

When (In)Consistency Matters: Racial Identification and Specification

By Jiannbin Lee Shiao, University of Oregon

Paper for presentation at the 2019 Population Association of America Annual Meeting

- Session 23, Racial/Ethnic Identity, Thursday, April 11, 8:30 AM
- Former title: “Mapping Multiplicity and Inconsistency in Racial/Ethnic Identification in Add Health”

Abstract:

Sociologists who rely on survey research have begun exploring the implications of racial inconsistency for not only multiracial identification but also other social outcomes. Using the National Longitudinal Study of Adolescent to Adult Health (Add 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 AIC and BIC to compare five methods of classifying multiple, changed, and other-race (MCO) 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.

Acknowledgements: The author thanks Aaron Gullickson and the participants in the 2016 Add Health Users Conference for comments on earlier versions of the manuscript. 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 website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

Author contact information: Jiannbin Lee Shiao, Department of Sociology, University of Oregon, Eugene, OR 97403-1291. Email: jshiao@uoregon.edu.

When (In)Consistency Matters: Racial Identification and Specification

“One Mexican, 25 year old, male...” I was getting pissed, but I knew I had to...reveal my true race because it’s stereotyped more favorably... I was arrested before because *that* cop thought I was black... I really have to...project different selves in order to survive in a society that treats brown and black men like shit.” (Williams 1996)

Kroft: Your mother was white. Your father was African... Yet, at some point, you decided you were black? (Schorn 2007)

I have brothers, sisters, nieces, nephews, uncles, and cousins of every race and every hue, scattered across three continents... It's a story that...has seared into my genetic makeup the idea that this nation is more than the sum of its parts — that out of many, we are truly one. (Obama 2008)

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 that 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 from 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 that the prevalence of racial *multiplicity* or the identification of individuals, either by themselves or by others, with multiple racial categories

instead of a single category (Brunsma and Rockquemore 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 themselves 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 other social outcomes (Campbell, Bratter, and Roth 2016; Saperstein, Kizer, and Penner 2016). How should quantitative sociologists measure race¹ 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 (NIH) and has been used in studies of both multiracial identities and racial identity change. Indeed it was one of the first datasets (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 calls into question the measurement of

¹ I do not make an analytic distinction between race and ethnicity, as both are social perceptions of ancestry in terms of typologies with particular social origins and consequences; however, in this article, I primarily use *race* instead of *race/ethnicity*, because this analysis examines the broader classifications (e.g. “Asian” instead of “Filipino”).

² In Bratter, Campbell, and Roth’s (2014) review of 14 datasets, they find only one dataset that predates Add Health’s collection of multiple responses for racial self-classification: the National Longitudinal Survey of Youth, initially administered in 1979.

racial identity as singular and stable. Second, focusing on Add Health's higher quality information, i.e. 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 dataset 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 (AMEA) in 1988, Project RACE (Reclassify All Children Equally) in 1991, and the Hapa Issues Forum (HIF) in 1992 (King-O'Riain 2006; King and DaCosta 1996; Spencer 1999). An early, pivotal event in the movement was the appointment of AMEA 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 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 continues to exclude 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 multiracials 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 regardless of whether they study racial identity.

Conceptualizing Racial Multiplicity and Inconsistency

The multiracial movement was accompanied by the rise of mixed race studies, an academic field examining the experiences of multiracials, 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 multiracials' 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 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 has comprehensively assembled in her typology of race dimensions (2016): 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, i.e. the race that one might choose from a menu of pre-established 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, i.e. as an indicator 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. subjective self-identification that may indicate salient meanings theorized to influence outcomes “such as political mobilization or voting patterns, residential decision-making, social network formation, or attitudes” (Roth 2016:5), i.e. socially and culturally embedded decisions, such as racial self-classification. Alternatively, researchers may use racial self-classification as a proxy for observed race, i.e. the race others believe you 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 that a person believes 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 their known ancestry based on family history or even their observed race based on genomic analysis.³ 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 further clarifies her dimensions of race by distinguishing between 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" (2016: 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" (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 vs. (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

³ I relabel Roth's racial ancestry and its sub-dimensions of known racial ancestry and known genetic ancestry as, respectively, known ancestry, known ancestry based on family history, and known ancestry based on genomic ancestry testing. The latter is distinct from observed race based on genomic ancestry testing, which only becomes "known" when the test result is shared with the test subject.

