Income and Education Disparities Track Genetic Ancestry

Meng Hu*
Independent Researcher

Emil O. W. Kirkegaard
Ulster Institute for Social Research, London, UK

John G. R. Fuerst
Cleveland State University
University of Maryland Global Campus, Bioinformatics

*Correspondence: mh19870410@gmail.com

Abstract

Structural racism has often been invoked to explain observed disparities in social outcomes, such as in educational attainment and income, among different American racial/ethnic groups. Theorists of structural racism typically argue that racial categories are socially constructed and do not correspond with genetic ancestry; additionally, they argue that social outcome differences are a result of discriminatory social norms, policies, and laws that adversely affect members of non-White race/ethnic groups. Since the examples of social norms and policies commonly provided target individuals based on socially-defined race/ethnicity, and not on genetic ancestry, a logical inference is that social disparities will be related to socially-defined race/ethnicity independent of genetically-identified continental ancestry. In order to evaluate this hypothesis, we employ admixture-regression analysis and examine the independent influences of socially-identified race/ethnicity and genetically-defined ancestry on the educational attainment and income of parents, using data from a large sample of US children. Our study focuses on self-identified Whites, Blacks, Hispanics, and East Asians in the United States. Analyses generally show that the association between socially-identified race/ethnicity and outcomes is confounded by genetic ancestry and that non-White race/ethnicity is unrelated to worse outcomes when controlling for genetic ancestry. For example, conditioned on European genetic ancestry, Americans socially-identified as Black and as Hispanic exhibit equivalent or better social outcomes in both education and income as compared to non-Hispanic Whites. These results are seemingly incongruent with the notion that social outcome differences are due to social policy, norms, and practices which adversely affect individuals primarily based on socially-constructed group status.

Key Words: Structural racism, education attainment, income, race, genetic ancestry

1. Introduction

In the United States (USA), there are well-documented socioeconomic status (SES) differences between socially (self and other)-defined race/ethnic groups. These groups, as defined by the USA Office of Management and Budget, include Hispanic/Latin American ethnicity and White, Black, American Indian, Asian, and Pacific Islander race or origin. Adults who self-identify as either Hispanic or as non-Hispanic Black, American Indian, and Pacific Islander typically have lower educational attainment and income than those who identify as non-Hispanic White. Conversely, those who identify as Asian have higher average educational attainment and income than Whites.

Racial/ethnic SES inequalities are often attributed to structural racism. Jones (2002, p. 10) gives one frequently cited definition of structural racism: "structures, policies, practices, and norms resulting in differential access to the goods, services, and opportunities of society by 'race'." Braveman et al. (2022) note, additionally, that structural racism encompasses laws, policies, institutional practices, and entrenched social norms, as distinct from individual acts of discrimination.[1]

According to Jones (2002) "race" is only based on a few phenotypic-related genes, not global genetic ancestry, since the few genes that determine skin color, hair texture, and facial features are not informative about other aspects of the genotype at the individual level. Advocates of the structural racism hypothesis frequently emphasize that race/ethnicity categorization "reflects neither biological nor cultural differences" and that race is "often conflated erroneously with biology and ancestry" (Adkins-Jackson et al., 2022, p. 540), that "race is a social construct and is distinct from ethnicity, genetic ancestry or biology" (O'Reilly, 2020, p. 2), that "social races bear little relationship to the reality of human biological diversity" (Smedley & Smedley, 2005, p. 22) and that race "is a social construct with no biological basis and stems from White supremacy" (Haeny et al., 2021, p. 889). While the phrase "race is a social construct" can have a range of meanings, a popular one, given by the American Sociological Association (2003), is that race is "a social invention that changes as political, economic, and historical contexts change"; this social invention is said to be important because "social and economic life is organized, in part, around race as a social construct."

Part of the reason for the emphasis on the socially constructed aspects of race is rhetorical, as it is believed that "a long history of work has attempted to link race to genetics in order to justify racial discrimination and race-based social stratification" (Gou et al., 2014, p. 2337). For example, Suyemoto et al. (2022, p .78) argue that the myth that race is "genetic or biological" is "important to unravel because it's the foundation of justifying racial hierarchies and therefore racism." Additionally, the assertion, "race is a social construct" is used to argue, fallaciously, that differences between socially-defined races are necessarily social or environmental in origin (e.g., Robbins, 2022).

[1] It should be noted that public policies don't have to be designed to affect the socio-economic outcomes of ethnic minorities. Policies which tend to increase, intentionally or not, SES inequalities will affect minorities more negatively since lower SES families are typically minorities (Noguera & Alicea, 2020).

Nonetheless, there are two substantive reasons for the adoption of social constructivism in context to discussion of Office of Management and Budget-defined race/ethnic groups and social inequalities.

First, laws, policies, institutional practices, and entrenched social norms must affect members of a group delineated in specific ways to affect social outcome differences. In the USA, race/ethnic groups are delineated based on "social and cultural characteristics" (Office of Management and Budget, 1997) in addition to, and not strictly based on, genetic ancestry.

