MEASURING RACE AND ANCESTRY IN THE AGE OF GENETIC TESTING

Sasha Shen Johfre, Stanford University[1]
Aliya Saperstein, Stanford University

Abstract:

Will the recent explosion in popularity of genetic ancestry tests (GAT) change how American adults respond to race and ancestry questions on censuses and demographic surveys? We draw on a unique survey of over 100,000 U.S. adults that inquired about respondents' racial and ancestral identities and genealogical knowledge. We find that people who have taken a GAT, compared to those who have not, are more likely to report multiple races as well as multiple regions of ancestral origin. Although reports of most race and ancestry categories increase among GAT takers, not all do; for example, we find intriguing declines in reporting American Indian ancestry among self-identified White respondents who have taken GATs. Our results likely foreshadow greater changes to come as more Americans embrace genetic ancestry testing. Current and future demographers must consider GATs in the development and interpretation of measures of race and ancestry, perhaps especially in health contexts.

[1] Primary email: sashajf@stanford.edu

With the increasing popularity and sophistication of consumer genetic ancestry tests (GATs), Americans have unprecedented access to new types of information about their family ancestry. The market for such information has grown exponentially: genetic testing went from being a $15 million industry in 2010 to a $173 million industry in 2016 (Keshavan 2016; Borry, Cornel, and Howard 2010). This historic change in the availability of genetic information may have implications for best practices in collecting and analyzing demographic data.

Will GATs change how American adults respond to race and ancestry questions on demographic questionnaires? Although previous research finds that taking a GAT can influence a person's racial identity, evidence is mixed on the magnitude and specific patterns of any such changes. For example, Roth and Lyon (2018) found that as many as 20% of GAT takers may update their racial identity based on GAT results; however, GAT-takers tend to be picky about which genetic ancestries they incorporate into their racial identities (see also Roth and Ivemark 2018). Research by Shim and colleagues (2018) found that people interpret GAT results as "just information" about their bodies that does not override lived experience as a meaningful source of racial identity. Other research argues that the increasing focus on genetic ancestry within medicine has contributed to the renewal of an essentialized notion of race as a form of genetically-based population difference (Fujimura and Rajagopalan 2011; Fullwiley 2008). This raises the possibility that any GAT-related changes in reporting could be particularly relevant in health contexts. Understanding whether (or when) Americans incorporate GAT results into their self-reported race or ancestry will be crucial for everything from population estimation and projection to understanding measured trends in racial disparities.

Previous studies in this area have been primarily qualitative, leaving it an open question whether any personal identity changes in people who have taken GATs will be reflected in demographic surveys or the upcoming 2020 census enumeration. To answer this question, we draw on a unique data source that includes information about the self-reported race and ancestry of over 100,000 U.S. adults who were registered as potential volunteer bone marrow donors with the National Marrow Donor Program (NMDP). The survey also inquired how much respondents knew about their family ancestry and how they came by that knowledge. These features, along with randomization in the questionnaire order, allow us to contrast the race and ancestry responses of people who reported taking a GAT with the responses of those who had not taken at GAT at the time of the survey.

We find that GAT-takers, compared to non-takers, are more likely to report multiple regions of ancestral origin, as well as multiple races. Although in most cases exposure to GAT results is associated with greater reporting of less commonly reported ancestries such as Scandinavian or Sub-Saharan African, among self-reported White respondents, GAT-takers are generally less likely to report American Indian ancestry than are non-GAT takers. Some, but not all, of the reporting differences between GAT takers and everyone else are attenuated if respondents weren't asked if they had a genetic ancestry test until after they had reported their race and ancestry. We expect that these results are harbingers of greater changes to come as more Americans embrace genetic ancestry testing. Current and future demographers must consider GATs in the development and interpretation of measures of race and ancestry, perhaps especially in health contexts.

Data and Methods

All registered NMDP donors with valid email addresses were invited to participate in a survey about race, ancestry and genetics between May-July 2015. Twenty percent of the nearly two million invitees opened the email, and five percent completed the survey. This low response rate is normal for email-based surveys and does not indicate low data quality (see Fan and Yan 2010), but it does mean that broad generalization from this sample is not advisable. However, our sample size affords us the

statistical power to assess patterns for sub-populations that are not well-represented in typical national surveys. For example, we have enough respondents to analyze smaller populations, such as people who racially identify as Asian alone (n=3,461) and American Indian alone (n=279), that are often overlooked, or lumped together, in survey research. The survey is therefore particularly well-suited to answer questions about potential differences in patterns of race and ancestry reporting.

