When (In)Consistency Matters: Racial Identification and Specification

Jiannbin Lee Shiao

Abstract
Sociologists who rely on survey research have begun exploring the implications of racial inconsistency for not only multiracial identification but also for other social outcomes. Using the National Longitudinal Study of Adolescent to Adult Health, the author explores the consequences of different specifications of race for modeling three outcomes: educational attainment, self-rated health, and interracial relationship history. The author uses the model-fit criteria of Akaike information criterion and Bayesian information criterion to compare five methods of classifying multiple, changed, and other-race respondents and of operationalizing heterogeneity among them. The results show substantial variation across outcomes in the specifications preferred and indications of a qualitative difference in the meaning of race for social contact versus for social stratification. The article concludes with recommendations for how quantitative researchers should measure race even when the categories primarily serve as covariates.

Keywords
race, identity, measurement, education, health, interracial relationships

The opening passages illustrate the complexity of racial/ethnic identity. In the first passage, an interview respondent in a study of biracial identity describes how he has negotiated the consequential assumptions police officers have made about his race/ethnicity. In the second, 60 Minutes reporter Steve Kroft questions then senator Barack Obama about his apparent choice of racial identification as inconsistent with his parental ancestry. In the third, U.S. presidential candidate Obama uses the racial diversity within his extended family as a metaphor for national unity.

Since the 1990s, sociologists have recognized the prevalence of racial multiplicity, or the identification of individuals, either by themselves or by others, using multiple racial categories instead of a single category (Rockquemore and Brunsma 2008; Doyle and Kao 2007; Harris and Sim 2002). More recently, researchers have focused on a critical correlate of multiplicity, the prevalence of inconsistency in racial identification across situations, regardless of whether an individual ever identifies himself or herself with multiple categories (Penner and Saperstein 2008). In particular, sociologists who rely on survey research have begun exploring the implications of racial inconsistency for not only multiracial identification but also for other social outcomes (Campbell, Bratter, and Roth 2016; Saperstein, Kizer, and Penner 2016). How should quantitative sociologists measure race in light of its relative multiplicity and inconsistency?

To address this question, I examine the scope and consequences of racial multiplicity and inconsistency in the National...
Longitudinal Study of Adolescent to Adult Health (Add Health). At its outset, Add Health was the largest social science project ever funded by the National Institutes of Health and has been used in studies of both multiracial identities and racial identity change. Indeed, it was one of the first data sets (1) to collect data on racial self-classification that allowed respondents to mark multiple responses rather than only one response, (2) to collect multiple types of racial information, and (3) to continue collecting this scope of racial information in subsequent waves with the same respondents.2

Beyond research on multiplicity and inconsistency, however, most Add Health users do not take advantage of its trailblazing collection of racial information, with unknown consequences for their findings. In this article, I first summarize recent sociological advances in conceptualizing racial multiplicity and inconsistency that call into question the measurement of racial identity as singular and stable. Second, focusing on Add Health’s higher quality information, namely, its data on racial self-classification, I explore the scope of multiplicity and inconsistency in Add Health’s racial information. Third, using model-fit criteria, I evaluate the consequences of different specifications of race for modeling three outcomes: educational attainment, self-rated health, and interracial relationship history.

Specifically, I develop two new methods for classifying multiple-race, changed-race, and other-race (MCO) respondents and compare them with three existing methods that ignore considerations of consistency. I also expand this comparison with additional measures for capturing heterogeneity among MCO respondents. In brief, I find substantial variation across outcomes in the specifications preferred and indications of a qualitative difference in the meaning of race for social contact versus for social stratification. I conclude with recommendations for sociologists of race, the limitations of self-classification data, and the scope of future data collection needed for interpretations that are more precise.

Add Health is a historically important data set for studying racial multiplicity and inconsistency. When it began collecting data in 1994, the multiracial movement had recently started, as indicated by the founding of the National Association of MultiEthnic Americans in 1988, Project RACE (Reclassify All Children Equally) in 1991, and the Hapa Issues Forum in 1992 (King and DaCosta 1996; King-O’Riaain 2006; Spencer 1999). An early, pivotal event in the movement was the appointment of National Association of MultiEthnic Americans president Ramona Douglas in 1995 to the 2000 U.S. Census Advisory Committee. This body ultimately recommended that the Census Bureau cease its practice of requiring Americans to register a single race in response to its racial self-identification question and instead allow Americans to “check one or more box” and “mark all that apply.” In brief, six years before the Census Bureau could implement this recommendation, Add Health had already begun permitting multiple responses on its race question.

Since then, however, Add Health has not continued to blaze new trails in collecting racial information, at least up through its Wave IV panel of 2007–2008. In particular, like the census, Add Health has excluded a Hispanic/Latinx response from its race question and continues to require a single response on its separate Hispanic/Latinx question, making it difficult to distinguish Latinx-descent multiracial persons from other Latinxs. In anticipation of Wave V’s full release and the planned transfer of Add Health to the National Institute on Aging, this is an opportune time for using Add Health to explore how quantitative sociologists should measure race.

**Conceptualizing Racial Multiplicity and Inconsistency**

The multiracial movement was accompanied by the rise of mixed-race studies, an academic field examining the experiences of multiracial individuals, anchored in the now classic anthologies edited by psychologist Maria Root (1992, 1996). Published before the availability of quantitative data on multiplicity and inconsistency, the early research in mixed-race studies focused almost exclusively on qualitative data collected by its pioneering investigators. To this day, this early work provides rich illustrations of the limits of simply allowing multiple responses on racial self-classification questions. A revealing example is the extended excerpt for the first passage introducing this article. In this study, Williams examines multiracial individuals’ experiences with the “What are you?” question, which is now commonly recognized as a racial faux pas that both positions the target as an exotic other and indicates the limitations of the asker’s “compass for navigating race relations” (Omi and Winant 1986:62). Below I have emphasized in boldface type phrases that illustrate distinct dimension of racial information:

> Yeah, I get asked What Are You? a lot. . . . I think society sees me and my friends as men of color. It can be good and it can be bad. I also see myself as a man of color. That’s the great part. I have a sense of belonging. It gives me a kind of masculinity that both Asians and whites don’t have. But the downside is like the other day, I was speeding on the freeway and—I get stopped by the cops all the time . . . the cop asked me for my ID. My name isn’t Spanish . . . it’s a common black or white last name so it’s kind of ambiguous. He was such an asshole. He asked if I was a gang member or if I had drugs in my truck. The cop starts calling into his headquarters, “One Mexican, 25 year old, male . . . ” I was getting pissed, but I knew I had to shift my speech patterns to standard
English and to reveal my true race because it’s stereotyped more favorably. . . . I know that sucks, but I was arrested before because that cop thought I was black. . . . After the cop was done, I said, “Excuse me sir,” real politely, “I’m not Mexican. My father is white and my mother is Japanese.” From that moment on, the cop changed. I don’t know if he was embarrassed because he misidentified me or because as a Japanese/white man—could he not see me as a drug dealer or gang member? My speech can also be considered nonethnic, which contrasts to my Latino-looking appearance so that’s one edge I have. I really have to manipulate myself and project different selves in order to survive in a society that treats brown and black men like shit. (Williams 1996:207)

In the present, theorists of race have developed a richer vocabulary for the complexities of racial information, which Wendy Roth (2016) has comprehensively assembled in her typology of race dimensions: racial identity, racial self-classification, observed race, reflected race, phenotype, and racial ancestry. Quantitative researchers, such as those who use Add Health, often have access to information about racial self-classification, that is, the race one might choose from a menu of preestablished options, such as a respondent’s choice of a racial category in response to a survey question. In research, racial self-classifications are most appropriate when used as covariates for modeling other population processes, that is, as indicators of exogenous processes for which the researcher needs statistical controls in models estimating the association between other variables of interest.

However, researchers also use racial self-classification as a proxy for other dimensions of race and, as a result, advance potentially idiosyncratic interpretations of its coefficients. These other dimensions include racial identity (i.e., one’s subjective self-identification), which is theorized to influence socially and culturally embedded decisions such as voting, residential decision making, social network formation, and indeed racial self-classification. Alternatively, researchers may use racial self-classification as a proxy for observed race or the race others believe someone to be. In contrast with racial identity, observed race is theorized to influence distinct outcomes such as racial profiling and discrimination, whether in public settings when based on appearance or in more organized settings such as workplaces and schools when based on interaction. In addition, researchers may use racial self-classification as a proxy for reflected race (i.e., the race people believe others assume them to be), which is theorized to influence both racial identity and perceived discrimination, while also being influenced by observed race.

