race, census race, and pentagon race coding schemes are drawn from Howell and Emerson (2017)

---

5 For the sake of brevity, we use the term “Asian” instead of “Asian or Pacific Islander.” Howell and Emerson (2017) keep the two categories separate, but we do not because the number of respondents specifying an Asian ancestry far outnumbers the number of respondents reporting a Pacific Islander ancestry, which results in too few cases for Pacific Islander respondents, particularly for our dynamic race variables.

6 We only use the “best race” measure to recode respondents to a single race (*N* = 370) for the pentagon race coding scheme, which does not allow for multiple race self-identification.
Dynamic Racial Operationalizations

Our dynamic operationalizations reflect the consistency of racial self-identification or interviewer classification across survey waves and involve a series of three binary indicators—always, ever, and never—for each of the racial categories offered in the survey. Respondents who report the same race in two waves are coded as “always” the same race (e.g., always Black). Respondents who report a race in at least one wave, but not both, are coded as “ever” that race (e.g., ever Black). Respondents who never report a given race are coded as “never” that race (e.g., never Black). Both of the dynamic race variables in our analysis—one based on self-identification and one based on interviewer classification—emerged from our sensitivity tests as the best-fitting among several options (see Online Appendix A).

Our dynamic self-identification measure draws on data from Waves I and III and is straightforward for monoracial responses, as described above. However, respondents who listed multiple races can count as both always [RACE] and ever [RACE]. For example, if a respondent reported their race as White and Asian in Wave I but reported their race as monoracial White in Wave III, they would be coded as always White—because they reported their race as White in both waves—and ever Asian—because they reported their race as Asian in at least one wave (cf. Shiao, 2019). This coding scheme was preferred over two alternate ways of accounting for changes in multiple race responses over time (see Online Appendix A). Our dynamic interviewer classification measure draws on data from Waves III and IV.

Outcome Variables

The six outcomes we measure—all drawn from Wave IV when the respondents were between 24 and 32 years old—are: arrest, college completion, employment, homeownership, per capita household income, and self-rated health. These outcomes were selected because they were common in published work that explicitly explored the contemporary racial hierarchy, as we reviewed earlier. Per capita household income is coded as a continuous variable that measures the respondent’s total household income divided by the number of residents in the respondent’s household. Self-rated health is an ordinal measure ranging from 1 (poor health) to 5 (excellent health). Our other four outcomes are binary indicators for: never (vs. ever) reporting an arrest, earning a bachelor’s degree, being employed at the time of the survey, and owning their own residence.

Control Variables

We account for a range of respondent and interviewer characteristics as controls in our models. Although there is some debate about the appropriateness of including other individual-level characteristics in an attempt to ‘isolate’ racial disparities from variation related to other measured factors (see, e.g., Stewart, 2008), we include analogous versions of the respondent characteristics used as controls by Howell and Emerson (2017). These characteristics are: self-reported sex (female vs. male), age (in years), nativity (U.S. born vs. foreign born), marital status (married vs. not married), whether the respondent lives in a rural area (rural vs. non-rural), and their current region of residence (Northeast, Midwest, South, West). We also include binary indicators for whether a respondent moved to a different region between Waves III and IV and whether their mother completed college (a measure of family background drawn from Wave I) because previous research finds these factors are related to changes in racial self-identification (see, e.g., Doyle & Kao, 2007). As our analysis draws on interviewer observations of categorical race and skin tone, we control for interviewer characteristics as well to account for differences in how interviewers classify respondents. The interviewer characteristics are: self-reported sex, self-reported race (White, Black, Asian, American Indian, and other) and Hispanic origin, age, college completion, and whether the interviewer speaks Spanish. All control variables are drawn from Wave IV data unless otherwise noted. We dropped respondents with missing data from our analytic sample, so our models include only respondents for whom we have complete data on all key variables (N = 10,169).

Models

To compare the different ways of operationalizing race, we estimated a series of regression models and assessed the amount of variation explained for each of our six outcomes. The binary outcome variables were modeled using logistic regression while per capita household income and self-rated health were modeled using ordinary least squares (OLS) regression. For each outcome variable, we estimated eight models (one for each racial operationalization) and all models included the full set of control variables. We used post-estimation commands in Stata 14.2 to retrieve the model Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which we then used to determine which operationalization of race explains the most variation in a given outcome of interest.8

7 We modeled self-rated health using both OLS and ordered logistic regression. The results (available upon request) were substantively similar, so we present OLS estimates for ease of interpretation.
8 Howell and Emerson (2017) use the Vuong test to compare model fit (Vuong, 1989). One of the benefits of the Vuong test is that it allows one to compare non-nested models and models with different functional forms. The drawbacks of the Vuong test, however, are that model fit does not penalize the inclusion of additional parameters and
Assessing Fit and Stratification

To assess model fit with BIC statistics we use Kass and Raftery (1995) as a guide to determine whether a particular model offers more explanatory power than a competing model. When comparing BIC (or AIC) statistics, the smaller the number the better the model fit. If the difference between BIC statistics is between zero and two points, we conclude there is no meaningful difference in model fit between the relevant models. BIC differences greater than two points suggests there is a meaningful difference in model fit, with differences between six to 10 points indicating a “strong difference” and differences greater than 10 points indicating a “very strong difference.” Burnham and Anderson (2004) offer a guideline for comparing AIC statistics that uses slightly different cutoffs (0–2, 4–7, and > 10). For consistency, we opted to apply the BIC guideline to compare both sets of statistics because the AIC guideline does not cover the full range of possible point differences. Here it is important to note that although we can directly compare the set of racial operationalization models within each outcome, our comparison of racial operationalization models across outcomes is more speculative because four of the six outcomes are modeled using logistic regression and the other two are modeled using OLS regression.