Second, structural racism is typically conceptualized as including many specific laws, policies, institutional practices, and social norms which did not discriminate or affect individuals strictly based on ancestry. Examples frequently given include: voter suppression of Blacks, political gerrymandering, predatory financial services, mass incarceration, police violence, sending American Indian children to boarding schools (Braveman et al., 2022), slavery, black code, Jim Crow laws, segregated housing, redlining (Bailey et al., 2021; Erikson et al., 2022). These policies and practices targeted individuals based on socially-defined race, not strictly based on genetic ancestry.

For example, Jim Crow impacted all individuals, in the Southern states, with a certain degree of recognizable African American ancestry. The impact of this policy was not apportioned according to genetic ancestry. Given that these policies and practices targeted individuals based on socially-defined race/ethnicity, not strictly genetic ancestry, it makes sense that proponents of the structural racism hypothesis emphasize the social construction of race. Discussing this point, Gichoya et al. (2022, p. 8) observe, "in the context of racial discrimination and bias, the vector of harm is not genetic ancestry but the social and cultural construct that of racial identity… biased decisions are not informed by genetic ancestry information, which is not directly available to medical decision makers in almost any plausible scenario.".

Strong social constructionist claims about self-identified racial categories, to the effect that they do not reflect biological differences or that they provide information only about genes related to conspicuous race-related phenotypes, are probably false. In the USA, at least, there is a moderate to high concordance between self/parental-identified race/ethnicity and continental-level lineage (Fang et al., 2019, Kirkegaard, 2021). Since continental-level populations - such as Sub-Saharan African, West Eurasian, East Asian, and Amerindian - are differentiated with respect to many morphological and physiological traits (Brues, 1990), socially-identified race can also be predicted from, for example, medical imaging data (Gichoya et al., 2022; Kirkegaard & Fuerst, 2023). Nonetheless, self-identified race, let alone ethnicity, in the USA, is also socio-politically constructed in that classifications are based on complex political and cultural considerations independent of continental-level genetic ancestry. For example: "Hispanic" is not defined by genetic or biogeographic ancestry, but refers to a "person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race" (Office of Management and Budget, 1997); "Black," while defined as "a person having origins in any of the black racial groups of Africa," includes individuals with a proportion of Black African ancestry that can range from 2% to 100% (Bryc et al., 2015).

Given the focus on the socially-constructed aspects of race/ethnicity in relation to specific laws, policies, and norms, one obvious scientific prediction is that socioeconomic differences will relate to socially-constructed aspects of race/ethnicity substantially above and beyond continental-genetic ancestry. This seems to be what, for example, Smedley and Smedley (2005, p. 22) predict when they state that "although the term race is not useful as a biological construct… social race remains a significant predictor of which groups have greater access to societal goods and resources and which groups face barriers - both historically and in the contemporary context - to full inclusion." Or, what Herd et al., (2021, p. 420) mean when they argue that it would be illogical to focus on "racial group genetic variation" and also argue that "[g]iven what is known about ancestry, it makes little sense to examine racial group genetic variation, especially in the American context. People share a social and political category, not a biological category…".

One method of assessing whether the socially constructed aspects of race/ethnicity have an ancestry-independent impact on outcomes is to use admixture-regression designs (Connor & Fuerst, 2022; Fuerst, 2021; Kirkegaard, 2019; Lasker et al., 2019). In these designs, recently-admixed populations are treated as natural experiments, and genetic admixture is used to disentangle various cultural, environmental, and genetic factors contributing to variation in a trait. Self-identified racial identity is treated as a "surrogate to an array of social, cultural, behavioral, and environmental variables" (Fang et al., 2019, p. 764) and included in the regression models alongside genetic ancestry variables. These designs can disentangle the effects of factors related to social racial/ethnic identity from the effects of factors related to genetic ancestry. In this paper, we test the hypothesis that self-identified race/ethnicity (this being an index of social race), is strongly related to two measures of SES -- income and education -- independent of continental-level genetic ancestry. We interpret the structural racism hypothesis, at least as commonly presented, as predicting that it will be since this is what many proponents of this hypothesis explicitly state.

2. Method

2.1. Data and Sample

The Adolescent Brain Cognitive Development (ABCD) study is a joint long-term initiative that includes 21 research sites throughout the US, focused on examining brain development and child health to investigate the psychological and neurobiological foundations of human growth. At baseline, around 11,000 children aged 9-10 years were sampled, using a probability-based sampling approach aimed at establishing a comprehensive and inclusive sample of US children within that age group. In this current investigation, we utilized the ABCD 3.01 baseline data.

