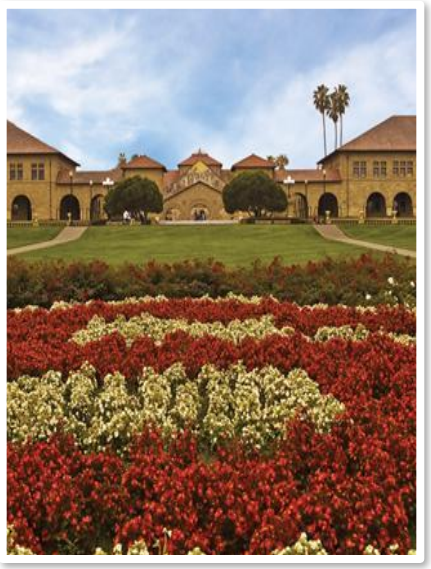
Schools & the Genetics of Educational Attainment: Evidence from Add Health

BEN DOMINGUE



Research Possibilities with Add Health Genetic Data

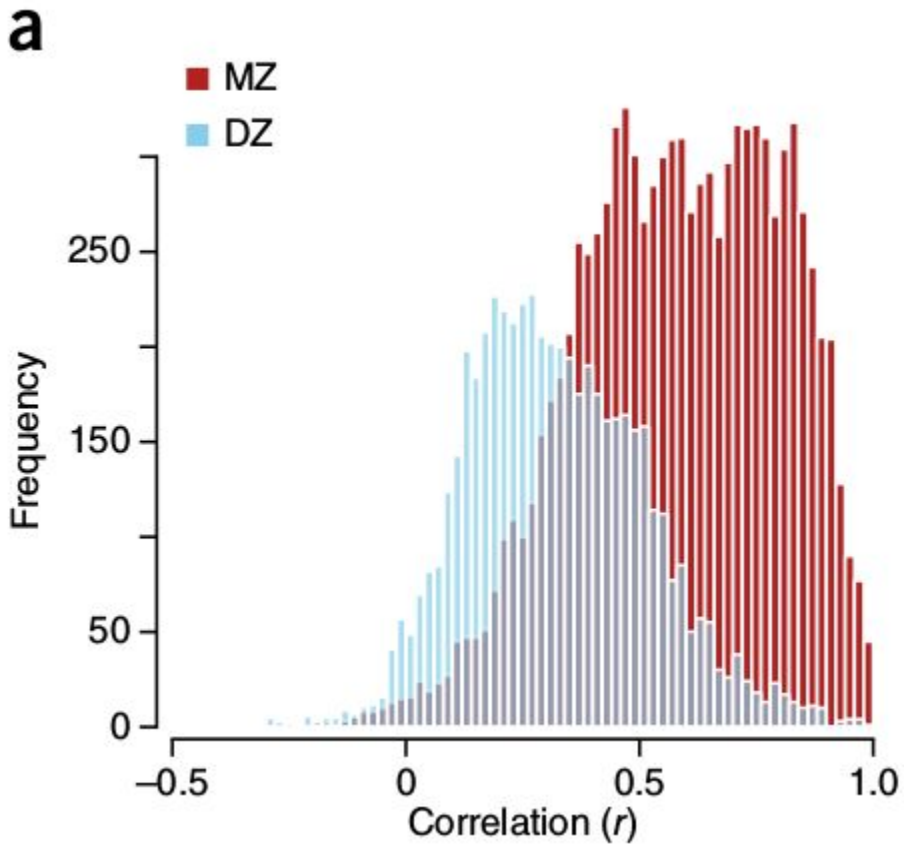
1. “Chip” heritability (GCTA)
2. Polygenic Scores
3. Mendelian Randomization

GOAL: Offer a structure for thinking about how genetically informed research can be done with Add Health.

1. “Chip” heritability with GCTA

1. “Chip” heritability with GCTA

- Twins
  - MZ twins should be more similar than DZ twins.

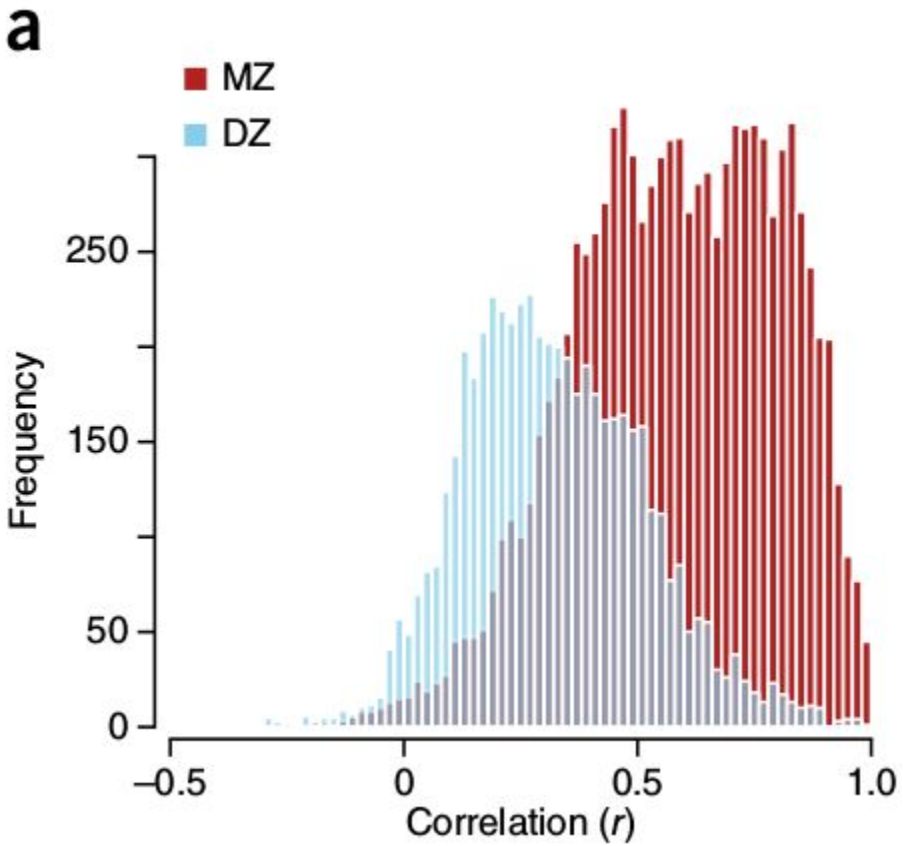


Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman[1,10], Beben Benyamin[2,10], Christiaan A de Leeuw[1,3], Patrick F Sullivan[4–6], Arjen van Bochoven[7], Peter M Visscher[2,8,11] & Danielle Posthuma[1,9,11]

1. “Chip” heritability with GCTA

- Twins
  - MZ twins should be more similar than DZ twins.
- Molecular gen
  - Genetic similarity →phenotypic similarity



Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman[1,10], Beben Benyamin[2,10], Christiaan A de Leeuw[1,3], Patrick F Sullivan[4–6], Arjen van Bochoven[7], Peter M Visscher[2,8,11] & Danielle Posthuma[1,9,11]

GCTA: A two-step process

- Step 1: Compute genetic similarities.
  - Many metrics for genetic similarity are available. GCTA uses something akin to a weighted correlation.
  - Need to restrict to genetically homogenous group due to way genetic similarities are computed.