Another particularly useful part of this survey design was question order randomization that allows for comparison of responses across experimental conditions. This is important because responses to race questions can vary based on the context, question type, and even the order of questions within a survey (e.g. Hirschman, Alba, and Farley 2000; Harris and Sim 2002). In our survey, respondents were randomly assigned to one of four question orders:

1. Knowledge, race, ancestry (knowledge first, race first)
2. Knowledge, ancestry, race (knowledge first, race second)
3. Race, ancestry, knowledge (knowledge second, race first)
4. Ancestry, race, knowledge (knowledge second, race second)

The knowledge section asked respondents about their interest in and exposure to genealogical information, including a question about whether they had taken a GAT. Overall, five percent of our sample indicated that they had taken a GAT (n = 5,319; 2,583 in the two knowledge-first [K1] conditions, and 2,736 in knowledge-second conditions [K2]). If differences between GAT takers and non-GAT takers are greater in the K1 conditions compared to K2, then it suggests GAT takers' race and ancestry responses are sensitive to priming. Either way, our study offers important scope conditions for future research measuring the effects of GATs in surveys.

Race and ancestry measures. The survey was originally designed to examine the relationship between multiple self-reported race measures and genetic measures (see Hollenbach et al. 2015 for more details). Therefore, it included an array of questions about race and ancestry, including racial self-identification and reflected race (how respondents think other people classify them; see Roth 2016). The racial self-identification item was based on a "combined" question format originally tested by the Census Bureau, in which the response option "Hispanic or Latino" was offered alongside other federally-recognized race categories. Respondents could offer multiple responses for racial self-identification, and 12% did so – a proportion considerably higher than the 2-3% found in nationally representative data collection. This may be the result of self-selection of people with multiracial ancestry into a survey that was about improving outcomes for transplant matching. However, the frequency of identifying as multiracial also varied by condition and previous exposure to a GAT, as we discuss below.

Respondents were also asked about ancestral origins. They were first offered a list of geographic regions, such as "Eastern Europe" or "Middle East and North Africa," and asked to "check all that apply" to "best describe your family origins or ancestry." The same geographic origin categories were used to inquire specifically about the ancestries of the respondents' four biological grandparents. We focus our preliminary analysis on respondents' self-reported ancestry, and expect to incorporate reported family ancestries into the final paper.

Of the 109,831 people who answered at least part of our survey, 103,252 answered all questions about race, ancestry and knowledge. We restrict all our analyses to this slightly smaller analytical sample.

Hypotheses. To gauge whether or not GAT takers and non-GAT takers differ in how they respond to questions about their race and ancestry, we consider the following outcomes: 1) frequency of reporting multiple races or multiple ancestries, 2) frequency of reporting particular race or ancestry categories, and 3) the "match" between one's race and ancestry responses. It is likely that, as a result of receiving genetic information that may indicate small proportions of a large number of ancestries, GAT takers

would report more ancestries than non-GAT takers; it is also possible that GAT results could contradict beliefs people had about family ancestries, leading to lower levels of reporting for particular responses. Similarly, it is possible that GAT results would lead people to harmonize their race and ancestry reporting because they interpret the genetic information as saying something about their "race." However, it is also possible that GAT takers may only change their ancestry reporting and continue to base their race on lived experiences (cf. Shim et al 2018). We explore each of these possibilities below.

Results

Preliminary analyses suggest that people who have taken a genetic ancestry test do differ in their responses to race and ancestry questions compared to people who have not taken a GAT.

Multiple race and ancestry reporting. GAT takers are significantly more likely than non-GAT takers to select more than one geographic origin to describe their family ancestry (70% vs. 59%, $p<0.001$). This is consistent with the interpretation that GAT takers have access to additional information about their ancestry (Roth and Ivemark 2018). We also find that GAT takers are more likely to select multiple races for self-identification than are non-GAT takers (14% vs. 11%, $p<0.001$). As our data are cross-sectional, we must exercise caution in attributing this reporting pattern to a result of taking GATs, rather than selection into who is most likely to take GATs (Horowitz and Saperstein 2018). People who saw the question about GATs *before* they were asked about their race or ancestry (i.e., the K1 conditions) were more likely to report multiple races and ancestries than people who were asked in the other order (see Figure 1). The difference between knowledge conditions is largest for GAT takers reporting multiple ancestries (72% in K1 vs. 68% in K2).[2]

