Polygenic scores, genetic engineering, validity of GWAS results across major racial groups and the Piffer method

EMIL O. W. KIRKEGAARD

A PDF of this paper without formatting errors can be downloaded here.

Abstract

I review recent findings in human behavioral genetics and their implications for selective breeding and estimation of genotypic racial differences in polygenic traits.

Key words: behavioral genetics, cognitive ability, GWAS, intelligence, IQ, race, selective breeding, embryo selection, genetic engineering, educational attainment

### 1. POLYGENIC SCORES FROM ALL SNPS VS. P<A SNPS

A recent paper (1) used polygenic scores derived from the Rietveld results (2) to score a non-overlapping sample of European Americans (EA) and African Americans (AA). They found that polygenic scores predicted educational outcomes for samples at r's = .18 and .11 for EAs and AAs respectively. In terms of variance, this corresponds to 3.24% and 1.21%, respectively. This is small, but not useless. They don't report confidence intervals, only p value inequalities, so it isn't so easy to see how precise these estimates are (3). The p value inequalities for the two results are p<.001 and <0.01. Note that sample sizes are different too. The main results table is shown below.

TABLE 3
Model Estimates of Polygenic Score on Educational Attainment

	EA respondents					AA respondents				
	Estimate	SE	pv	N	Model r^2	Estimate	SE	pv	N	Model r^2
Model 1: $\widehat{\beta_U}$	0.37	0.08	7.5E-07	917	0.06	0.20	0.09	2.1E-02	677	0.02
Model 2A: $\widehat{\beta_{U'}}$	0.30	0.07	3.1E-05	901	0.16	0.22	0.09	1.0E-02	671	0.04
Model 2B: $\widehat{\beta_{U'}}$	0.29	0.08	1.4E-04	762	0.23	0.14	0.09	1.2E-01	556	0.12
Model 2: $\widehat{\beta_{U'}}$	0.26	0.07	5.4E-04	752	0.26	0.14	0.09	1.2E-01	555	0.12
Model 3: $\widehat{\beta_W}$	0.35	0.11	2.3E-03	808	0.74					

Note. Model 1 captures the estimated effect of the polygenic score in unrelated individuals (where the standard error is adjusted for the clustering) adjusting for age and the top 10 PCs. Model 2 adds controls related to parental education and neighborhood disadvantage to Model 1. Model 2A and Model 2B are, respectively, restricted versions of Model 2 where the parental education coefficient and neighborhood coefficients are forced to zero, respectively. Model 3 focuses within family using sibling fixed effects (still adjusting for the age of the respondents).

These findings are interesting because they use polygenic scores instead of scores derived from just the findings that surpass the NHST threshold, i.e. those that have a p value below the alpha value (p<α).[1] Using the full set of betas instead of just the set with p<α results in better predictions. It has even been found that differential weighting of the SNPs does not have a major effect of the predictive power of the deriving polygenic scores (4).
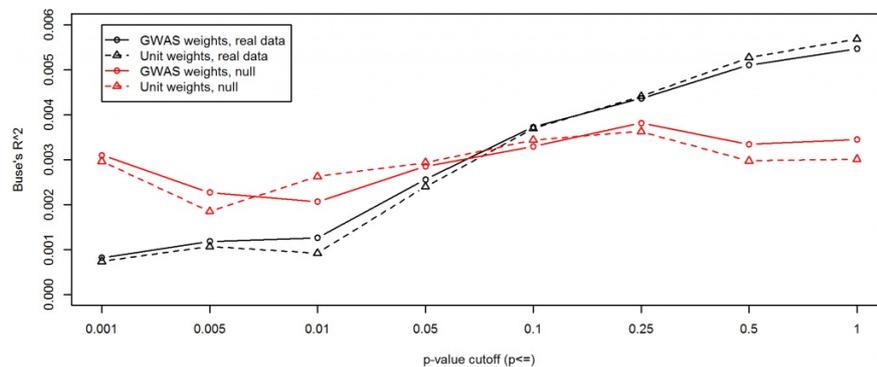
This should be seen in light of conceptually related results in psychometrics where it has been shown that it doesn't matter much if one uses g factor scores, simple sums or even randomly weighted subtests (5). The general mathematical explanation for this is that when one creates a linear combination (i.e. adds together) many variables, the common variance ('the signal') adds up while the unshared variance ('the noise') does not. Thus, the more variables one averages, the more more signal in the noise (simplifying a bit). The general idea goes back at least to 1910, when Spearman and Brown independently derived a formula for it (6). Their papers were published in the same journal, even in the same issue (7,8). Another example of multiple discovery/invention.