GCTA: A two-step process

- Step 2. Estimate

$$y_i \sim \mathrm{N}[X_i'\beta + \gamma_i, \sigma^2]$$

where

$$\gamma \sim \mathrm{MVN}[0, \sigma_G^2 A]$$

- Covariance matrix for gamma is based on the genetic similarities estimated in step 1.
- Heritability: $\dfrac{\sigma_G^2}{\sigma_G^2 + \sigma^2}$

GCTA is popular.

Common SNPs explain a large proportion of the heritability for human **height**

J **Yang**, B Benyamin, BP McEvoy, S Gordon… - Nature …, 2010 - nature.com

SNPs discovered by genome-wide association studies (GWASs) account for only a small fraction of the genetic variation of complex traits in human populations. Where is the remaining heritability? We estimated the proportion of variance for human height ...

Cited by 1723 Related articles All 28 versions Web of Science: 1163 Cite Saved More

GCTA is popular.

Common SNPs explain a large proportion of the heritability for human **height**
J **Yang**, B Benyamin, BP McEvoy, S Gordon… - Nature …, 2010 - nature.com
SNPs discovered by genome-wide association studies (GWASs) account for only a small fraction of the genetic variation of complex traits in human populations. Where is the remaining heritability? We estimated the proportion of variance for human height ...
Cited by 1723 Related articles All 28 versions Web of Science: 1163 Cite Saved More

Biodemography and Social Biology, 62:1–18, 2016

ISSN: 1948-5565 print / 1948-5573 online
DOI: 10.1080/19485565.2015.1068106

Routledge
Taylor & Francis Group

Genome-Wide Estimates of Heritability for Social Demographic Outcomes

BENJAMIN W. DOMINGUE,[1] ROBBEE WEDOW,[2] DALTON CONLEY,[3] MATT MCQUEEN,[4] THOMAS J. HOFFMANN,[5] AND JASON D. BOARDMAN[2]

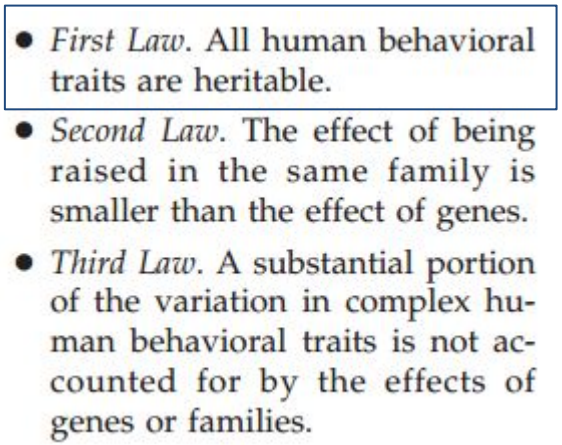
[1]Graduate School of Education, Stanford University, Stanford, California, USA
[2]Department of Sociology, Institute of Behavioral Science & Institute for Behavioral Genetics, University of Colorado–Boulder, Boulder, Colorado, USA
[3]Department of Integrative Physiology, Institute of Behavioral Science, University of Colorado-Boulder, Boulder, Colorado, USA
[4]Department of Sociology & Center for Genomics and Systems Biology, New York University, New York, New York, USA
[5]Department of Epidemiology & Biostatistics, and Institute for Human Genetics, University of California, San Francisco, San Francisco, California, USA

Turkheimer's Laws

- GCTA has been used to estimate univariate heritabilities for a great many traits.

Turkheimer's Laws

- GCTA has been used to estimate univariate heritabilities for a great many traits.



- *First Law*. All human behavioral traits are heritable.
- *Second Law*. The effect of being raised in the same family is smaller than the effect of genes.
- *Third Law*. A substantial portion of the variation in complex human behavioral traits is not accounted for by the effects of genes or families.

Three Laws of Behavior Genetics and What They Mean

Eric Turkheimer[1]

- Need to think carefully about a paper focusing simply on univariate heritability.

More intriguing possibilities

- Genetic correlation (rG)
  - Estimating pleiotropy, or the extent to which common genetics are influencing multiple traits.

- GxE Analyses
  - Examining changes in genetic influences over time.

Table 3
Bivariate genome wide covariance estimates for education and three health outcomes.

	Body mass index	Depression	Self-rated health
Genetic variance			
Health	10.668	0.007	0.128
Education	2.139	2.173	2.142
Cov (health,education)	−0.159	−0.089	−0.477
Environmental variance			
Health	14.677	0.028	0.576
Education	4.059	4.025	4.055
Cov (health,education)	−0.788	−0.003	−0.178
Phenotypic variance			
Health	25.345	0.034	0.704
Education	6.197	6.198	6.198
Heritability			
Health	0.421	0.193	0.181
Education	0.345	0.351	0.346
rG	−0.033	−0.746	−0.912
95% CI (rg)	(−0.297, .331)	(−1.0, −0.201)	(−1.0, −0.374)
logL	−14860.413	−837.535	−7089.678
logL0 (rG = 0)	−14860.429	−841.035	−7094.394
LRT	0.032	6.999	9.432
df	1	1	1
pr. <	0.4	0.004	0.001

Note: Data come from the Health and Retirement Study; $n = 4233$.

Genes as potential confounder.