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 their 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 based on 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 the one hand, Liebler, Bhaskar, and Porter (2016) have found significant demographic differences between (1) Americans who only occasionally self-classify as American Indian and (2) those who consistently self-classify as American Indian while also occasionally self-classifying by additional races. In other words, relative consistency indicates a substantive divide among multiracial American Indians, i.e. 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) find that inconsistency does not significantly distinguish multiracials in models predicting depression, school belonging, school engagement, and GPA. Instead they argue that observed race is a more important indicator of heterogeneity among multiracials, based on 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 dataset for exploring racial multiplicity and inconsistency, but it also omits important dimensions of racial information. Like most quantitative datasets, 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 typology, and where Add Health excels is its data on racial self-classification.

In all panels except Wave IV, Add Health first asks respondents whether they are of "Hispanic or Spanish/Latino"⁴ origin, accepting only a single response, before then asking respondents for their race, i.e. their self-classification as white, African American or black, Asian American or Pacific Islander, American Indian or Alaskan Native,⁵ or other race, with an explicit prompt for multiple responses. In the Wave I in-home parent panel, the Wave I in-home adolescent panel, and the Wave III in-home panel, Add Health asks an additional "best race" question to respondents who make multiple responses to the race item, about the "one category [that] best describes your racial background." Thus for three panels, Add Health includes three pieces of racial information on self-classification.

However, besides the absence of racial self-classification items in Wave IV and the conflation of multiracial Latinxs with other Latinxs in every panel, there are other limitations: The absence of the "best race" item from the in-school panel means that researchers cannot fully

⁴ From this point forward, I only use the gender neutral "Latinx" instead of Latino to refer to Latinas and Latinos without distinction.

⁵ For 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."

combine the item with the race question for Add Health’s largest panel (n=83,783 with non-missing race information), which also houses Add Health’s most extensive data on friendship networks. There is also the sudden disappearance 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 their current spouse or partner is of Hispanic/Latinx origin and what their 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,⁶ which means that its information is actually observed race based on a very particular form of interaction, a request for self-classification, rather than observed race based on appearance alone (c.f. 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/partners, much less clarify whether their answers should be based on their partner’s

⁶ In 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.

appearance or the respondent's knowledge of their partner's racial identity, self-classification, or known ancestry⁷ or (2) any race information for non-residential biological parents.

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?”⁸ 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 2009). In sum, Add Health includes racial information for four of Roth's six dimensions, with the higher quality information coming from its multiple observations of racial self-classification⁷ 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 multiracials 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 which places multiracials into the other-race category, along with Native Americans, and

⁷Add 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.

⁸ 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 pre-constructed indicators for observed race based on genomic analysis are not currently available in the restricted-use dataset.

3. A specification that employs the “best race” question, variations of which are available in other datasets (Bratter, Campbell, and Roth 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 multiple, changed, or other-race information (MCO).

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 et al 2009). In the first wave of data collection in 1994–1995, the study administered in-school interviews 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), 2001–2002 (Wave III), and 2007–2008 (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% (Wave III in-home) to 6.01% (Wave I in-school). In aggregate, 6.68% of

⁹ I exclude the Wave II panel, whose race information was simply preloaded from the Wave I in-home panel.

respondents gave multiple responses in at least one panel, among the 89,463 respondents who had non-missing 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 non-missing 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 $_i = 1 - \text{Consistency}_i$, i.e. the proportion of valid pairs with inconsistent racial information for respondent i .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% 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 “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

¹⁰ Appendix Table 1 illustrates the inconsistency index with sample values for respondent-level information.

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, i.e. 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 pre-calculated 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 comprised 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. As 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 responders into a separate category. Second, respondents are reclassified to a multiple, changed, or other race 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 data¹¹, and the Wave III in-home data.¹²

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 they mark "Native American" and "white" in one panel and "Native American" in another.

¹¹ 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.

¹² 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.

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” multiracials (c.f. 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, e.g. the Entirely Consistent specification plus Degree Inconsistency, Entirely Consistent plus Never-white, and Entirely Consistent plus Ever-black, adding altogether another 14 specifications.¹³

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. 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, i.e. 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.

¹³ 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.

To measure educational attainment after Wave I, I used Add Health's questions on highest degree completed in Wave III or, if available, Wave IV, and I reduced the options to four ordinal categories: Postgraduate degree, Bachelors 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 (Jylha 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 a respondent as having had interracial relationship if (1) they are classified as white, and they identify any romantic partner as being nonwhite, or (2) they are classified as nonwhite, and they identify any romantic 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 15.1 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 only includes 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 (c.f. 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,¹⁴ group size, any interracial intimacy in adolescence,¹⁵ number of romantic relationships after adolescence, and any same-sex relationships (Qian 1997; Shiao 2018).