We used parents' variables, with the exception of the child's genetic ancestry. As we only had access to the children's genetic ancestry, we limited the sample to cases where both parents were the biological parents of the child. In families with multiple children, we used the genetic ancestry estimates of the first biological child. We then selected cases with no missing data in our variables, which included income, educational attainment, parents' age, and the child's genetic ancestry. Additionally, we excluded parents who identified as Pacific Islander, South Asian, or Other Asian so as to focus on individuals primarily of European, African, East Asian, and Amerindian ancestry. These restrictions resulted in a final sample of 5,073 parent dyads.

2.2. Variables

Several variables were computed for the purpose of the present study. The list of these variables is provided below.

2.2.1 Genetic ancestry

Genetic ancestry of the biological children was computed using the National Institutes of Health (NIH) estimates. The process of imputing and genotyping was carried out by the ABCD Research Consortium. To determine genetic ancestry, the ABCD Research Consortium utilized a k = 4 solution (European, African, Amerindian, and East Asian). The researchers used the 1000 Genomes populations as reference samples and *fastStructure* as the algorithm for this purpose (Hatton, 2018). We divided the ancestry estimates by the sum of European, African, Amerindian, and East Asian ancestries so that the sum of the four ancestries is equal to 1. For this reason, we have to drop one ancestry variable in the regression equation for the estimation of the other variables. We selected European ancestry as the reference category.

2.2.2 Educational attainment and income

Two outcome variables, parental education attainment and parental income, were computed for analyses. The parental education variable was calculated as the average of both parents' education level, which originally had 22 categories. To create an interval variable, we recoded the variable to have 11 categories, where 0 represented never attended school, 1-12 represented 1st-12th grade, 12 represented High school graduate, 14 represented Some college or an Associate degree (Occupational or Academic Program), 16 represented Bachelor's degree, and 18 represented Professional School or Doctoral degree.

Parental income was an interval variable that reflected the total combined family income of the parents in the past 12 months, which was originally coded as 1-10. The categories were recoded as follows: 1 represented Less than \$5,000, 2 represented \$5,000 through \$11,999, 3 represented \$12,000 through \$15,999, 4 represented \$16,000 through \$24,999, 5 represented \$25,000 through \$34,999, 6 represented \$35,000 through \$49,999, 7 represented \$50,000 through \$74,999, 8 represented \$75,000 through \$99,999, 9 represented \$100,000 through \$199,999, and 10 represented \$200,000 or more.

Detailed descriptions of the educational and income variables are provided by Fuerst, Hu, & Connor (2021) and Fuerst, Shibaev, & Kirkegaard (2023) who used the same coding scheme.

Both the education attainment and income variables were centered around the mean and standardized. However, some extreme values (i.e., lower than -3 standard deviations (*SDs*) below the mean) were identified in both variables. To retain these cases while minimizing the impact of outliers, we winsorized the data using a 3 standard deviation threshold.

2.2.3 Age

Age represents the age of the responding parent. For the regression analyses, the age variable was mean-centered and standardized.

2.2.4 Immigrant status and English

In the survey, parents were asked whether anyone in the child's family, including maternal or paternal grandparents, was born outside of the United States, and this variable was assigned a value of "1" for "Yes" and "0" otherwise. Additionally, the parent who provided the response was asked if their native language was English, and this variable was also coded as "1" for "Yes" and "0" otherwise.

2.2.5 Self-identified race/ethnicity

The responding parent was asked 18 questions about his or her race/ethnicity but not the race/ethnicity of his or her spouse. Based on these responses, we created six dummy-coded variables: Hispanic ethnicity and White, Black, East Asian, Native American, and Other race. The East Asian category included individuals identifying as Chinese, Japanese, Korean, Filipino and Vietnamese, while the Native American category included individuals identifying as American Indian and Alaska Native. Because the race of the nonresponding parent is unknown, we supplied a supplementary analysis by creating parallel sets of race/ethnicity dummy variables based on the biological child's race/ethnicity, as indicated by the parents. If the parents are intermarried, this would be reflected in the child belonging to two race categories.

2.2.6 State racism and xenophobia

ABCD calculated state-level indicators of both racism and anti-immigrant bias (xenophobia) for the 18 states in which the recruitment sites were located. These were based on both implicit bias measures and state-level structural variables. The two indicators correlated at $r = .34$ ($p < 0.05$, $N = 5{,}073$). Both variables were centered around the mean and standardized. These variables were only used in supplementary analyses.

2.3. Description of Analyses

2.3.1 Main analyses

To examine the effect of race/ethnic identity on income and education we ran a series of admixture regressions. We used a multilevel regression model, specifying research sites as a random effect and the following fixed effects: dummy coded race/ethnicity (Black, Native American, Hispanic, Other Race), native English fluency, genetic ancestry (African, Amerindian and East Asian), and finally parent's age. Both White racial identity and European ancestry are used as benchmark variables and the associated variables are dropped from the regression models. Because we are only interested in the fixed effects estimates, Full Maximum Likelihood (FML), instead of Restricted Maximum Likelihood (REML), is used as the estimation method. FML produces more accurate fixed effects whereas REML produces more accurate random effects. Regression analyses are carried out using lme4 package for R (Bates et al., 2009).