What can genes tell us about the relationship between education and health?☆

Jason D. Boardman [a, *], Benjamin W. Domingue [a], Jonathan Daw [b]

[a] University of Colorado, Boulder, United States
[b] University of Alabama, Birmingham, United States

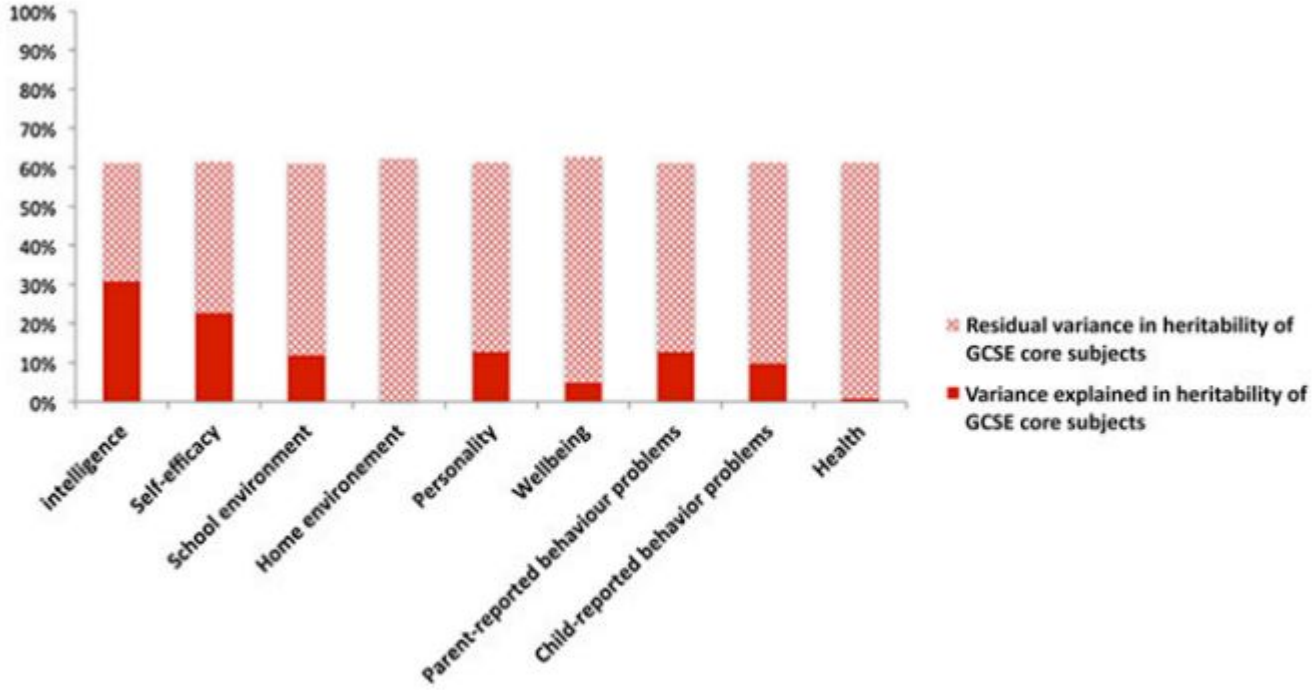
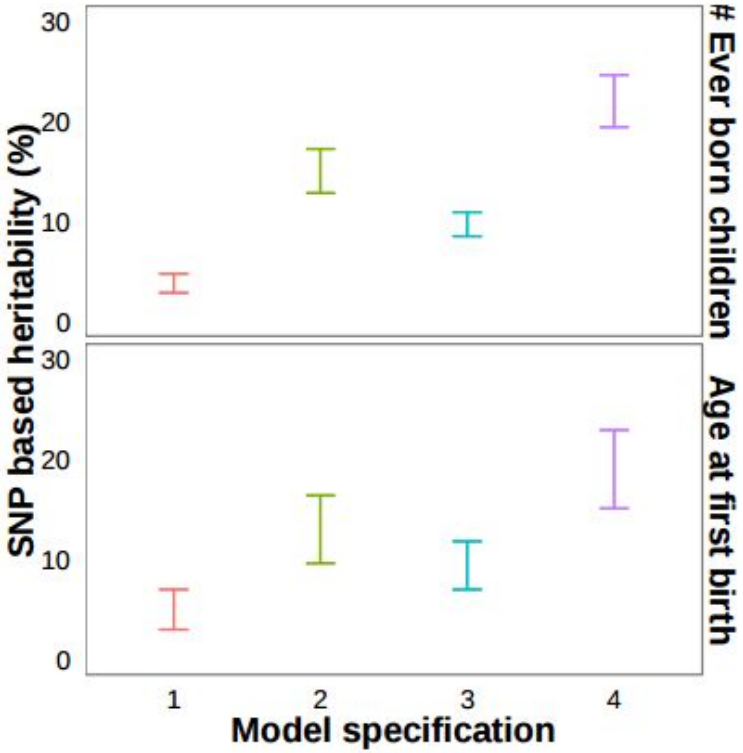


Fig. 3. Bivariate estimates of the extent to which the heritability of GCSE can be accounted for by each of the nine predictors, respectively (path a_{12} from the Cholesky decomposition; Fig. S1).

The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence

Eva Krapohl[a,1], Kaili Rimfeld[a,1], Nicholas G. Shakeshaft[a], Maciej Trzaskowski[a], Andrew McMillan[a], Jean-Baptiste Pingault[a,b], Kathryn Asbury[c], Nicole Harlaar[d], Yulia Kovas[a,e,f], Philip S. Dale[g], and Robert Plomin[a,2]

Figure 2. Bar Charts of the SNP-heritability estimates in number of children ever born (NEB) and age at first birth (AFB) for the different model specifications from Ta 1.



Note: SNP-heritability as the sum of genetic variance over the total variance in Model specification 1 = amongst all individuals, 2 = amongst individuals living within the same population, 3 = amongst individuals living within the same demographic birth cohort bc either before or after fertility postponement, 4 = amongst individuals living in the same population and demographic birth cohort, dots = estimate, lines = estimate ± 1 SE, The corresponding table to Figure 2 an be found in Supporting Table S4.

"Our findings imply that the environment strongly modifies genetic effects on the tempo and quantum of fertility, that currently ongoing natural selection is heterogeneous across environments, and that gene-environment interactions may partly account for missing heritability in fertility."