¹⁴ My indicator for interracial friendships is a modification of Add Health's precalculated variable for network racial heterogeneity, restricting the number of groups to two (white and nonwhite), with the result that its range contracts from 0.0-0.8 to 0.0-0.5 and its interpretation becomes the probability that two randomly selected persons in a respondent's network will be a white person and a nonwhite person.

¹⁵ In the Wave I in-home panel, respondents were permitted to identify their romantic partners with multiple categories. I classified partners who identified as both white and nonwhite as interracial for both white and nonwhite respondents.

Model 4 adds interaction terms to control for intersectionality¹⁶ (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 X each racial category in the specification in question, whereas for interracial intimacy, it is interracial friendship X 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 classification across specifications. In the rows listing the independent variables, **Table 1** shows the relative overlap of covariates across outcomes.

TABLE 1 ABOUT HERE

Method for Comparing Nested and Non-nested Models

To evaluate the racial specifications, I implement a variation of Howell and Emersen's approach to comparing methods of measuring race on their effectiveness in predicting social inequality (2017). 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 non-nested models for each race specification, e.g. the models using the Add Health specification vs. those using the Moody specification to estimate

¹⁶ My approach is consistent with intercategory or process-centered forms of intersectionality theory (McCall 2005) which is better suited for quantitative analysis than Choo and Ferree's preferred form of systemic intersectionality (2010).

educational attainment. Both AIC and 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 the 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, i.e. 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 characterize as “considerably [more] support” for AIC scores (2004) and solidly within the range of what Raftery characterizes as “strong” evidence for BIC scores (1995). I use this threshold to construct three levels of preference: (1) **worse models** whose fit values are 7+ points larger than the model with the minimum value, i.e. the best fitting model, (2) **preferred models** whose fit values are both less than 7 points larger than the best fitting model and 7+ points smaller than every worse model, and (3) **less-preferred models** whose fit values are also less than 7 points larger than the best model but not 7+ points smaller than every worse model.

Results

Descriptive Statistics

Table 2 summarizes the scope of inconsistency across racial categories and Add Health panels within the matched sample (Total N=91,040).¹⁷ 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 those who self-classify with multiple responses (0.503-0.573). The latter suggests that most of those associated with multiple responses in one panel are associated with single-race responses in another. In brief, using only a single panel's racial information may severely undercount multiracial respondents, e.g. N=949 with multiple-race responses in the Wave I in-school panel vs. N=2,916 with changed responses across panels.

TABLE 2 ABOUT HERE

¹⁷ Appendix Table 2 reports the distribution of respondents across racial/ethnic categories and panels, including the percentages of respondents associated with multiple responses. Appendix Table 3 reports the means of select variables both for each outcome's analytic sample and by racial groupings with the groups ordered in columns from left to right by subsample size, except for the bottom panel (interracial-relationship history) where the black sample is larger than the MCO sample.

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 = 5397), whereas the "Best Race" specification with no MCO extension has an AIC value 246 points higher. In the table, the preferred fits are shaded, in solid boxes, and in bold, whereas the less-preferred fits are only shaded, the worse fits are unformatted, and the worst fits among the five main specifications (without MCO extensions) are in dashed boxes.

TABLE 3 ABOUT HERE

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, Kizer, and Penner 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). The complete absence of less-preferred specifications among these models indicates the strength of these preferences.

Among the remaining nested models, the best AIC and BIC show much less convergence. For education (Panel B), among the outcome-contingent models for education (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 almost every model. The “Best Race” specification actually has the worst fit for eight of the twelve nested models, i.e. 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 provides a less-preferred fit in Model 4. Instead, the preferred specifications for health across Models 2-4 are split between the

Mostly Consistent specification, with or without the Never-white extension (on the basis of AIC), and the Moody specification (on the basis of BIC). 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 4-6 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 1, **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% 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 non-white. This pattern suggests that much of the estimated effect of interracial friendships in the “Best Race” model is associated with the MCO respondents who never identify as black.

TABLE 4 ABOUT HERE

Regarding educational attainment by Add Health respondents after Wave 1, **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% 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.

TABLE 5 ABOUT HERE

Regarding self-rated health by Add Health respondents after Wave 1, **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% decrease in the odds of a higher health rating. In fact, the Mostly Consistent specification is a preferred specification primarily among the 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% decrease in the odds of a higher health rating.