All of the analyses we present are unweighted. We do this for several reasons. First, as noted above, we rely on post-estimation tests that report AIC and BIC statistics as measures of model fit. We cannot estimate our regression models with weights and calculate the measures of model fit that provide key evidence for our research questions. Second, Add Health’s weighting scheme uses self-identified race categories while our analyses account for multiple measures of race, including those recorded by the interviewer. Third, sampling weights are used to present estimates that are representative of the national population. Our analysis is not focused on establishing such population parameters; rather, it is focused on comparing how estimates differ by model specification and outcome. Finally, Bollen et al. (2016) concluded in a recent review that there is no cross-disciplinary consensus or clear-cut guidance on when post-stratification survey weights are necessary nor how to choose the appropriate ones; this is particularly the case for logistic regression models, which we use for four of our six outcomes. Nevertheless, our inclusion of controls for respondent’s sex, age, and place of residence does help to account for some of the ways our analytic sample is unrepresentative and avoids attributing such differences to racial disparities in our outcomes of interest.

To visualize the racial hierarchy implied by these data, we order racial categories according to their estimated coefficients and whether differences between categories were statistically significant. We used post-estimation tests to determine whether the estimated coefficient for each racial category was statistically significantly different from the estimated coefficients for other racial categories in the model as well as the reference category.9 We used the resulting patterns of statistical significance to determine the number of strata to present in our summary tables (Tables 7 and 8). A blank line in the tables marks the boundary between strata, where estimates for categories above the blank line were statistically significantly different from estimates for categories shown below the blank line. Where results showed estimates were not statistically significantly different from one another, we grouped those racial categories in the same stratum to signal they had more similar average experiences on the outcome of interest. In cases of mixed stratum, where some groups were statistically significantly from each other and some were not, we also considered the magnitude of the estimated coefficients (see Online Appendix B) to determine where to split strata. Although this approach errs on the side of presenting a greater number of strata in the observed hierarchies, we think this better represents the substance of the estimates. Further, this type of decision generally affected hierarchies with three or more strata, so it does not affect our conclusion about the relative prevalence of binary hierarchies in our data.

Results

We begin with an overview that highlights which racial operationalizations explain the most variation across our six outcomes according to both fit statistics. We then examine how substantive conclusions about the racial hierarchy vary across outcomes and across racial operationalizations, showing how and why methodological choices can produce divergent understandings of the racial hierarchy. Finally, we highlight how dynamic measures of race can help improve studies of the racial hierarchy by capturing variation that static measures cannot. For ease of comparison, we summarize the resulting racial hierarchies and relative model fits in the tables presented below. (See Online Appendix B

Footnote 8 (continued)

9 The reference category varies across models, from “White,” monoracial White, or monoracial and non-Hispanic White in the static models to “always White” in the dynamic models. (See Online Appendix B for full results).
for the full regression model results for each outcome and racial operationalization).

**Best-Fitting Racial Operationalizations**

There is no universally best-fitting racial operationalization across the outcomes we examined. Of the eight racial categorization schemes we compared, six are preferred by at least one fit statistic for at least one outcome (see Table 4). Generally, we find that operationalizations based on self-identification offer better fits compared to interviewer classification across our outcomes. This finding is consistent across both static and dynamic racial operationalizations and holds for both interviewer-recorded skin tone and interviewer-classified race. However, given the variation in model fit not only across outcomes and racial operationalizations but also across fit statistics, we caution against aiming for a single gold-standard racial categorization scheme in studies of racial stratification. Instead, we encourage exploring when and why certain operationalizations fit the observed data best.

Among the racial categorization schemes that are based on self-identification, we find the strongest support for combined race. For three of the six outcomes we measure (home ownership, college completion, and per capita household income), combined race has the lowest BIC statistic (see Table 5), meaning its racial categorization scheme strikes the best balance between explaining the most variation in those outcomes of interest with the fewest parameters. For example, for home ownership, combined race has a BIC statistic of 12,231 while the next best models are census race (12,234) and skin color (12,236). According to the rubric outlined above, the three- and five-point differences in model fit are both “meaningful.” We see greater variation in model fit when considering combined race against the second and third best-fitting operationalizations for college completion and per capita household income. For household income, the five-point difference in BIC statistics between combined race (229,532) and pentagon race (229,537) is “meaningful” while the six-point difference between combined race and skin color (229,538) is considered “strong.” For college completion, the five-point difference in BIC statistics between combined race (11,740) and pentagon race (11,745) is “meaningful” while the 29-point difference between combined race and census race (11,769) is “very strong.”

The evidence from BIC differences that combined race is consistently favored over skin color is particularly striking given that operationalizing combined race requires four additional parameters. Although the continuous measure of skin tone is the most parsimonious of all the

---

**Table 4** Preferred racial operationalization by BIC and AIC

| Outcome                      | BIC     | AIC     |
|------------------------------|---------|---------|
| Never arrested               | Pentagon| Dynamic Self-ID |
| Home ownership               | Combined| Combined |
| College                      | Combined| Combined |
| Employment                   | Skin Color| Interviewer (W3) |
| Per capita household income  | Combined| Dynamic Self-ID |
| Self-rated health            | Triracial| Dynamic Self-ID |

*Source* Add Health Waves I, III, and IV

The racial operationalizations shown are the best-fitting categorization schemes according to each fit statistic. For the full array of Bayesian Information Criterion (BIC) statistics, see Table 5 and Online Appendix B for the Akaike Information Criterion (AIC) statistics.

---

**Table 5** BIC statistics by outcome and racial operationalization

| Static operationalizations | Never arrested | Home ownership | College | Employment | Per capita household income | Self-rated health | Parameters |
|----------------------------|----------------|----------------|---------|------------|-----------------------------|-------------------|------------|
| Skin color                 | 10,924         | 12,236         | 11,787  | 3501       | 229,538                     | 26,945            | 21         |
| Triracial                  | 10,922         | 12,247         | 11,824  | 3505       | 229,556                     | 26,921            | 22         |
| Interviewer (W3)           | 10,923         | 12,237         | 11,785  | 3511       | 229,573                     | 26,965            | 23         |
| Census                     | 10,918         | 12,234         | 11,769  | 3523       | 229,546                     | 26,964            | 24         |
| Pentagon                   | 10,917         | 12,238         | 11,745  | 3524       | 229,537                     | 26,927            | 24         |
| Combined                   | 10,932         | 12,231         | 11,740  | 3532       | 229,532                     | 26,939            | 25         |

| Dynamic operationalizations| Never arrested | Home ownership | College | Employment | Per capita household income | Self-rated health | Parameters |
|----------------------------|----------------|----------------|---------|------------|-----------------------------|-------------------|------------|
| Interviewer classification | 10,940         | 12,262         | 11,804  | 3546       | 229,594                     | 26,994            | 27         |
| Self-identification        | 10,973         | 12,287         | 11,796  | 3582       | 229,569                     | 26,973            | 32         |