Mega-analysis of 31,396 individuals from 6 countries uncovers strong gene-environment interaction for human fertility

Felix C. Tropf[1,2*], Renske M. Verweij[2], Peter J. van der Most[3], Gert Stulp[4], Andrew Bakshi[5], Daniel A. Briley[6], Matthew Robinson[5], Anastasia Nyman[7], Tõnu Esko[8,9], Andres Metspalu[8], Sarah E. Medland[10], Nicholas G. Martin[10], Harold Snieder[3], S. Hong Lee [5,11], Melinda C. Mills[1]

A word of caution

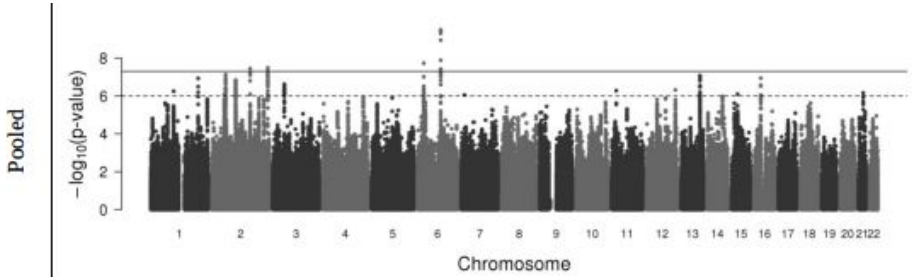
- There is a substantial back-and-forth about the statistical properties of GCTA estimates.
  - Kumar et al. Limitations of GCTA as a solution to the missing heritability problem. PNAS 2015
  - bioRxiv responses:
    - Gamazon & Park
    - Yang et al.
      - Kumar et al. have responsed to the response!

A word of caution

- There is a substantial back-and-forth about the statistical properties of GCTA estimates.
  - Kumar et al. Limitations of GCTA as a solution to the missing heritability problem. PNAS 2015
  - bioRxiv responses:
    - Gamazon & Park
    - Yang et al.
      - Kumar et al. have responsed to the response!
- Key Question: How sensitive are GCTA estimates to population stratification in relatively homogeneous samples?

2. Polygenic Scores (PGS)

- Like a credit score



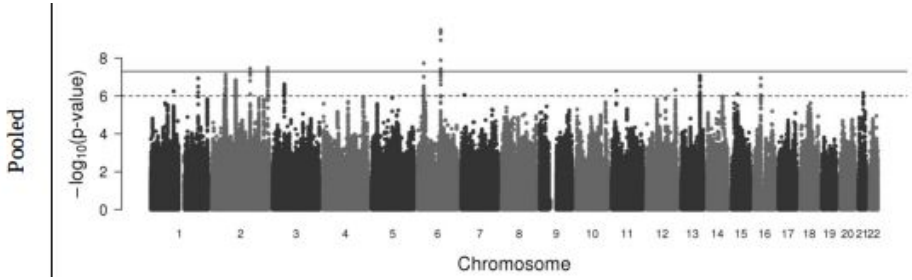
+

	SNP 1	SNP 2	$\cdots$	SNP 1,000,000
P1	0	1	$\cdots$	2
P2	1	0	$\cdots$	0
P3	1	2	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
P1000	2	1	$\cdots$	2

$1,000 \times 1,000,000$ matrix; each cell $\in \{0, 1, 2\}$.

2. Polygenic Scores (PGS)

- Like a credit score



+

	SNP 1	SNP 2	⋯	SNP 1,000,000
P1	0	1	⋯	2
P2	1	0	⋯	0
P3	1	2	⋯	1
⋮	⋮	⋮	⋱	⋮
P1000	2	1	⋯	2

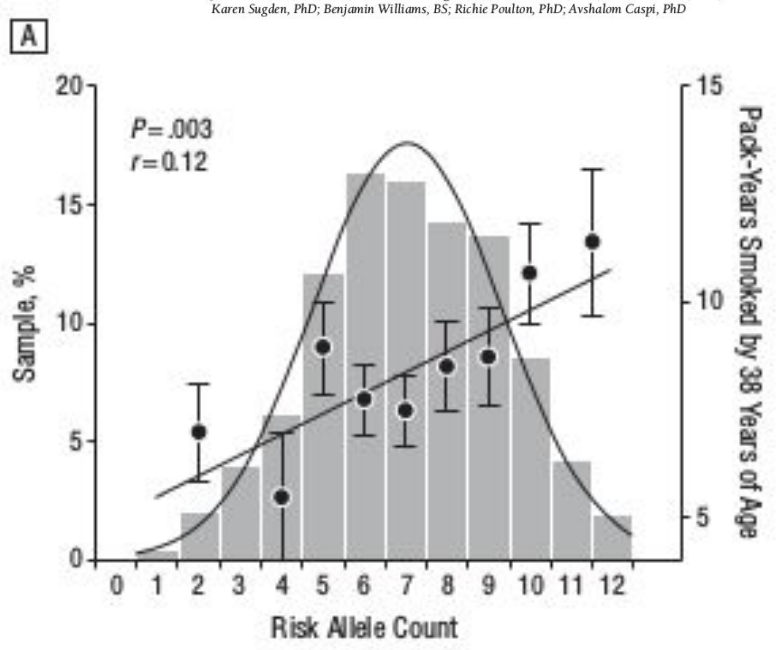
$1,000 \times 1,000,000$ matrix; each cell $\in \{0, 1, 2\}$.

ORIGINAL ARTICLE

Polygenic Risk and the Developmental Progression to Heavy, Persistent Smoking and Nicotine Dependence

Evidence From a 4-Decade Longitudinal Study

Daniel W. Belsky, PhD; Terrie E. Moffitt, PhD; Timothy B. Baker, PhD; Andrea K. Biddle, PhD; James P. Evans, MD, PhD; HonaLee Harrington, BA; Renate Houts, PhD; Madeline Meier, PhD; Karen Sugden, PhD; Benjamin Williams, BS; Richie Poulton, PhD; Avshalom Caspi, PhD



- They are predictors of the type we are accustomed to.
- Prediction >> Description

Examples

- Will go through examples of this type of research with
  - educational attainment
  - BMI
- First question: how robust is prediction in context of educational attainment PGS?