Researchers may also share the popular assumption that an individual’s racial self-classification should not deviate from his or her known ancestry on the basis of family history or even his or her observed race on the basis of genomic analysis. Similarly, another popular assumption that may influence researchers is that racial self-classification should follow phenotype, whether based on skin color or other features, and indeed this very assumption is what produces the “What are you?” situation. In actuality, both known ancestry and phenotype are more precisely important inputs to racial identity, which in turn influences racial self-classification. In brief, when researchers use racial self-classification as more than a statistical control, they may be tempted to interpret its coefficients as proxies for a variety of distinct social processes.

Roth (2016) further clarified her dimensions of race by distinguishing among forms of possible inconsistency between pieces of racial information, principally between fluidity, which “refers to fluctuation in [a single] dimension as opposed to inconsistency across different dimensions” (p. 20), which she terms multidimensionality. She also distinguishes between contextual fluidity, or “changes across contexts within a fairly limited period of time,” and temporal fluidity, or “changes over time within the same context” (p. 20). An example of contextual fluidity in the qualitative interview excerpt is the implied contrast between the respondent’s race as observed (1) by the police officer versus (2) by the respondent’s friends or, more precisely, between the situation of the traffic stop and the situation of a social setting, which conceivably could involve the same person observing the respondent through the distinct schema appropriate for each context (i.e., as an officer of the law vs. as a friend of a friend). In contrast, an example of temporal fluidity is the hypothetical development of the respondent’s racial identity in adulthood as a man of color from an earlier racial identity in adolescence as Asian American, Japanese, or even white. That said, most of the excerpt concerns types of multidimensionality: (1) the equation of the respondent’s racial identity with his reflected race, which suggests the influence of the latter on the former; (2) the contrast between his observed race as assigned by police officers and his self-classification to one police officer; and (3) the contrast between his known ancestry on the basis of family history and his phenotype or reflected phenotype as “Latino looking.”

Quantitative researchers have begun exploring the significance of these distinctions for the study of various social outcomes. On one hand, Liebler, Bhaskar, and Porter (2016) found significant demographic differences between (1) Americans who only occasionally self-classify as American Indian and (2) those who consistently self-classify as American

---

1 I relabel Roth’s racial ancestry and its subdimensions of known racial ancestry and known genetic ancestry as, respectively, known ancestry, known ancestry on the basis of family history, and known ancestry on the basis of genomic ancestry testing. The latter is distinct from observed race on the basis of genomic ancestry testing, which becomes “known” only when the test result is shared with the test subject.
Indian while also occasionally self-classifying by additional races. In other words, relative consistency indicates a substantive divide among multiracial American Indians between those who do and do not report tribal affiliations, live in American Indian areas, and attain higher levels of education. In contrast, Kramer, Burke, and Charles (2015) found that inconsistency does not significantly distinguish multiracial persons in models predicting depression, school belonging, school engagement, and grade point average. Instead they argued that observed race is a more important indicator of heterogeneity among multiracial individuals, on the basis of the better outcomes of those who ever self-classify as white. In brief, does inconsistency actually indicate an experience of social marginality, or does it mask one’s relative exposure to racial status hierarchy? More generally, how does the relationship between racial multiplicity and inconsistency matter across the outcomes of interest to quantitative researchers?

**Mapping Racial Multiplicity and Inconsistency in Add Health**

Add Health is a uniquely comprehensive data set for exploring racial multiplicity and inconsistency, but it also omits important dimensions of racial information. Like most quantitative data sets, there is no open-ended self-identification question to capture racial identity, much less a transcript of a discussion about respondent’s subjective identification and its salience in their lives. There is also no item for reflected race to capture how respondents perceive they are observed by most others, much less by particular groups. However, Add Health has data on the other four dimensions in Roth’s (2016) typology, and where Add Health excels is its data on racial self-classification.4

In all panels except Wave IV, Add Health first asks respondents whether they are of “Hispanic or Spanish/Latinx” origin, accepting only a single response, before then asking respondents for their race, that is, their self-classification, or known ancestry,8 or (2) any race beyond the scope of this article, for respondents who respond as Native American, Asian American, black, or white, though they could have been coded as other race in the Wave IV self-classification item. Furthermore, in no panel did Add Health instruct its interviewers to code the Hispanic/Latinx appearance of the respondent, which means that “Latinx-looking” respondents were instead coded as Native American, Asian American, black, or white, though they could have been coded as other race in the Wave I in-home panels. Add Health also did not ask its parent respondents for (1) the “best race” of their current spouse or partner is of Hispanic/Latinx origin and what his or her race is, explicitly allowing multiple responses, as for the self-classification items. Unfortunately, in almost every panel, the observed race item immediately follows the self-classification items,7 which means that its information is actually observed race on the basis of the very particular form of interaction, a request for self-classification, rather than observed race on the basis of appearance alone (cf. Kramer et al. 2015). In addition, there is the absence of the “other race” category from the Wave III race questions, as well as the phrasing of the Hispanic/Latinx question and the “best race” question in terms of, respectively, “origin” and “background,” both of which may evoke known ancestry instead of self-classification alone.

In addition to self-classification, Add Health includes multiple pieces of information on observed race. In all panels except the Wave I in-school panel, Add Health instructed interviewers to code the race of the respondent from observation alone into a single category, and in the parent panel, it asks the parent respondent whether his or her current spouse or partner is of Hispanic/Latinx origin and what his or her race is, explicitly allowing multiple responses, as for the self-classification items. Unfortunately, in almost every panel, the observed race item immediately follows the self-classification items,7 which means that its information is actually observed race on the basis of the very particular form of interaction, a request for self-classification, rather than observed race on the basis of appearance alone (cf. Kramer et al. 2015). In addition, there is the absence of the “other race” category from the Wave III observed-race item, consistent with its absence from the Wave III self-classification item. Furthermore, in no panel did Add Health instruct its interviewers to code the Hispanic/Latinx appearance of the respondent, which means that “Latinx-looking” respondents were instead coded as Native American, Asian American, black, or white, though they could have been coded as other race in the Wave I in-home panels. Add Health also did not ask its parent respondents for (1) the “best race” of their current spouses or partners, much less clarify whether their answers should be based on their partners’ appearance or the respondents’ knowledge of their partners’ racial identity, self-classification, or known ancestry,8 or (2) any race information for nonresidential biological parents.

---

4Table A1 in the Appendix shows the measures of race available in Add Health for each dimension in each panel.

5From this point forward, I use only the gender-neutral Latinx instead of Latino to refer to Latinas and Latinos without distinction.

6For the sake of brevity, I refer in the remainder of this article to the category “Asian American or Pacific Islander” as “Asian American” and to the category “American Indian or Alaskan Native” as “Native American.”

7In contrast, the absence of self-classification items in the Wave IV panel means that its interviewer-coded information is based more on appearance than are its counterparts in other panels.

8Add Health includes additional self-classification questions, beyond the scope of this article, for respondents who respond as Hispanic/Latinx on that item or as Asian American on the race questions in all panels except Wave IV.
Add Health also asks respondents in Wave III for their family origins in terms of “as many as four countries, groups, or geographic areas. What are your family ancestries?”

Given the unknown generational scope of these responses, however, a better indicator for known ancestry is arguably the parent self-classification and the parent-partner observed-race items (Campbell 2000). In sum, Add Health includes racial information for four of Roth’s (2016) six dimensions, with the higher quality information coming from its multiple observations of racial self-classification and its collection of parent self-classification and parent-partner observed race, from which known ancestry can be inferred, all of which permitted multiple responses.

Despite the available range of racial information, most Add Health researchers have largely chosen to follow the lead of other social scientists in classifying multiracial respondents by only using information from a single panel. The three major methods for allocating multiracial persons are

1. the official Add Health specification that distributes them to the single-race categories for nonwhites, creating a uniquely monoracial white category;
2. the specification that is the basis for Add Health’s race-contingent network variables and that places multiracial persons into the other-race category, along with Native Americans; and
3. a specification that uses the “best race” question, variations of which are available in other data sets (Bratter et al. 2014) to distribute them to the single-race category of their choice.