Focusing on the number of SNPs with p<α for a trait is the wrong metric to think of. One should instead think of the found correlation (or other effect size measure if outcome data is categorical) between polygenic scores and outcomes for cross-validation samples. Thinking of SNPs where p<α is dichotomous thinking instead of continuous thinking. When using dichotomous models for phenomena that really is continuous, one will get threshold effects that bias the effect sizes d\nownwards.

2. INCONVENIENT RESULTS CAN BE MADE TO GO AWAY (MAYBE)

Since the study contained both a EA and an AA sample with mean IQs of 105.1 and 94.3, it should be possible to derive polygenic scores for members of both groups and compare the mean of the groups. This would be a test of the genetic hypothesis for the well known cognitive ability difference between the groups (9–11). There are two things worth noting, however.

First, the group difference is only 10.8 IQ, smaller than the usual gap found. There is some question as to whether the gap has been changing over time (12,13). Some newer samples find smaller gaps especially those based on WORDSUM scores (14), while others find standard (~1 SD, 15 IQ) sized gaps (15). The smaller than expected gap in the samples may result from selection bias in the AA sample (presumably it is difficult to recruit very low S inner city AAs for scientific studies) Note that results are generally weaker for this sample, which is expected given restriction of range.

Second, not all persons in the groups were genotyped. Those that were had lower mean IQs of 103.9 and 91.6, respectively. This gives a gap of 12.3 points. Note that the reason the EA score is not 100, is that the overall Add Health sample mean is set to ~100 (100.6).

Despite these caveats, the polygenic scores would be interesting to see. However, the authors decided to standardize the results within each group, such that the mean of the polygenic scores was 0 for both groups. The of course makes any group difference impossible to see. They provide the following rationale:

The 917 European Americans (EAs) in our analytic sample are in 386 sibling pairs and 12 sibling trios, with an additional 109 singletons. The 677 African Americans (AAs) are in 100 sibling pairs and four trios, with an additional 465 singletons. Table 1 shows characteristics of the EA and AA sibling pairs study participants who provided genetic data and constitute our analytic sample. The table also shows characteristics of the full Add Health EA and AA samples for comparison. The EAs in our analytic sample are largely comparable to the full population of EA respondents in the Add Health study. The AAs in our sample are less educated, have less educated parents, and score lower on the verbal intelligence measure as compared to all AA Add Health participants. The bulk of our analysis is focused on the EA sample because the original Rietveld et al. (2013) GWAS was conducted on European-descent individuals. Replication of polygenic scores discovered in EA samples among AA samples may be compromised because LD differences in the groups lead to less precision among AA samples. Accordingly, large-scale GWASs of educational attainment in African Americans will be needed to better quantify genetic influences on attainment in this population. Nevertheless, in the interest of testing the extent to which findings made in European-descent individuals replicate in a different population, we conduct several analyses of the AA sample. Due to the small number of AA sibling pairs in the data, sibling analyses are conducted only in EAs.

The rationale is not entirely unreasonable, but not sufficient reason not to standardize the polygenic

scores for both samples together. In my opinion, the reason they provide should be taken into account when *interpreting* the results, but is not sufficient for not *showing* the results. My guess is that they did calculate the scores for both groups and compared them. Upon finding that the AA sample had a lower mean polygenic score than the EA sample, they decided that result was too toxic to publish. Reverse publication bias in effect. See also this post. A respected academic acquaintance of mine contacted the authors but they refused to share the results.

Lastly, one can use the combined sample to investigate whether the data shows a Simpson's paradox pattern. The lack of a such pattern is a central finding of Fuerst's and my upcoming paper (16). Jensen's default hypothesis (17) predicts the absence of such a pattern since the same genetic causes are postulated to be involved in the within race differences as those between them.

3. POLYGENIC SCORES AND SIBLING PAIRS

Another interesting aspect of the study is that they have sibling data. Since siblings receive a random mix of genes from their parents, they will differ in their genotypic for polygenic traits. This was also found in this sample: "The mean sibling difference in polygenic scores in the EA sample was 0.8." (they did not calculate this for the AA sample, stating that it was too small). In other words, the difference between siblings is nearing the size of the mean difference in the whole population. The same result is known to be true for siblings and IQ scores. The mean difference is about 11 IQ compared to a full sample mean difference of 17 IQ (17). This gives a ratio of 11/17 = .65. Since the educational attainment data is standardized, we know that the mean difference in scores is 1.13 (Fuerst posted the formula here, but I'm not sure about the source). This gives a ratio of 1.13/.8 = .71. These ratios are pretty close as they should be.