Reporting specific racial categories. We find marked differences in the particular reporting of multiple races among GAT takers, as well as differences in race reporting based on question order (see Table 1). When the race questions came before the ancestry questions (regardless of the knowledge condition) all respondents were more likely to report: 1) White in combination with another race or 2) three or more races (see "Race first" columns in Table 1). However, differences between GAT takers and everyone else are especially striking: 5% of GAT takers reported three or more races even before they were reminded about their genealogical knowledge, compared to 1.5% of non-GAT takers in the same survey condition. Other results suggest that reminding GAT takers about their knowledge immediately before they self-identify their race results in especially high frequencies of selecting two races including White (13%) and self-identifying only as American Indian (0.4%). Note that as with our comparisons above, these results also likely reflect selection bias in who takes GATs in the first place; for example, the lower frequency of self-identifying as Asian among GAT takers is likely because self-identified Asian Americans have low overall interest in GATs (Horowitz and Saperstein 2018). Nevertheless, these findings suggest GAT takers may be especially sensitive to question ordering, when multiple race and ancestry questions are offered, perhaps because they are eager to share all their "information" on the first question they see.

Race and ancestry "matching." Given speculation that GATs could reinforce essentialized notions of race, we are especially interested in the extent of alignment between the lists of races and ancestries given by respondents. To address this aim, we rely on the definitions of racial categories used by the Office of Management and Budget (OMB 1997), which link ancestral origins in particular regions of the world to federally-tracked racial and ethnic categories. If a person reports ancestry from Western

[2] Throughout, we report statistical significance using standard levels (*p<0.05; **p<0.01; ***p<0.001) because when we compare among GAT takers, or other smaller sub-groups, we have sample sizes in the low thousands. However, we urge caution when interpreting such significance tests among non-GAT takers because differences may be statistically significant at conventional levels, but may not be particularly substantively significant.

Europe and selects White as their race, we consider this a "match," while someone who reports ancestry from East Asia, but does not include "Asian" in their race responses would not match. Using these criteria, we found that 21% of respondents had race and ancestry lists that did not match. This was most often because they listed ancestry categories that, according to official definitions, do not correspond to their racial self-identification (81% of those without a match or 17% of total).

We found no significant differences in overall levels of race-ancestry matching either by condition or between GAT takers and non-GAT takers. However, we did find differences in the specific combinations of races and ancestries. For example, overall GAT takers were more likely to report American Indian, sub-Saharan African, and Scandinavian ancestry, and less likely to report Unknown ancestry, than people who have not taken a GAT (see Table 2). However, among respondents who identified as White, GAT takers are significantly *less* likely to report American Indian ancestry than non-GAT takers (14% vs. 16%; see Panel A). This result holds across three of the four survey conditions and runs counter to trends toward increasing American Indian ancestry and race reporting among White Americans over the past several decades (see, e.g., Nagel 1995).

Some "mismatch" between race and ancestry responses is also evident when comparing racial self-identification as Black with self-reported sub-Saharan African ancestry. Although tracing ancestry to the original peoples of sub-Saharan Africa is the official definition of the "Black or African American" racial category in the U.S., many descendants of former slaves know little about their pre-slavery geographic origins (Nelson 2008). To acknowledge this, we offered both "Sub-Saharan Africa" and "African American" categories among our ancestry responses. Among respondents who identified as Black alone or in combination with other race categories, 67% reported African American ancestry; 16% reported sub-Saharan African ancestry; and 38% selected "Unknown" as one of their ancestries.

We found that the sub-Saharan African ancestry response resonated most with two different types of respondents: 1) those who had taken a genetic ancestry test and 2) those who were born outside the United States. Overall, among respondents who identified as Black, 57% reported sub-Saharan African ancestry if they had taken a GAT compared to 13% of those who had never taken an ancestry test. However, US-born Black Americans were very unlikely to list sub-Saharan African ancestry unless they had taken a GAT (see Figure 2). These patterns of self-reporting do not vary across survey condition and may partly reflect understandings of ancestry at different time-scales (e.g., recent and known relatives vs. more distant lineage), with genetic ancestry tests making one's distant lineage more salient than it would be otherwise (see Zeruvabel 2012).

