Research notebook for biracial identification project

Contents

Crosstabs of Racial Identification

First, lets look at the cross-tabulation of father's and mother's race from the roster reports.

Table 1: Father's reported race (row) by mother's reported race (column)

	White	Black	AmIndian	Asian	Hispanic
White	3441	8	27	27	138
Black	64	1630	1	4	32
AmIndian	21	3	18	0	2
Asian	13	0	0	106	0
Hispanic	147	35	8	5	1143

Obviously a lot of action on the diagonals. The cell-sizes get pretty small when you move off the white father or white mother lines. Some biracial groups are not captured at all.

Now lets look at the full table of racial self-identification in 2002 based on the parental race identified from the roster data.

Table 2: Cross-tabulation of parentally-defined race (row) by racial self-identification in 2002 with consistent racial self-identification in bold.

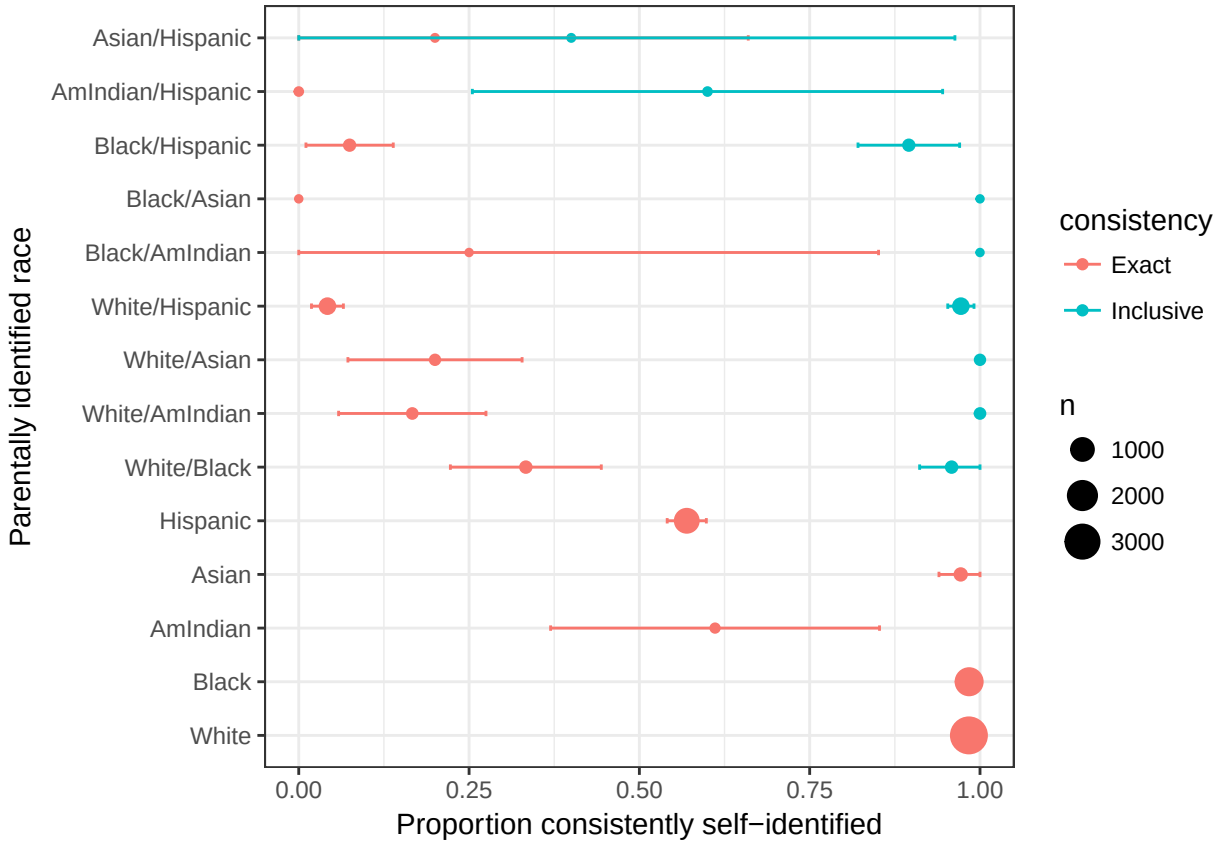
	W	B	I	A	H	WB	WI	WA	WH	BI	BA	BH	AH	WBI	WIH
W	***3385***	4	8	5	12	5	16	3	3	0	0	0	0	0	0
B	10	***1604***	0	1	0	6	0	0	0	3	1	3	0	2	0
I	2	0	***11***	3	0	0	2	0	0	0	0	0	0	0	0
A	2	0	0	***103***	0	0	0	0	0	0	0	0	1	0	0
H	447	9	16	6	***651***	0	8	0	5	0	0	0	1	0	0
WB	*3*	*42*	0	0	1	***24***	0	0	0	1	1	0	0	***0***	0
WI	*30*	0	*10*	0	0	0	***8***	0	0	0	0	0	0	***0***	***0***
WA	*16*	0	0	*16*	0	0	0	***8***	0	0	0	0	0	0	0
WH	*212*	0	1	2	*53*	0	5	0	***11***	0	0	0	0	0	***1***
BI	0	*3*	*0*	0	0	0	0	0	0	***1***	0	0	0	***0***	0
BA	0	*3*	0	*1*	0	0	0	0	0	0	***0***	0	0	0	0
BH	5	*46*	0	0	*9*	2	0	0	0	0	0	***5***	0	0	0
IH	3	0	*5*	0	*1*	0	0	1	0	0	0	0	0	0	***0***
AH	0	0	0	*1*	*0*	0	0	3	0	0	0	0	***1***	0	0