*Source* Add Health Waves I, III, and IV

The best-fitting racial operationalizations for each outcome, those with the lowest Bayesian Information Criterion (BIC) statistic, for each outcome are noted in **bold**. When the BIC difference is not meaningful (i.e., less than two points), we note similarly well-fitting operationalizations in *italics*. See Online Appendix B for full model results.
operationalizations we tested, it yielded the best model fit in only one case: employment (see Tables 4 and 5). For employment, the single continuous measure of skin tone best captured the variation other measures summarized as a binary Black/nonblack hierarchy. For most other outcomes the skin color measure provided worse fit in part because observed average experiences did not differ monotonically from white to black skin color or vice versa (see Online Appendix A for further discussion).

When combined race yields better model fit, it points to the importance of including separate categories for Hispanic and multiracial respondents. One outcome that helps to demonstrate what distinguishes combined race from the other static approaches is home ownership. As noted above, according to the BIC, combined race is followed by census race in offering the best-fitting racial operationalization. Also recall, the difference between these two classification schemes is that combined race includes a separate category for people who identify as Hispanic, while census race categorizes respondents by their race response only, regardless of whether they also reported a Hispanic origin (see Table 3). This distinction turns out to be important for home ownership as non-Hispanic White respondents and Hispanic respondents have statistically significantly different outcomes (see Table B4 in Online Appendix B).

By contrast, the results for self-rated health show why combined race might not always be the preferred operationalization. Overall, the models for self-rated health indicate a binary hierarchy with White respondents reporting significantly better health than all others and with the combined race, Wave III interviewer classification, and triracial measures all pointing to this same White/nonwhite divide. However, because triracial is more parsimonious than combined race—it uses three categories compared to six to capture a binary hierarchy—it is the preferred racial operationalization for self-rated health. These examples foreshadow a theme we turn to in the next section—variation across outcomes.

Our dynamic race categorization schemes require the most parameter estimates of all the racial operationalizations we examine at 27 and 32, for interviewer classification and self-identification, respectively. So, perhaps unsurprisingly, our dynamic race variables do not perform well according to the BIC model fit statistics (see Table 5). However, the AIC model fit statistic does not penalize more complex models, and when we use that as a (less conservative) benchmark, we find that dynamic self-identification is favored as the best-fitting operationalization for three of our six outcomes: arrest, household income and self-rated health (see Table B1 in Online Appendix B). We interpret this to mean that, although they may be more trouble than they are worth for some purposes, measures that account for the possibility that Americans may not identify consistently with the same racial categories over time are an important addition to the toolkit of researchers interested in better understanding U.S. racial inequality (see also Shiao, 2019).

**Variation Across Outcomes**

Across our six outcomes, there is broad agreement between our results and previous research. Scholarship on racial inequality has shown that White and Asian Americans tend to have better socioeconomic outcomes than Black and Indigenous Americans and our results support that conclusion. What previous studies have not done, however, is to show how varied conclusions about the racial hierarchy can be even when using the same racial categorization scheme to model racial stratification. To focus on variation across outcomes, we visualize the racial hierarchy implied by our model results for just one racial operationalization (see Table 6). We chose to focus on combined race for this illustration because of its performance in terms of overall model fit, as discussed above.

Our combined race results for home ownership, self-rated health, and employment all indicate a binary hierarchy, but in importantly different ways. The home ownership and self-rated health results both exhibit a White/nonwhite hierarchy, but for home ownership Black respondents are at the lowest position in the nonwhite stratum while for self-rated health American Indian respondents are at the lowest position in the nonwhite stratum. In contrast, the employment results show a Black/nonblack rather than a White/nonwhite hierarchy. Although these three outcomes yield the same number of strata, they are binary in varied ways, thus, demonstrating the importance of specifying an outcome of interest when discussing the racial hierarchy. Moreover, the variation in binary hierarchies is a subset of the variation across outcomes. As the results in Table 6 show, the same racial operationalization also can yield hierarchies that are tri-partite (i.e., college completion) or even more variegated (i.e., arrest).

In addition to variation in the amount of stratification (i.e., binary, tri-partite, etc.), there is great variation in hierarchical ordering when viewed across outcomes (Table 6). Although White and Asian respondents are generally at the top of the hierarchy, their order often switches between one group or the other occupying a solo spot atop the hierarchy (i.e., arrest, home ownership, college completion, and self-rated health) or a joint spot atop the hierarchy (i.e., employment and per capita household income). At the other end of the hierarchy we often see Black and American Indian respondents together at the bottom, but we also see rather interesting deviations. For college completion and self-rated health we see Black respondents in the middle of the hierarchy; most notably for self-rated health, Black respondents
report better health, on average, than Asian respondents. For employment, we see an anomalous pattern for American Indian respondents where they are positioned in the middle of the hierarchy above multiple race, Hispanic, and Black respondents. Taken together, this variation shows the hierarchy order is far from uniform and complicates the stylized hierarchies as drawn from the literature (Table 1). Allusion to a racial hierarchy with White individuals consistently at the top and Black individuals consistently at the bottom is not supported by the results presented here. The racial hierarchy—both the level of stratification and its order—is often outcome specific and more attention must be paid to such differences.

**Variation Across Operationalizations**

We now consider the full array of results across operationalizations and outcomes (see Table 7). Overall, there is more consistency in terms of both the ordering of the racial hierarchy and the estimated number of strata when comparing between the various static racial categorization schemes (reading down the columns in Table 7) than when comparing between outcomes (reading across the rows). As with the variation across outcomes for only combined race discussed above, we find that White and Asian respondents are generally at the top of the hierarchy, and Black and American Indian respondents are at the bottom. However, there are several instances when the patterns of stratification differ in unexpected ways. To highlight the differences in the hierarchy across racial operationalizations, let us take the employment and college completion results as examples.