Additional analysis. Before PAA 2019, we plan to model our outcomes of interest controlling for additional factors that we know to be important, including: knowledge-seeking behaviors beyond taking a GAT; demographic characteristics; and differences in survey administration. Through multivariate analysis, we will be able to further untangle the relationships between GAT taking, question order, and other characteristics such as gender, age, region and educational attainment. We will also examine two more types of potential "matching" patterns: 1) matching between racial self-identification and reflected race and 2) matching between ancestry lists that respondents generated for themselves vs. lists they generated for their biological relatives. Although our data are cross-sectional, if reporting patterns between GAT takers and non-GAT takers continue to differ across all survey conditions after controlling for other factors (and very preliminary models suggest they do), then we would have relatively strong evidence that taking a GAT is related to a change in response.

References

Borry, Pascal, Martina C. Cornel, and Heidi C. Howard. 2010. "Where Are You Going, Where Have You Been: A Recent History of the Direct-to-Consumer Genetic Testing Market." *Journal of Community Genetics 1*(3):101–6.

Fan, Weimiao and Zheng Yan. 2010. "Factors Affecting Response Rates of the Web Survey: A Systematic Review." *Computers in Human Behavior 26*(2):132–39.

Fujimura, J.H, and Rajagopalan, R. 2011. "Different Differences: the Use of "Genetic Ancestry" versus Race in Biomedical Human Genetic Research." *Social Studies of Science 41*(1): 5-30

Fullwiley, Duana. 2008. "The Biologistical Construction of Race: Admixture Technology and the New Genetic Medicine." *Social Studies of Science, 38*(5), 695-735.

Harris, David R. and Jeremiah Joseph Sim. 2002. "Who is Multiracial? Assessing the Complexity of Lived Race." American Sociological Review 67:614-27

Hirschman, Charles, Richard Alba, and Reynolds Farley. 2000. "The meaning and measurement of race in the US census: Glimpses into the future." *Demography*, 37(3), 381-393.

Horowitz, Adam L., and Aliya Saperstein 2018. "Consumer (Dis-)Interest in Genetic Ancestry Testing: The Roles of Race, Immigration, and Ancestral Certainty." Presention at the *American Sociological Association Annual Meeting*, Philadelphia, PA, August 11.

Keshavan, Meghana. 2016. "20 Key Players in the Direct-to-Consumer Lab Testing Market." *MedCityNews*, January 20.

Nagel, Joane. 1995. "American Indian Ethnic Renewal: Politics and the Resurgence of Identity." *American Sociological Review 60*:947-65

Nelson, A. 2008. "Bio science: Genetic genealogy testing and the pursuit of African ancestry." *Social Studies of Science, 38*(5), 759-783.

Office for Management and Budget. 1997. Office of Management and Budget. 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity." *Federal Register* 62(210):58782–90.

Roth, Wendy D., & Ivemark, B. 2018. "Genetic Options: The Impact of Genetic Ancestry Testing on Consumers' Racial and Ethnic Identities." *American Journal of Sociology, 124*(1), 150-184.

Roth, Wendy D., and Lyon, K.A. 2018. "Genetic Ancestry Tests and Race: who takes them, why and how do they affect racial identities?" In: Reconsidering Race, ed. By Suzuki, K., & Von Vacano, D.A. New York: Oxford University Press.

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):519-537.

Shim, Janet K. et al. 2018. "Knowing Something versus Feeling Different: The Effects and Non-Effects of Genetic Ancestry on Racial Identity." *New Genetics and Society*:1–23.

Zerubavel, E. 2012. *Ancestors and relatives: Genealogy, identity, and community*. Oxford University Press USA.

	GAT				Non-GAT			
	Knowledge first		Knowledge second		Knowledge first		Knowledge second	
	Race first	Race second	Race first	Race second	Race first	Race second	Race first	Race second
American Indian Only	0.4	0.3	0.1	0.2	0.2	0.3	0.3	0.3
Asian Only	2.4	2.4	1.8	2.4	3.3	3.3	3.5	3.5
Black Only	2.3	3.3	3.7	3.3	2.7	3.1	3.0	2.9
Hispanic Only	2.8	2.6	2.6	3.2	4.7	4.8	5.0	4.9
NHOPI Only	0.3	0.1	0.2	0.0	0.1	0.1	0.2	0.2
Other Only	1.2	1.7	1.4	1.1	1.2	1.1	1.1	1.0
White Only	72.5	76.6	74.5	77.7	73.8	77.0	75.1	77.2
2 races incld. White	12.7	9.6	9.5	8.3	10.9	7.8	9.2	7.6
2 races not incld. White	0.9	1.7	1.2	1.5	1.2	1.2	1.2	1.2
3 or more races	4.5	1.8	5.0	2.3	2.0	1.2	1.5	1.2
	Pearson chi2(27) = 66.4, p<0.001				Pearson chi2(27) = 308.8, p<0.001			