TABLE 6 ABOUT HERE

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 multiple-race, changed-race, and other-race (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, Kizer, and Penner 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 monoracials, 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, so 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

findings, I find that the differences among racial specifications are smaller for self-rated health than for educational attainment or interracial relationship history (2017). 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, consistent with Bratter and Gorman's (2011) and Liebler et al's (2016) findings.

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. 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? 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 impact their health. 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, Kizer, and Penner 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 (c.f. Feliciano 2016); (2) a reflected race item that includes Latinx as a classification; (3) a racial self-classification item that allows for 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 non-resident biological parents and resident non-biological parents. Ultimately, researchers would also need multiple panels of these items to assess Liebler et al’s 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” (2016:536).

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." [Online database.] Retrieved (https://kinder.rice.edu/sites/g/files/bxs1676/f/documents/Multiple%20Components%20of%20Race%20Data%20Library_0.pdf).
- Burnham, Kenneth P. and David R. Anderson. 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods & Research* 33(2):261–304.
- Campbell, Mary E. 2009. "Multiracial Groups and Educational Inequality: A Rainbow or a Divide?" *Social Problems* 56(3):425–46.
- Campbell, Mary E., Jenifer L. Bratter, and Wendy D. Roth. 2016. "Measuring the Diverging Components of Race: An Introduction." *American Behavioral Scientist* 60(4):381–89.
- Choo, Hae Yeon and Myra Marx Ferree. 2010. "Practicing Intersectionality in Sociological Research: A Critical Analysis of Inclusions, Interactions, and Institutions in the Study of Inequalities." *Sociological Theory* 28(2):129–49.
- Domina, Thurston, Andrew Penner, and Emily Penner. 2017. "Categorical Inequality: Schools As Sorting Machines." *Annual Review of Sociology* 43(1):311–30.
- Doyle, Jamie Mihoko and Grace Kao. 2007. "Friendship Choices of Multiracial Adolescents: Racial Homophily, Blending, or Amalgamation?" *Social Science Research* 36(2):633–53.
- Feliciano, Cynthia. 2016. "Shades of Race: How Phenotype and Observer Characteristics Shape Racial Classification." *American Behavioral Scientist* 60(4):390–419.

- Gullickson, Aaron and Ann Morning. 2011. "Choosing Race: Multiracial Ancestry and Identification." *Social Science Research* 40(2):498–512.
- Guo, Guang and Hongxin Zhao. 2000. "Multilevel Modeling for Binary Data." *Annual Review of Sociology* 26(1):441–62.
- Harris, David R. and Jeremiah Joseph Sim. 2002. "Who Is Multiracial? Assessing the Complexity of Lived Race." *American Sociological Review* 67(4):614.
- Harris, Kathleen Mullan, Carolyn Tucker Halpern, Eric Whitsel, Jon Hussey, Joyce Tabor, Pamela Entzel, and J. Richard Udry. 2009. "The National Longitudinal Study of Adolescent to Adult Health: Research Design" [WWW document]. URL: <http://www.cpc.unc.edu/projects/addhealth/design>.
- Harris, Kathleen Mullan. 2009. *The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007–2009* [Machine-readable database file and documentation]. Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill.
- Howell, Junia and Michael O. Emerson. 2017. "So What 'Should' We Use? Evaluating the Impact of Five Racial Measures on Markers of Social Inequality." *Sociology of Race and Ethnicity* 3(1):14–30.
- Jylhä, Marja. 2009. "What Is Self-Rated Health and Why Does It Predict Mortality? Towards a Unified Conceptual Model." *Social Science & Medicine* 69(3):307–16.
- King, Rebecca Chiyoko and Kimberly McClain DaCosta. 1996. "Changing Face, Changing Race: The Remaking of Race in the Japanese American and African American Communities." Pp. 227-244 in Maria Root, ed. *The Multiracial Experience: Racial Borders as the New Frontier*. Thousand Oaks, California: Sage.