We care about sibling comparisons because they by design control for shared environment effects, so that we don't need to control them statistically (18). The authors found that results held within sibling pairs, an important finding. The table below is from their paper:

TABLE 2
Correlations Between Educational Attainment, Polygenic Score, and Other Key Variables in EA and AA Samples

	EA Respondents		AA Respondents	
	Correlation with W4 education	Correlation with polygenic score	Correlation with W4 education	Correlation with polygenic score
Years of education at W4		**.18*****		**.11****
Verbal intelligence	.36***	.14***	.36***	.15***
Years of maternal education	.42***	.05	.32***	.12**
Neighborhood disadvantage	−.35***	−.13***	−.14***	.00
Polygenic score	.18***		.11**	

Note. EA = European American; AA = African American; W4 = Wave 4.
† $p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

As we will see below, this has another important practical implication.

4. GENETIC ENGINEERING AND CAUSAL VARIANTS

Since socially valued outcomes have non-zero heritability (19), it means that it is in theory possible to improve the outcomes by genetic means, just as we have done for animals. I see two main routes to do this: selection among possible children and direct editing.

The first method is widely used but so far only for a small number of traits. When two persons want a child, the usual method involves having sex and producing a fetus. As mentioned above, this fetus will have a random combination of genes from the parents. If the same parents produce a different combination we call it a sibling.

Selective abortion involves screening fetuses in the womb for anomalies and aborting ones sufficiently undesirable. Probably the most common target for this is Down's syndrome, which is substantially reduced due to the high rate of abortions when it is detected (20,21). For Denmark, the abortion rate given detection is 99%.

Selective abortion is better than nothing but it is not a good method. Not only is it painful for the woman, but it is inefficient because one has to wait until one can perform a prenatal screening. At that point, the fetus is many weeks old. Furthermore, abortions can result in infertility.

Embryo selection is the natural extension of the same idea. Instead of selectively aborting fetuses, we select between embryos (fertilized eggs). Essentially, we choose an embryo and implant it. This illustration shows how this works.

The second and best option for genetic engineering is to edit the genes directly. In that we one could potentially create a genome free of known flaws. This would involve using something like CRISPR.

The problem with direct editing is that we need to know the actual causal variants. This is not required for selection among possible children. Here is it sufficient that we can make predictions. The difference

here is that the SNPs we know are in most cases probably not the causal variants. Instead, they are proxies for the causal variants because the are in linkage disequilibrium (LD) with them. In simple terms, the reason for this is that the mixing of gene variants from sexual reproduction (meiosis) happens at random, but *in chunks*. Thus, gene variants that are located closer to each other in the genome tend to travel together during splits. This means that they get correlated, which we call LD.

Since practical use of embryo selection requires working on sibling embryos, it is necessary that we can make genomic predictions among siblings that work. The new paper showed that we can do this for educational attainment.

5. REPLICABILITY OF GWAS RESULTS ACROSS RACIAL GROUPS

There are two matters. The first is to which degree the genetic architecture of polygenic traits is similar across racial groups, i.e. if the same genes cause traits across populations or if there is substantial race-level gene-gene interaction (epistasis). The second is the degree to which SNP betas derived from one race can be used to make valid predictions for another race.

For polygenic traits that have been under the selection for many thousands of years (e.g. cognitive ability or height (22)), I think substantial race-level gene-gene interaction is implausible. They are however plausible for traits that involve a small number of genes and show substantial race differences, such as those for hair, eye and skin color.

LD patterns change over time. Since the LD patterns change independently and randomly in each population, they will tend to become different with time.

If the GWAS SNPs owe their predictive power to being actual causal variants, then LD is irrelevant and they should predict the relevant outcome in any racial group. If however they owe wholly or partly their predictive power to just being statistically related to causal variants, they should be relatively worse predictors in racial groups that are most distantly related. One can investigate this by comparing the predictive power of GWAS betas derived from one population on another population. Since there are by now 1000s of GWAS, meta-analyses have in fact made such comparisons, mostly for disease traits. Two reviews found substantial cross-validity for the Eurasian population (Europeans and East Asians), and less for Africans (usually African Americans) (23,24). The first review only relied on SNPs with p<α and found weaker results. This is expected because using only these is a threshold effect, as discussed earlier.

The second review (from 2013; 299 included GWAS) found much stronger results, probably because it included more SNPs and because they also adjusted for statistical power. Doing so, they found that: ~100% of SNPs replicate in other European samples when accounting for statistical power, ~80% in East Asian samples but only ~10% in the African American sample (not adjusted for statistical power, which was ~60% on average). There were fairly few GWAS for AAs however, so some caution is needed in interpreting the number. Still, this throws some doubt on the usefulness of GWAS results from Europeans or Asians used on African samples (or reversely).

