

The Multiracial Complication: The 2020 Census and the Fictitious Multiracial Boom

Paul Starr, Christina Pao

Princeton University

Abstract: The Census Bureau set off reports of a “multiracial boom” when it announced that, according to the 2020 census, multiracial people accounted for 10.2 percent of the U.S. population. Only the year before, the bureau’s American Community Survey had estimated their share as 3.4 percent. We provide evidence that the multiracial boom was largely a statistical illusion resulting from methodological changes that confounded ancestry with identity and mistakenly equated national origin with race. Under a new algorithm, respondents were auto-recoded as multiracial if, after marking a single race, they listed an “origin” that the algorithm did not recognize as falling within that race. However, origins and identity are not the same; confounding the two did not improve racial statistics. The fictitious multiracial boom highlights the power of official statistics in framing public and social-science understanding and the need to keep ancestry and identity distinct in both theory and empirical practice.

Keywords: mixed race; racial classification; racial statistics; ethnicity; identity; origins

Reproducibility Package: R code for replication is available on the Open Science Framework (OSF), https://osf.io/8ebup/?view_only=67a953b996684d128c9384d4841ed1c5. Data are available from IPUMS USA (Ruggles et al. 2024): <https://usa.ipums.org/usa/index.shtml>

Citation: Starr, Paul, and Christina Pao. 2024. “The Multiracial Complication: The 2020 Census and the Fictitious Multiracial Boom” *Sociological Science* 11: 1107-1123.

Received: September 17, 2024

Accepted: October 21, 2024

Published: December 3, 2024

Editor(s): Ari Adut, Michael Rosenfeld

DOI: 10.15195/v11.a40

Copyright: © 2024 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

THE rapid growth in multiracial Americans reported in the 2020 census, up from 2.9 percent in 2010 to 10.2 percent of the U.S. population 10 years later, led to a rash of news stories about a “multiracial boom” in the United States (Bahrampour and Mellnik 2021; Galvan and Schneider 2021; Jones et al. 2021). At the time, neither social scientists nor journalists subjected the new census multiracial count to serious scrutiny, although they might have noticed how large and sudden the increase was in 2020. The bureau’s American Community Survey (ACS) had put multiracial individuals at a five-year average of only 3.3 percent of the population for the years 2015–2019, with only slightly more, 3.4 percent, in 2019 itself. In other words, the multiracial population, when measured by a consistent method, was growing, but only slowly. Could their share of the population have tripled to 10.2 percent in one year as measured in the 2020 decennial census?

The Census Bureau introduced two major changes in 2020 affecting its racial data: a new question design and a new reclassification algorithm. For the first time, the 2020 census asked respondents who checked off white or black on the race question to also list their “origins.” The bureau then used information on up to six listed origins to recode individuals’ race if any one of the origins, according to a computerized algorithm, did not fit into the single-race box or boxes they had checked off. These changes led the bureau to reclassify as multiracial many

people who had checked off only one race. For example, a respondent who marked only white but listed Argentina among his or her origins would be auto-recoded as multiracial because Argentina was not coded as a white, European origin (this individual was instead classified as both “white” and “some other race,” as were others who checked white and listed a Latin American national origin). People who marked “black” and listed the Dominican Republic as an origin were also recoded as multiracial.

In a background post on the 2020 census (Marks and Ríos-Vargas 2021), the bureau acknowledged that it had adopted new methods for classifying people by race that boosted the multiracial count, but it did not clarify what the counts would have been without the changes in methods. Since 2020, it has also applied the new methods to the ACS, but it has not published any studies to determine how many people who have checked off only one race on either the decennial census or ACS have been auto-recoded as multiracial based on a listed origin.

In addition to leaving unanswered how much of the increased multiracial count is a statistical artifact, the changes in census methods raise two important questions: Were the new reclassification methods “improvements,” as the bureau contends? And what does the entire episode tell us about the relation of identity and ancestry and the role of official statistics in shaping public understanding of race?

After providing evidence about how much of the multiracial increase is an artifact of changed methods, we argue that the reclassifications have been misconceived: first, because they have confounded identity with ancestry, and, second, because they have mistakenly equated national origin with race. Identity and origins are not the same; people may write in an “origin,” perhaps a distant ancestry, even a minute one according to a DNA test, which is not part of their identity (Roth and Ivemark 2018; Waters 1990). “Origins” may be understood as ancestry in national as well as racial or ethnic terms, but a national origin does not necessarily indicate a race. Not everyone with origins in South Africa is black; not everyone with origins in France is white. Multiple origins do not necessarily multiply racial identities or indicate that people think of themselves, much less are regarded by others, as mixed race.

The Census Bureau’s changed methods raise several serious problems for both social scientists using census data and for policies relating to legislative redistricting and other legal issues. As recent headlines show, these methods also shape media representations and public understanding of America’s changing racial landscape. In addition, the new census practices are inconsistent with the commitment to racial self-identification to which the bureau itself has long been committed. We argue that the Census should return to its tradition of respecting racial and ethnic self-identification and it should keep that commitment in mind as it addresses the new challenges of a government-wide change in race-ethnic data (OMB Interagency Technical Working Group on Race & Ethnicity 2024).

How the U.S. Government Frames Official Race-Ethnic Statistics

With the full adoption of a mail-in census in 1970, the Census Bureau began relying on respondents to identify their own and other household members’ race, instead of

having census enumerators make racial assignments, as the government had done previously. In 1994, the bureau declared, “The concept of race as used by the Census Bureau reflects self-identification . . . The data for race represent self-classification by people *according to the race with which they most closely identify*” (quoted in Perlmann and Waters 2002:10, italics added). However, by recoding individuals in 2020, the bureau reverted to its earlier practice of making racial assignments, though this time using an algorithm.

The U.S. Census Bureau produces race-ethnic statistics in conformity with statutory requirements and the standards that the Office of Management and Budget (OMB) sets for all federal data collection. Aiming to achieve consistency in racial classifications and data, OMB originally issued uniform standards in 1977 in Statistical Policy Directive 15, which it revised in 1997 (and again in 2024). Within those standards, the Census Bureau still has considerable flexibility in the methods it adopts for putting the standards into practice. However, similar to other agencies, it needs to keep one eye on Congress and the White House and another on the interest groups that follow its decisions most closely. In the case of racial and ethnic statistics, those interest groups are chiefly racial and ethnic advocacy and civil rights organizations. Preemptively satisfying those groups, who might otherwise protest agency practices and take their case to politicians, is one way for census officials to maintain as much bureaucratic autonomy as possible.