In terms of the order of the hierarchy, when we look across racial operationalizations for a given outcome, we see little variation in rank order indicating this aspect of the hierarchy is relatively robust to the inclusion of new categories (e.g., separating out Hispanic respondents when going from census race to combined race). An exception to this pattern is the static interviewer classification result for employment. For employment, the static self-identification measures (i.e., census race, combined race, and pentagon race) show White respondents having better employment outcomes than American Indian respondents. However, when we turn to static interviewer-classified race, we see that respondents perceived as American Indian have better employment outcomes than respondents perceived as White. Notably, employment is the only outcome for which static interviewer-classified race is found to be among the better fitting operationalizations (see Tables 4 and 5).

Turning to the degree of stratification, we see that there are some outcomes for which the hierarchy is robust to the racial operationalization used in analysis but there are other outcomes for which the choice of racial categorization scheme affects substantive conclusions. For example, the results for employment (Table 7) point to a binary racial hierarchy, regardless of racial operationalization. All the static, categorical measures (census, combined, pentagon, interviewer classified, and triracial) show White respondents having better employment outcomes than American Indian respondents. However, when we turn to static interviewer-classified race, we see that respondents perceived as American Indian have better employment outcomes than respondents perceived as White. Notably, employment is the only outcome for which static interviewer-classified race is found to be among the better fitting operationalizations (see Tables 4 and 5).

| Source | Add Health Waves III and IV. Summary illustration based on series of estimated models (see Online Appendix B) |
|---|---|
| **Table 6** Racial hierarchy across outcomes for overall best-fitting operationalization | **Table 6** Racial hierarchy across outcomes for overall best-fitting operationalization |

| Source | Add Health Waves III and IV. Summary illustration based on series of estimated models (see Online Appendix B) |
|---|---|
| **Table 6** Racial hierarchy across outcomes for overall best-fitting operationalization | **Table 6** Racial hierarchy across outcomes for overall best-fitting operationalization |

| Combined race | Asian | White | Asian | Asian | Asian | White | White |
|---|---|---|---|---|---|---|---|
| White | Asian | Hispanic | White | American Indian | Hispanic | Multiple races | Black | Hispanic |
| Hispanic | Multiple races | Black | Hispanic | Multiple races | American Indian | Hispanic | Black |
| Multiple races | American Indian | Hispanic | Black | American Indian | Hispanic | Black | Black |
| Black | American Indian | Hispanic | Black | American Indian | Hispanic | Black | Black |

| Never arrested | Home ownership | College | Employment | Per capita household income | Self-rated health |
|---|---|---|---|---|---|
| Combined race | Asian | White | Asian | Asian | White |
| White | Asian | Hispanic | White | American Indian | Hispanic | Multiple races | Black | Hispanic |
| Hispanic | Multiple races | Black | Hispanic | Multiple races | American Indian | Hispanic | Black | Hispanic |
| Multiple races | American Indian | Hispanic | Black | American Indian | Hispanic | Black | Black |
| Black | American Indian | Hispanic | Black | American Indian | Hispanic | Black | Black |

*Black and Hispanic are statistically significantly different from each other, but the remaining race combinations depicted in the same strata are not statistically significantly different from one another.

*Black and American Indian are not statistically significantly different from one another, but Black is statistically significantly different from all other racial categories in the model.
Table 7  Comparing racial hierarchy across outcomes, with static race operationalization

| Source | Add Health Waves III and IV. Summary illustration based on series of estimated models (see Online Appendix B) |
|--------|----------------------------------------------------------------------------------------------------------|
| a      | Black and Hispanic are statistically significantly different from each other, but the remaining race combinations depicted in the same strata are not statistically significantly different from one another |
| b      | Black and American Indian are not statistically significantly different from one another, but Black is statistically significantly different from all other racial categories in the model |
| c      | Hispanic and American Indian are not statistically significantly different from one another, but Hispanic is statistically significantly different from all other racial categories in the model |
| d      | American Indian and White are statistically significantly different, but American Indian is not statistically significantly different from Asian |

racial categories the odds of graduating are statistically significantly different from all the others. If we were to estimate college completion using only a triracial coding of race, we would conclude that the racial hierarchy is stratified along a Black/nonblack divide. If we were to use a pentagon race coding scheme, however, our results would suggest an
entirely different conclusion about racial stratification in college completion.

**Value Added by Dynamic Operationalization**

Given the larger number of parameters employed, our dynamic racial categories may introduce more complexity than is strictly necessary. Nevertheless, we do find the dynamic operationalizations helpful for specific purposes. For example, the added value of using dynamic self-identification is clearest when comparing findings regarding the racial hierarchy in self-rated health. The best-fitting operationalization according to the BIC statistic was triracial (see Table 5); it suggests a White/nonwhite binary in which White respondents are at the top and honorary White and collective Black respondents are grouped together at the bottom with honorary White respondents reporting the worst average self-rated health of the three groups (see Table 7). Dynamic self-identification, on the other hand, yields a tripartite hierarchy where the bottom stratum is comprised of respondents who always identify as Asian or American Indian, or ever identify as Black (see Table 8). The middle stratum is comprised of respondents who always identify as Black, ever identify as American Indian, and respondents who identify as Hispanic regardless of whether they do so consistently or not. In the top stratum are respondents who always identify as White or always identify as multiracial along with respondents who sometimes identify as White, Asian, or multiracial. Broadly speaking, the order revealed using dynamic self-identification is similar to the order found using the three static self-identification measures—White respondents at the top, Black and Hispanic respondents in the middle, and Asian and American Indian respondents at the bottom. However, if we only used race measures from one survey wave, we would miss the grouping of many of the “ever” categories at the top of the self-rated health hierarchy. This aligns with previous research suggesting that racial fluidity, in and of itself, is not a disadvantaged status (Kramer et al., 2015; Saperstein & Penner, 2012).