Which brings us back to…

6. LOW CROSS-VALIDITY OF GWAS BETAS AND POLYGENIC SCORES FOR EDUCATIONAL ATTAINMENT IN AAS

Despite the relatively weak evidence for European sample derived GWAS betas in Africans, the study mentioned in the beginning of this review (1) still found a reliable polygenic correlation of .11 in AAs. However, AAs are an admixed group that are about 75-85% African and 25-15% European (25,26). The exact admixture proportions depend on the selectivity of the sample. Bryc et al used the 23andme database which represents individuals willing to pay to have their genomes sequenced. Since this requires both money (price is about 100$ for US citizens) and interest in genetic results, this will lead to selection for S (27) and cognitive ability. Both traits are known to correlate with European admixture at the individual, region and country levels (16), which would then result in higher proportions of European admixture in AA sample. Shriver et al's sample is more representative and found mean proportions of 78.7% and 18.6% for African and European ancestry respectively.

If we make the assumption that the polygenic correlation for educational attainment in the AA sample is purely due to the European admixture, we can make a prediction for the effect size, namely that it should be about 20% of the size of that for Europeans. I'm not sure but I think that in this case one should use the proportion of variance, not correlation coefficient. Recall that these were 3.24% and 1.21% (r's .18 and .11), which gives a ratio of .37. This is higher than the expected value of .186. This means that there is an excess validity of .187 in the African part of their genome under the null model. We can use this to make an estimation of the cross-racial validity. Since we have accounted for AAs European admixture, the rest of the predictive power must come from the African admixture (ignoring Native American admixture for simplicity), which constitutes 78.7%. This gives an estimated cross-racial validity ratio of about .24 (0.187/.787). In a pure African sample, this corresponds to an estimated correlation coefficient of .09 (sqrt(.18$_2$ * .24)). Future studies will reveal how far off these estimates are,

but most importantly, they are *quantitative* predictions, not merely qualitative (directional) (28).

7. POOR AFRICAN-EURASIAN CROSS-VALIDITY AND THE PIFFER METHOD

The findings related to the relatively poor, but non-zero cross-validity of GWAS betas between
European and African samples throw some doubt on the SNP evidence found by Piffer in his studies
of the population/country IQ and cognitive ability SNP factors (29). If the betas for the SNPs identified
in European sample GWAS do not work well as predictors for Africans, they would be equally
unsuitable for estimating mean genotypic cognitive ability from SNP frequencies. Thus, further
research is needed to more precisely estimate the cross-racial validity of GWAS betas, especially with
regards to African vs. Eurasian samples.

References

1. Domingue BW, Belsky DW, Conley D, Harris KM, Boardman JD. Polygenic Influence on Educational
Attainment. AERA Open. 2015 Jul 1;1(3):2332858415599972.

2. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559
Individuals Identifies Genetic Variants Associated with Educational Attainment. Science. 2013 Jun
21;340(6139):1467–71.

3. Cumming G. The New Statistics Why and How. Psychol Sci. 2014 Jan 1;25(1):7–29.

4. Kirkpatrick RM, McGue M, Iacono WG, Miller MB, Basu S. Results of a "GWAS Plus:" General
Cognitive Ability Is Substantially Heritable and Massively Polygenic. PLoS ONE. 2014 Nov
10;9(11):e112390.

5. Ree MJ, Carretta TR, Earles JA. In Top-Down Decisions, Weighting Variables does Not Matter: A
Consequence of Wilks' Theorem. Organ Res Methods. 1998 Oct 1;1(4):407–20.

6. Carroll JB. Human cognitive abilities: A survey of factor-analytic studies [Internet]. Cambridge
University Press; 1993 [cited 2015 Jun 3]. Available from: www.google.com/books?
hl=en&lr=&id=i3vDCXkXRGkC&oi=fnd&pg=PR7&dq=Carroll,+1993+human+co\ngnitive+abilities&ots=3b3O4R_IKc&sig=wOss3EHXu37Q3

7. Spearman C. Correlation Calculated from Faulty Data. Br J Psychol 1904-1920. 1910 Oct
1;3(3):271–95.

8. Brown W. Some Experimental Results in the Correlation of Mental Abilities1. Br J Psychol 1904-
1920. 1910 Oct 1;3(3):296–322.

9. Fuerst J. Ethnic/Race Differences in Aptitude by Generation in the United States: An Exploratory
Meta-analysis. Open Differ Psychol [Internet]. 2014 Jul 26 [cited 2014 Oct 13]; Available from:
openpsych.net/ODP/2014/07/ethnicrace-differences-in-aptitude-by-generation-in-the-unit\ned-states-
an-exploratory-meta-analysis/

10. Rushton JP, Jensen AR. Thirty years of research on race differences in cognitive ability. Psychol
Public Policy Law. 2005;11(2):235–94.