These institutional and political realities entangle the Census Bureau and the social sciences in a contradictory relationship. The government not only collects racial statistics for political and bureaucratic reasons but also explicitly disavows any claim that these are scientific. Announcing Directive 15 with its definition of approved racial categories, OMB declared in 1977 in the preamble that “these classifications should not be interpreted as being scientific or anthropological in nature” (Nobles 2000:179). Nor do political choices end with the designation of official races and ethnicities. Despite their seemingly technical nature, the practices the census adopts in enumerating, tabulating, and projecting populations have political implications of which advocacy groups and census officials are fully aware. However, in a modern equivalent of medieval alchemy, the agency is supposed to generate scientific numbers from a process that is shaped politically from the beginning (Prewitt 2018).

The adoption and subsequent history of the multiracial category illustrate how federal data depend on political influences. The original Directive 15 allowed individuals to pick one single-race category alone. However, by the 1990s, a movement had emerged calling for federal recognition of a multiracial category. This movement involved only a few thousand people, led mostly by white women who were married to black men and wanted to identify their children not with one race alone but with both (Williams 2008). Civil rights groups initially opposed a multiracial option for fear it would “drain” people from single-race categories and undermine civil rights enforcement, but the multiracial movement got the attention and support of Rep. Newt Gingrich and other Republicans (Williams 2008). Eager to appear racially sensitive but not wishing to offend civil rights constituencies, federal officials came up with the solution: Give individuals the chance to “mark one or more” on the census and then count all white-minority individuals as minor-

ity for legal purposes. The initial impact of “mark one or more” seemed minimal because only about 2 percent of the population marked more than one race on the 2000 census (Williams 2008, 2017). Although the census did not list “multiracial” as a category on the census form, the term nonetheless became the one that the bureau itself used for the “two or more” category, that is, people who marked more than one race.

In recent years, ethnic advocacy groups have pressed federal statistical authorities to gather data not only on the broad official racial categories but also on “detailed categories” within them. The response to that pressure led to the changed question design on the 2020 census, but the bureau decided on its own not only to change the question design and use the term “origins” but also to introduce a new recoding algorithm raising the multiracial count. Without explicitly defending the algorithmic recoding, the bureau explained all the changes it made to racial and ethnic data as an effort “to reflect more fully and accurately the complex details of how people identify their race and ethnicity” (Marks and Ríos-Vargas 2021). It evidently decided to override single-race identifications in the belief that the “complex details” individuals listed about origins were a more accurate measure of racial and ethnic identity.

In 2005, the political scientist Kenneth Prewitt, who had been a director of the Census Bureau during the introduction of “mark one or more,” wrote that there had been “scattered grumbling” that allocating white-minority individuals to the minority race “echoed the one-drop rule of the nation’s racist past,” but it didn’t become a public issue because the 2000 census had so few multiple-race respondents. Prewitt continued: “Until the multiple-race population grows to a significant percent of the population, *10 percent or so*, ‘mark one or more’ will remain a change that points to the future without disrupting the present” (Prewitt 2005:200, italics added).

In the 2020 census, the multiracial population hit Prewitt’s 10 percent threshold far earlier than he expected. Moreover, the Census Bureau had given the multiracial share a sudden boost by recoding people as multiracial who checked off white or black alone but listed at least one origin coded as being different from their single-race identification. Yet, when the census published its 2020 census results, hardly anyone seemed to notice there might be a problem.

Why the Multiracial Count Jumped in 2020

Hypothetically, the 2020 increase in the multiracial count could have had several causes. First, there could have been a demographic increase from more multiracial births or in-migration of multiracial individuals in the United States. Births from multiracial or multiethnic parents are indeed on the rise (Alba 2020; Livingston 2017). Second, changes in self-identification could have produced the higher multiracial count in the 2020 census. Liebler et al. (2017) have shown that some respondents change their race and ethnicity responses from one census to the next, creating a phenomenon they term “churning races.”

If, however, either demographic increase or changed identification caused the increase in the multiracial count, the increase should have occurred more among

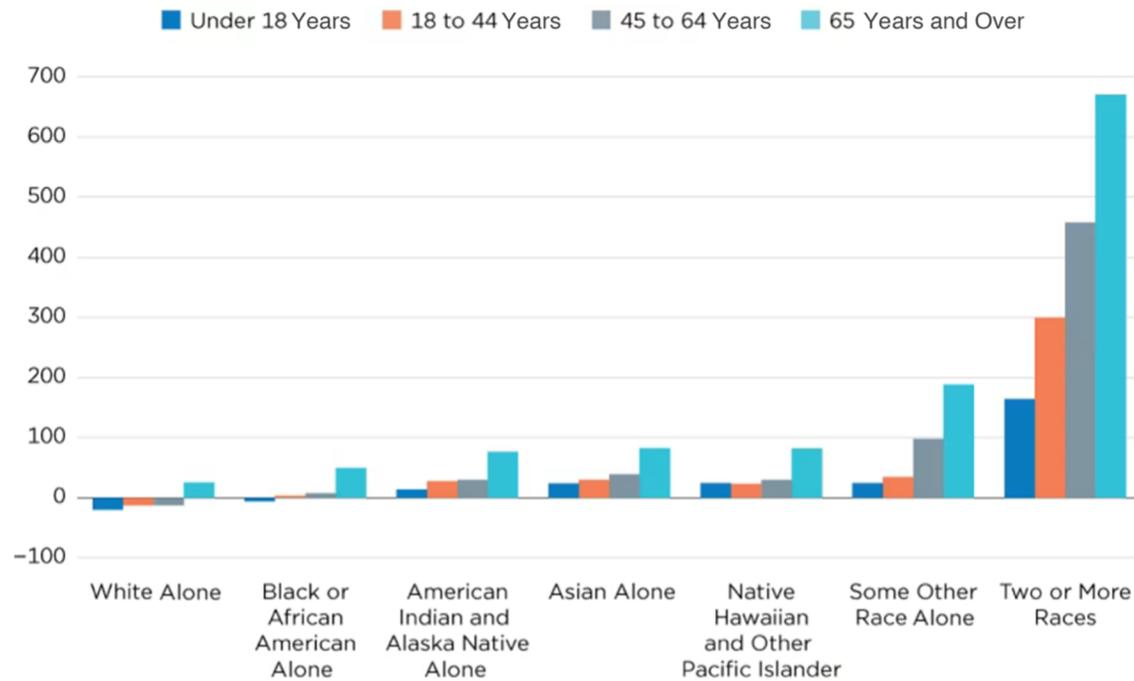


Figure 1: Percentage change in selected age categories by race, 2010–2020.

the young than among the old. This was, in fact, what happened between 2000 and 2010, when the multiracial count rose by more than 50 percent, taking into account a belated discovery of an overcount in the 2000 census (Jones and Bullock 2013). Detailed analyses indicate that the increase was higher at younger ages (see Currie and Schwandt 2016, Figure 4a).