2.3.2 Supplementary analyses

In order to supplement our main analyses and further explore the robustness of our findings, we conducted several additional analyses. Firstly, as a robustness check, we reran the analyses excluding all cases with values of education and income 3 standard deviations (*SD*s) or more below the mean. This was done in order to ensure that our results were not primarily driven by the extremely low values of education and income. Secondly, we used an alternative approach to determine race/ethnicity. Specifically, we used the children's race/ethnicity as reported by the responding parent, instead of the responding parent's race/ethnicity. This approach was taken because the parent-reported race/ethnicity of the child may provide a better representation of the average race/ethnicity identity of both biological parents, since the race/ethnicity of the nonresponding parent is unknown. In a third set of analyses, we restricted the sample to families with at least one biological parent (as opposed to having strictly two biological parents), which yielded a larger sample size due to the higher rate of single parenthood among Blacks and Hispanics (N=7,652). In a fourth set of analyses, we added state-level indicators of racism and anti-immigrant bias to the regression models to explore the potential impact of these variables on our findings. A fifth set of analyses were conducted using weighted regressions with the survey package (Lumley, 2020). This approach was used to take into account selection bias, with weights based on the propensity-based weight of children provided by the ABCD. It is important to note that we did not use these weight variables in our main analysis because we primarily used parents' variables. Finally, a sixth set of analyses was conducted within the combined Hispanics, Black, and Native American subsamples. This approach was taken to explore potential differences within disadvantaged non-White groups.

2.4. Data

The complete data set is available to qualified researchers at: https://nda.nih.gov/abcd

3. Result

3.1 Descriptive statistics

Table 1 presents the descriptive statistics for the variables used in our study. In this table, age is presented in the unstandardized form for ease of interpretation. Black, Hispanic, and Native American self-identified race/ethnicity are associated with lower levels of income and educational attainment, while White and East Asian self-identified race/ethnicity are associated with higher levels. It is noteworthy that the Hispanic and East Asian groups are characterized by a higher percentage of immigrant families and also low rates of English fluency. To note, the total case number in Table 1 (N = 5,162) is larger than the dyads (N = 5,073) because some non-Hispanic individuals self-identified as more than one race. This overlap does not impact our regression analyses because in these analyses we control for race dummy variables. The high level of European admixture among East Asian respondents reflects the relatively large number of White-East Asian couples in this sample.

Table 1. Variable Means (and Standard deviations) by the Race/Ethnicity of The Responding Parent.

	non-Hispanic White	non-Hispanic Black	non-Hispanic East Asian	non-Hispanic Native American	Hispanic
Age	42.51 (5.41)	40.34 (6.49)	43.79 (5.10)	41.37 (6.64)	40.57 (6.14)
Education	0.20 (0.80)	-0.53 (0.93)	0.38 (0.68)	-0.35 (1.02)	-0.82 (1.21)
Income	0.20 (0.75)	-0.67 (1.22)	0.37 (0.71)	-0.31 (1.11)	-0.74 (1.20)
English %	0.97 (0.18)	0.93 (0.26)	0.48 (0.50)	0.99 (0.12)	0.28 (0.45)
Immigrant %	0.22 (0.41)	0.23 (0.42)	0.84 (0.37)	0.16 (0.37)	0.83 (0.38)
*European%	0.95 (0.11)	0.25 (0.17)	0.39 (0.27)	0.80 (0.25)	0.61 (0.21)
*African%	0.02 (0.07)	0.70 (0.18)	0.05 (0.10)	0.09 (0.20)	0.09 (0.11)
*Amerindian%	0.01 (0.04)	0.01 (0.03)	0.01 (0.04)	0.08 (0.14)	0.27 (0.20)
*East Asian%	0.02 (0.07)	0.03 (0.04)	0.55 (0.26)	0.03 (0.09)	0.03 (0.06)
N	3861	326	182	67	726

Note: Standard deviations are reported in parentheses; *Genetic ancestry is the average of both biological parents.

3.2 Main regression results

Table 2 displays the results for the analysis involving education. In the first model, excluding ancestry variables, the unstandardized coefficients of Black, Asian, Native American, Hispanic, and Other race/ethnicity are b = -0.55, 0.22, -0.14, -0.64, and -0.42, respectively. However, in the second model which includes ancestry variables, the unstandardized race/ethnicity coefficients are b = 0.24, -0.01, -0.04, 0.01, and -0.17, respectively. This finding indicates that ancestry statistically explains the effect of socially-defined race/ethnicity and, more importantly, that minority race/ethnic categories, apart from Other race, are not associated with lower educational attainment levels as compared to White identity once genetic ancestry is controlled for. In fact, Black is associated with higher educational attainment than White race/ethnicity. When examining the coefficients of the genetic ancestry variables, the African and Amerindian ancestry variables are negatively related to education (b = -1.25 and b = -3.27, respectively), while East Asian ancestry shows a weak positive association (but non-significant) with education as compared to European ancestry (b = 0.13, p = .363).