In this article, I compare these existing methods with two new methods that use racial consistency to classify respondents associated with MCO information.

**Data and Methods**

Add Health includes a nationally representative sample of schools, from which it constructed a sampling frame of more than 100,000 students (Harris 2009). In the first wave of data collection in 1994 and 1995, the study administered an in-school survey that students completed on their own, with an original sample of 90,118 students when they were 11 to 20 years of age, followed by in-home interviews in 1995 with a subsample of 20,745 students and 17,700 parents. The study returned to the in-home sample of students for three additional waves of interviews in 1996 (Wave II), 2001 and 2002 (Wave III), and 2007 and 2008 (Wave IV) when subjects reached 24 to 32 years of age. In my analysis, I use Add Health’s restricted-use data from the Wave I in-school, Wave I in-home, Wave III in-home, and Wave IV in-home panels, as well as the Wave I in-home parent panel.\(^9\)

I matched Add Health respondents across these five panels into a combined sample with 91,040 unique respondents. The percentage of respondents who gave multiple responses to the racial self-classification question, excluding those identified as Hispanic/Latinx, ranges across panels from 3.56 percent (Wave III in-home) to 6.01 percent (Wave I in-school). In aggregate, 6.68 percent of respondents gave multiple responses in at least one panel, among the 89,463 respondents who had nonmissing racial information in any of the four panels with information on racial self-classification.

To measure the level of inconsistency, I construct an index of consistency among the valid pairs of racial information available for any given respondent. The number of possible pairs for comparison is \(\frac{n(n-1)}{2}\), where \(n\) is the number of nonmissing observations (e.g., four valid observations mean six pairs for comparison). I calculate consistency for each respondent \(i\) as Consistency\(_i\) = \(\frac{\sum \text{pairs with consistent information}}{\sum \text{pairs with non−missing information}}\), and I reverse-code inconsistency as Degree of Inconsistency\(_i\) = 1 − Consistency\(_i\), or the proportion of valid pairs with inconsistent racial information for respondent \(i\). For example, if a respondent’s race is completely consistent across panels, his or her consistency score is 1 and inconsistency score is 0.\(^{10}\) Among the 19,587 Add Health respondents who were associated with at least one valid pair of racial information, the mean score is 0.107, but among the 2,916 respondents with any inconsistent information, the mean score is 0.716. In combination with the 218 respondents who were consistently associated with multiple responses, there are 3,134 respondents, or 16.0 percent, who have multiple and/or inconsistent racial information, a percentage larger than the percentage of black respondents in two panels.

**Alternative Racial Specifications**

The official Add Health specification is available as a variable supplied with the Wave I in-home data. If respondents answer the Hispanic/Latinx origin question with

\(^9\)Add Health also requested and received consent for DNA samples in the Wave I in-home, Wave III in-home, and Wave IV in-home panels, but preconstructed indicators for observed race on the basis of genomic analysis are not currently available in the restricted-use data set.

\(^{10}\)I exclude the Wave II panel, whose race information was simply preloaded from the Wave I in-home panel.

\(^{11}\)Table A2 illustrates the inconsistency index with sample values for respondent-level information.
“yes,” they are classified as Latinx regardless of their answer to the race questions. Among the remaining respondents, if they mark “black or African American” in the race questions, they are classified as black, even if they mark another race category. This procedure is then repeated iteratively with the remaining respondents in the following order: Asian, Native American, other, and white, that is, a sequential one-drop specification that reduces the white subsample to those who only mark “white” in response to the race questions. I implement a different version of this specification by applying these procedures to the largest panel, the Wave I in-school data.

The Moody specification is based on procedures used by James Moody (personal communication, February 26, 2017) to construct race-contingent variables for Add Health’s network data in its Wave I in-school panel, such as the precalculated variable for the racial diversity of respondents’ friendship nominations. Again, if respondents answer the Hispanic/Latinx origin question with “yes,” they are classified as Latinx regardless of their answer to the race questions. Among the remaining respondents, they are classified as white, black, and Asian if they exclusively mark the corresponding race category (e.g., the white subsample is composed of those who answer “no” on Hispanic/Latinx origin and only mark “white” on the race question). All remaining non-Latinx respondents, including those who mark multiple race categories or Native American, are classified as “other race.” I construct the variable for this specification by replicating these procedures on the in-school data, which does not actually include this variable.

The “best race” specification uses the “best race” question to allocate multiple-race respondents who did not answer the Hispanic/Latinx origin question with “yes,” in which case they are classified as Latinx, regardless of their race question answers. Because this question does not appear in the in-school data, I construct the variable for this specification using the Wave I in-home data.

I expand on these major methods with the entirely consistent specification, which creates a separate category for MCO respondents and reserves the remaining race categories for respondents who consistently and exclusively marked that category across panels (i.e., they are entirely consistent single-race identifiers). It uses a two-stage procedure for combining the Hispanic/Latinx and race questions across multiple panels. First, respondents are classified within each panel as Asian American, black, Latinx, Native American, white, other race, and multiple races, by applying the customary one-drop rule for Hispanic/Latinx origin and placing the remaining non-Latinx, multiple-race respondents into a separate category. Second, respondents are reclassified to an MCO category if they are classified as multiple-race responders in any panel or if they change their responses between any panels. I construct the variable for this specification using the Wave I in-school data, the Wave I in-home respondent data, the Wave I in-home parent data, and the Wave III in-home data.

I also modify the entirely consistent specification into the mostly consistent specification, extending Liebler et al.’s (2016) analysis of mostly consistent Native Americans to all MCO respondents by reclassifying them to a single race category if they consistently marked a single category, while marking other categories inconsistently. For example, I reclassify a respondent from MCO to Native if he or she marked “Native American” and “white” in one panel and “Native American” in another.

I explore additional specifications by extending these five specifications with three measures of MCO heterogeneity: (1) my degree of inconsistency index for the proportion of valid pairs of racial information that had inconsistent information and also two indicators of particular racial identification combinations: whether an MCO respondent (2) never marks “white” (never-white) and (3) ever marks “black” (ever-black). The latter two extensions measure, respectively, the distinctiveness of entirely “interminority” multiracial persons (cf. Kramer et al 2015) and the persistence of the one-drop rule historically applied to African Americans (Gullickson and Morning 2011). I add each extension separately, for example, the entirely consistent specification plus degree inconsistency, entirely consistent plus never-white, and entirely consistent plus ever-black, adding altogether another 14 specifications.

**Outcome Variables**

To best capture the effects of racial multiplicity and inconsistency, I considered three outcomes: educational attainment, self-rated health, and interracial relationship history.

---

12 To incorporate the parent data, I first restricted the panel to parent respondents who were the biological parents of the adolescent respondents and who had current partners who were also biological parents of the same adolescent respondents, before applying the two-stage procedure to each parent. Next, I combined the parental information into a single variable for known ancestry with categories comparable with the self-classification categories: (1) biracial parentage or any parent associated with multiple responses, (2) both parents Asian American only, (3) both parents African American or black only, (4) any Hispanic/Latinx parent, (5) both parents Native American only, (6) both parents white only, and (7) both parents other race only, which was lastly combined with the biracial/multiracial parentage category for comparability with the MCO self-classifications.

13 To address the absence of the other-race category from the Wave III race questions, I treat the Wave III information as a missing observation if in any other panel, the respondent is associated with “other race” but not with Latinx.

14 I do not examine the mostly consistent specification with the degree inconsistency extension, as it is redundant with the entirely consistent specification with the same extension.
Researchers have long recognized education to be a critical component of contemporary racial stratification (Wilson 1980), an importance consequence of which are racial disparities in health (Bratter and Gorman 2011; Williams and Sternthal 2010), whereas interracial relationships are an important indicator of social contact, specifically interpersonal interactions across the social boundaries that divide individuals into groups (Shiao 2018). This range of outcomes permits me to observe potential variation in the consequences of multiplicity and inconsistency across distinct social processes.