By distinguishing fluid and consistent categorizations, our dynamic schemes also offer clarity when other operationalizations compete in the estimates. For example, when we consider the racial hierarchy for household income, the best-fitting static schemes (i.e., combined and pentagon) are split on whether American Indian or Black respondents have the worst outcomes (see Table 7). However, both dynamic self-identification and dynamic interviewer classification indicate that respondents who consistently self-identify as Black or consistently self-identify as American Indian share the bottom rung of the income hierarchy while their fluid counterparts are relatively better off (see Table 8). Similarly, for arrest, respondents who consistently self-identify as American Indian or are classified as such are more likely to report having been arrested while their fluid counterparts are less likely to have had such contact with the criminal justice system.

By considering the relative positions of the “ever” and “always” categories for each racial category across all the outcomes in Table 8, the results for our dynamic variables offer broader insight into the U.S. racial hierarchy as well. Across all six socioeconomic outcomes, we find that respondents who “ever” self-identify as American Indian have consistently better average outcomes than respondents who “always” self-identify as American Indian. At the other end of the spectrum, we find that respondents who “always” self-identify as White tend to be equally well off (or better off) than respondents who “ever” self-identify as White. Similarly, respondents who “always” self-identify as multiracial tend to be better off, on average, than respondents who “ever” self-identify as multiracial. In between, we find that for respondents who self-identify as Asian or Hispanic, “ever” and “always” are evenly split between better and worse socioeconomic outcomes, while people who “ever” self-identify as Black tend to have better average outcomes than people who “always” self-identify as Black. Thus, by comparing socioeconomic status for respondents who are solidly in a category with respondents whose category membership is more contingent, we reveal a more variegated hierarchy—one that is more frequently observed as tri-partite or even more stratified—but one that still points towards White and Asian respondents having better outcomes on indicators of socioeconomic wellbeing than Hispanic and multiracial respondents, who in turn have better outcomes than Black and American Indian respondents.

**Discussion**

With this analysis we sought to answer three questions: (1) whether we observe a consistent racial hierarchy across several socioeconomic outcomes; (2) whether conclusions about the racial hierarchy depend on how race is operationalized; and (3) whether accounting for multiple measures of race and racial fluidity offers different conclusions about the racial hierarchy? Overall, the results presented here support the broad conclusion that average outcomes tend to put White and Asian Americans at the top of the U.S. racial hierarchy and Black Americans and American Indians at the bottom of the racial hierarchy, but we find notable exceptions that vary in response to methodological decisions and complicate how we think about racial stratification in the United States.

In answer to our first research question, the results for combined race—the racial operationalization most often preferred relative to other coding schemes in our data—show that depending on the outcome, the hierarchy could
can be characterized as White/nonwhite or Black/nonblack, but not necessarily with White individuals and Black individuals alone at the extremes (see results for employment and self-rated health in Table 6). Moreover, there are outcomes for which the degree of stratification stretches beyond a simple binary divide (see results for arrest, college completion and household income in Table 6).

In answer to our second research question, we find that although some categorization schemes offer markedly better statistical fits than others, substantive conclusions about the racial hierarchy are less varied by operationalization than by outcome. There are instances, though, when the choice of operationalization is also substantively meaningful—as when adjudicating between *triracial* and *pentagon* for college completion. Ironically, when our operationalization of Bonilla-Silva’s (2004) triracial hierarchy offers the best statistical fit, it highlights a binary White/nonwhite divide in self-rated health (Table 7) rather than the expected tri-partite division. So, not only do we find that the racial hierarchy differs across outcomes and operationalizations, but also the racial hierarchies most often referenced in the literature lack consistent empirical support from Add Health data.

And in answer to our third research question, when we turn our attention to dynamic operationalizations, we see that they do not perform well when the benchmark is parsimony, but by not reifying racial categories into static identities or “groups” (Brubaker, 2004), they offer a more nuanced picture of the racial hierarchy. This ability to highlight contingency or proximity to standard categories helps to underscore which consistent racial categorization

| Source | Add Health Waves I, III, and IV. Summary illustration based on series of estimated models (see Online Appendix B) |
|---------|------------------------------------------------------------------------------------------------------------------|
| AIAN    | American Indian (and Alaska Native). Italics indicate that the outcome is not statistically significantly different from always White (the reference category) |

In Table 8, we compare racial hierarchy across outcomes, with dynamic race operationalization. The table illustrates how different Operationalizations of race impact self-identification and categories. For example, Always Asian and Ever Asian are depicted in a separate stratum because Ever White is statistically significantly different from Ever Asian. Ever White is depicted in a separate stratum because it is statistically significantly different from Always Asian.

### Table 8 Comparing racial hierarchy across outcomes, with dynamic race operationalization

| Source | Always Asian | Always Multiple | Always Hispanic | Always White | Ever Asian | Ever Hispanic | Always Black | Always Multiple | Ever Black | Always AIAN | Always American | Always Hispanic | Always Asian | Always Black | Ever White | Always Asian | Always Black | Ever Black | Always AIAN | Always American | Ever Hispanic | Always Hispanic | Always Black | Always AIAN |
|---------|--------------|-----------------|-----------------|-------------|------------|--------------|--------------|-----------------|------------|-------------|----------------|----------------|--------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|--------------|-------------|--------------|
| Interview classification | Always Asian | Always White | Always Hispanic | Always Black | Ever Black | Always AIAN | Always Black | Always AIAN | Ever White | Always Hispanic | Always Black | Ever AIAN | Always Black | Ever White | Always AIAN | Always Black | Ever AIAN | Always Black | Always American | Ever Hispanic | Always Hispanic | Always Black | Always American |
| Always Asian | Ever Asian | Always Asian | Ever Hispanic | Always Black | Ever Black | Ever AIAN | Ever White | Ever Black | Ever AIAN | Ever Hispanic | Always Black | Always AIAN | Always Black | Ever Asian | Ever AIAN | Always Black | Ever AIAN | Ever Black | Always American | Ever Hispanic | Always Hispanic | Always Black | Always American |
| Always Black | Always Hispanic | Always Black | Ever Black | Always White | Always Black | Ever Hispanic | Always AIAN | Ever Black | Always AIAN | Ever Hispanic | Always Black | Always AIAN | Always Black | Ever Asian | Ever AIAN | Always Black | Ever AIAN | Ever Black | Always American | Ever Hispanic | Always Hispanic | Always Black | Always American |
| Always Hispanic | Ever White | Always White | Ever Hispanic | Always Black | Ever Black | Ever AIAN | Ever White | Ever Black | Ever AIAN | Ever Hispanic | Always Black | Always AIAN | Always Black | Ever Asian | Ever AIAN | Always Black | Ever AIAN | Ever Black | Always American | Ever Hispanic | Always Hispanic | Always Black | Always American |
| Always Black | Always AIAN | Always Black | Ever Hispanic | Ever AIAN | Ever Hispanic | Ever Hispanic | Always Black | Ever AIAN | Ever Hispanic | Always Black | Always AIAN | Always Black | Ever Asian | Ever AIAN | Always Black | Ever AIAN | Ever Black | Always American | Ever Hispanic | Always Hispanic | Always Black | Always American |