- King-O'Riain, Rebecca Chiyoko. 2006. *Pure Beauty: Judging Race in Japanese American Beauty Pageants*. Minneapolis: University of Minnesota Press.
- Kramer, Rory, Ruth Burke, and Camille Z. Charles. 2015. "When Change Doesn't Matter: Racial Identity (In)Consistency and Adolescent Well-Being." *Sociology of Race and Ethnicity* 1(2):270–86.
- Liebler, Carolyn A., Renuka Bhaskar, and Sonya R. Porter. 2016. "Joining, Leaving, and Staying in the American Indian/Alaska Native Race Category Between 2000 and 2010." *Demography* 53(2):507–40.
- McCall, Leslie. 2005. "The Complexity of Intersectionality." *Signs: Journal of Women in Culture and Society* 30(3):1771–1800.
- Obama, Barack. 2008. "A More Perfect Union." Philadelphia, PA: National Public Radio. Retrieved March 18, 2008.
<http://www.npr.org/templates/story/story.php?storyId=88478467>).
- Omi, Michael and Howard Winant. 1986. *Racial Formation in the United States: From the 1960s to the 1980s*. New York: Routledge & Kegan Paul.
- Penner, Andrew and Aliya Saperstein. 2008. "How Social Status Shapes Race." *Proceedings of the National Academy of Sciences* 105(50):19628-30.
- Phelan, Jo C. and Bruce G. Link. 2015. "Is Racism a Fundamental Cause of Inequalities in Health?" *Annual Review of Sociology* 41(1):311–30.
- Qian, Zhenchao. 1997. "Breaking the Racial Barriers: Variations in Interracial Marriage Between 1980 and 1990." *Demography* 34(2):263.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111.

- Reardon, Sean F. and Ann Owens. 2014. "60 Years After *Brown* : Trends and Consequences of School Segregation." *Annual Review of Sociology* 40(1):199–218.
- Rockquemore, Kerry, David L. Brunnsma, and Joe R. Feagin. 2008. *Beyond Black: Biracial Identity in America*. 2nd ed. Lanham, Md: Rowman & Littlefield Publishers.
- Root, Maria P. P., ed. 1992. *Racially Mixed People in America*. Newbury Park, Calif: Sage Publications.
- Root, Maria P. P., ed. 1996. *The Multiracial Experience: Racial Borders as the New Frontier*. London [Enk] ; Thousand Oaks, Cal: Sage.
- Roth, Wendy D. 2016. "The Multiple Dimensions of Race." *Ethnic and Racial Studies* 39(8):1310–38.
- Saperstein, Aliya, Jessica M. Kizer, and Andrew M. Penner. 2016. "Making the Most of Multiple Measures: Disentangling the Effects of Different Dimensions of Race in Survey Research." *American Behavioral Scientist* 60(4):519–37.
- Schorn, Daniel. 2007. "Transcript Excerpt: Sen. Barack Obama [with Steve Kroft]." CBS News. February 11. Retrieved March 6, 2019 (<https://www.cbsnews.com/news/transcript-excerpt-sen-barack-obama/>).
- Shiao, Jiannbin Lee. 2018. "It Starts Early: Toward a Longitudinal Analysis of Interracial Intimacy." *Sociology of Race and Ethnicity* 4(4):508–26.
- Spencer, Rainier. 1999. *Spurious Issues: Race and Multiracial Identity Politics in the United States*. Boulder, Colo: Westview Press.
- Williams, David R. and Michelle Sternthal. 2010. "Understanding Racial-Ethnic Disparities in Health: Sociological 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 Maria Root, ed. *The Multiracial Experience: Racial Borders as the New Frontier*. Thousand Oaks, California: Sage.

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

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 1 James Moody procedure, applied to In-School responses of Wave 1 Use of " Best Race " item to assign multiple-category In-Home responses of Wave 1 Entirely Consistent responses across 4 panels with separate category for Multiple, Changed, and Other responses (MCO) Mostly Consistent responses across 4 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, Bachelors, HS, Less than HS	Self-rated health after Wave I: Excellent, Very Good, Good, Fair, Poor	Any interracial relationship that crosses white/non-white boundary after Wave I (Yes, No)
Independent variables	Race specification (reference=white) MCO extension		
Shared covariates (Models 2-4)	Age; Female; Married parents Parental educational attainment (highest) Non-English dominant language at home Any religious affiliation; Region		
Outcome-contingent covariates (Models 3-4)	Cumulative GPA Median family income at school Percent black students at school	Educational attainment after Wave I (highest) Ever stopped by police before age 18	Network racial heterogeneity (in adolescence) Proportion same-race students at school Any interracial relationship in adolescence (white/nonwhite) Number of romantic relationships after Wave I Any same-sex relationship
Interaction covariates (Model 4)	Race X Female Race X Parent education	Race X Female Race X Parent education	Race X Female Nonwhite X Network racial heterogeneity