However, this was not the pattern of change from 2010 to 2020. As shown in Figure 1, the older respondents (especially those more than 65) were more likely to shift to a multiracial identification (the “two or more” category) than younger respondents (reprinted from Rico, Jacobs, and Coritz 2023, Figure 3). The census multiracial count rose by almost 700 percent among those aged 65 or more years of age and more than 450 percent among those 45–64 but by smaller percentages among younger groups (Rico et al. 2023). It seems implausible that older people were adopting a multiracial identity for the first time at a rate many times greater than younger people. The more probable explanation lies in the changed census methods that overrode respondents’ single-race self-identification according to their listed “origins.”

We hypothesize that the new question design and recoding algorithm were largely responsible for the multiracial increase. The example of the individual from Argentina who checked only “white” but was coded as multiracial is typical of what happened with both whites and blacks with any Latin American heritage. The census algorithm did not recognize Latin American national origins or ethnicities as a race. When anyone who checked off “white” or “black” alone indicated a Latin American origin, they were reclassified as multiracial—that is, as white or black

plus “some other race.” The number of Hispanics classified as multiracial jumped from 3.0 million in 2010 to 20.3 million in 2020.

The reclassification of whites as multiracial was not limited to self-identified whites with Latin American origins. Among non-Hispanics, the biggest jump in the multiracial population was in the “white and American Indian” category—an increase of 2.3 million (National Academies of Sciences 2023:275). If someone who marked white and no other race listed “Cherokee” as one of several origins along with, say, English and German, they were recoded as multiracial. Perhaps not coincidentally, Oklahoma was the state with the greatest percentage increase in the multiracial population in the 2020 census; Alaska and Arkansas were next. A former census employee points out another example of recoding self-identified whites: “[T]hose who wrote South African in the white write-in line were recoded as both Black and white because the bureau considers South African to be indicative of a Black place of origin, despite the fact that most South Africans in the U.S. are white” (Lowe 2023). Whites with any family origins from outside Europe were now counted as multiracial apparently because the census algorithm was unfamiliar with the history of colonialism.

The 2020 Census Bureau also recoded people as Hispanic even if they said they were *not* Hispanic on the “ethnicity” question but listed a Latin American origin on the race question (National Academies of Sciences 2023:269). In doing so, the bureau erased information about ethnic attrition—that is, the intergenerational decline of Hispanic self-identification among people who have some Latin American ancestry but who, for example, may not speak Spanish or retain Hispanic cultural connections (Duncan and Trejo 2011). In this respect as in others, the 2020 census conflated ancestry and identity, indeed subordinated self-identification to ancestry. The algorithm biased the census results against the recognition of assimilation into the white non-Hispanic population.

According to an OMB working group on race and ethnicity standards, “Any revision to the categories should provide for a crosswalk at the time of adoption between the old and the new categories so that historical data series can be statistically adjusted and comparisons can be made” (OMB Interagency Technical Working Group on Race & Ethnicity 2024). The Census Bureau did issue a “crosswalk” for racial codes in the 2010 and 2020 censuses. However, it did not publish tables showing what the racial counts would have been under the previous procedures. Nor has it released public data from the 2020 census with full race responses of respondents that would make it possible to determine who was algorithmically recoded as multiracial. As a result, researchers have to use indirect methods to analyze the effect of recoding.

Data and Methods

Here, we follow the approach of Lowe et al. (2024) and use one-year microdata from the 2019 and 2021 ACS available from the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. 2024). Similar to Lowe et al., we use the 2019 and 2021 ACS data instead of the 2020 ACS data for comparability due to the unique

13 What is this person's ancestry or ethnic origin?

(For example: Italian, Jamaican, African Am., Cambodian, Cape Verdean, Norwegian, Dominican, French Canadian, Haitian, Korean, Lebanese, Polish, Nigerian, Mexican, Taiwanese, Ukrainian, and so on.)

Figure 2: American Community Survey ancestry question.

data collection challenges in 2020 with the onset of the COVID-19 pandemic (Villa Ross, Shin, and Marlay 2021).

Using the 2019 and 2021 ACS has two key advantages: (1) it makes possible a direct pre- and post-2020 comparison and (2) the ACS, unlike the decennial census, provides ancestry write-in information. Using 2019 ACS as the reference point instead of the 2010 decennial census mitigates the impact of both natural increase and cultural change in multiracial identification; we should observe less change at a population level between 2019 and 2021, making the ACS one-year estimates particularly useful in isolating the effects of methodological changes. The ACS ancestry question looks as shown in Figure 2. We use the two ancestry responses coded by the Census as a proxy for the potential write-in responses that individuals gave when listing “origins” in response to the race question. The data are publicly available from IPUMS, and all code to produce the analyses is available online.

We provide three case studies of change by using descriptive statistics from the ACS: (1) shifts of Hispanic respondents from “white alone” to “white and some other race,” (2) the increase in “black and white” responses from colonized African countries, and (3) the increase in “native in combination.” Although other work in progress shows recoding affected every race group’s composition (Arias et al. 2024; Lowe et al. 2024), we pick these three specific case studies to show how recoding generated a higher multiracial count.

Results

Case Study 1: White Hispanic Respondents

In the 2019 and 2021 ACS as well as the 2020 decennial census, the census continued to use a two-question format for ethnicity and race: first, an “ethnicity” question asking respondents whether they were of Hispanic or Latino origin or descent and, second, a race question without a Hispanic response option. Since the 2020 census algorithm did not recognize Latin American origins as falling in any of its racial categories, it treated them as “some other race.” We anticipate that many Hispanics who answered white on the race question, while listing a Latin American origin as one of their origins, were recoded from “white alone” to “white and some other race,” and this is shown in Figure 3. The white-alone population among Hispanics