Predict out of sample

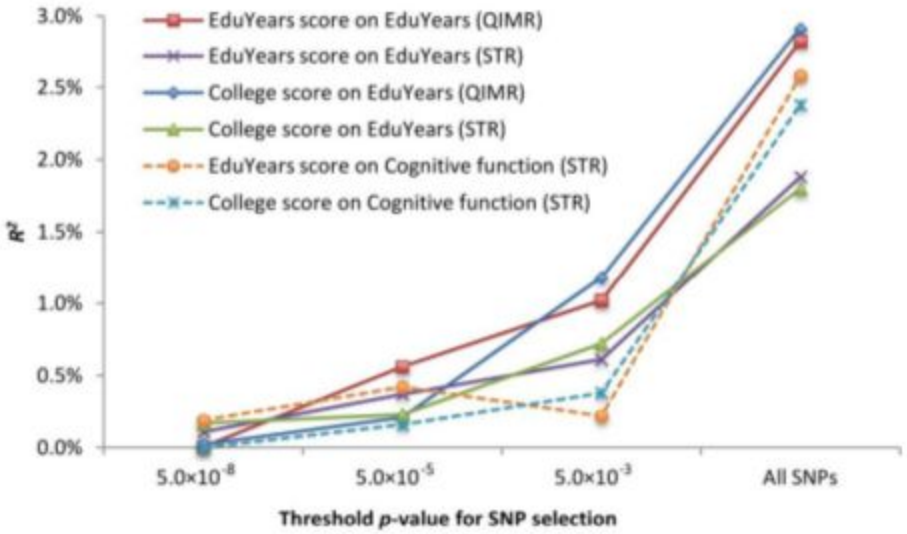


Fig. 2. Solid lines show results from regressions of *EduYears* on linear polygenic scores in a set of unrelated individuals from the QIMR (N = 3526) and STR (N = 6770) cohorts. Dashed lines show results from regressions of *Cognitive function* on linear polygenic scores in a sample from STR (N = 1419). The scores are constructed from the meta-analysis for either *EduYears* or *College*, excluding the QIMR and STR cohorts.

Important point given the failures of candidate gene studies!

Predicts net of mom's education

Table 4: Regression Models of Respondent's Total Years of Completed Education with Standard Errors Robust to Clustering on Family ID, by Sample

Health and Retirement Study	(1)	(2)	(3)
Female sex	−0.32†	−0.40†	−0.30†
	(0.06)	(0.03)	(0.06)
Age	−0.01*	−0.04†	−0.01†
	(0.00)	(0.00)	(0.00)
Survey year	0.02	0.00	0.02
	(0.01)	(0.01)	(0.01)
Mother's highest grade completed	0.30†		0.28†
	(0.01)		(0.01)
Resp. educational genetic score, std.		0.41†	0.33†
		(0.03)	(0.03)
Constant	−34.61	−84.04†	−43.16
	(26.71)	(28.04)	(26.44)
R^2	0.14	0.05	0.16
R^2 for score w/out other controls		0.03	

sociological science

Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?

Dalton Conley,[a] Benjamin W. Domingue,[b] David Cesarini,[a] Christopher Dawes,[a] Cornelius A. Rietveld,[c] Jason D. Boardman[b]

a) New York University; b) University of Colorado, Boulder; c) Erasmus University

Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits

Cornelius A. Rietveld[1,2], Dalton Conley[3], Nicholas Eriksson[4], Tõnu Esko[5], Sarah E. Medland[6], Anna A. E. Vinkhuyzen[7], Jian Yang[7], Jason D. Boardman[8,9], Christopher F. Chabris[10], Christopher T. Dawes[11], Benjamin W. Domingue[8], David A. Hinds[4], Magnus Johannesson[12], Amy K. Kiefer[4], David Laibson[13], Patrik K. E. Magnusson[14], Joanna L. Mountain[4], Sven Oskarsson[15], Olga Rostapshova[13], Alexander Teumer[16], Joyce Y. Tung[4], Peter M. Visscher[7,17], Daniel J. Benjamin[18], David Cesarini[19,20], Philipp D. Koellinger[1,2,21], and the Social Science Genetics Association Consortium

Predicts within families

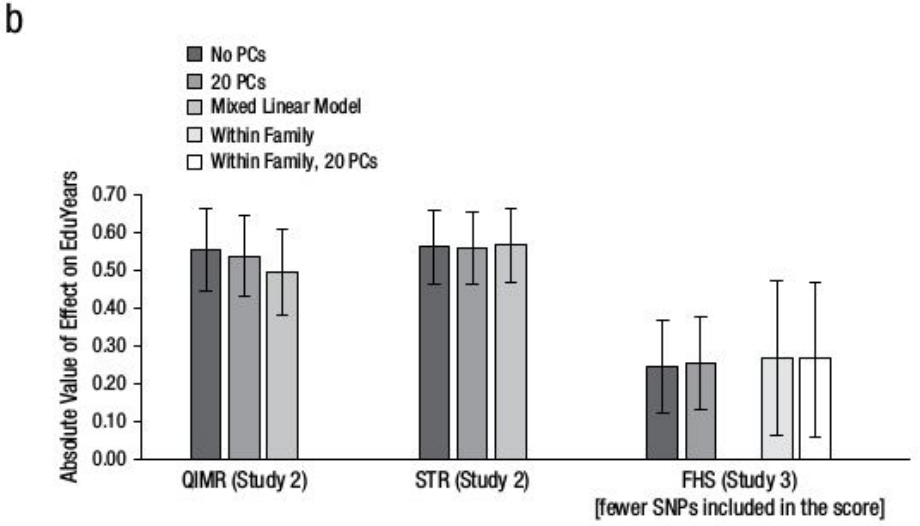


Fig. 1. Absolute value of the effect on years of schooling (EduYears) of a change in one reference allele for each of the three individual single nucleotide polymorphisms (SNPs) as a function of SNP and sample (a) and absolute value of the effect on years of schooling of a 1-*SD* change in polygenic score (including all SNPs) as a function of study and type of analysis (b). Because rs9320913 was unavailable in the 23andMe data, we used rs12206087 as a (very reliable) proxy (R^2 = .99; see Section 2.3 in the Supplemental Material). Error bars show 95% confidence intervals. QIMR = Queensland Institute of Medical Research; STR = Swedish Twin Registry; FHS = Framingham Heart Study; PC = principal component.

Predicts within families.