To measure educational attainment after Wave I, I used Add Health’s question on highest degree completed in Wave IV, supplemented by the same question in Wave III if the Wave IV information was not available.15 I reduced the options to four ordinal categories: postgraduate degree, bachelor’s degree only, high school degree or GED only, and less than high school. To measure self-rated health after Wave I, I used Add Health’s questions in Wave III or, if available, Wave IV, and I used their original, five ordinal categories: excellent, very good, good, fair, and poor. Self-rated health has been found to be a persistent predictor of mortality even after non-subjective measures of health have been included in the same models, arguably because self-assessments summarize bodily sensations not measured by direct indicators, not expressed to health care professionals, and even outside of the rater’s own consciousness (Jylhä 2009).

To measure interracial relationship history, I combined Add Health’s inventory of post–Wave I “romantic relationships” in Wave III with its Wave IV inventory of marriages, cohabitation partners, unions resulting in pregnancy, current romantic partners, and partners since 2001 (approximately Wave III). I operationalize interracial relationship history as having had any relationship in either Wave III or Wave IV that crosses the white/nonwhite boundary. Thus I classify respondents as having had interracial relationships if (1) they are classified as white, and they identify any romantic partner as being non-white, or (2) they are classified as non-white, and they identify any romantic partners as white. For example, in models using the official Add Health specification, if a respondent is classified as Asian and also identifies a partner as white, he or she is classified as having had an interracial relationship; in contrast, the same respondent might be classified as not having had an interracial relationship in other models using a different specification (e.g., if the mostly consistent specification classifies the respondent as white and the respondent classifies all of his or her partners as white).

To estimate these outcomes, I use hierarchical logistic regression models (Guo and Zhao 2000) with two levels (respondents and schools) with random intercepts for each school. Specifically, I use the meologit procedure in Stata to estimate levels of educational attainment and self-rated health as ordered categorical dependent variables, and I use the meqrlogit procedure to estimate the likelihood of having an interracial relationship as a binary outcome.

### Analytic Strategy and Covariates

I included both shared and distinctive covariates for each outcome across four nested regression models. For all three outcomes, model 1 includes only the racial specification in question with whites as the reference group, whereas model 2 adds a shared set of covariates for each outcome to control for differences across groups in age, gender, natal family structure, socioeconomic status (parental education), acculturation, religion, and region (cf. Howell and Emerson 2017; Kramer et al. 2015).

Model 3 adds outcome-contingent covariates that sociologists of education, health, and interracial intimacy have found to be important predictors of racial disparities. For education, I add indicators of academic performance, school-level affluence, and racial segregation (Domina, Penner, and Penner 2017; Reardon and Owens 2014), and for health, I add respondents’ educational attainment and whether they were ever been stopped by police before the age of 18 (Phelan and Link 2015). For interracial intimacy, I add indicators of interracial friendships,16 group size, any interracial intimacy in adolescence,17 number of romantic relationships after adolescence, and any same-sex relationships (Qian 1997; Shiao 2018).

Model 4 adds interaction terms to control for intersectionality18 (1) between race and gender and (2) between race and a theoretically important predictor for each outcome. For education and health, the latter interaction is parental education × each racial category in the specification in question, whereas for interracial intimacy, it is interracial friendship × nonwhite category, with whites as the reference group. In each nested model, I code the variables that are contingent on the race of the respondent, using the racial specification in question (e.g., when estimating interracial relationship history, the variable for group size is the proportion of same-race students at school, which may vary for the same respondent depending on their

---

15Using Wave III adds 744 respondents to the analytic sample for the education models, including some who could not have finished college by that time. In separate analysis using only Wave IV educational outcomes (not shown), I confirmed that the results remained the same.
Table 1. Variables in Nested Models for Educational Attainment, Self-Rated Health, and Interracial Relationship History.

| Variables in Models | Education                     | Health                                  | Interracial                                      |
|--------------------|-------------------------------|-----------------------------------------|--------------------------------------------------|
| Race specifications| Official Add Health procedure, applied to in-school responses of Wave I | Moody procedure, applied to in-school responses of Wave I | Use of “best race” item to assign multiple-category in-home responses of Wave I |
|                    | Entirely consistent responses across four panels with separate category for MCO | Mostly consistent responses across four panels with separate category for remaining MCOs | |
| MCO extensions to five specifications | Degree inconsistency | Never-white MCO | Ever-black MCO |
| Dependent variables | Educational attainment after Wave I: postgraduate degree, bachelor’s degree, HS, less than HS | Self-rated health after Wave I: excellent, very good, good, fair, poor | Any interracial relationship that crosses white/nonwhite boundary after Wave I (yes, no) |
| Independent variables | Race specification (reference: white) | MCO extension | Age, female, married parents |
|                    | Parental educational attainment (highest) | Non-English dominant language at home | Any religious affiliation, region |
|                    | Cumulative GPA | Median family income at school | Number of romantic relationships after Wave I |
|                    | Percentage black students at school | Educational attainment after Wave I (highest) | Any same-sex relationship |
|                    | Race × female | Race × parent education | Race × nonwhite × network racial heterogeneity | Race × female |
| Shared covariates (models 2–4) | Parental educational attainment (highest) | Non-English dominant language at home | Network racial heterogeneity (in adolescence) |
|                      | Any religious affiliation, region | Any religious affiliation, region | Proportion same-race students at school |
| Outcome-contingent covariates (models 3 and 4) | Cumulative GPA | Educational attainment after Wave I (highest) | Number of romantic relationships after Wave I |
|                      | Percentage black students at school | Ever stopped by police before age 18 | Any same-sex relationship |
| Interaction covariates (model 4) | Race × female | Race × parent education | Race × female |
|                      | Race × parent education | Nonwhite × network racial heterogeneity |

Note: Add Health = National Longitudinal Study of Adolescent to Adult Health; GPA = grade point average; HS = high school; MCO = multiple race, changed race, and other race. 

Method for Comparing Nested and Nonnested Models

To evaluate the racial specifications, I implement a variation of Howell and Emerson’s (2017) approach to comparing methods of measuring race on their effectiveness in predicting social inequality. Instead of using a single model test, I use two model-fit criteria to confirm their reliability. Specifically, I use both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to make comparisons (1) between the nested models (models 1–4) for each outcome and (2) between the nonnested models for each race specification (e.g., the models using the Add Health specification vs. those using the Moody specification to estimate educational attainment). Both the AIC and the BIC are penalized likelihood model-selection criteria that are intended to estimate, respectively, how closely a model comes to the true model (AIC) or how likely a model is the true model (BIC), penalized by its number of variables, with a lower or smaller score indicating a better model for both criteria. I use the comparison of AICs and BICs to identify which racial specifications are preferred for each nested model, which nested models are preferred for each outcome, and which outcomes are the most sensitive to racial specifications.

As a threshold for preference, I use a difference of 7 points for comparing among AIC and among BIC; specifically, a model is preferred over another if its AIC or BIC is at least 7 points lower than another model’s. This threshold is at the upper bound of what Burnham and Anderson (2004) characterized as “considerably [more] support” for AIC scores and solidly within the range of what Raftery (1995) characterized as “strong” evidence for BIC scores. I use this threshold to highlight four levels of preference: (1) worse models whose fit values are 7 or more points larger than (2) the best fitting model (i.e., the model with the minimum value), (3) preferred models whose fit values are both less than 7 points larger than the best fitting model and at least 7 points smaller than every worse model, and (4) the worst fitting model (i.e., the model with the maximum value).

Results

Descriptive Statistics

Table 2 summarizes the scope of inconsistency across racial categories and Add Health panels within the matched sample
Table 2. Race Inconsistency Scores in the National Longitudinal Study of Adolescent to Adult Health.

| Racial responses       | Wave I In-School | Wave I In-Home Parents | Wave I In-Home | Wave III In-Home | Respondents with Valid Pairs of Race Information across Four Panels |
|------------------------|------------------|------------------------|----------------|------------------|-------------------------------------------------------------------|
| Asian American         | .111             | .092                   | .123           | .116             | .000                                                              |
| Black/African American | .064             | .078                   | .117           | .106             | .000                                                              |
| Latinx                 | .165             | .062                   | .050           | .059             | .000                                                              |
| Native American        | .585             | .139                   | .404           | .494             | .000                                                              |
| White                  | .031             | .057                   | .076           | .066             | .000                                                              |
| Other race             | .583             | .475                   | .624           | —                | .000                                                              |
| Multiple responses     | .573             | .537                   | .531           | .503             | .000                                                              |
| Inconsistent responses | —                | —                      | —              | —                | .716                                                              |
| n with any valid pairs | 15,086           | 8,402                  | 19,572         | 15,132           | 19,587                                                            |

Note: In the rightmost column, almost all of the inconsistency scores are zero because the top six cells include only respondents who exclusively and consistently chose those race categories, and the seventh cell (multiple responses) includes only those who consistently marked multiple categories. Wave III did not include “other race” as a response option.