Table 2. Admixture Regression Results for Parental Educational Attainment.

	Model 1		**Model 2**	
Predictors	*b*	*P*	*b*	*p*
(Intercept)	-0.06 (0.06)	0.355	0.20 (0.06)	**0.001**
Black	-0.55 (0.05)	**<0.001**	0.24 (0.10)	**0.012**
East Asian	0.22 (0.07)	**0.001**	-0.01 (0.09)	0.957
Native American	-0.14 (0.09)	0.145	-0.04 (0.09)	0.656
Hispanic	-0.64 (0.05)	**<0.001**	0.01 (0.06)	0.864
Other Race	-0.42 (0.06)	**<0.001**	-0.17 (0.06)	**0.003**
English Fluency	0.21 (0.05)	**<0.001**	0.00 (0.05)	0.993
Age	0.22 (0.01)	**<0.001**	0.18 (0.01)	**<0.001**
Immigrant Family	0.10 (0.03)	**0.002**	0.15 (0.03)	**<0.001**
African ancestry			-1.25 (0.13)	**<0.001**
Amerindian ancestry			-3.27 (0.14)	**<0.001**
East Asian ancestry			0.13 (0.14)	0.363
Random Effects				
σ^2	0.69		0.61	
τ_{00}	0.03 $_{\text{site\_id\_l}}$		0.02 $_{\text{site\_id\_l}}$	
ICC	0.04		0.03	
N	22 $_{\text{site\_id\_l}}$		22 $_{\text{site\_id\_l}}$	
Observations	5073		5073	
Marginal R^2 / Conditional R^2	0.207 / 0.239		0.297 / 0.319	

Note: standard errors are reported in parentheses

Table 3 displays the result for the analysis involving income. In the first model, excluding ancestry variables, the unstandardized coefficients of Black, Asian, Native American, Hispanic, and Other race/ethnicity are b = -0.70, 0.20, -0.21, -0.53, and -0.32, respectively. However in the second model including ancestry variables, the unstandardized race/ethnicity coefficients are b = 0.25, 0.18, -0.16, -0.04, and -0.11, respectively. Once more, these findings indicate that ancestry statistically explains the effect of race/ethnicity categories, except in the case of Asian and Native American race/ethnicity. When examining the coefficients of the genetic ancestry variables, African and Amerindian ancestry variables are negatively related to income (b = -1.48 and b = -2.51 respectively). East Asian ancestry also shows a negative association (but non-significant) with income as compared to European ancestry (b = -0.23, p = .102).

Table 3. Admixture Regression Results for Parental Income.

	Model 1		**Model 2**	
Predictors	*b*	*P*	*b*	*p*
(Intercept)	-0.17 (0.07)	**0.013**	0.06 (0.07)	0.421
Black	-0.70 (0.05)	**<0.001**	0.25 (0.09)	**0.010**
East Asian	0.20 (0.07)	**0.003**	0.18 (0.09)	0.051
Native American	-0.21 (0.09)	**0.019**	-0.16 (0.09)	0.076
Hispanic	-0.53 (0.05)	**<0.001**	-0.04 (0.06)	0.519
Other Race	-0.32 (0.06)	**<0.001**	-0.11 (0.06)	0.067
English Fluency	0.35 (0.05)	**<0.001**	0.17 (0.05)	**<0.001**
Age	0.18 (0.01)	**<0.001**	0.15 (0.01)	**<0.001**
Immigrant Family	0.08 (0.03)	**0.005**	0.14 (0.03)	**<0.001**
African ancestry			-1.48 (0.13)	**<0.001**
Amerindian ancestry			-2.51 (0.14)	**<0.001**
East Asian ancestry			-0.23 (0.14)	0.102
Random Effects				
σ^2	0.66		0.61	
τ_{00}	0.05 $_{\text{site\_id\_l}}$		0.05 $_{\text{site\_id\_l}}$	
ICC	0.08		0.08	

N	22 site_id_l	22 site_id_l
Observations	5073	5073
Marginal R^2 / Conditional R^2	0.196 / 0.256	0.259 / 0.319

Note: standard errors are reported in parentheses

3.3 Results from the Robustness analyses

We ran a series of robustness tests using variations of the main models, displayed in Tables 2 and 3. Detailed results from these additional analyses are provided in the supplementary file.[2] First, we exclude all cases with values of education and income 3 (or more) *SD* below the mean so as to ensure that our results are not driven mainly by extreme values. The results from these analyses are very similar to those from the main analyses except that, in the models without ancestry, Hispanic and Black identity is associated with slightly less worse socioeconomic outcomes and in the models with genetic ancestry both African and Amerindian ancestries are somewhat less negatively associated with outcomes.