Polygenic Influence on Educational Attainment

New Evidence From the National Longitudinal Study of Adolescent to Adult Health

Benjamin W. Domingue, Daniel W. Belsky, Dalton Conley, Kathleen Mullan Harris, Jason D. Boardman

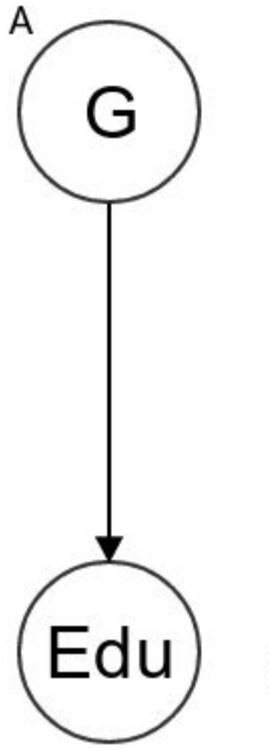
TABLE 3

Model Estimates of Polygenic Score on Educational Attainment

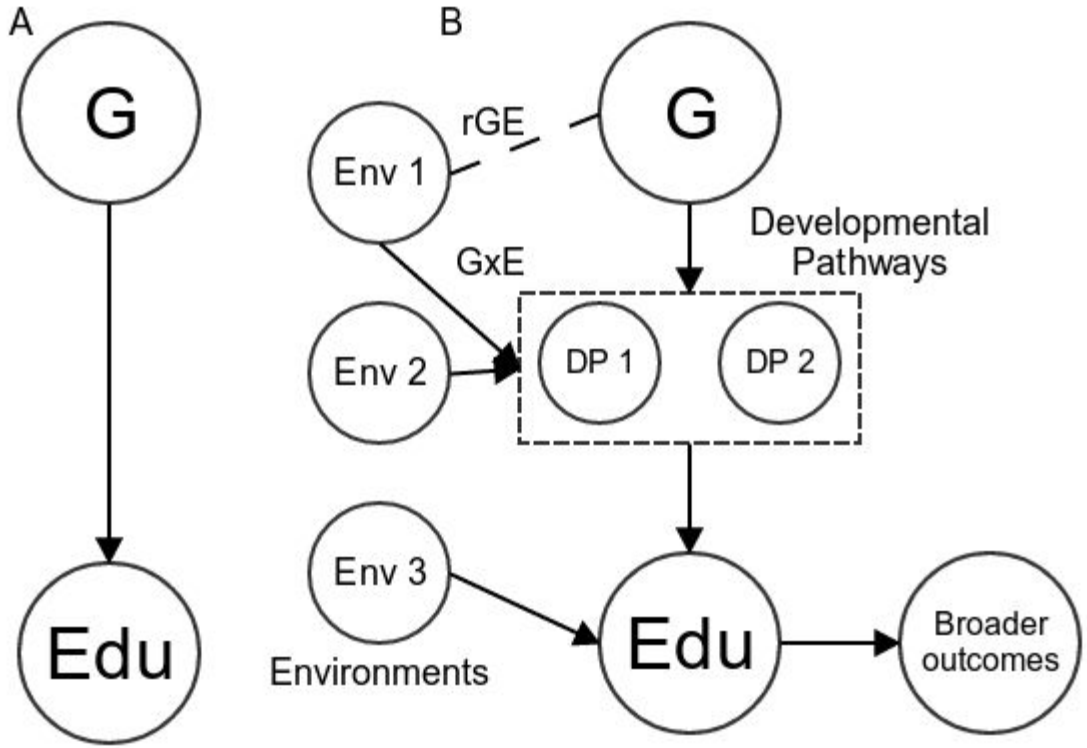
	EA respondents					AA respondents				
	Estimate	*SE*	pv	*N*	Model r^2	Estimate	*SE*	pv	*N*	Model r^2
Model 1: $\widehat{\beta_U}$	0.37	0.08	7.5E-07	917	0.06	0.20	0.09	2.1E-02	677	0.02
Model 2A: $\widehat{\beta_{U'}}$	0.30	0.07	3.1E-05	901	0.16	0.22	0.09	1.0E-02	671	0.04
Model 2B: $\widehat{\beta_{U'}}$	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).

Robust prediction, now what?



Robust prediction, now what?



And nothing is necessarily constant over time!

Possibilities

- Developmental pathways & broader outcomes
- Association and moderation by environment
- Dynamics (as function of birth cohort)

OPEN ACCESS Freely available online

PLOS ONE

Genetic Variation Associated with Differential Educational Attainment in Adults Has Anticipated Associations with School Performance in Children

Mary E. Ward[1], George McMahon[1], Beate St Pourcain[1,2,3], Davi
Daniel J. Benjamin[7], Philipp D. Koellinger[5,6,8], David Cesarini[9,
Association Consortium, George Davey Smith[1], Nicholas J. Tim

Predicts academic performance

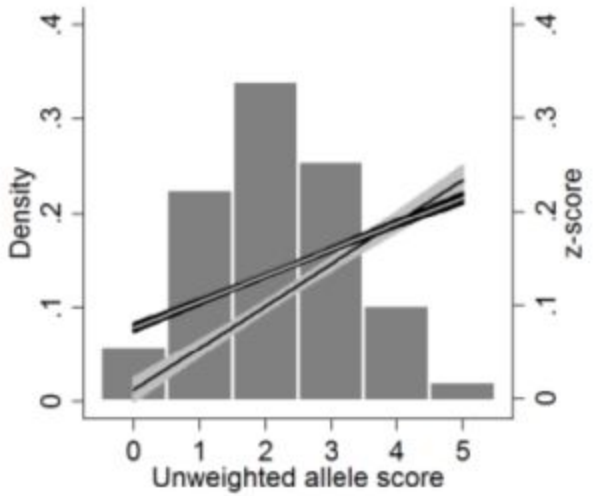


Figure 2. Histogram of allele score, with linear relationships between SATS z-scores and the allele score superimposed. The unweighted allele score is created from three SNPs rs9320913, rs11584700 and rs4851266. Each unit increase in the allele score corresponds to an individual having an additional educational attainment increasing allele. The density for the allele score taking the value 6 is 0.0016, which is too small to be visible in this figure. The linear relationships with 95%CIs from our regressions of SATS z-scores on allele score are superimposed. The English regression is represented by a black line with grey 95%CI, and mathematics by a grey line with black 95%CI.

doi:10.1371/journal.pone.0100248.g002

One obvious pathway from genetics to educational attainment is cognition.