11. Fuerst J. The facts that need to be explained [Internet]. Unwelcome Discovery. 2012 [cited 2015
Aug 31]. Available from: z139.wordpress.com/2012/06/10/the-facts-that-need-to-be-explained/

12. Fuerst J. Secular Changes in the Black-White Cognitive Ability Gap [Internet]. Human Varieties.
2013 [cited 2015 Aug 31]. Available from: humanvarieties.org/2013/01/15/secular-changes-in-the-
black-white-cognitive-ability-gap\n/

13. Malloy J. The Onset and Development of B-W Ability Differences: Early Infancy to Age 3 (Part 1)
[Internet]. Human Varieties. 2013 [cited 2015 Aug 31]. Available from:
humanvarieties.org/2013/05/26/the-onset-and-development-of-b-w-ability-differences-ear\nly-infancy-
to-age-3-part-1/

14. Hu M. An update on the secular narrowing of the black-white gap in the Wordsum vocabulary test
(1974-2012) [Internet]. 2014 [cited 2015 Aug 31]. Available from: osf.io/hiuzk/

15. Frisby CL, Beaujean AA. Testing Spearman's hypotheses using a bi-factor model with WAIS-
IV/WMS-IV standardization data. Intelligence. 2015 Jul;51:79–97.

16. Fuerst J, Kirkegaard EOW. Admixture in the Americas. In London, UK.; 2015. Available from:
docs.google.com/presentation/d/1hjhOiitk0MnqMHgVthyj8j7qa4qcAqPUDNaTT8rpetg/e dit?
pli=1#slide=id.p

17. Jensen AR. The g factor: the science of mental ability. Westport, Conn.: Praeger; 1998.

18. Murray C. IQ and income inequality in a sample of sibling pairs from advantaged family
backgrounds. Am Econ Rev. 2002;339–43.

19. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015 May 18;47(7):702–9.

20. Natoli JL, Ackerman DL, McDermott S, Edwards JG. Prenatal diagnosis of Down syndrome: a systematic review of termination rates (1995-2011): Prenatal diagnosis of down syndrome: systematic review. Prenat Diagn. 2012 Feb;32(2):142–53.

21. de Graaf G, Buckley F, Skotko BG. Estimates of the live births, natural losses, and elective terminations with Down syndrome in the United States. Am J Med Genet A. 2015 Apr;167A(4):756–67.

22. Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T, et al. Directional dominance on stature and cognition in diverse human populations. Nature. 2015 Jul 23;523(7561):459–62.

23. Ntzani EE, Liberopoulos G, Manolio TA, Ioannidis JPA. Consistency of genome-wide associations across major ancestral groups. Hum Genet. 2011 Dec 20;131(7):1057–71.

24. Marigorta UM, Navarro A. High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. PLoS Genet [Internet]. 2013 Jun [cited 2015 Aug 31];9(6). Available from: www.ncbi.nlm.nih.gov/pmc/articles/PMC3681663/

25. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet. 2015 Jan 8;96(1):37–53.

26. Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet. 2003 Feb 11;112(4):387–99.

27. Kirkegaard EOW, Fuerst J. Educational attainment, income, use of social benefits, crime rate and the general socioeconomic factor among 71 immigrant groups in Denmark. Open Differ Psychol [Internet]. 2014 May 12 [cited 2014 Oct 13]; Available from: openpsych.net/ODP/2014/05/educational-attainment-income-use-of-social-benefits-crime-r\nate-and-the-general-socioeconomic-factor-among-71-immmigrant-groups-in-denmark/

28. Velicer WF, Cumming G, Fava JL, Rossi JS, Prochaska JO, Johnson J. Theory Testing Using Quantitative Predictions of Effect Size. Appl Psychol Psychol Appl. 2008 Oct;57(4):589–608.

29. Piffer D. A review of intelligence GWAS hits: their relationship to country IQ and the issue of spatial autocorrelation [Internet]. 2015 [cited 2015 Aug 2]. Available from: figshare.com/articles/A_review_of_intelligence_GWAS_hits_their_relationship_to_country\n_IQ_and_the_issue_of_spatial_autocorrelation_

Notes

1For GWAS the alpha value is usually set at $5*10^{-8}$. The number comes from correcting the standard $\alpha=.05$ (95% theoretical true positive rate) for multiple testing when using SNP data: $.05 * 1e^{-6} = 5e^{-8}$.