Table 2. Race Inconsistency Scores in Add Health

Racial responses	Wave 1 In-School	Wave 1 In-Home Parents	Wave 1 In-Home	Wave 3 In-Home	Entirely Consistent Specification
Multiple responses	0.573 (N=949)	0.537 (371)	0.531 (754)	0.503 (537)	0 (N=218)
Asian American	0.111	0.092	0.123	0.116	0 (1,044)
Black/Afr. Amer.	0.064	0.078	0.117	0.106	0 (3,389)
Latinx	0.165	0.062	0.050	0.059	0 (3,064)
Native American	0.585	0.139	0.404	0.494	0 (54)
White	0.031	0.057	0.076	0.066	0 (8,879)
Other race	0.583	0.475	0.624	-	0 (23)
Missing	0.071 (N=4,501)	0.123 (11,185)	0.067 (15)	0.145 (4,455)	0 (0)
Changed responses (inconsistent)	-	-	-	-	0.716 (N=2,916)
N with any valid pairs	15,086	8,402	19,572	15,132	19,587
N with one observation only	68,697	1	1,152	47	69,876

Notes: The right-most column combines respondents across all 4 panels with at least one valid pair of racial information across 2+ panels (Total N=19,587), for which the mean inconsistency score is 0.107. In this column, almost all inconsistency scores are zero because only respondents who always responded with the same single-race category or always responded with multiple responses are included in the first seven rows, and the remaining are in the separate category of "Changed responses."

Table 3. Fit Criteria across Race Specifications, with MCO Extensions, in Nested Models Estimating Three Outcomes (Panels A-C)

Panel A: Interracial Relationship History

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
Official	None	+166	+160	+160	+153	+122	+116	+114	+107
Add Health (In-School, Wave 1)	Deg-Inconsistency	+157	+157	+154	+154	+124	+124	+116	+116
	Never-white	+146	+146	+135	+135	+100	+100	+97	+97
	Ever-black	+168	+168	+161	+161	+118	+118	+113	+113
Moody (In-School, Wave 1)	None	+193	+180	+190	+177	+153	+140	+139	+120
	Deg-Inconsistency	+192	+186	+190	+184	+155	+148	+141	+128
	Never-white	+133	+127	+124	+118	+107	+100	+105	+92
	Ever-black	+134	+128	+125	+118	+95	+88	+99	+86
"Best Race" (In-Home, Wave 1)	None	+246	+239	+238	+232	+190	+184	+212	+205
	Deg-Inconsistency	+218	+218	+213	+213	+187	+187	+207	+207
	Never-white	+246	+246	+238	+238	+190	+190	+213	+213
	Ever-black	+243	+243	+236	+236	+192	+192	+214	+214
Entirely Consistent (4 panels)	None	+205	+199	+200	+193	+134	+127	+89	+83
	Deg-Inconsistency	+205	+205	+201	+201	+135	+135	+91	+91
	Never-white	5397	5449	5361	5485	+29	+29	+16	+16
	Ever-black	+19	+19	+10	+10	4563	4720	4476	4652
Mostly Consistent (4 panels)	None	+235	+228	+227	+220	+184	+178	+181	+174
	Never-white	+123	+123	+118	+118	+122	+122	+130	+130
	Ever-black	+171	+171	+159	+159	+127	+127	+139	+139

Notes for all panels (A-C): All models are multilevel, mixed-effect models. Each column includes the full fit-value for each model's best fitting specification(s) and the deviations of other specification from that value. **Preferred fits, including the best fit**, are shaded, in solid boxes, and in bold; these have fit values both less than 7 points larger than the best fitting model and 7+ points smaller than every worse model. **Worse fits** are unformatted and have fit values 7+ points larger than the best fit. **Less-preferred fits** are only shaded and have fit values less than 7 points larger than the best model but not 7+ points smaller than every worse model. **Worst fits** among the five main specifications (without MCO extensions) are in dashed boxes.