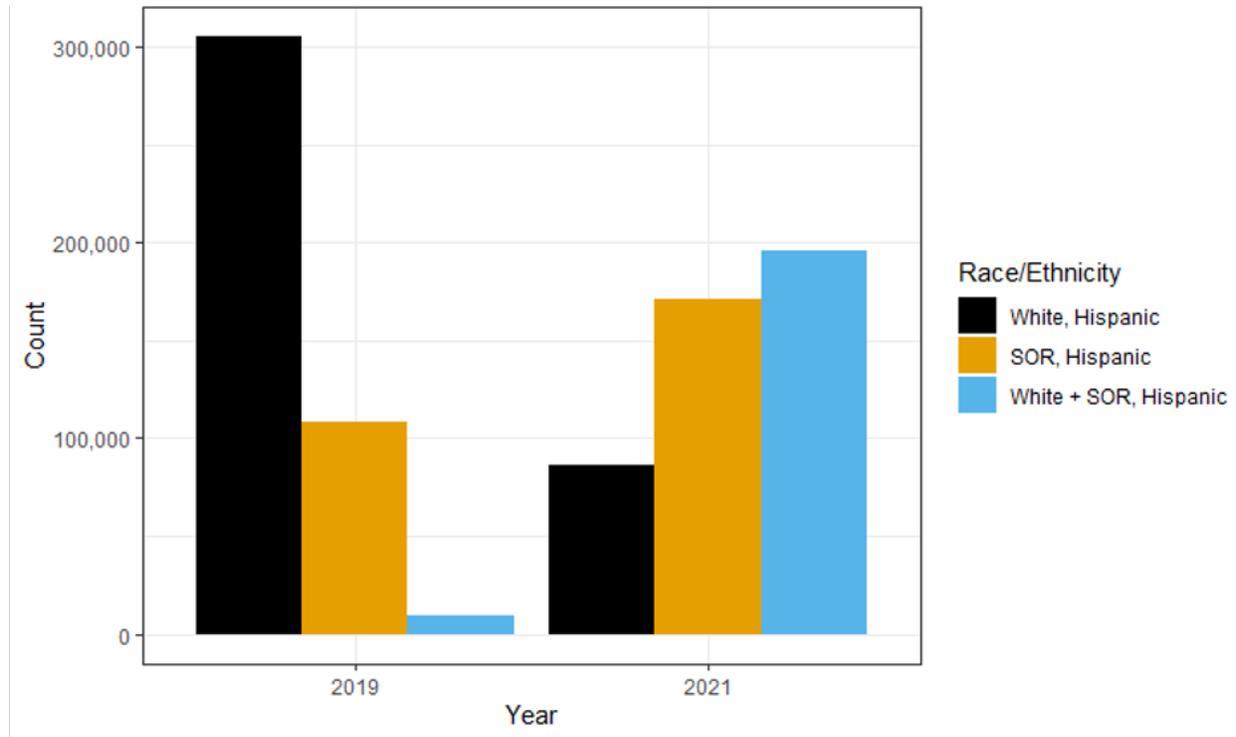


Figure 3: Changes in white Hispanic responses in 2019 and 2021 ACS.

was cut by around two thirds between 2019 and 2021, whereas the “white and some other race” jumped by approximately a factor of eight.

This shift is highly consequential. Much of the “shrinking” white population seen in headlines has come from the algorithmic reclassification of the “white-alone” Hispanic population as multiracial.

Case Study 2: Increased “Black and White” Responses from Specific African Countries of Origin

The ancestry data from the ACS on people with origins in Africa illuminate how the racial reclassification algorithm may affect racial composition estimates of immigrant groups. Figure 4 shows the differences in frequencies of “white alone” and “black and white” from the 2019 and 2021 ACS for three countries—Kenya, South Africa, and Zimbabwe—which have relatively large white populations to this day, stemming from European colonization. Although the absolute size of the white African population in the United States is relatively small, we nonetheless see evidence again of the movement from “white alone” to a multiracial, “white in combination” group, in this case “black and white.”

Under the new reclassification algorithm, which deemed all African origins to be indicative of blackness, respondents who marked white and listed an African country as an origin were recoded as both “white” and “black.” For example, if Elon Musk filled out the 2020 census and entered “South Africa” (where he was

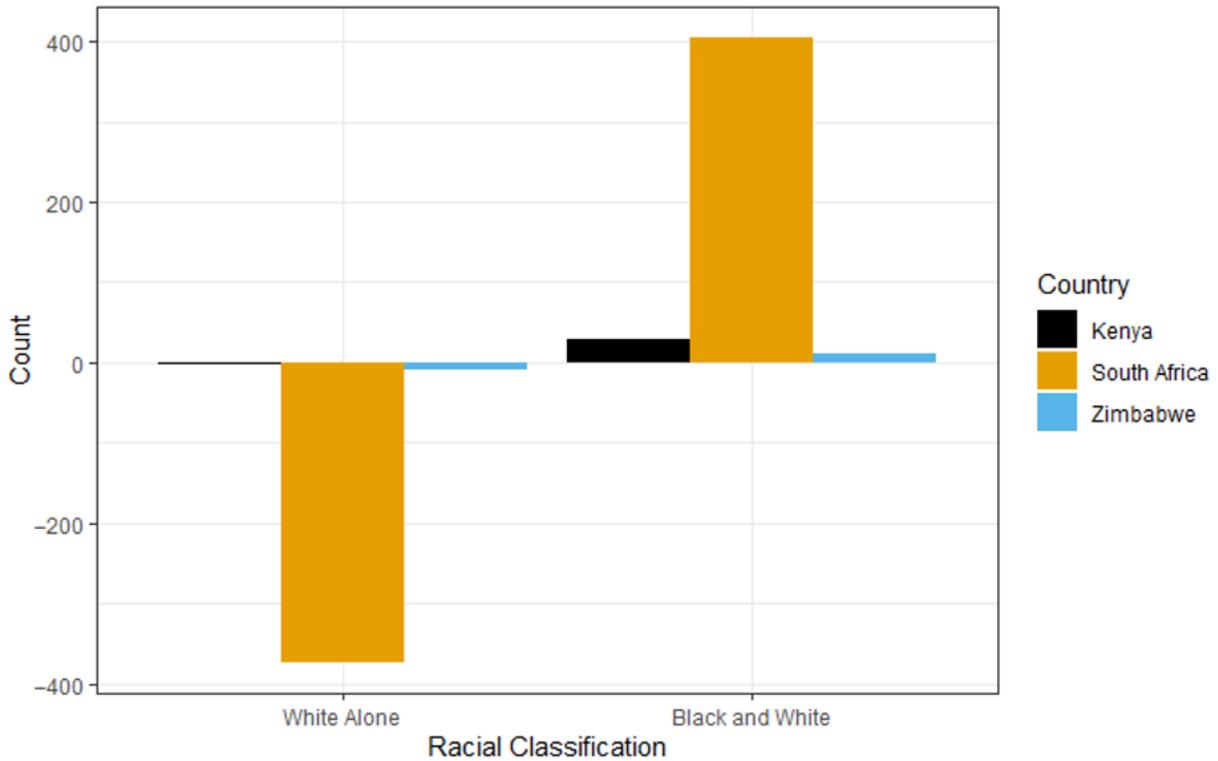


Figure 4: Change in “white alone” and “black and white” frequencies from 2019 and 2021 ACS.

born) as an origin, he would have been recorded as white and black and therefore multiracial. Figure 4 shows the ACS data for those listing ancestries in Kenya, South Africa, and Zimbabwe. Although the absolute numbers in the ACS are small, the percentage changes are not. The measured white-alone populations collapsed, and the multiracial (white in combination) population soared.