(rows and columns)

(total N = 91,040). The lowest inconsistency scores are among white respondents (0.031–0.076) and among respondents living with two biological parents who are both associated with the same single racial category (0.057–0.092), with the exception of the Native and other-race categories (0.139–0.624). In contrast, higher levels of inconsistency are concentrated among respondents who self-classify as Latinx in school (0.165), respondents who self-classify exclusively as Native (0.404–0.585), respondents who self-classify exclusively as other race (0.475–0.624), and particularly those who self-classify with multiple responses (0.503-0.573), which suggests that many of those associated with multiple responses in one panel are associated with single-race responses in another. For example, in the Wave I in-school panel, only 7.6 percent self-classify as either multiracial or other race, whereas twice as many, 16.0 percent, are associated with multiple, changed, or other responses among respondents with at least one valid pair of racial information.

Which Racial Specifications Provide Better Fits?

Table 3 examines which race specifications lead to better fit values in the four nested models for each outcome, in three separate panels. Each column presents the minimum AIC and BIC values for a model’s best fitting specification(s) and the deviations of every other specification from those minimum values. For example, panel A shows that for interracial relationship history, the best fitting specification among the race-only models (model 1) is the entirely consistent specification with the never-white MCO extension (AIC = 5,397), whereas the “best race” specification with no MCO extension has an AIC value 246 points higher. In the table, the best fits are highlighted in bold and in solid boxes, and the worst fits among the five main specifications (without MCO extensions) are in dashed boxes.

Across the three outcomes, the AIC and BIC consistently converge on some version of the entirely consistent specification among the race-only models (model 1, panels A–C): for education, the entirely consistent specification with the degree inconsistency extension, and for health, the entirely consistent specification without an MCO extension. This indicates the value of tracking consistency when estimating the “gross” patterns of social contact and stratification before controlling for other factors that differ between populations (Saperstein et al. 2016). For one outcome, interracial relationship history (panel A), both criteria also converge on the entirely consistent specification among the remaining nested models. Among the race-only models (model 1) and the shared-covariate models (model 2), the best fit is provided by the entirely consistent specification with the never-white MCO extension, whereas the entirely consistent specification with the ever-black extension provides the best fit among the outcome-contingent models (model 3) and the intersectionality models (model 4).

Among the remaining nested models, the best AIC and BIC show much less convergence. For education (panel B), among the outcome-contingent models (model 3), they converge on both the “best race” specification with the degree inconsistency extension and the entirely consistent specification with the ever-black extension. However, they do not converge on any specifications among the shared-covariate models or the intersectionality models (models 2 and 4) for education, and for health (panel C), they do not converge for any models beyond the race-only models (model 1).

Focusing on the five specifications without MCO extensions, the AIC and BIC converge on a worst specification for
Table 3. Fit Criteria across Race Specifications, with MCO Extensions, in Nested Models Estimating Three Outcomes.

| Specifications                                      | MCO Extensions | Model 1: Race Only | Model 2: Shared Covariates | Model 3: Outcome Contingent | Model 4: Intersectional |
|-----------------------------------------------------|----------------|--------------------|----------------------------|----------------------------|-------------------------|
|                                                     | AIC            | BIC                | AIC            | BIC                | AIC            | BIC                | AIC            | BIC                |
| Panel A: Interracial relationship history           |                |                    |                |                    |                |                    |                |                    |
| Official Add Health (in-school, Wave I)             | None           | +166              | +160           | +153              | +122          | +116              | +114          | +107              |
|                                                     | Degree inconsistency | +157          | +157           | +154              | +124          | +124              | +116          | +116              |
|                                                     | Never white    | +146              | +146           | +135              | +100          | +100              | +97           | +97               |
|                                                     | Ever black     | +168              | +168           | +161              | +118          | +118              | +113          | +113              |
| Moody (in-school, Wave I)                           | None           | +193              | +180           | +190              | +153          | +140              | +139          | +120              |
|                                                     | Degree inconsistency | +192          | +186           | +190              | +155          | +148              | +141          | +128              |
|                                                     | Never white    | +133              | +127           | +124              | +118          | +107              | +100          | +105              |
|                                                     | Ever black     | +134              | +128           | +125              | +118          | +95               | +88           | +96               |
| “Best race” (in-home, Wave I)                       | None           | +246              | +239           | +238              | +232          | +190              | +184          | +212              |
|                                                     | Degree inconsistency | +218          | +218           | +213              | +187          | +187              | +207          | +207              |
|                                                     | Never white    | +246              | +246           | +238              | +190          | +190              | +213          | +213              |
|                                                     | Ever black     | +243              | +243           | +236              | +192          | +192              | +214          | +214              |
| Entirely consistent (four panels)                   | None           | +205              | +199           | +200              | +193          | +134              | +127          | +89               |
|                                                     | Degree inconsistency | +205          | +205           | +201              | +135          | +135              | +91           | +91               |
|                                                     | Never white    | 5,397             | 5,449          | 5,361             | 5,485         | 29                | 29            | +16               |
|                                                     | Ever black     | +19               | +19            | +10               | +10           | 4,563             | 4,720         | 4,476             | 4,652             |
| Mostly consistent (four panels)                     | None           | +235              | +228           | +227              | +220          | +184              | +178          | +181              |
|                                                     | Degree inconsistency | +246          | +246           | +238              | +190          | +190              | +213          | +213              |
|                                                     | Never white    | +123              | +123           | +118              | +118          | +122              | +122          | +130              |
|                                                     | Ever black     | +171              | +171           | +159              | +159          | +127              | +127          | +139              |
| Panel B: Educational attainment                      | Official add health (in-school, Wave I) | None | +20 | +13 | +15 | +11 | +9 | +2 | +12 | +21 |
|                                                     | Degree inconsistency | +10 | +10 | +6 | +8 | +2 | +2 | +7 | +22 |
|                                                     | Never white    | +21 | +21 | +16 | +18 | +11 | +11 | +14 | +29 |
|                                                     | Ever black     | +22 | +22 | +17 | +19 | +11 | +11 | +14 | +29 |
| Moody (in-school, Wave I)                           | None           | +20 | +6 | +15 | +3 | +18 | +5 | +25 | 9,503 |
|                                                     | Degree inconsistency | +9 | +2 | +5 | 11,376 | +12 | +5 | +20 | +1 |
|                                                     | Never white    | +22 | +15 | +16 | +12 | +20 | +13 | +27 | +8 |
|                                                     | Ever black     | +19 | +12 | +15 | +11 | +14 | +7 | +21 | +3 |
| “Best race” (in-home, Wave I)                       | None           | +55 | +49 | +32 | +27 | +10 | +3 | +10 | +18 |
|                                                     | Degree inconsistency | +20 | +20 | +8 | +10 | 9,229 | 9,391 | 9,222 | +15 |
|                                                     | Never white    | +55 | +55 | +32 | +34 | +11 | +11 | +11 | +26 |
|                                                     | Ever black     | +57 | +57 | +34 | +36 | +11 | +11 | +12 | +27 |
| Entirely consistent (four panels)                   | None           | +13 | +6 | +6 | +1 | +24 | +17 | +27 | +36 |
|                                                     | Degree inconsistency | 11,847 | 11,915 | 11,236 | +2 | +19 | +19 | +23 | +38 |
|                                                     | Never white    | +14 | +14 | +7 | +9 | +20 | +20 | +23 | +39 |
|                                                     | Ever black     | +10 | +10 | +3 | +5 | 9,229 | 9,391 | +5 | +20 |
| Mostly consistent (four panels)                     | None           | +32 | +26 | +17 | +12 | +10 | +4 | +8 | +16 |
|                                                     | Never white    | +34 | +34 | +19 | +21 | +12 | +12 | +9 | +24 |
|                                                     | Ever black     | +34 | +34 | +19 | +21 | +5 | +5 | +2 | +17 |
| Panel C: Self-rated health                           | Official Add Health (in-school, Wave I) | None | +8 | +8 | +7 | +8 | +6 | +9 | +5 | +35 |
|                                                     | Degree inconsistency | +9 | +16 | +9 | +16 | +8 | +18 | +6 | +44 |
|                                                     | Never white    | +9 | +15 | +8 | +15 | +7 | +17 | +6 | +43 |
|                                                     | Ever black     | +9 | +16 | +9 | +16 | +8 | +18 | +6 | +43 |
| Moody (in-school, Wave I)                           | None           | +8 | +1 | +6 | 16,113 | +4 | 15,986 | +3 | 16,079 |
|                                                     | Degree inconsistency | +8 | +8 | +8 | +8 | +6 | +9 | +5 | +9 |
|                                                     | Never white    | +9 | +9 | +8 | +8 | +5 | +8 | +4 | +8 |