Obviously there is some heterogeneity in racial self-ID for all groups, but more for multiracial groups. For all

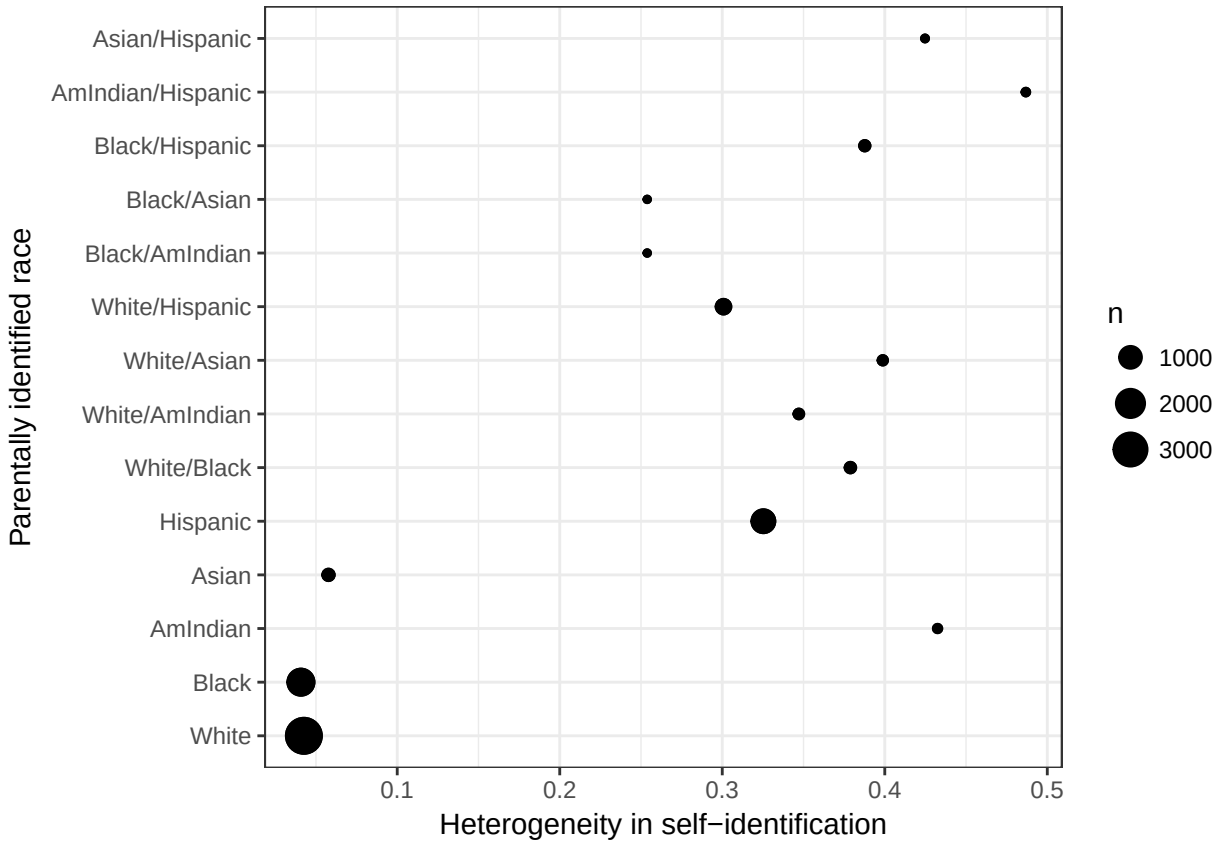
multiracial groups, consistent ID is not the modal category whereas it is the modal category for single-race groups.