In the table, we see a variety of outcomes and operationalizations, each depicting different aspects of the racial hierarchy. For instance, in the case of employment, Always Hispanic and Always Black are depicted in separate strata because of statistical significance differences. This table provides a comprehensive view of how different operationalizations of race affect various outcomes, highlighting the nuances and complexities inherent in the racial hierarchy.
is associated with positive or negative outcomes and when. Using dynamic self-identification (Table 8), we also see several outcomes for which the estimated number of strata extend beyond three (arrest, college completion and per capita household income). If the U.S. is moving toward a more variegated racial hierarchy, measuring fluidity between categories may reflect those dynamics better than a measure of skin color, or a concept of “honorary”-ness can do alone. In that sense, Bonilla-Silva’s prediction of increasing complexity could be correct, but more work should be done to link theories of racial hierarchy to measures that best test those theories.

Our findings further build on previous research by suggesting that, in the contemporary United States, White Americans do not always have the best socioeconomic outcomes and Black Americans do not always have the worst socioeconomic outcomes. In particular, our results point to American Indian respondents as most consistently at the bottom of the racial hierarchy in terms of observed outcomes. The finding that American Indian respondents have some of the worst socioeconomic outcomes in the United States echoes some previous research (e.g., Bailey et al., 2014), but this observation has generally not been incorporated into broader debates about the U.S. racial hierarchy, perhaps because so few studies focus on relative outcomes for Indigenous people (but see Huysers et al., 2017, 2018). Responses to the COVID-19 pandemic further illustrate how Indigenous people are often situated at the periphery of discussions about racial inequality. National news coverage was slow to acknowledge racial disparities in both infection and mortality rates and continues to overlook the outsized toll on American Indian tribal communities (Akee & Reber, 2021). Their relatively small population size and the geographic remoteness of many reservations serves to both literally and metaphorically remove Indigenous peoples from the sight of most Americans. This lack of attention is compounded by “glaring data inequities” (Rodriguez-Lonebear, 2016). To understand the racial hierarchy and redress racial inequalities, scholars must do better to study the U.S. population in its entirety—for racial categories large and small—which means paying greater attention to socioeconomic outcomes for Indigenous peoples.

Rather than universal Black disadvantage, as often portrayed in the U.S., our findings for Black respondents show that their relative position in the racial hierarchy fluctuates based on the outcome being measured. In terms of college completion and self-rated health, Black respondents have outcomes that place them in the middle of the U.S. racial hierarchy, not the bottom as many people might expect. That we see some positive outcomes for Black respondents should not come as a complete surprise, however, and underscores the call by Pattillo (2021) for a new approach to studying U.S. racial inequality: “Black Advantage Vision.” As Pattillo argues, scholars should aim to disrupt the cycle of stigma that informs and is informed by research premised on Black “deficits.” While Black Americans continue to suffer the cumulative effects of historical oppression and structural racism, as our findings for Black respondents suggest, there is also room to rethink our understanding of when Black Americans are advantaged or disadvantaged (and why) to better understand contemporary racial inequality in the United States.

Asian Americans also exhibit a combination of outcomes and experiences that challenge both the idea of a consistent, singular racial hierarchy and expectations of White dominance. Our results for Asian American respondents align with previous research that highlights particular contexts in which Asian Americans have better socioeconomic outcomes than White Americans (e.g., Jiménez & Horowitz, 2013). Indeed, observed Asian American gains in educational attainment, employment, and income have often been framed as a debate about whether Asian Americans are “becoming White” (Kim, 2007; Lee & Bean, 2007; Zhou, 2004), though the “racial triangulation” of Asian Americans is also more complex than the terms of such a debate imply (Kim, 1999; Xu & Lee, 2013). This is further demonstrated in our data by the much lower average self-rated health of Asian American respondents and, more broadly, by the uptick in anti-Asian hate crimes during the COVID-19 pandemic including the Atlanta area spa shooting in which six of the eight victims were Asian American women (Fausset & Vigdor, 2021). Asian Americans are better off, on average, than other racial minorities on many socioeconomic indicators, but allusions to being welcomed into whiteness must account for the fact that these socioeconomic gains are neither evenly distributed among Asian-origin Americans nor consistently protective against discrimination and racial violence.

As future scholars build on this analysis, they will have to contend with a challenge that also limited our approach: the racial categories we use gloss over internal heterogeneity by national origin, immigrant generation, immigration status, and other potentially stratifying factors. For example, not only does previous research indicate that high observed educational attainment for Asian Americans does not necessarily hold across all national-origin groups (Vartanian et al., 2007), there also is evidence of a sharp divide in socioeconomic outcomes between foreign-born and U.S.-born Black Americans with significantly higher levels of educational attainment among Black immigrants (Hamilton, 2014, 2020; Waters et al., 2014). Other research highlights that conclusions about the racial hierarchy can also vary according to immigrant generation (Kim, 2015), deportation status (Rugh, 2020), and region (Kwon & Kposowa, 2017). As scholars continue to study the racial hierarchy, it will be important to heed calls to attend to these dimensions of
intra-group heterogeneity (Ford & Harawa, 2010; Jiménez et al., 2015) and to be mindful about how methodological decisions affect the research process and, ultimately, knowledge production.