A second set of analyses repeats the main regression models using child's race/ethnicity (as reported by the parents) instead of the responding parent's race/ethnicity. The results of these analyses are also very similar to the main results except that Black identity shows a slightly lower positive coefficient in the regression analysis for educational attainment. A third set of analyses restricts the sample to families with at least one (as opposed to two) biological parent. Doing so yields a much larger sample (N = 7,652) due to the high rate of single-parenthood among Black and Hispanic identifying individuals. The results are similar to those from the main analyses except that Black identity is less positively associated with outcomes and that Native American identity is more negatively associated with outcomes.

A fourth set of analyses adds state-level racism and anti-immigrant bias. Both racism and anti-immigrant bias are weak predictors of outcomes and so had little effect on the results. The results for the second through fourth analyses are summarized in Table 4, alongside those from Tables 2 and 3.

As seen in Table 4, when ancestry is controlled for, Black identity is statistically significantly positively associated with outcomes in the majority of models. Since a large portion of the structural racism literature focuses on discrimination against individuals socially identified as Black it would be worthwhile to see if these results replicate on other samples. Given these counterintuitive results, before such replication is done (our Black sample is small in our analysis), we are reluctant to speculate on possible causes.

Finally, we replicate the full model of our main analyses within the combined Hispanic, Black, and Native American sample. In the educational attainment model, the

[2] In the supplementary file, we also provided the bivariate correlation of European ancestry with income and with educational attainment for two biological Black families (respectively, .19 and .18) and for two biological Hispanic families (respectively, .43 and .54).

coefficients for Black, Asian, Native American, and Hispanic identity are, respectively, *b* = 0.56 (*p* = .034), 0.24 (*p* = .592), and 0.35 (*p* = .134), whereas the coefficients for African, Amerindian, and East Asian ancestry are, respectively, *b* = -1.55 (*p* < .001), -3.21 (*p* < .001), and -1.01 (*p* = .104). In the income model, the coefficients for Black, Asian, Native American, and Hispanic identity are, respectively, *b* = 0.05 (*p* = .872), 0.10 (*p* = .836), and -0.20 (*p* = .440), whereas the coefficients for African, Amerindian, and East Asian ancestry are *b* = -1.59 (*p* <.001), -2.54 (*p* <.001), and -0.23 (*p* = .733), respectively. As seen, the effect of race/ethnicity is similar to that in the main sample. Additionally, Black identity is not negatively associated with socioeconomic outcomes after controlling for genetic ancestry.

Table 4. Summary of Admixture Regression Results for Parental Education and Income.

	Main models		Models with child race		Models with => 1 biological parent		Models with state-level racism variables	
	Educ.	Income	Educ.	Income	Educ.	Income	Educ.	Income
Black dummy	**0.24**	**0.25**	0.11	**0.25**	0.11	0.08	**0.24**	**0.24**
Asian dummy	-0.01	0.18	0.00	**0.19**	0.11	0.15	0.00	**0.18**
Native American dummy	-0.04	-0.16	-0.07	**-0.17**	**-0.13**	**-0.17**	-0.04	-0.15
Hispanic dummy	0.01	-0.04	-0.04	0.00	0.01	-0.05	0.01	-0.04
Other Race dummy	**-0.17**	-0.11	-0.09	-0.05	**-0.15**	**-0.17**	**-0.17**	-0.11
African ancestry	**-1.25**	**-1.48**	**-1.13**	**-1.55**	**-1.06**	**-1.59**	**-1.24**	**-1.48**
Amerindian ancestry	**-3.27**	**-2.51**	**-3.23**	**-2.57**	**-2.84**	**-1.92**	**-3.28**	**-2.52**
East Asian ancestry	0.13	-0.23	0.13	-0.22	0.02	-0.26	0.13	-0.24
N	5073	5073	5073	5073	7652	7652	5073	5073

4. Discussion

The research examined whether racial differences in educational attainment and income are associated with socially-identified race/ethnicity independent of genetic ancestry. We found that, after controlling for genetic ancestry, Hispanic and Native American identities, as compared to White identity, are not associated with lower educational attainment and that Native American identity is associated with only slightly lower income levels. Additionally, when genetic ancestry is held constant, Black identity is associated with higher education and income than White identity. The results hold when the children's race/ethnicity are used instead of those of the responding parents. Furthermore, the results hold when subsetting to only Hispanic, Black, and Native American individuals, a finding which is consistent with the findings of a meta-analysis of American studies (Kirkegaard et al., 2017).

The structural racism hypothesis, as commonly formulated, clearly predicts that socially-defined race/ethnicity will have an effect on educational attainment and income independent of genetic ancestry. This is because most of the specific laws, policies, institutional practices, and entrenched social norms in the USA, which could have adversely affected non-White groups, did not target individuals based on genetic ancestry, but rather did so based on socially defined race/ethnicity (Gichoya et al., 2022; Smedley & Smedley, 2005). An alternative is the cognitive meritocracy hypothesis. According to this, educational and income differences are mostly due to general cognitive ability and other human capital differences. Since a number of studies have shown that general cognitive ability tracks genetic ancestry better than socially-defined race and genetically predicted color (e.g., Kirkegaard et al., 2019), this model would predict that social outcome differences, being antecedent to cognitive ability ones, also follows genetic ancestry.