(continued)
almost every model. The “best race” specification actually has the worst fit for 8 of the 12 nested models, that is, every interracial-relationship model, as well as for the race-only models for both education and health (model 1), the shared-covariates models for education (model 2), and the intersectionality models for health (model 4). The entirely consistent specification has the worst fit for the outcome-contingent and intersectionality models for education (models 3 and 4), and the official Add Health specification has the maximum fit values for the shared-covariate and outcome-contingent models for health (models 2 and 3). Surprisingly, the official Add Health specification does not provide a preferred fit in any of the health models, though it comes close in model 4 on the basis of the AIC. Instead, the preferred specifications for health across models 2 to 4 are split between the mostly consistent specification, with or without the never-white extension (on the basis of the BIC), and the Moody specification (on the basis of the AIC). Because BIC penalizes additional parameters more heavily, it may be biased in favor of the Moody specification, which has one fewer category than the other specifications. That said, the Moody specification is also never associated with the worst AIC in any model for any outcome.

What Difference Does Using the “Right” Specification Make?

Tables 4 to 6 highlight the main differences between the best and worst fitting specifications of race by comparing parameter estimates for each outcome on a select model. For interracial relationship history, Table 4 reports estimates from the intersectionality models (model 4), which clearly have the best fit among the nested models. In contrast, for both education and health, the AIC values prefer the intersectionality models (model 4) whereas the BIC values prefer the outcome-contingent models (model 3), consistent with BIC’s greater penalization of additional parameters. For education, Table 5 reports estimates from the outcome-contingent models (model 3), which are preferred by the more conservative BIC values. For health, however, Table 6 reports estimates from the intersectionality models (model 4) because the BIC preferred models actually hide important findings.

Regarding the interracial relationship history of Add Health respondents after Wave I, Table 4 shows that using the (worst fitting) “best race” specification instead of the entirely consistent specification with the ever-black extension would hide how identifying as black even occasionally is, for MCO respondents, associated with greater romantic isolation from whites: an 88.0 percent decrease in the odds of reporting a relationship with a white partner. The latter effect is also associated with a decrease in the (positive) main effect of network heterogeneity and increases in the (positive) main effects of identifying as MCO and the interaction of network heterogeneity and identifying as nonwhite. This pattern suggests that much of the estimated effect of interracial friendships in the “best race” model is associated with the never-black MCO respondents that the “best race” specification distributes among its single-race categories.

Regarding educational attainment by Add Health respondents after Wave I, Table 5 shows that using the (worst fitting) entirely consistent specification with no MCO
extension, instead of the “best race” specification with the degree inconsistency extension hides how relative consistency in racial identification is associated with substantial variation in the educational attainment of MCO respondents: a 28.1 percent decrease in the odds of completing more education, at the MCO mean inconsistency of 0.716. Indeed, the degree inconsistency extension is associated with preferred specifications almost exclusively among the models of educational attainment.

Regarding self-rated health by Add Health respondents after Wave I, Table 6 shows that using the (worst fitting) “best race” specification instead of the mostly consistent specification hides how identifying consistently as Native, even if also identifying occasionally as non-Native is associated with poorer health: a 71.6 percent decrease in the odds of a higher health rating. In fact, the mostly consistent specification is a preferred specification primarily among the

| Variable                                      | "Best Race" Specification | Entirely Consistent Specification with Ever-Black Extension |
|-----------------------------------------------|----------------------------|------------------------------------------------------------|
| Race (reference: white)                       |                            |                                                            |
| MCO                                           | .33                        | 1.08****                                                  |
| Network racial heterogeneity (interracial friendships) | .91**                      | .54*                                                       |
| Nonwhite × network heterogeneity              | 1.09*                      | 2.45*                                                      |
| MCO × ever black (extension)                  |                            |                                                            |
| Intercept (fixed-effects equation)             | -.97***                    | -.88****                                                  |
| Across-schools variance (random intercept)     | .049**                     | .013**                                                     |
| Individual-level n                            | 5,171                      | 5,171                                                      |
| School-level n                                | 113                        | 113                                                        |
| AIC                                           | 4,687                      | 4,476                                                      |
| BIC                                           | 4,858                      | 4,652                                                      |

Note: The sample comes from National Longitudinal Study of Adolescent to Adult Health restricted-use data. Estimates are from multilevel mixed-effects logistic regression models of any white/nonwhite relationships. Models also include fixed effects for identification as Native American, Asian American, black, and Latinx, age, female, married parents, parental educational attainment, non-English dominant language at home, any religious affiliation, region, proportion same race at school, any interracial relationship in adolescence (white/nonwhite), number of romantic relationships after Wave I, any same-sex relationships, and race × female. AIC = Akaike information criterion; BIC = Bayesian information criterion; MCO = multiple race, changed race, and other race.

* p < .10. ** p < .05. *** p < .01.

| Variable                                      | Entirely Consistent Specification | "Best Race" Specification with Inconsistency Extension |
|-----------------------------------------------|----------------------------------|-------------------------------------------------------|
| Race (reference: white)                       |                                  |                                                        |
| MCO                                           | -.03                             | .45                                                   |
| Degree of inconsistency (extension)           |                                  | -.46****                                               |
| Thresholds                                    |                                  |                                                        |
| Cut 1 (HS degree vs. less than HS)            | -2.97***                        | -3.02****                                              |
| Cut 2 (bachelor’s degree vs. HS)              | 2.97***                         | 2.96***                                                |
| Cut 3 (postgraduate degree vs. bachelor’s degree) | .10***                     | 5.31****                                               |
| Across-schools variance (random intercept)     | .10***                          | .10***                                                |
| Individual-level n                            | 6,305                            | 6,305                                                  |
| School-level n                                | 113                              | 113                                                    |
| AIC                                           | 9,253                            | 9,229                                                  |
| BIC                                           | 9,408                            | 9,391                                                  |

Note: The sample comes from National Longitudinal Study of Adolescent to Adult Health restricted-use data. Estimates are from multilevel mixed-effects ordered logistic regression models. Models also include fixed effects for identification as Native American, Asian American, black, and Latinx, age, female, married parents, parental educational attainment, non-English dominant language at home, any religious affiliation, region, cumulative grade point average, median family income at school, and percentage black students at school. AIC = Akaike information criterion; BIC = Bayesian information criterion; MCO = multiple race, changed race, and other race.

*** p < .01.
models of self-rated health. That said, the BIC-preferred nested model is actually the outcome-contingent model (model 3), but among these models (not shown), it is the worst-fitting official Add Health specification that reveals how identifying as Native is associated with poorer health, whereas using the best-fitting (BIC) Moody specification would hide a 13.1 percent decrease in the odds of a higher health rating.

In sum, the classification of MCO respondents is more central to the substantive differences between the best and worst fitting models of interracial relationships, in comparison with the best and worst models of educational attainment and self-rated health models. This is consistent with the substantially larger range of AIC and BIC values among the models for interracial relationships (>200 points) than those for education (<60 points) and health (almost all <20 points).