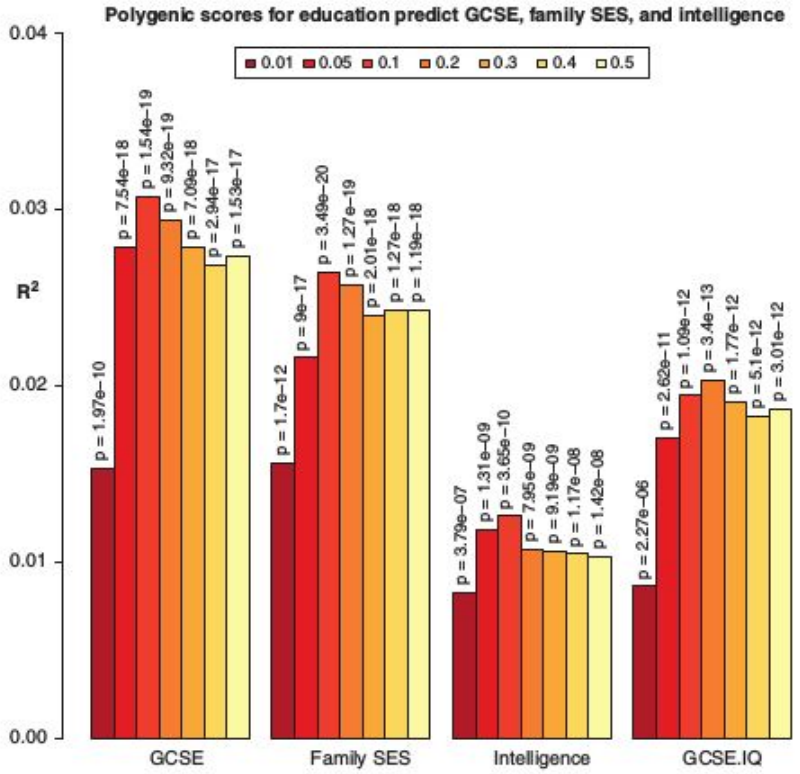
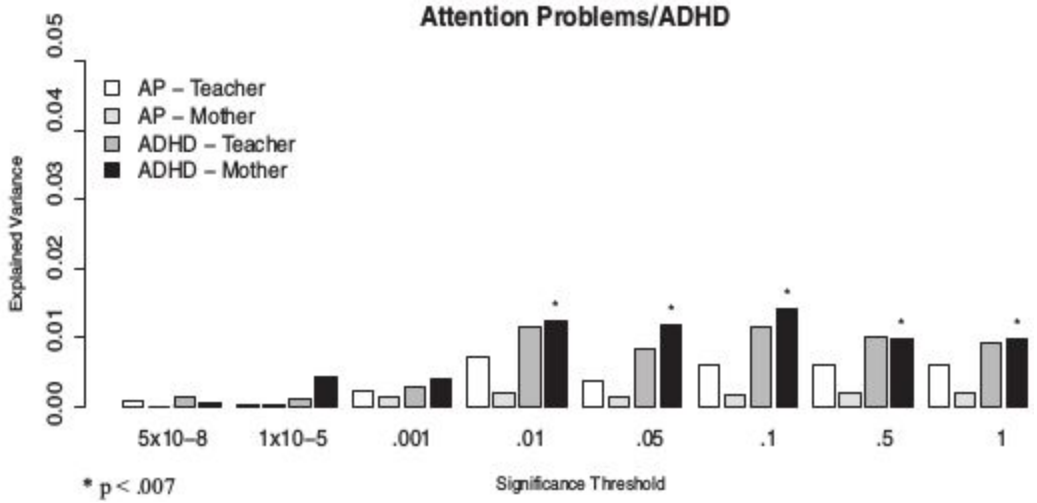


Figure 2. Genome-wide polygenic scores (GPS) for years of schooling in adults (Rietveld *et al.*[13]) predict variance (R^2) in children's educational achievement (General Certificate of Secondary Education (GCSE)), family socioeconomic status (SES), intelligence and educational achievement after controlling for intelligence (GCSE.IQ). GPS were created using different significance thresholds for inclusion of variants for years of education, ranging from $P = 0.01$ to 0.50, indicated by heat colors. The uncorrected *P*-values above each bar indicate the statistical significance of the observed association between the GPS and the respective trait.

But there are other candidates.

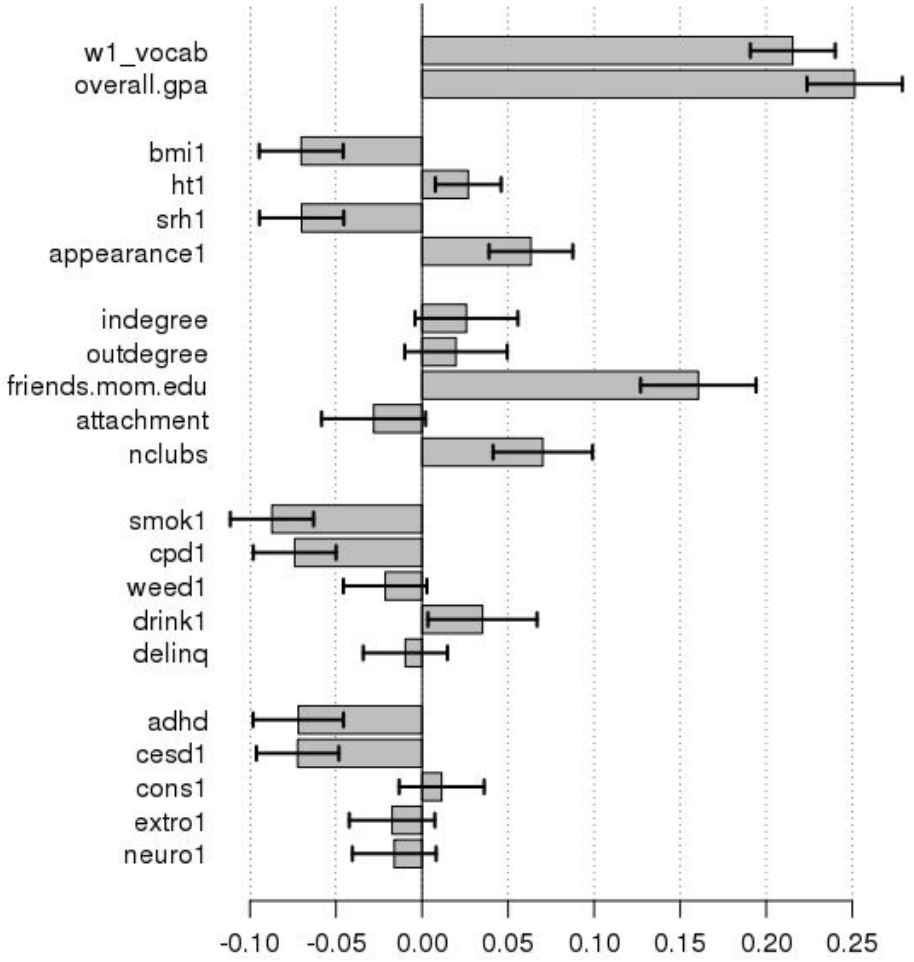
Polygenic Scores Associated With Educational Attainment in Adults Predict Educational Achievement and ADHD Symptoms in Children