Case Study 3: American Indian and Alaska Native Self-Identification

Figure 5 shows the frequencies of Native American Indian and Alaska Native (AIAN) identification in 2019 and 2021 ACS data via the race question. Although the consequences of these changes for Native identification have been discussed in depth in other working articles (Arias et al. 2024), we add further details using ancestry data.

Between 2019 and 2021, the “white and Native” group surpassed “Native alone” identification. If respondents wrote in a Native tribal origin under white or black, the reclassification algorithm changed that individual’s reported race to multiracial, even if that does not align with the respondent’s stated racial self-identification.

This pattern becomes even more apparent when we incorporate the ACS ancestry write-in information. Figure 6 shows that the largest increase in “Native and white” responses is among those who did *not* include a Native ancestry as one of their two ancestry write-ins. In 2019, the largest Native racial subgroup consisted

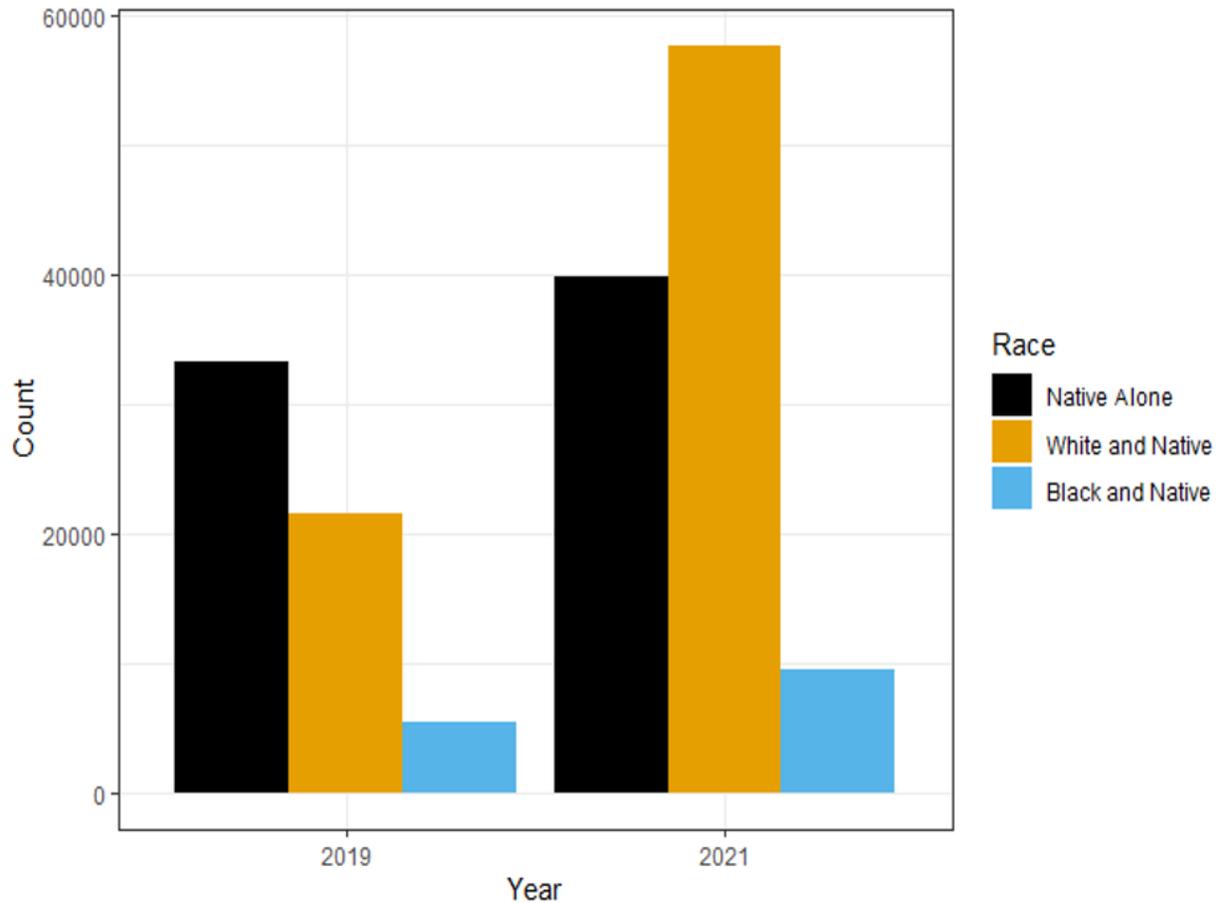


Figure 5: Frequency of American Indian and Alaska Native Reporting in 2019 and 2021 ACS.

of those who also included Native ancestry; but, in 2021, the largest subgroup consisted of those who were coded as “Native and white” but did *not* list even one Native ancestry.

Discussion

The idea that the United States is experiencing a dramatic increase in multiracial identity has become received wisdom. For example, when the *New York Times* reported in August 2024 that Vice President Kamala Harris’s identity as both black and Asian reflected the “jump in multiracial Americans,” it reported census data on the counts of multiracial Americans from 2000 to 2023 as though the Census Bureau had measured multiracial identity the same way over that period (Bosman, Gebeloff, and Singer 2024). No doubt there has been an upward trend, but it has been a much more slowly growing trend than recent census data and the *Times* suggest.