### Conclusions and Limitations

My analysis suggests that quantitative researchers should take more advantage of the scope of racial information that is available in their survey data, even when racial categories primarily serve as covariates. Using Add Health to examine the scope and consequences of racial multiplicity and inconsistency, I have explored different methods for (1) classifying MCO respondents and (2) capturing heterogeneity among them. I find that racial specifications that do not track inconsistency may mischaracterize more than half of multiracial respondents, which in Add Health amounts to roughly one in six respondents, a fraction that rivals its numbers of African American and Latinx respondents.

For all three outcomes, the new entirely consistent specification outperforms the three existing methods for classifying MCO respondents among the race-only models that estimate the “gross” effects of race (Saperstein et al. 2016). Beyond model 1, however, there is substantial variation across outcomes in the specifications that are preferred, with certain specifications revealing important differences among MCO respondents. Indeed, the value of using more racial information rather than relying on a single panel depends on the outcome under consideration. It has clear value for the interracial relationship models but makes a more modest difference among the models for educational attainment and self-rated health, suggesting a qualitative difference in the contemporary role of race for social contact versus for social stratification.

The difference between the best and worst fitting models for interracial relationships suggest that for modeling social contact, researchers should (1) strictly separate respondents with inconsistent information from those with consistent information, unlike the official Add Health, Moody, “best race,” and even the mostly consistent specifications; (2) distinguish respondents associated with multiple-race categories from monoracial respondents, unlike the Add Health and “best race” specifications; and (3) distinguish among MCO respondents along the black/nonblack boundary, rather than along the white/nonwhite boundary or by relative level of inconsistency. The latter in particular suggests a critical divide among MCO respondents in their interpersonal contact with whites along the black/nonblack social boundary. I strongly recommend that quantitative researchers of interracial contact specify
race using something similar to the entirely consistent specification supplemented by an indicator for whether the MCO respondents are ever classified as black (ever-black extension). In addition, Add Health researchers should consider recalculating the study’s preconstructed network variables using the entirely consistent specification.

In comparison, the best and worst fitting models for educational attainment suggest that for modeling processes of stratification, researchers should (1) classify multiple-race respondents by a single category of their choice, unlike the entirely consistent specification, and (2) distinguish respondents associated with inconsistency by their degree of inconsistency. There appears to be less support for an independent MCO experience with education, as there is for interracial intimacy, and more evidence for a penalty for inconsistency relative to their most salient single category. I recommend that quantitative researchers of stratification processes specify race using something similar to the “best race” specification supplemented by an indicator of the level of inconsistency within the racial information associated with MCO respondents. That said, this recommendation is tempered by the rival fit of models using the ever-black extension with the entirely consistent and mostly consistent specifications. Thus I also suggest that researchers should (3) compare the recommended specification with alternative specifications that separate respondents with multiple or inconsistent information from those with single-race and consistent information, as long as they are supplemented by something like the ever-black extension.

In contrast, the best and worst fitting models for health have even more modest implications. For modeling the consequences of stratification for individuals, researchers should (1) not assign MCO respondents to single-race categories using follow-up questions about their “best race” and (2) consider classifying respondents with consistent Native information as Native American, regardless of their other racial information. Similar to Howell and Emerson’s (2017) findings, I find that the differences among racial specifications are smaller for self-rated health than for educational attainment or interracial relationship history. Thus I cannot recommend a single specification for quantitative researchers of health inequality, and instead, I recommend that they compare the relative fit of alternative racial specifications while paying special attention to the consequences for Native respondents. In sum, when researchers only use a single source of race information when more is available, they undermine their ability to more fully model the distinctiveness and heterogeneity of the multiracial experience and by extension the complexity of race-related processes in their data.

That said, there remain important limitations to the generalizability of my analysis of primarily self-classification patterns to the analytically distinct processes of (1) racial identity, (2) observed race, (3) reflected race, and (4) known ancestry (Roth 2016). With respect to the interracial relationship results, it is unclear whether ever-black MCO respondents have a lower likelihood of interracial relationships with whites because of, respectively, their subjective identification with blacks, their exclusion by whites, their anticipation of exclusion, or their desire for “ancestral continuity” (i.e., some sense of endogamy that excludes whites) (cf. Childs 2005). With respect to the educational attainment results, it is unclear why the less consistent MCO respondents have a lower likelihood of completing more schooling. For example, is it because racial fluidity interferes with, respectively, their school engagement (via their racial identity), with their inclusion in racially bounded networks for academic support, with their anticipation of inclusion, or with their fulfillment of “ancestral expectations” for educational achievement (i.e., ethnically oriented), or otherwise ethnically bounded, family expectations (cf. Lee and Zhou 2015)? With respect to the health results, it is unclear whether “mostly consistent” Native respondents have a lower likelihood of better health because of, respectively, their subjective association of being Native with certain health-related behaviors, their exclusion from resources associated with alternative behaviors, or their expectation of exclusion, or because their desire for “ancestral survival” places them in disadvantaged circumstances that negatively affect their health (cf. Snipp 1992). In brief, my findings show that relative consistency in self-classification matters for a range of racial outcomes but cannot identify their specific mechanism in the absence of additional information, thereby confirming “the importance of looking beyond self-identification to understand how categorical racial perceptions shape people’s life outcomes” (Saperstein et al. 2016:521).

To address these limitations, I recommend that future survey collections include the following items: (1) an observed race item collected before self-classification questions and inclusive of Latinx as a classification (cf. Feliciano 2016), (2) a reflected race item that includes Latinx as a classification, (3) a racial self-classification item that allows multiple responses and includes Latinx as a classification, and (4) a “best race” self-classification item that does not use the phrasing of background or ancestry as synonymous with personal salience. If possible, I would also recommend a supplemental parent survey with (5) a similar self-classification item and (6) an observed-race item for parent-partners inclusive of both nonresident biological parents and resident non-biological parents. Ultimately, researchers would also need multiple panels of these items to assess Liebler et al.’s (2016) claim that

analysts from all fields would benefit from conceptualizing and operationalizing a person’s race as having a past, present, and future (as is done for marital status, work, and residence, for example), rather than acting as if race were an unchanging trait. (p. 536)
Appendix

Table A1. Race Information Available in the National Longitudinal Study of Adolescent to Adult Health by Dimension and Panel.

| Approximate Race Question or Topic | Race Dimension (Roth 2016) | Wave I In-School | Wave I In-Home Parents | Wave I In-Home | Wave III In-Home | Wave IV In-Home |
|------------------------------------|-----------------------------|-----------------|------------------------|---------------|-----------------|----------------|
| Are you of Hispanic origin?        | Self-classification          | X               | X                      | X             | X               | X              |
| What is your race? You may give more than one answer. | Self-classification          | X               | X                      | X             | X               | X              |
| Which one category best describes your racial background? | Self-classification          | X               | X                      | X             | X               | X              |
| Please code the race of the respondent from observation alone. | Observed race               | X               | X                      | X             | X               | X              |
| Family origins or ancestries (up to four countries, groups or geographic areas) | Known ancestry              | X               | X                      | X             | X               | X              |
| Is your current (spouse/partner) of Hispanic origin? | Observed race based on interaction | X               | X                      | X             | X               | X              |
| What is the race of your current (spouse/partner)? You may give more than one answer. | Observed race based on interaction | X               | X                      | X             | X               | X              |

Table A2. Inconsistency Scores for Simulated Respondents.

| Respondent | Wave I In-School | Wave I In-Home Parents | Wave I In-Home | Wave III In-Home | Consistent Pairs | Valid Pairs | Inconsistency Score |
|------------|------------------|-------------------------|---------------|------------------|------------------|-------------|---------------------|
| R1         | Black            | Multiple                | Asian         | Latinx           | 0                | 6           | 1.00                |
| R2         | Asian            | Multiple                | —             | Native           | 0                | 3           | 1.00                |
| R3         | Native           | Native                  | Native        | Native           | 6                | 6           | .00                 |
| R4         | Multiple         | Multiple                | —             | —                | 1                | 1           | .00                 |
| R5         | Latinx           | Multiple                | Multiple      | Asian            | 1                | 6           | .83                 |
| R6         | Black            | Black                   | Multiple      | Multiple         | 2                | 6           | .67                 |
| R7         | Asian            | —                       | Multiple      | Multiple         | 1                | 3           | .67                 |
| R8         | Latinx           | Multiple                | Multiple      | Multiple         | 3                | 6           | .50                 |
| R9         | Other            | Other                   | —             | Blacka           | 1                | 1           | .00                 |
| R10        | Other            | Multiple                | White         | Blacka           | 0                | 3           | .00                 |
| R11        | Other            | Multiple                | White         | Blacka           | 1                | 3           | .67                 |

Note: Missing observations reduce the number of valid pairs and constrain the range of values that inconsistency can take: four observations allow the index to take the values 0.00, 0.50, 0.67, 0.83, and 1.00, whereas three observations constrain the index to 0.00, 0.67, and 1.00, and two observations make the index dichotomous: 0.0 and 1.00. One limitation is that the score 0.67 represents two potentially distinct situations: (1) “two pairs,” as in a poker hand, among four observations (respondent R6) and (2) one pair plus one distinct single among three observations (respondent R7).

a. Because Wave III excludes the other-race category, I treat the Wave III information as a missing observation if in any other panel, the respondent is associated with “other race” but not with Latinx (respondents R9, R10, and R11).