To look at this more closely, lets construct a couple of figures. First, lets construct a figure that shows the proportion of consistently identified individuals by parental race group with error bars.

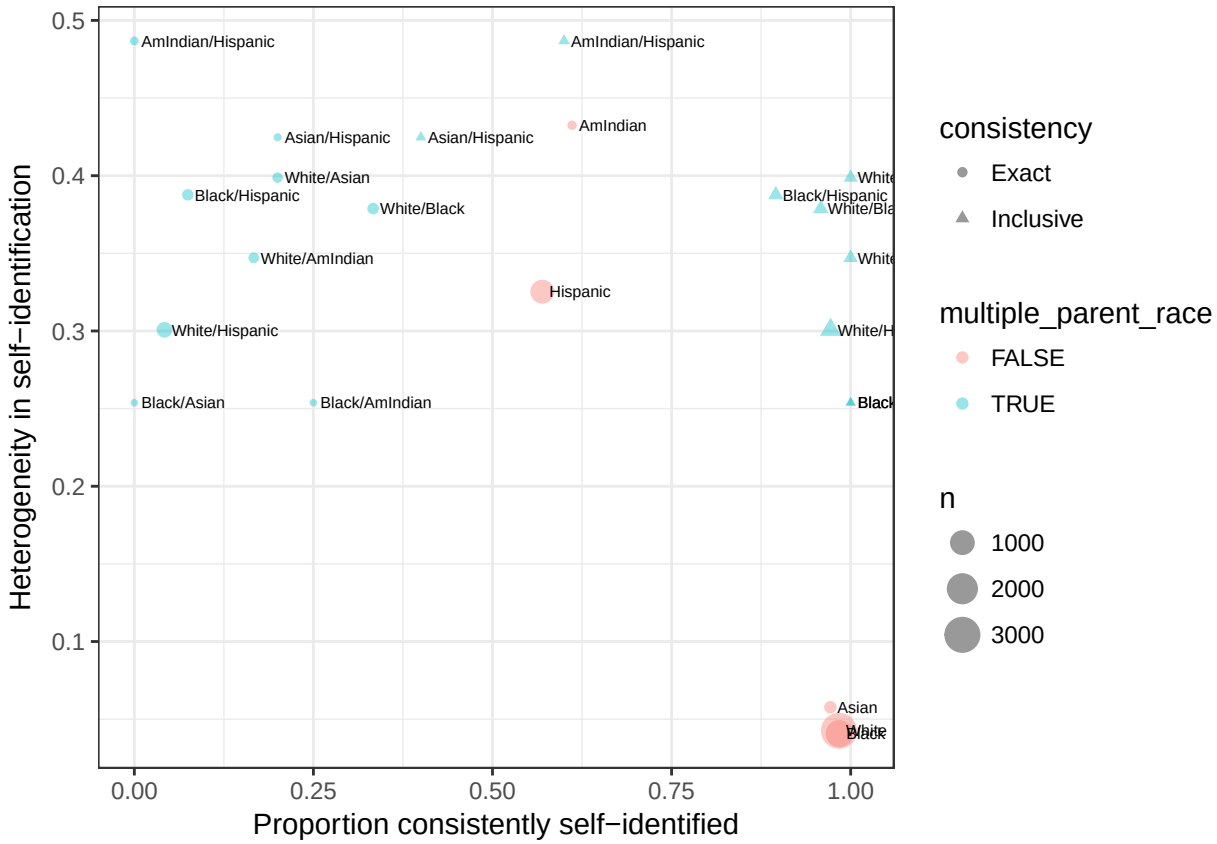
Figures showing racial consistency and heterogeneity of self ID



Now lets consider a measure of entropy which gets more at overall heterogeneity regardless of consistency.