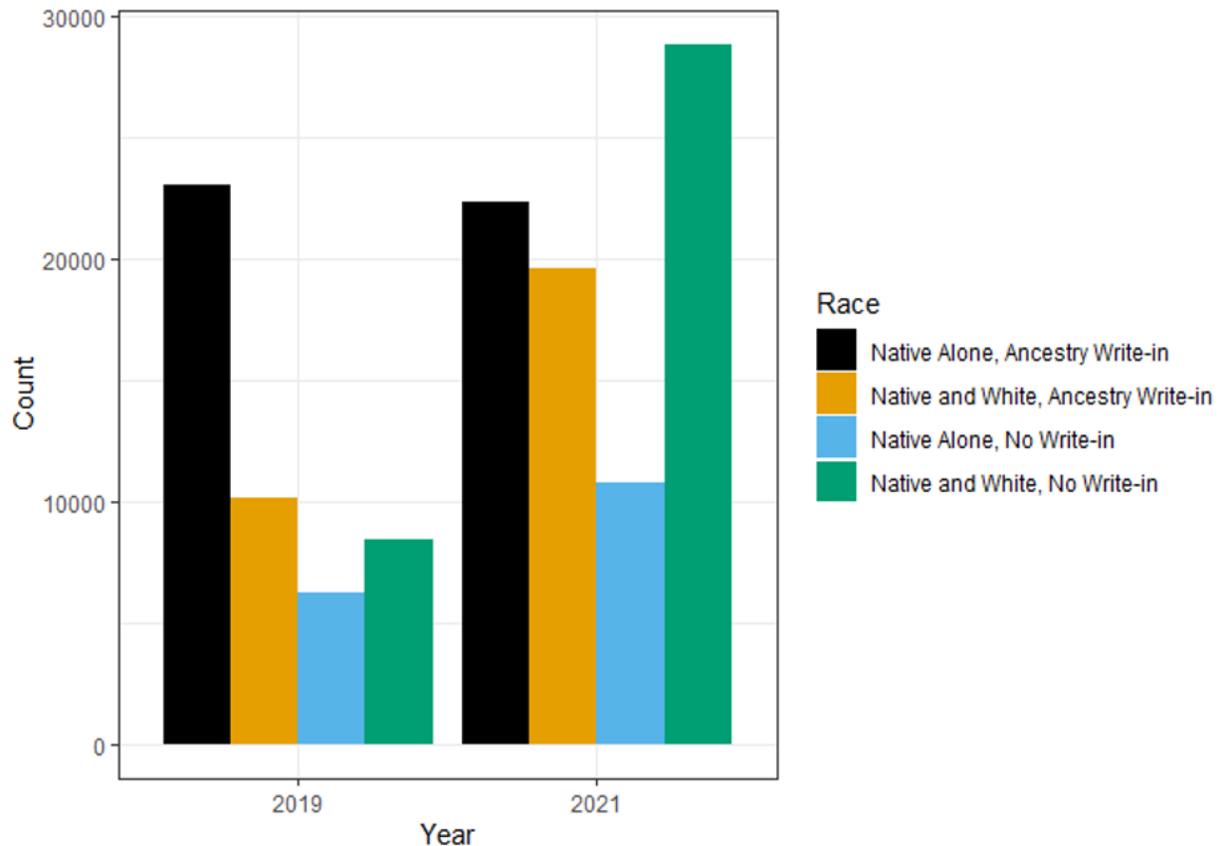


Figure 6: Frequency of American Indian and Alaska Native Reporting by ancestry write-in responses in 2019 and 2021 ACS.

In its main press release on racial data for the 2020 census, the Census Bureau announced that the white-alone population had “decreased by 8.6 percent since 2010,” whereas the multiracial population had increased by 276 percent (Jones et al. 2021). However, the census was not making an apples-to-apples comparison between 2010 and 2010; its racial recoding had inflated the multiracial count. The *Washington Post* declared, “The report marks the first time the absolute number of people who identify as White alone has shrunk since a census started being taken in 1790” (Bahrapour and Mellnik 2021). However, the decline in the white-alone count also resulted primarily from the bureau’s recoding. These and other reports have encouraged the public to believe that white people are in precipitous decline.

The recoding practices adopted by Census Bureau since 2020 have created problems for social scientists who remain mostly unaware of the methodological change in racial statistics. Census data are widely used to weight survey results, but survey researchers are not typically generating racial data the same way as the census is—that is, by sweeping single-race identifiers into the multiracial category based on their listed origins. Data users may also not be aware of the changing composition of population subgroups under different methodological regimes. Algorithmic

overwriting of racial self-identification can inflate estimates of subpopulations, such as Native groups, which can in turn reduce the comparability across years and create conceptual confusion between populations defined by racial self-identification and algorithmically recoded race.

The enlarged multiracial count also exacerbates a fundamental difficulty that social analysts have had since the census adopted the “two or more” category: how to analyze or interpret a category that is a mix of radically different mixes. A category that includes both Asian-whites and multiple-race blacks is not a coherent category for most analytical purposes (Alba 2020; Lee and Bean 2004); it is even less coherent if people who identify as white are shifted into it because they list as one of their “origins” a place or group that does not match the census list for white origins. Since 2000, many social-science studies and surveys have just ignored the roughly 2–3 percent in the multiracial category, but it is hard to ignore or to make sense of the double-digit multiracial percentages reported since 2020.

Extensive Census Bureau racial recoding also raises questions about the use of census results for civil rights enforcement. For legal purposes, the federal government has since 2000 allocated individuals identifying with “one minority race and white” to the minority race (Williams 2008). Under its algorithmic recoding system, however, many individuals who marked *only* white will have first been reassigned to the multiracial category because of listed “origins” and then, for legal purposes, reallocated to a minority race. (To take Elon Musk again as an example, he would have been reallocated to the black population if he listed South Africa as an origin.) This double switch risks creating legally troublesome questions for legislative redistricting and the benchmarks set for civil rights enforcement by agencies such as the Equal Employment Opportunity Commission.

The new combined race-ethnicity question adopted by OMB in 2024 presents a host of complications. The earlier two-question format assumed that Hispanic heritage is not racial and that Hispanics could be “of any race.” However as census tests in 1996, 2010, and 2015 showed, a combined race-ethnicity question including a Hispanic option and a Middle Eastern and North African (MENA) option reduced nonresponse on the race question and nearly eliminated the “some other race” responses. These effects were the basis for the bureau’s conclusion that the combined question was “optimal” (Mathews et al. 2017).

However, a combined question has other effects. Tests showed that the combined question produced a significantly lower white-alone count, a slightly lower black-alone count, and an ambiguous effect on the “two or more” category. The latter effect is ambiguous because when race and ethnicity are combined into one question, it is unclear who should be included in the “two or more” category or even what “two or more” means. Does it refer only to two or more of the pre-existing race categories or does it also include those categories plus Hispanic and MENA? For example, should the two-or-more category include respondents who choose both Hispanic and white? Or should Hispanic identity still be treated separately as an “ethnicity,” limiting the “two or more” category to respondents who select two or more of the previously designated official races? The same ambiguity applies to MENA. Is “two or more” still “two or more races” or now two or more groups regardless of whether they were previously conceptualized as racial or ethnic?