Fuerst, Wang, & Kirkegaard (2017) conducted a large meta-analysis of genetic epidemiological studies on the relation between European, African, and Amerindian ancestry and indices of SES. The authors found that, with a high degree of consistency, European ancestry was positively associated with better outcomes and that African and Amerindian ancestry was negatively associated with better outcomes. For the most part, these studies did not examine the independent effect of socially-defined race on outcomes. However, the authors report results from one study from Brazil which examines the effects of both interviewer and participant-reported color/race ("cor") on household assets, schooling, and income. Independent of European ancestry, racial/color identification was not statistically significantly associated with outcomes, while European ancestry was strongly associated with better SES outcomes. These results from Brazil, then, are consistent with the ones reported in this paper.

One major issue related with our data is that, as we only had admixture estimates for children, we have limited the sample to dyads who were biological parents. As a result the non-White sample sizes were modest. Moreover, since dual parenthood is positively related to socioeconomic status, the samples used are not representative of American populations. This should, therefore, be seen as an exploratory study. There are large datasets containing genetic data and adult educational attainment and income on which admixture regression analyses could be run in the future to determine if socially-defined race/ethnicity has predictive validity independent of genetic ancestry. Since many researchers very clearly argue that the constructive aspects of race/ethnicity are strongly related to socioeconomic outcomes, this research is worth pursuing to better understand the nature of race/ethnic related socioeconomic disparities in the USA.

A major issue with the study design is that factors other than discrimination, such as assortative mating and selective ethnic attrition, can potentially induce associations between social outcomes and socially defined race. For example, mixed-race couples could be socioeconomically selective, and this selectivity could lead to social race being correlated with SES independent of genetic ancestry. Were these processes influencing our results, we would generally expect socially identified races to be associated with outcomes independent of ancestry. However, we generally do not find this to be the case. Nonetheless, it is theoretically possible that the effects of discrimination could be moderated by countervailing effects of assortative mating and selective ethnic attrition. This leaves open the question as to whether the pattern of mating could have explained the positive coefficient of Black identity on socio-economic outcomes. It might be argued that Black mothers who intermarry typically achieve higher SES levels than White mothers who intermarry, therefore weighing up the impact of Black identity. This requires a rate high enough to create a positive effect of b=0.24 (as found in this study). At the same time, not having information about the racial identification of the spouse of the responding parent does not help to address this issue. Whether this is an empirically plausible scenario depends on the pattern of mating and identification in American race/ethnic groups and is a subject for future research.

Author contributions: J.F. conceived of the idea. Analyses used the ABCD sample were conducted by J.F. under the supervision of B. J. Pesta while at Cleveland State University (2020-2021). J. F, M.H, and E.O.W. K edited and revised the manuscript. Both authors discussed the results and contributed to the final manuscript.

Acknowledgments: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data, but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 3.01. The raw data are available at https://nda.nih.gov/edit_collection.html?id=2573. Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc. NDA waved the requirement of creating a NDA data project (NDA, Oct 26, 2022).

References

Adkins-Jackson, P. B., Chantarat, T., Bailey, Z. D., & Ponce, N. A. (2022). Measuring structural racism: a guide for epidemiologists and other health researchers. *American Journal of Epidemiology*, *191(*4), 539-547. doi: 10.1093/aje/kwab239

American Sociological Association. 2003. *The Importance of Collecting Data and Doing Social Scientific Research on Race.* Washington, DC: American Sociological Association.

Bailey, Z. D., Feldman, J. M., & Bassett, M. T. (2021). How structural racism works—racist policies as a root cause of US racial health inequities. *New England Journal of Medicine*, *384*(8), 768-773. doi: 10.1056/nejmms2025396

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Grothendieck, G. (2009). Package 'lme4'. Retrieved from: https://lme4.r-forge.r-project.org/.

Bowdler, J. & Harris, B. (July, 21, 2021). Racial Inequality in the United States. U.S. Department of the Treasury. Retrieved from: https://home.treasury.gov/news/featured-stories/racial-inequality-in-the-united-states

Braveman, P. A., Arkin, E., Proctor, D., Kauh, T., & Holm, N. (2022). Systemic And Structural Racism: Definitions, Examples, Health Damages, And Approaches To Dismantling: Study examines definitions, examples, health damages, and dismantling systemic and structural racism. *Health Affairs*, *41*(2), 171-178. doi: 10.1377/hlthaff.2021.01394

Brues, A. M. (1990). *People and races*. Prospect Heights, IL: Waveland.