How well does heterogeneity correlate with inconsistency?



Models

Now I will focus on models predicting self-identification consistent with parentally identified race. The models will be HLM models with random intercepts for each race group and potentially random slopes for some individual characteristics. Group-level variables will also be entered and may be interacted with individual level variables.

All models are based on a multiple imputation with five complete datasets.

Random Intercept Models

The structure of the basic model is as follows:

$$logit(p_{ij}) = \beta_{0j} + \sum_{k=1}^{p} \beta_k (x_{ijk} - \bar{x}_{..k})$$

$$\beta_{0j} = \alpha_0 + \alpha_1(\texttt{hispanic parent}_j) + \alpha_2(\texttt{mixed race parentage}_j) + \upsilon_j$$

The dependent variable is the the logit of the probability of consistent reporting of race (p_ij) for the ith respondent belonging to the jth parental race group. This outcome is then predicted by a set of individual level characteristics. Importantly, the model also includes a random intercept that allows the probability to vary by parental race group. This random intercept is then predicted in a second level equation by a combination of dummies for having a hispanic parent and being of mixed race parentage. The remaining υ_j random intercepts provide important information about differences in the reporting of racial consistency after accounting for a general lower level of consistency among individuals of hispanic and mixed-race parentage,

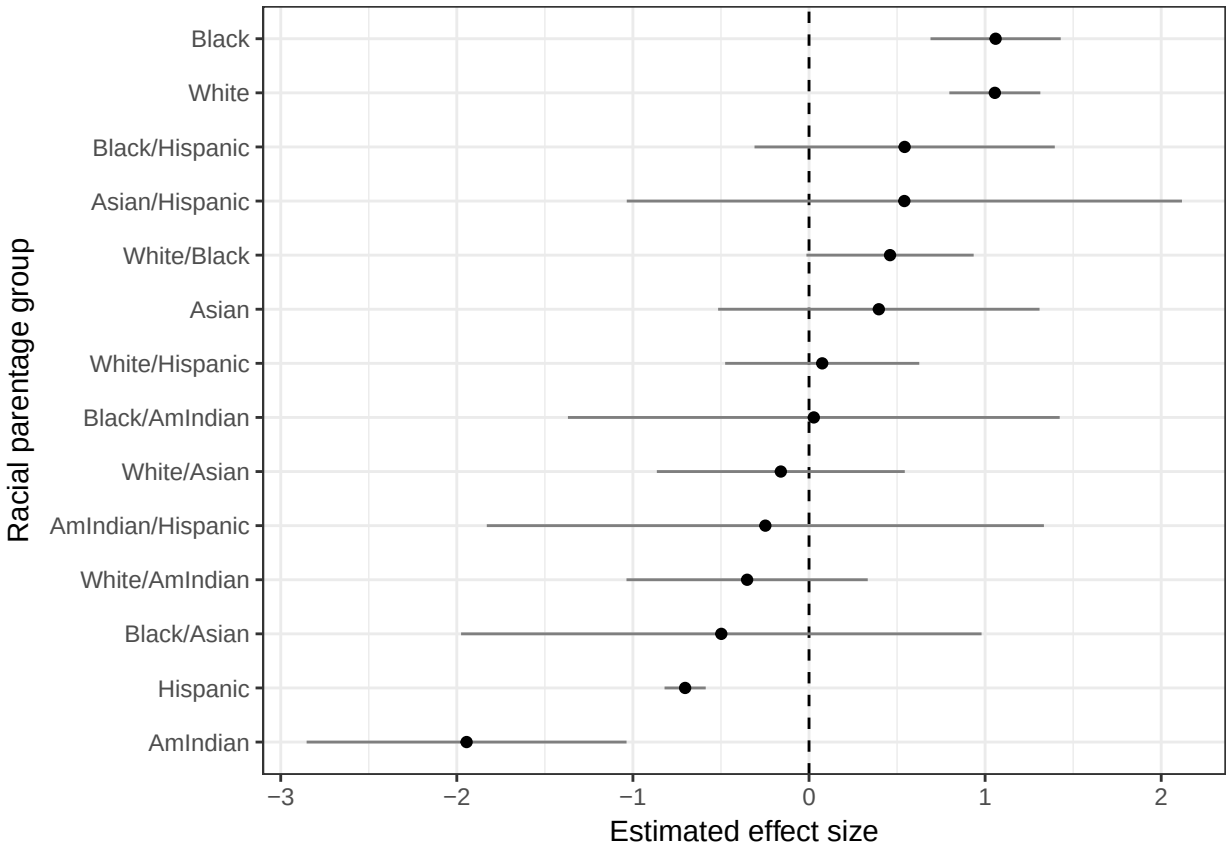


Figure 1: Random intercepts of self-identification consistency for racial parentage groups, after accounting for a general effect of mixed race parentage and hispanic parentage.

some of which may be driven by survey instrument bias. All individual level independent variables are grand-mean centered, so the random intercepts "control" for these individual level characteristics.