TABLE 1: Racial self-identification, by question order and whether the respondent took a genetic ancestry test (GAT). Total n=103,252.

Panel A. American Indian ancestry reporting (%)

	GAT	No GAT	Diff.	*p*
American Indian	85.3	81.7	3.6	0.137
Asian	8.4	4.4	4.0	0.002
Black	31.6	25.9	5.8	0.018
Hispanic	33.3	14.0	19.3	0.000
NHOPI	23.9	13.2	10.7	0.043
Other	21.5	15.2	6.3	0.160
White	13.6	15.9	-2.3	0.000

Panel B. Unknown ancestry reporting (%)

	GAT	No GAT	Diff.	*p*
American Indian	15.4	27.8	-12.4	0.000
Asian	7.7	7.0	0.7	0.680
Black	20.5	39.3	-18.8	0.000
Hispanic	13.4	16.2	-2.8	0.117
NHOPI	17.4	13.8	3.6	0.501
Other	19.6	18.3	1.3	0.643
White	10.9	16.8	-5.9	0.000

Panel C. Sub-Saharan African ancestry reporting (%)

	GAT	No GAT	Diff.	*p*
American Indian	14.7	1.8	12.9	0.000
Asian	4.7	0.5	4.2	0.000
Black	57.3	13.0	44.2	0.000
Hispanic	7.8	0.9	6.9	0.000
NHOPI	13.0	0.1	12.9	0.000
Other	12.0	3.1	8.9	0.000
White	3.3	0.2	3.0	0.000

Panel D. Scandinavian ancestry reporting (%)

	GAT	No GAT	Diff.	*p*
American Indian	25.9	13.6	12.4	0.000
Asian	12.0	4.7	7.3	0.000
Black	19.7	2.4	17.3	0.000
Hispanic	14.7	4.2	10.5	0.000
NHOPI	21.7	9.6	12.1	0.090
Other	25.4	9.0	16.4	0.000
White	31.5	17.3	14.2	0.000

TABLE 2: Frequency of ancestry reporting by racial self-identification and whether the respondent has taken a genetic ancestry test (GAT). Percentages include people who chose the category alone or in combination with other race responses. “Diff.” column represents the percentage of GAT takers minus the percentage of non-GAT takers who selected the given race response. P-values are from a two-tailed test of the difference in proportions. Total n = 103,252.