Table A3. Responses to Racial Identification Items across National Longitudinal Study of Adolescent to Adult Health Panels (N = 91,040 Unique Respondents).

| Racial Responses | Wave I In-School | Wave I In-Home Parents | Wave I In-Home | Wave III In-Home | Valid Race Information in Any of Four Panels |
|------------------|------------------|-------------------------|---------------|------------------|---------------------------------------------|
| Multiple         | 6.01%            | 4.42%                   | 3.87%         | 3.56%            | 6.68%                                       |
| Asian American   | 4.78%            | 6.07%                   | 6.34%         | 6.60%            |                                             |
| Black/African American | 15.88%       | 11.41%                  | 20.85%        | 20.77%           |                                             |
| Latinx           | 17.49%           | 18.27%                  | 17.01%        | 16.32%           |                                             |
| Native American  | 0.94%            | 0.14%                   | 0.55%         | 0.76%            |                                             |
| White            | 53.26%           | 59.47%                  | 50.45%        | 51.98%           |                                             |
| Other race       | 1.64%            | 0.23%                   | 0.93%         | —                |                                             |

(continued)
Table A3. (continued)

| Valid Responses | Wave I In-School | Wave I In-Home Parents | Wave I In-Home | Wave III In-Home | Valid Race Information in Any of Four Panels |
|-----------------|------------------|------------------------|---------------|------------------|---------------------------------------------|
| Valid n         | 83,783           | 8,403                  | 20,724        | 15,179           | 89,463                                      |
| Missing         | 7,257            | 82,637                 | 70,316        | 75,861           | 1,577                                       |

Note: Total N = 91,040 unique respondents across these four panels plus Wave IV, which does not include racial identification items. The valid n is the number of respondents in a panel with nonmissing racial information. The Latinx row includes all who answered “yes” to the Hispanic ethnicity item, whatever their race item responses, and includes partial ancestry Latinx because the Hispanic item did not permit multiple responses. The Wave I in-home parent column includes only respondents with residential biological parents, and its first row is the percentage of respondents whose parents are classified in different categories or are multiracial themselves. Wave III did not include “other race” as a response option.

Table A4. Group-Specific Means for Educational Attainment, Self-Reported Health, and Interracial Relationship History in the National Longitudinal Study of Adolescent to Adult Health, along with Select Independent Variables.

| Racial Identification (Entirely Consistent Race Specification) | Analytic Sample | White | MCO | Black/African American | Latinx | Asian American | Native American |
|---------------------------------------------------------------|----------------|-------|-----|------------------------|--------|----------------|----------------|
| n (educational attainment models)                             | 6,305          | 3,182 | 1,080 | 983                    | 762    | 284            | .14            |
| Educational attainment after Wave I (0–3)                     | 1.41           | 1.47  | 1.30 | 1.37                   | 1.29   | 1.65           | .86            |
| Parental educational attainment in Wave I (0–3)               | 1.43           | 1.53  | 1.39 | 1.36                   | 1.96   | 1.81           | 1.36           |
| Age in Wave I                                                 | 14.91          | 14.87 | 14.73 | 14.93                  | 15.20  | 15.28          | 15.36          |
| Female                                                        | .53            | .53   | .52  | .58                    | .50    | .49            | .43            |
| Cumulative GPA (high school)                                  | 2.64           | 2.81  | 2.51 | 2.33                   | 2.35   | 3.14           | 1.88           |
| Median family income at school in Wave I (thousands)          | 39.24          | 43.19 | 38.77 | 31.71                  | 32.44  | 42.19          | 16.43          |
| n (self-rated health models)                                  | 6,264          | 3,166 | 1,076 | 973                    | 755    | 280            | 14             |
| Self-rated health after Wave I (1–5)                          | 3.75           | 3.80  | 3.68 | 3.73                   | 3.68   | 3.74           | 3.07           |
| Parental educational attainment in Wave I (0–3)               | 1.43           | 1.54  | 1.39 | 1.36                   | .95    | 1.81           | 1.36           |
| Age in Wave I                                                 | 14.92          | 14.87 | 14.73 | 14.93                  | 15.20  | 15.28          | 15.36          |
| Female                                                        | .53            | .53   | .52  | .58                    | .50    | .49            | .43            |
| Educational attainment after Wave I (0–3)                     | 1.41           | 1.47  | 1.30 | 1.36                   | 1.29   | 1.66           | .86            |
| Ever stopped by police before age 1B                          | .08            | .08   | .09  | .07                    | .11    | .07            | .14            |
| n (interracial relationship models)                           | 5,171          | 2,631 | 775  | 832                    | 650    | 260            | 23             |
| Interracial relationship after Wave I                         | .30            | .25   | .63  | .12                    | .35    | .28            | .13            |
| Network racial heterogeneity (send network, 0.0–0.5)          | .15            | .17   | .25  | .04                    | .11    | .08            | .11            |
| Age in Wave I                                                 | 14.87          | 14.78 | 14.65 | 14.86                  | 15.34  | 15.33          | 14.43          |
| Female                                                        | .56            | .55   | .56  | .63                    | .56    | .47            | .57            |
| Proportion same race at school in Wave I                      | .55            | .75   | .12  | .45                    | .47    | .23            | .22            |
| Interracial relationship in Wave I                           | .12            | .07   | .36  | .04                    | .15    | .08            | .00            |

Note: Race groupings are ordered from left to right by subsample size, except for the bottom panel (interracial relationship sample), where the black sample is larger than the MCO sample. GPA = grade point average; MCO = multiple race, changed race, and other race.

Acknowledgments

I thank Aaron Gullickson and the participants in the 2016 Add Health Users Conference for comments on earlier versions of this article. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health Web site (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

References

#Bratter, Jenifer L., and Bridget K. Gorman. 2011. “Does Multiracial Matter? A Study of Racial Disparities in Self-Rated Health.” *Demography* 48(1):127–52.
#Bratter, Jenifer L., Mary E. Campbell, and Wendy D. Roth. 2014. “Multiple Components of Race Data Library.” Retrieved April 23, 2019. https://kinder.rice.edu/sites/g/files/bxs1676/f
Contributions.” *Journal of Health and Social Behavior* 51(1 Suppl.):S15–27.

#Williams, Teresa Kay. 1996. “Race as Process: Reassessing the ‘What Are You?’ Encounters of Biracial Individuals.” Pp. 191–210 in *The Multiracial Experience: Racial Borders as the New Frontier*, edited by Maria Root. Thousand Oaks, California: Sage.

#Wilson, William J. 1980. *The Declining Significance of Race: Blacks and Changing American Institutions*. 2nd ed. Chicago: University of Chicago Press.

**Author Biography**

**Jiannbin Lee Shiao** is an associate professor and associate head of sociology at the University of Oregon. He is the author of *Identifying Talent, Institutionalizing Diversity: Race and Philanthropy in Post–Civil Rights America* (2005) and *Choosing Ethnicity, Negotiating Race: Korean Adoptees in America* (2011, with Mia Tuan). He has published articles in the *American Journal of Sociology, Sociological Theory, Asian American Policy Review, Comparative Sociology, Contexts, Du Bois Review, Race and Society, and Sociology of Race and Ethnicity.*