I ran into some convergence problems with some of the more complex models here. I found that if I updated the models with the starting parameters from the last fit, as suggested here, the models generally converged, although in some cases it required doing this update multiple times.

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, Mar 07, 2018 - 15:13:48

Models of white/black identification

Ok, now try some models trying to predict whiteness/blackness among subset of respondents with some white/black ancestry

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, Mar 07, 2018 - 15:14:30

Table 3: Multilevel logistic regression model predicting whether self-identified race is the same as parentally-identified race.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Multiple race parentage		−4.209**	−4.251**	−4.188**	−4.210**	−4.219**	−4.213**
		(0.588)	(0.573)	(0.591)	(0.585)	(0.576)	(0.576)
Hispanic parent		−2.020**	−2.033**	−2.104**	−2.009**	−2.123**	−2.132**
		(0.635)	(0.619)	(0.639)	(0.632)	(0.622)	(0.622)
Informant, biological parent		0.251	0.254	0.257	0.247	0.262	0.267
		(0.194)	(0.196)	(0.195)	(0.195)	(0.198)	(0.198)
Informant, other		0.183	0.159	0.178	0.195	0.180	0.181
		(0.225)	(0.226)	(0.226)	(0.225)	(0.229)	(0.229)
Age			−0.006			−0.011	−0.011
			(0.036)			(0.039)	(0.039)
Two-biological parent household			−0.224*			−0.199	−0.197
			(0.106)			(0.121)	(0.120)
Moved out of county			0.312*			0.319*	0.326*
			(0.147)			(0.148)	(0.149)
Moved out of parental home			−0.274*			−0.260*	−0.260*
			(0.112)			(0.115)	(0.115)
Highest parental yrs of educ				−0.021		−0.035*	−0.033
				(0.017)		(0.018)	(0.018)
Household income (logged)				−0.130*		−0.134*	−0.131*
				(0.058)		(0.065)	(0.066)
No high school diploma					−0.079	−0.201	−0.204
					(0.129)	(0.135)	(0.149)
College degree or enrollment					0.131	0.196	0.204
					(0.122)	(0.127)	(0.138)
Enrolled in high school					−0.019	−0.133	−0.143
					(0.179)	(0.197)	(0.198)
CAT-ASVAB score							−0.063
							(0.085)
High school GPA							0.060
							(0.170)
Female							
Urban area							
Constant	−0.484	2.787**	2.823**	2.797**	2.786**	2.825**	2.817**
	(0.745)	(0.490)	(0.479)	(0.491)	(0.488)	(0.481)	(0.481)
SD of random intercept	2.67	0.9	0.87	0.91	0.9	0.88	0.88
Number of racial parentage groups	14	14	14	14	14	14	14
N	6,873	6,873	6,873	6,873	6,873	6,873	6,873

Reference categories are non-two biological parent household,
respondent as parent race informant, stayed in the same county,
and high school diploma not enrolled.
All individual level variables are mean-centered.
All models include random intercepts for racial parentage groups.

Table 4: Multilevel logistic regression model predicting whether self-identified race is white or black among racial parentage groups with white or black parent.

	White alone	Black alone
	Model 1	Model 2
Multiple race parentage	-5.100^{**}	-3.698^{**}
	(1.268)	(0.324)
Hispanic parent	2.113	0.062
	(1.268)	(0.380)
Informant, biological parent	-0.437	-1.160
	(0.529)	(1.081)
Informant, other	-1.001	-1.873
	(0.576)	(1.104)
Age	0.029	0.098
	(0.069)	(0.108)
Two-biological parent household	-0.431^{*}	0.106
	(0.208)	(0.330)
Moved out of county	0.309	-0.505
	(0.238)	(0.330)
Moved out of parental home	0.350	-0.0001
	(0.204)	(0.313)
Highest parental yrs of educ	0.045	-0.119
	(0.037)	(0.066)
Household income (logged)	0.283^{*}	-0.111
	(0.114)	(0.190)
No high school diploma	-0.395	0.894^{*}
	(0.246)	(0.397)
College degree or enrollment	0.133	0.018
	(0.219)	(0.324)
Enrolled in high school	0.304	-0.027
	(0.380)	(0.566)
Constant	4.534^{**}	5.441^{**}
	(1.200)	(1.087)
SD of random intercept	1.07	0
Number of racial parentage groups	5	5
N	3,886	1,777