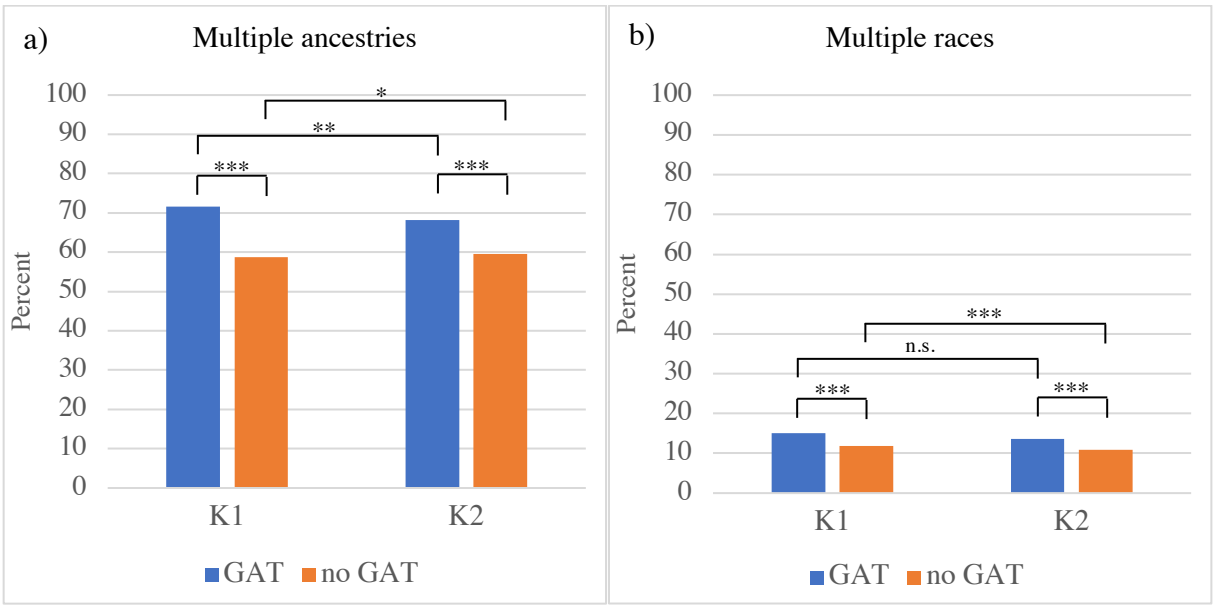

FIGURE 1: Frequency of selecting multiple ancestries (a) or multiple races (b) by knowledge question order and whether respondents took a genetic ancestry test (GAT). Total n=103,252.

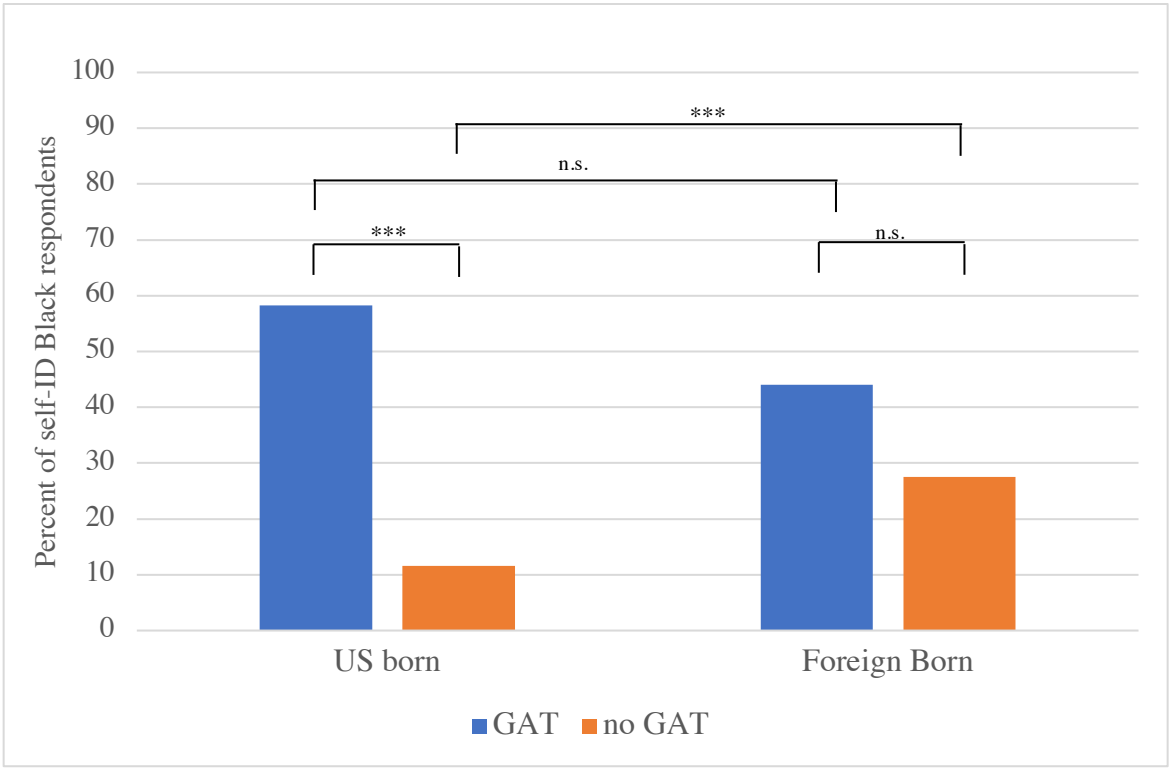

FIGURE 2: Sub-Saharan African ancestry reporting among respondents who identify as Black, by nativity and whether the respondent took a genetic ancestry test (GAT). Total n=4,789