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of african americans, latinos, and european Americans across the United States. *The American Journal of Human Genetics*, *96*(1), 37-53. doi: 10.1016/j.ajhg.2014.11.010

Connor, G. & Fuerst, J.G.R. (2022). Linear and partially linear models of behavioral trait variation using admixture regression. In: R. Lynn (ed.), *Intelligence, Race and Sex: Some Controversial Issues. A Tribute to Helmuth Nyborg at 85*. Budapest, Hungary: Arktos. doi: 10.1101/2021.05.14.444173

Erikson, C. E., Dent, R. B., Park, Y. H., & Luo, Q. (2022). Historic redlining and contemporary behavioral health workforce disparities. *JAMA Network Open*, *5*(4), e229494-e229494. doi: 10.1001/jamanetworkopen.2022.9494

Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T. L., Huang, J., ... & Tang, H. (2019). Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *The American Journal of Human Genetics*, *105*(4), 763-772. doi: 10.1016/j.ajhg.2019.08.012

Fuerst, J. G. (2021). Robustness analysis of African genetic ancestry in admixture regression models of cognitive test scores. *Mankind Quarterly, 62*(2). doi: 10.46469/mq.2021.62.2.12

Fuerst, J. G., Hu, M., & Connor, G. (2021). Genetic ancestry and general cognitive ability in a sample of American youths. Mankind Quarterly, 62(1), 186-216. doi: 10.46469/mq.2021.62.1.11

Fuerst, J. G., Shibaev, V., & Kirkegaard, E. O. (2023). A Genetic Hypothesis for American Race/Ethnic Differences in Mean g: A Reply to Warne (2021) with Fifteen New Empirical Tests Using the ABCD Dataset. Mankind Quarterly, 63(4). doi: 10.46469/mq.2023.63.4.2

Gichoya, J. W., Banerjee, I., Bhimireddy, A. R., Burns, J. L., Celi, L. A., Chen, L. C., ... & Zhang, H. (2022). AI recognition of patient race in medical imaging: a modelling study. *The Lancet Digital Health*, *4*(6), e406-e414. doi: 10.1016/s2589-7500(22)00063-2

Guo, G., Fu, Y., Lee, H., Cai, T., Li, Y., & Harris, K. M. (2014). Recognizing a small amount of superficial genetic differences across African, European and Asian Americans helps understand social construction of race. *Demography*, *51*(6), 2337-2342. doi: 10.1007/s13524-014-0349-y

Haeny, A. M., Holmes, S. C., & Williams, M. T. (2021). The need for shared nomenclature on racism and related terminology in psychology. *Perspectives on Psychological Science*, *16*(5), 886-892. doi: 10.1177/17456916211000760

Herd, P., Mills, M. C., & Dowd, J. B. (2021). Reconstructing sociogenomics research: Dismantling biological race and genetic essentialism narratives. *Journal of Health and Social Behavior*, *62*(3), 419-435. doi: 10.1177/00221465211018682

Jones, C. P. (2002). Confronting institutionalized racism. *Phylon (1960-)*, 7-22. doi: 10.2307/4149999

Kirkegaard, E. O. W. (2021). Genetic ancestry and social race are nearly interchangeable. *OpenPsych*. doi: 10.26775/op.2021.12.22

Kirkegaard, E. O., & Fuerst, J. G. (2023). A Multimodal MRI-based Predictor of Intelligence and Its Relation to Race/Ethnicity. *Mankind Quarterly*, *63*(3). doi: 10.46469/mq.2023.63.3.2

Kirkegaard, E.O.W.; Wang, M.; Fuerst, J. (2017). Biogeographic Ancestry and Socioeconomic Outcomes in the Americas: A Meta-Analysis. *Mankind Quarterly*, 57(3), 398–427. doi: 10.46469/mq.2017.57.3.10

Kirkegaard, E. O., Williams, R. L., Fuerst, J., & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych*, *1*(1), 1-25. doi: 10.3390/psychology1010001

Lasker, J., Pesta, B. J., Fuerst, J. G., & Kirkegaard, E. O. (2019). Global ancestry and cognitive ability. *Psych, 1*(1), 431-459. doi: 10.3390/psych1010034

Lumley, T. (2020). Package 'survey'. Retrieved from: https://cran.r-project.org.

Noguera, P. A., & Alicea, J. A. (2020). Structural racism and the urban geography of education. *Phi Delta Kappan, 102*(3), 51-56. doi: 10.1177/0031721720970703

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register*, *62*(210), 58782-58790.

O'Reilly, K. B. (2020). AMA: Racism is a threat to public health. *American Medical Association.*

Robbins, N. M., Charleston, L., Saadi, A., Thayer, Z., Codrington, W. U., Landry, A., ... & Hamilton, R. (2022). Black Patients Matter in Neurology: Race, Racism, and Race-Based Neurodisparities. *Neurology*, *99*(3), 106-114. doi: 10.1212/wnl.0000000000200830

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American Psychologist*, *60*(1), 16.

Suyemoto, K. L., Donovan, R. A., & Kim, G. S. (2022). *Unraveling assumptions: A primer for understanding oppression and privilege*. Taylor & Francis. doi: 10.4324/9780429059599