Panel B: Educational Attainment

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
Official	None	+20	+13	+15	+11	+9	+2	+12	+21
Add Health (In-School, Wave 1)	Deg-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 1)	None	+20	+6	+15	+3	+18	+5	+25	9503
	Deg-Inconsistency	+9	+2	+5	11376	+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 1)	None	+55	+49	+32	+27	+10	+3	+10	+18
	Deg-Inconsistency	+20	+20	+8	+10	9229	9391	9222	+15
	Never-white	+55	+55	+32	+34	+11	+11	+11	+26
	Ever-black	+57	+57	+34	+36	+11	+11	+12	+27
Entirely Consistent (4 panels)	None	+13	+6	+6	+1	+24	+17	+27	+36
	Deg-Inconsistency	11847	11915	11236	+2	+19	+19	+23	+38
	Never-white	+14	+14	+7	+9	+20	+20	+23	+39
	Ever-black	+10	+10	+3	+5	9229	9391	+5	+20
Mostly Consistent (4 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

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
Official	None	+8	+8	+7	+8	+6	+9	+5	+35
Add Health (In-School, Wave 1)	Deg-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 1)	None	+8	+1	+6	16113	+4	15986	+3	16079
	Deg-Inconsistency	+8	+8	+8	+8	+6	+9	+5	+9
	Never-white	+9	+9	+8	+8	+5	+8	+4	+8
	Ever-black	+8	+8	+8	+8	+5	+8	+5	+8
"Best Race" (In-Home, Wave 1)	None	+11	+11	+7	+7	+3	+6	+15	+46
	Deg-Inconsistency	+9	+15	+7	+14	+4	+14	+17	+54
	Never-white	+11	+18	+7	+14	+4	+14	+17	+54
	Ever-black	+12	+19	+9	+16	+5	+15	+17	+54
Entirely Consistent (4 panels)	None	16082	16149	+2	+2	+2	+5	+7	+38
	Deg-Inconsistency	+1	+8	+4	+11	+4	+14	+9	+46
	Never-white	+1	+8	+3	+10	+3	+13	+8	+45
	Ever-black	16082	+7	+2	+9	+3	+13	+8	+45
Mostly Consistent (4 panels)	None	+1	+1	15971	+1	15820	+3	15806	+31
	Never-white	16082	+7	15971	+7	15820	+10	15806	+37
	Ever-black	+3	+9	+2	+9	+2	+12	+2	+39

Table 4. Select Parameter Estimates of Interracial Relationships after Wave 1 in Log Odds, Comparing Race Specifications with Best and Worst Fits among the Intersectionality Models (Model 4)

Variables	"Best Race" Specification	Entirely Consistent Specification, with Ever-Black Extension
Race (reference = white)		
Multiple/Changed/Other (MCO)	0.33	1.08***
Network racial heterogeneity (interracial friendships)	0.91**	0.54*
Nonwhite x Network heterogeneity	1.09*	2.45*
MCO x Ever Black (Extension)		-2.12***
Intercept (fixed effects equation)	-0.97***	-0.88***
Across-schools variance (random intercept)	0.048**	0.013**
Individual-level n	5,171	5,171
School-level n	113	113
AIC	4687	4476
BIC	4858	4652

Notes: Sample comes from Add 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 X female. *** p<0.01, ** p<0.05, * p<0.1

Table 5. Selected Parameter Estimates of Educational Attainment after Wave 1 in Log Odds, Comparing Race Specifications with Best and Worst Fits among the Outcome-contingent Models (Model 3)

Variables	Entirely Consistent Specification	"Best Race" Specification, w/Inconsistency Extension
Race (reference = white)		
Multiple/Changed/Other (MCO)	-0.03	0.45
Degree of Inconsistency (Extension)		-0.46***
Thresholds		
Cut 1 (HS degree vs. Less than HS)	-2.97***	-3.02***
Cut 2 (Bachelors degree vs. HS)	2.97***	2.96***
Cut 3 (Postgraduate degree vs. Bachelors)	0.10***	5.31***
Across-schools variance (random intercept)	0.10***	0.10***
Individual-level n	6,305	6,305
School-level n	113	113
Test of parallel lines chi square	38.51***	42.89***
AIC	9253	9229
BIC	9408	9391

Notes: Sample comes from Add 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 GPA, median family income at school, and percent black students at school. ***

p<0.01, ** p<0.05, * p<0.10

Table 6. Selected Parameter Estimates of Self-rated Health after Wave 1 in Log Odds, Comparing Race Specifications with Best and Worst Fits among the Intersectionality Models (Model 4)