Although we underscore the importance of attending to intra-group variation, we also recognize that not all data are amenable to such analysis. Studying both within and between group differences requires data that maintain categories large enough for statistical analysis; in this study we faced that exact challenge. For example, among our static race models, there were several instances when there were too few American Indian respondents to find statistically significant results even with observed disparities that would be substantively meaningful. One such instance can be found in Table 6 where we see a Black/nonblack hierarchy in employment with American Indian respondents in the middle of the nonblack stratum. Although the difference in estimates suggest that Black Americans and American Indians have meaningfully different employment outcomes (log odds = −0.895 and −0.103, respectively), the difference between them is not statistically significant in this sample (see Table B8 in Online Appendix B). The challenge of maintaining sufficient sample size across categories is exacerbated when we turn to our dynamic race measures, where the “ever” category was often too small to make claims of statistical significance (see Table A8 in Online Appendix A). Because of these smaller cell sizes, we expect our results underestimate stratification and are, thus, conservative in their predictions about the relative prevalence of binary hierarchies in these data. We call on future scholars to both improve on our work and confirm the patterns we find using datasets that offer both larger sample sizes and allow for racial fluidity, such as linked data across censuses or other administrative records.

Conclusion

We hope our analysis has underscored the importance of carefully attending to issues of operationalization and measurement in studies of racial inequality. In practice, this means not taking the relevant racial categories for granted and fully utilizing the measures of race available in one’s data to test whether and how various racial categorization schemes affect the results. It also means being explicit about why a particular outcome was chosen and cautious about generalizing conclusions beyond that case. Recognizing that such decisions are not simply semantic could go a long way toward improving our understanding of the nature and extent of racial stratification in the United States.

Today’s official racial categories are pan-ethnic political constructions forged in part to draw attention to historical disparities (e.g., Mora, 2014; Okamoto, 2014). Most research simply assumes these standard categorizations are the most relevant for understanding contemporary inequality without testing them against alternative approaches. Our operationalizations are also limited by the current vision of race in the United States, even in their dynamic forms. For example, at present, the official U.S. racial classification scheme does not mandate a separate category for Americans of Middle Eastern and North African origins; thus, whether and when their outcomes might affect conclusions about the degree of racial stratification or the rank ordering of categories is beyond the scope of our analysis. It is also worth recalling that not long before Du Bois made his prediction about the color line the U.S. census was still enumerating “mulattos,” “quadroons,” and “octoroons” separately from “Black” Americans and classifying Indigenous people by blood quantum (Snipp, 2003). Applying those classification schemes to today’s socioeconomic outcomes would likely produce very different results than the ones we present here. Thus, our call for a more explicit accounting of methodological decisions in research on racial inequality also extends to the initial decision of who and how to count in the first place.

We see careful attention to methodological rigor as one way, among many, that scholars can honor the legacy of Du Bois in social science research (see Itzigsohn & Brown, 2020). As one of the first quantitative sociologists, Du Bois was committed to thorough analysis guided by detailed empirical data collection (see, e.g., Du Bois, [1899] 2007; Morris, 2017). Although our statistical tools and techniques may have changed, our overall aims and approach should remain the same: to produce meticulous research that can support and inform efforts at social change. The foundation for any such effort is understanding the full scope and nature of racial inequality in the United States. To assume that contemporary disparities in one context or with respect to one outcome necessarily translate across contexts or outcomes does not square with our findings presented here. Who is considered a member of a given racial category can also change both across contexts and over time. As we continue to grapple with the evolution of the “color line” in the United States, both our theories and our methodology should evolve as well.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12552-021-09351-2.

Acknowledgements This research was previously presented at the Columbia University Race, Ethnicity and Migration Workshop and the Stanford Migration, Ethnicity, Race, and Nation (MERN) workshop. The authors wish to thank the workshop participants for their feedback as well as Jeremy Freese for comments and suggestions on the modeling approach. Add Health is directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01 AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Waves I-V data are from the Add Health Program Project, grant P01 HD31921 (Harris) from Eunice Kennedy Shriver National Institute of Child Health.
and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Funding  N/A.

Data Availability  The data used for this project are Restricted Use Add Health data accessed via the Stanford Institute for Research in the Social Sciences (IRIiSS). The Public Use data are available through Add Health (https://addhealth.cpc.unc.edu/) or ICPSR (https://www.icpsr.umich.edu/). For information on how to obtain access to the Restricted Use Add Health data visit https://addhealth.cpc.unc.edu/.

Code Availability  Upon request.

Conflict of interest  The authors declared that they have no conflict of interest.