The name for people with more than one ethnic or racial heritage has long been a dilemma not only in the United States but also elsewhere. The combined question may now require a shift from “multiracial” to another term, perhaps “mixed” or “multiethnic” to avoid the controversy that the use of “multiracial” might create.

As the census tests indicated, the change in Hispanic responses on the combined question drives a fall in white- and black-alone counts. On a combined question, according to the Census Bureau’s 2015 National Content Test, about 70 percent of Hispanics chose Hispanic alone, reducing to 19 percent the share who identified as white alone (compared to about half in the two-question format) (Mathews et al. 2017). The combined question drove down to 1.3 percent the share of Hispanics who identified as black alone. Because 70 percent chose Hispanic alone, the remaining 30 percent would go into the two-or-more category if that category is defined as including all Hispanics marking a second race or ethnic category. However, if the two-or-more category is limited to those identifying with two traditional race categories, only 5.5 percent of Hispanics would fall in that category (Mathews et al. 2017, Table 6).

In its 2015 study, the census tested the combined question with a new instruction that called for respondents to “mark all that apply” instead of “mark one or more” and found, unsurprisingly, that the “all that apply” instruction increased the share of respondents who checked off more than one category. In that study, the share of total respondents who reported “two or more race/ethnicity groups” rose from 12.1 percent to 13.1 percent with the new instruction as opposed to the old one (here the census appears to have counted individuals marking any two race or ethnicity groups, including, for example, people checking off both “Hispanic” and “white”) (Mathews et al. 2017, Table 29).

The Census Bureau could adopt three measures together: the combined race-ethnicity question, the instruction to “mark all that apply,” and the recoding of individuals based on listed origins as it has done since the 2020 census. If the bureau did all three, the changes might drive the “two or more” category (depending on how it is conceptualized) to the highest level ever. However, the 2015 National Content Test also illustrated another possibility: The report hardly mentioned the term “multiracial” and generally did not break out the “two or more” category in its tables. Instead, the report highlighted data on each of the race and ethnic categories “alone or in combination,” with no separate display of a “two or more” group. Presented this way, the shift from separate race and ethnicity questions to a combined question appeared to have relatively little effect on the racial composition of the U.S. population. With the combined question, “white alone or in combination” dropped from 78.6 percent only to 75.2 percent and “black alone or in combination” dropped from 8.3 percent only to 8.1 percent. “Hispanic alone or in combination” rose from 11.3 percent to 12.5 percent and “some other race alone or in combination” fell from 10.2 percent to 1.5 percent (Mathews et al. 2017:47). If the Census Bureau follows the model in that report, it could effectively drop the “multiracial” category altogether.

In short, the various steps the Census Bureau has already undertaken (using “origins” for recoding) or has used in its tests (displaying racial data without the two-or-more category) raise the multiracial complication to a new level of perplexity.

Depending on the policies it adopts, the bureau could increase the multiracial count beyond the 2020 level, leave it about that level, bring it back down toward 2010 levels, or remove it from its principal tables. These drastically different possibilities underline how susceptible the counting of the multiracial population is to bureaucratic and political choices.

Conclusion

At the outset, we posed two primary questions. Regarding the first—whether the algorithmic recoding of individuals as multiracial was an improvement in racial statistics—we think the answer is that it was not. The Census Bureau should abandon recoding race and ethnicity responses based on listed “origins.” Responses about race and ethnicity should again be clearly understood as forms of self-identification that do not necessarily correspond to ancestry or to national origin. Even if the Census Bureau stops using the term “origins” in asking for detailed categories in its racial self-identification question, the detailed information provided by respondents should not be a basis for recoding their race or ethnicity. People’s ethnic and racial self-identifications co-exist with national origins and genealogies that do not necessarily line up neatly nested under OMBs official races. As a result of the history of colonialism, migration, and the intermixing of groups, many people come from countries where they are minorities (e.g., whites from South Africa, Asians from Latin America, and blacks from Canada) or have different strands to their heritage but nonetheless place themselves within one ethnoracial category. If they report only one ethnic or racial identification, the Census Bureau should not “correct” them by reclassifying them as multiracial.

The second question we raised concerns problems of general sociological interest: the relation of identity and ancestry and the role of official statistics in shaping public understanding. The idea that the mixing of peoples in the United States is necessarily producing a rise in the multiracial population is an illusion. Whether people see themselves as belonging to a separately conceived multiracial category does not depend only on their genealogies. If that were the case—if people knew their full genealogies extending back centuries—the multiracial population would be staggeringly large and sociologically meaningless. Racial identity depends on cultural and political forces, including changes in law and policy, which have historically tended to reduce multiple origins to single-race identities. Whether that process continues in the future is something that social scientists do not have the tools to predict.

Sociologists have long recognized that individuals’ ideas about their ancestry are not necessarily accurate or stable but vary considerably according to a number of factors, including surnames that highlight their patrilineage and the positive or negative evaluations attached to different origins (Waters 1990). Understanding the relationship of ancestry to identity requires keeping the two concepts separate in both theory and data. However, there are also opportunities for bringing ancestry and identity together, especially because of the increased availability of genealogical data and the advent of large-scale studies merging data from the census with genealogical sources (see Abramitzky and Boustan 2022). Mixedness

can be of different kinds: mixed genealogies, mixed self-identifications, and social attributions of mixed identity. Maintaining clarity on these different dimensions of mixedness will be of increasing importance in a society where a great deal of mixing has occurred and will continue to occur, even if self-identification as mixed does not suddenly jump in the way the 2020 census made it appear.