Variables	“Best Race” Specification	Mostly Consistent Specification
Race (reference = white)		
Multiple/Changed/Other (MCO)	0.13	0.23
Native American	0.07	-1.26***
Thresholds		
Cut 1 (Health rating Fair vs. Poor)	-5.08***	-5.10***
Cut 2 (Health rating Good vs. Poor)	-2.76***	-2.77***
Cut 3 (Health rating Very Good vs. Good)	-0.72***	-0.73***
Cut 4 (Health rating Excellent vs. Very Good)	1.14***	1.13***
Across-schools variance (random intercept)	0.03***	0.03***
Individual-level n	6,264	6,264
School-level n	113	113
Test of parallel lines chi square	17.12***	17.29***
AIC	15822	15807
BIC	16125	16110

Notes: Sample comes from Add 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, educational attainment after Wave I, ever stopped by police before age 18, race X female, and race X parental educational attainment. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 1. Inconsistency Scores for Simulated Respondents

Respondent	Wave 1	Wave 1	Wave 1	Wave 3	Consistent Pairs	Valid Pairs	Inconsistency Score
	In-School	In-Home Parents	In-Home	In-Home			
R1	Black	Multiple	Asian	Latinx	0	6	1.00
R2	Asian	Multiple	-	Native	0	3	1.00
R3	Native	Native	Native	Native	6	6	0.00
R4	Multiple	Multiple	-	-	1	1	0.00
R5	Latinx	Multiple	Multiple	Asian	1	6	0.83
R6	Black	Black	Multiple	Multiple	2	6	0.67
R7	Asian	-	Multiple	Multiple	1	3	0.67
R8	Latinx	Multiple	Multiple	Multiple	3	6	0.50
R9	Other	Other	-	Black*	1	1	0.00
R10	Other	Multiple	White	Black*	0	3	0.00
R11	Other	Multiple	White	Black*	1	3	0.67

Notes: Missing observations reduce the number of valid pairs and constrain the range of values that inconsistency can take: Four observations allow for 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). *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).

Appendix Table 2. Responses to Racial Identification Items across Add Health Panels (N=91,040 unique respondents)

Racial responses	Wave 1 In-School	Wave 1 In-Home Parents	Wave 1 In-Home	Wave 3 In-Home	Valid Race information in any of 4 panels
Multiple	6.01%	4.42%	3.87%	3.56%	6.68%
Asian American	4.78%	6.07%	6.34%	6.60%	(Remaining responses require decision rules for allocating inconsistent responses)
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%	-	
Valid N	83,783	8,403	20,724	15,179	89,463
Missing	7,257	82,637	70,316	75,861	1,577

Notes: Total N=91,040 unique respondents across these 4 panels plus Wave IV, which does not include racial identification items. The Valid N is the number of respondents in a panel with non-missing 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 only include 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 for its race item.

Appendix Table 3. Group-specific Means for Educational Attainment, Self-reported Health, and Interracial-Relationship History in Add Health, along with Select Independent Variables

	Racial identification (Entirely-consistent race specification)						
	Analytic sample	White	Multiple, Changed, or Other	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 1 (0-3)	1.41	1.47	1.30	1.37	1.29	1.65	0.86
Parental educational attainment in wave 1 (0-3)	1.43	1.53	1.39	1.36	0.96	1.81	1.36
Age in wave 1	14.91	14.87	14.73	14.93	15.20	15.28	15.36
Female	0.53	0.53	0.52	0.58	0.50	0.49	0.43
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 1 (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 1 (1-5)	3.75	3.80	3.68	3.73	3.68	3.74	3.07
Parental educational attainment in wave 1 (0-3)	1.43	1.54	1.39	1.36	0.95	1.81	1.36
Age in wave 1	14.92	14.87	14.73	14.93	15.20	15.28	15.36
Female	0.53	0.53	0.52	0.58	0.50	0.49	0.43
Educational attainment after wave 1 (0-3)	1.41	1.47	1.30	1.36	1.29	1.66	0.86
Ever stopped by police before age 18	0.08	0.08	0.09	0.07	0.11	0.07	0.14
N (Interracial-relationship models)	5,171	2,631	775	832	650	260	23
Interracial relationship after wave 1	0.30	0.25	0.63	0.12	0.35	0.28	0.13
Network racial heterogeneity (send network, 0.0-0.5)	0.15	0.17	0.25	0.04	0.11	0.08	0.11
Age in wave 1	14.87	14.78	14.65	14.86	15.34	15.33	14.43
Female	0.56	0.55	0.56	0.63	0.56	0.47	0.57
Proportion same-race at school in wave 1	0.55	0.75	0.12	0.45	0.47	0.23	0.22
Interracial relationship in wave 1	0.12	0.07	0.36	0.04	0.15	0.08	0.00

Notes: Racial 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.