References
Agadjanian, A., & Lacy, D. (2021). Changing votes, changing identities? Racial fluidity and vote switching in the 2012–2016 US Presidential Elections. Public Opinion Quarterly. https://doi.org/10.1093/poq/nfab045
Akee, R., Jones, M. R., & Porter, S. R. (2019). Race matters: Income shares, income inequality, and income mobility for all U.S. races. Demography, 56(3), 999–1021.
Akee, R., & Reber, S. (2021, February 18). American Indians and Alaska natives are dying of COVID-19 at shocking rates. Brookings Institution. https://www.brookings.edu/research/american-indians-and-alaska-natives-are-dying-of-covid-19-at-shocking-rates/
Bailey, S. R., Saperstein, A., & Penner, A. M. (2014). Race, color, and income inequality across the Americas. Demographic Research, 31(24), 735–756.
Blackmon, D. A. (2009). Slavery by another name: The re-enslavement of Black Americans from the Civil War to World War II. Anchor Books.
Bollen, K. A., Biemer, P. P., Karr, A. F., Tueller, S., & Berzofsky, M. E. (2016). Are survey weights needed? A review of diagnostic tests in regression analysis. Annual Review of Statistics and Its Application, 3, 375–392.
Bonilla-Silva, E. (2002). We are all Americans!: The Latin Americanization of racial stratification in the USA. Race & Society, 5(1), 3–16.
Bonilla-Silva, E. (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, E., & Dietrich, D. R. (2008). The Latin Americanization of racial stratification in the US. In R. E. Hall (Ed.), Racism in the 21st century: An empirical analysis of skin color (pp. 151–170). Springer.
Branigan, A. R., Freese, J., Patir, A., McDade, T. W., Liu, K., & Kief, C. I. (2013). Skin color, sex, and educational attainment in the Post-Civil Rights Era. Social Science Research, 42(6), 1659–1674.
Bratter, J. L., & Gorman, B. K. (2011). Is discrimination an equal opportunity risk? Racial experiences, socioeconomic status, and health status among Black and White adults. Journal of Health and Social Behavior, 52(3), 365–382.
Brubaker, R. (2004). Ethnicity without groups. Harvard University Press.
Burch, T. (2015). Skin color and the criminal justice system: Beyond Black-White disparities in sentencing. Journal of Empirical Legal Studies, 12(3), 395–420.
Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods & Research, 33(2), 261–304.
Charles, K. K., & Hurst, E. (2006). The transition to homeownership and the Black-White wealth gap. The Review of Economics and Statistics, 84(2), 281–297.
Couch, K. A., & Fairlie, R. (2010). Last hired, first fired? Black-White unemployment and the business cycle. Demography, 47(1), 227–247.
Davenport, L. (2020). The fluidity of racial classifications. Annual Review of Political Science, 23, 221–240.
Doyle, J. M., & Kao, G. (2007). Are racial identities of multiracials stable? Changing self-identification among single and multiple race individuals. Social Psychology Quarterly, 70(4), 405–423.
Du Bois, W. E. B. ([1899] 2007). The Philadelphia Negro: A social study. Oxford University Press.
Du Bois, W. E. B. ([1903] 1994). The souls of Black folk. Dover.
Emeka, A. (2009). Race and unemployment amidst the new diversity: More evidence of a Black/non-black divide. Race and Social Problems, 1(3), 157–170.
Fausset, R., & Vigdor, N. (2021, March 2021). 8 people killed in Atlanta-area massage parlor shootings. The New York Times. https://www.nytimes.com/2021/03/16/us/atlanta-shootings-massage-parlor.html.
Feliciano, C. (2016). Shades of race: How phenotype and observer characteristics shape racial classification. American Behavioral Scientist, 60(4), 390–419.
Fischer, C. S., & Hout, M. (2006). Century of difference: How America changed in the last one hundred years. Russell Sage Foundation.
Ford, C. L., & Harawa, N. T. (2010). A new conceptualization of ethnicity for social epidemiologic and health equity research. Social Science & Medicine, 71(2), 251–258.
Freeman, L. (2005). Black homeownership: The role of temporal changes and residential segregation at the end of the 20th century. Social Science Quarterly, 86(2), 403–426.
Gans, H. J. (1999). The possibility of a new racial hierarchy in the twenty-first-century United States. In M. Lamont (Ed.), The cultural territories of race: Black and White boundaries (pp. 371–390). University of Chicago Press and Russell Sage Foundation.
Gans, H. J. (2012). “Whitening” and the changing American racial hierarchy. Du Bois Review: Social Science Research on Race, 9(2), 267–279.
Garcia, J. A., Sanchez, G. R., Sanchez-Youngman, S., Vargas, E. D., & Ybarra, V. D. (2015). Race as lived experience: The impact of multi-dimensional measures of race/ethnicity on the self-reported health status of hispanics. Du Bois Review: Social Science Research on Race, 12(2), 349–373.
Hamilton, T. (2014). Selection, language heritage, and the earnings trajectories of Black immigrants in the United States. Demography, 51(3), 975–1002.
Hamilton, T. (2020). Black immigrants and the changing portrait of Black America. Annual Review of Sociology, 46, 295–313.
Hannon, L., DeFina, R., & Bruch, S. (2013). The relationship between skin tone and school suspension for African Americans. Race and Social Problems, 5(4), 281–295.
Harris, K. M. (2009). The national longitudinal study of adolescent to adult health (Add Health). Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007–2009 [machine-readable data file and
Stewart, Q. T. (2008). Chasing the race effect: An analysis of traditional quantitative research on race. In C. A. Gallagher (Ed.), *Racism in post-race America: New Theories, New directions* (pp. 285–304). Chapel Hill: Social Forces.

Stokes-Brown, A. K. (2012). America’s shifting color line? Reexamining determinants of Hispanic racial self-identification. *Social Science Quarterly, 93*(2), 309–332.

Sue, C. (2009). An assessment of the Latin Americanization thesis. *Ethnic and Racial Studies, 32*(6), 1058–1070.

Vargas, N., & Kingsbury, J. (2016). Racial identity contestation: Mapping and measuring racial boundaries. *Sociology Compass, 10*(8), 718–729.

Vartanian, T. P., Karen, D., Buck, P. W., & Cadge, W. (2007). Early factors leading to college graduation for Asians and non-Asians in the United States. *The Sociological Quarterly, 48*(2), 165–197.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307–333.

Warren, J. W., & Twine, F. W. (1997). White Americans, the new minority? Non-blacks and the ever-expanding boundaries of whiteness. *Journal of Black Studies, 28*(2), 200–218.

Waters, M. C., Kasinitz, P., & Asad, A. L. (2014). Immigrants and African Americans. *Annual Review of Sociology, 40*, 369–390.

White, K. M. (2015). The salience of skin tone: Effects on the exercise of police enforcement authority. *Ethnic and Racial Studies, 38*(6), 993–1010.

Xu, J., & Lee, J. C. (2013). The marginalized “model” minority: An empirical examination of the racial triangulation of Asian Americans. *Social Forces, 91*(4), 1363–1397.

Zhou, M. (2004). Are Asian Americans becoming “White?” *Contexts, 3*(1), 29–37.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.