References

- Abramitzky, Ran, and Leah Boustan. 2022. *Streets of Gold: America's Untold Story of Immigrant Success*. New York: PublicAffairs.
- Alba, Richard. 2020. *The Great Demographic Illusion: Majority, Minority, and the Expanding American Mainstream*. Princeton: Princeton University Press. <https://doi.org/10.1515/9780691202112>
- Arias, Elizabeth, Carolyn Liebler, Marc A. Garcia, and Rogelio Sáenz. 2024. "Data Impacts of Changes in Census Bureau Procedures for Race and Ethnicity Data." Presented at the PAA, Columbus, OH.
- Bahrampour, Tara and Ted Mellnik. 2021. "Census Data Shows Widening Diversity; Number of White People Falls for First Time." *Washington Post*, August 12.
- Bosman, Julie, Robert Gebeloff, and Ethan Singer. 2024. "Trump Remarks on Harris's Multiracial Identity Overlook a Growing Trend." *The New York Times*, August 2.
- Currie, Janet, and Hannes Schwandt. 2016. "Mortality Inequality: The Good News from a County-Level Approach." *The Journal of Economic Perspectives: A Journal of the American Economic Association* 30(2):29–52. <https://doi.org/10.1257/jep.30.2.29>
- Duncan, Brian and Stephen J. Trejo. 2011. "Tracking Intergenerational Progress for Immigrant Groups: The Problem of Ethnic Attrition." *American Economic Review* 101(3):603–8. <https://doi.org/10.1257/aer.101.3.603>.
- Galvan, Astrid, and Mike Schneider. 2021. "Multiracial Boom Reflects US Racial, Ethnic Complexity." *AP News*. Retrieved May 20, 2024, <https://apnews.com/article/lifestyle-race-and-ethnicity-census-2020-7c10fabbb71b1a0aaf0db08a75b6fc6b6>.
- Jones, Nicholas A. and Jungmiwha J. Bullock. 2013. "Understanding Who Reported Multiple Races in the U.S. Decennial Census: Results from Census 2000 and the 2010 Census." *Family Relations* 62(1):5–16. <https://doi.org/10.1111/j.1741-3729.2012.00759.x>
- Jones, Nicholas A., Rachel Marks, Roberto Ramirez, and Merarys Ríos-Vargas. 2021. "2020 Census Illuminates Racial and Ethnic Composition of the Country." *Census.Gov*. Retrieved September 5, 2024, <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>.
- Lee, Jennifer and Frank D. Bean. 2004. "America's Changing Color Lines: Immigration, Race/Ethnicity, and Multiracial Identification." *Annual Review of Sociology* 30(1):221–42. <https://doi.org/10.1146/annurev.soc.30.012703.110519>
- Liebler, Carolyn A., Sonya R. Porter, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis. 2017. "America's Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census." *Demography* 54(1):259–84. <https://doi.org/10.1007/s13524-016-0544-0>
- Livingston, Gretchen. 2017. "The Rise of Multiracial and Multiethnic Babies in the U.S." *Pew Research Center*. Retrieved September 5, 2024, <https://www.pewresearch.org/short-reads/2017/06/06/the-rise-of-multiracial-and-multiethnic-babies-in-the-u-s/>.

- Lowe, Ricardo Henrique, Jr. 2023. "Census Bureau's Proposal Threatens Integrity of Race and Ethnicity Data." *UT News*. Retrieved September 5, 2024, <https://news.utexas.edu/2023/12/08/census-bureaus-proposal-threatens-integrity-of-race-and-ethnicity-data/>.
- Lowe, Ricardo Henrique Jr., Yasmiyn Irizarry, Shania Montufar, Edward Vargas, and Nancy López. 2024. "'Black and Some Other Race?': Examining Shifts in the Black Latino Population in the Census Bureau's Modified Race Question." *SocArXiv*. <https://doi.org/10.31235/osf.io/5g6vs>
- Marks, Rachel and Merarys Ríos-Vargas. 2021. "Improvements to the 2020 Census Race and Hispanic Origin Question Designs, Data Processing, and Coding Procedures." *Census.Gov*. Retrieved September 5, 2024, <https://www.census.gov/newsroom/blogs/random-samplings/2021/08/improvements-to-2020-census-race-hispanic-origin-question-designs.html>.
- Mathews, Kelly, Jessica Phelan, Nicholas A. Jones, Sarah Konya, Rachel Marks, Beverly M. Pratt, Julia Coombs, and Michael Bentley. 2017. *2015 National Content Test: Race and Ethnicity Analysis Report*. U.S. Census Bureau.
- National Academies of Sciences, Engineering, and Medicine. 2023. *Assessing the 2020 Census: Final Report*. Washington, DC: The National Academies Press.
- Nobles, Melissa. 2000. *Shades of Citizenship: Race and the Census in Modern Politics*. Stanford: Stanford University Press.
- OMB Interagency Technical Working Group on Race & Ethnicity. 2024. "About OMB's Interagency Technical Working Group on Race & Ethnicity." *OMB Statistical Policy Directive 15 on Race and Ethnicity Data Standards*. Retrieved September 5, 2024, <https://spd15revision.gov/content/spd15revision/en/about.html>.
- Perlmann, Joel and Mary C. Waters. 2002. *The New Race Question: How the Census Counts Multiracial Individuals*. New York: Russell Sage Foundation.
- Prewitt, Kenneth. 2005. "Racial Classification in America: Where Do We Go from Here?" *Daedalus* 134(1):5–17. <https://doi.org/10.1162/0011526053124370>
- Prewitt, Kenneth. 2018. "The Census Race Classification: Is It Doing Its Job?" *The ANNALS of the American Academy of Political and Social Science* 677(1):8–24. <https://doi.org/10.1177/0002716218756629>
- Rico, Brittany, Paul Jacobs, and Alli Coritz. 2023. "2020 Census Shows Increase in Multiracial Population in All Age Categories." *Census.Gov*. Retrieved September 5, 2024, <https://www.census.gov/library/stories/2023/06/nearly-a-third-reporting-two-or-more-races-under-18-in-2020.html>.
- Roth, Wendy D. and Biorn Ivermark. 2018. "Genetic Options: The Impact of Genetic Ancestry Testing on Consumers' Racial and Ethnic Identities." *American Journal of Sociology* 124(1):150–84. <https://doi.org/10.1086/697487>
- Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024. <https://doi.org/10.18128/D010.V15.0>
- Villa Ross, Ceci A., Hyon B. Shin, and Matthew C. Marlay. 2021. "Pandemic Impact on 2020 American Community Survey 1-Year Data." *Census.Gov*. Retrieved September 5, 2024, <https://www.census.gov/newsroom/blogs/random-samplings/2021/10/pandemic-impact-on-2020-ac-s-1-year-data.html>.
- Waters, Mary C. 1990. *Ethnic Options: Choosing Identities in America*. Berkeley: University of California Press. <https://doi.org/10.1525/9780520354616>

Williams, Kim M. 2008. *Mark One or More: Civil Rights in Multiracial America*. Ann Arbor: University of Michigan Press.

Williams, Kim M. 2017. "The Recursive Outcomes of the Multiracial Movement and the End of American Racial Categories." *Studies in American Political Development* 31(1):88–107. <https://doi.org/10.1017/S0898588X17000074>.

Acknowledgments: Presentations by Ricardo Lowe and colleagues helped inform the empirical strategy in our article. We also thank participants in the May 2024 Conference of the American Association for Public Opinion Research for their responses to our research.

Paul Starr: Sociology Department, Princeton University. E-mail: starr@princeton.edu.

Christina Pao: Sociology Department, Princeton University. E-mail: christina.pao@princeton.edu.