Reference categories are non-two biological parent household, respondent as parent race informant, stayed in the same county, and high school diploma not enrolled.
All individual level variables are mean-centered.
All models include random intercepts for racial parentage groups.

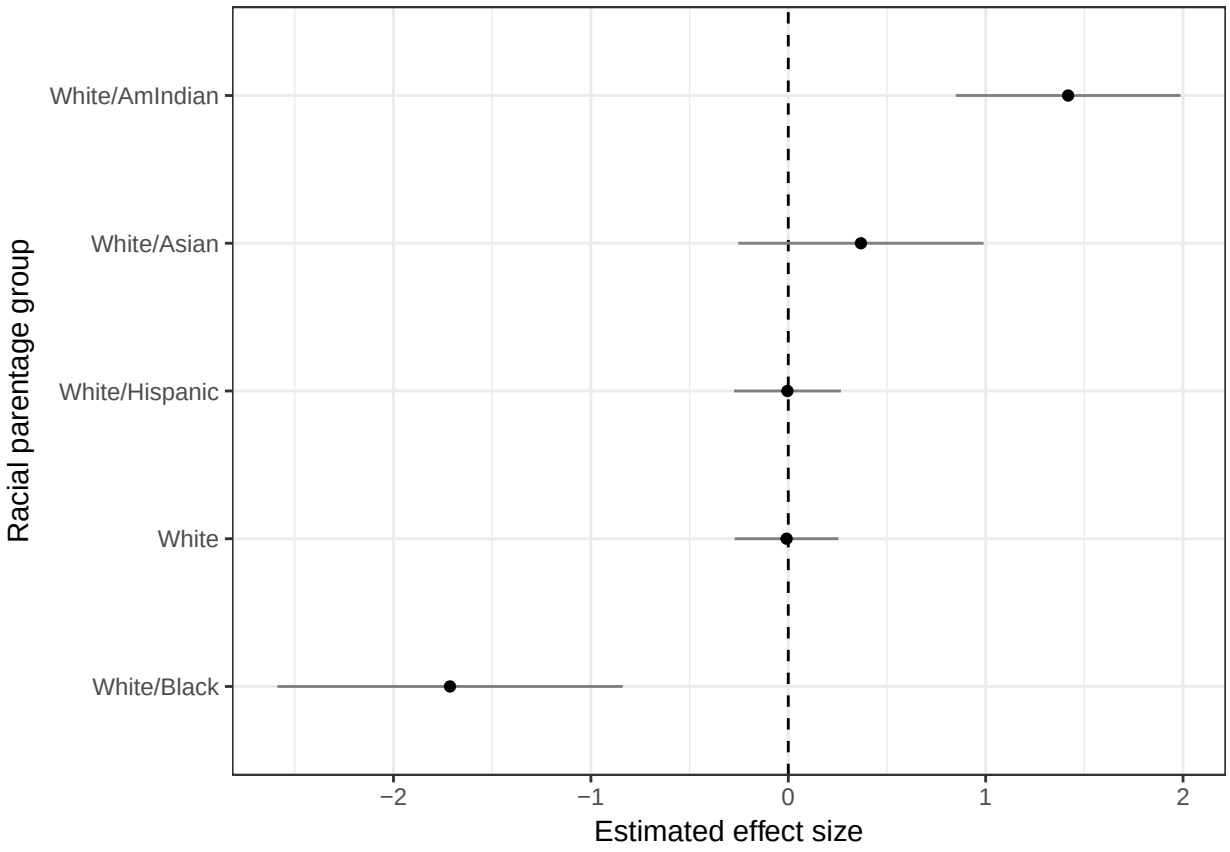


Figure 2: Catterpillar plot of racial parentage random intercepts predicting white identification for individuals with at least one white parent.