Eveline L. de Zeeuw,[1,2]* Catharina E.M. van Beijsterveldt,[1,2] Tina J. Glasner,[1] M. Bartels,[1,2] Erik A. Ehli,[3] Gareth E. Davies,[3] James J. Hudziak,[4] Social Science Genetic Association Consortium, Cornelius A. Rietveld,[5] Maria M. Groen-Blokhuis,[1,2] Jouke Jan Hottenga,[1] Eco J.C. de Geus,[1,2] and Dorret I. Boomsma[1,2]



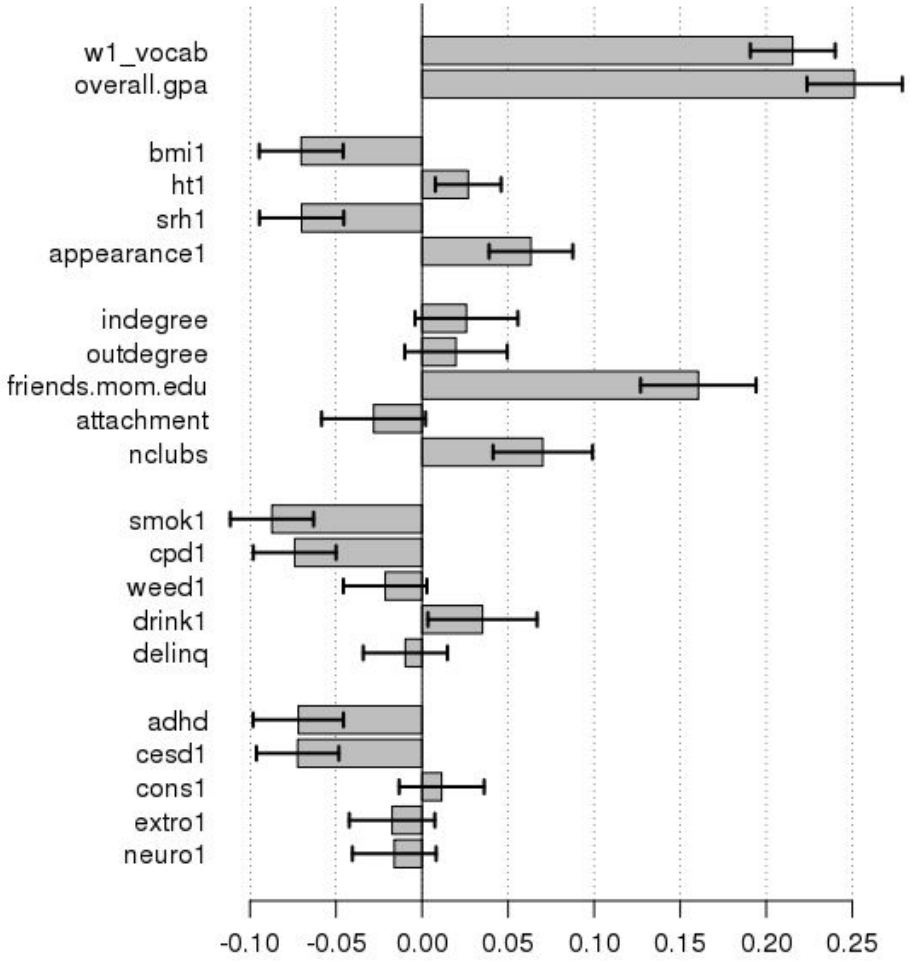
Preliminary Evidence regarding developmental pathways from Add Health

- Strong associations with GPA and verbal ability.



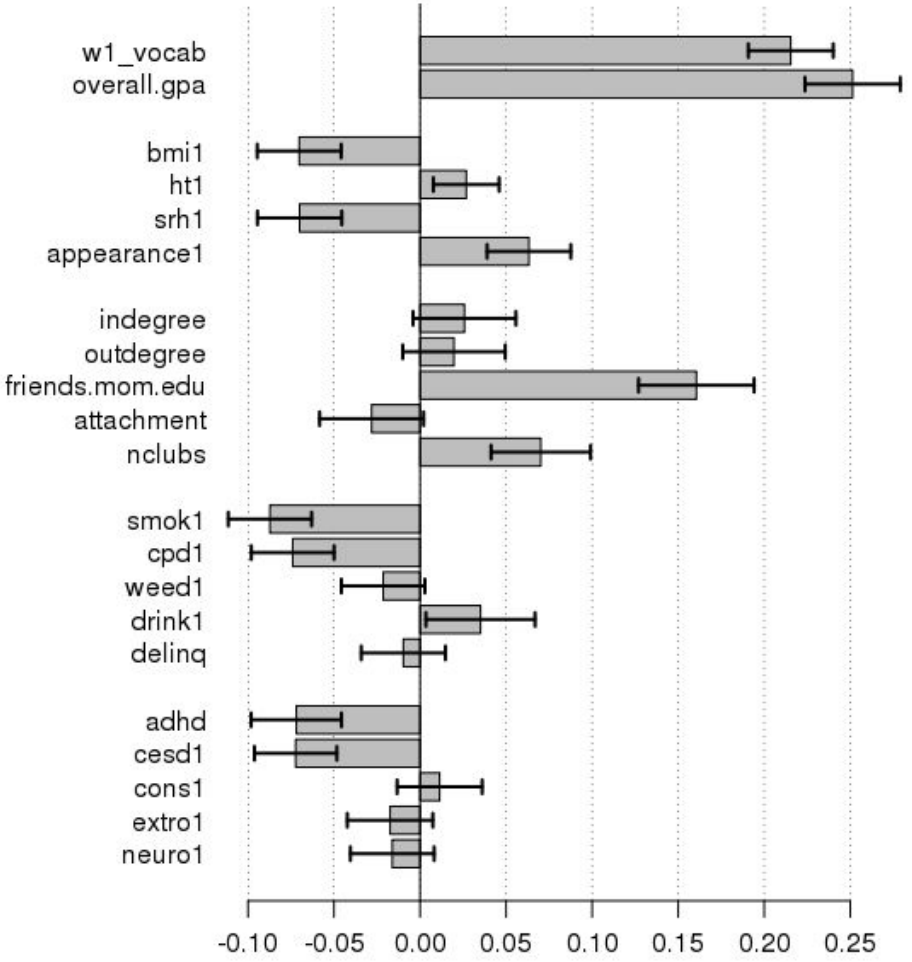
Preliminary Evidence regarding developmental pathways from Add Health

- Strong associations with GPA and verbal ability.
- Appearance!



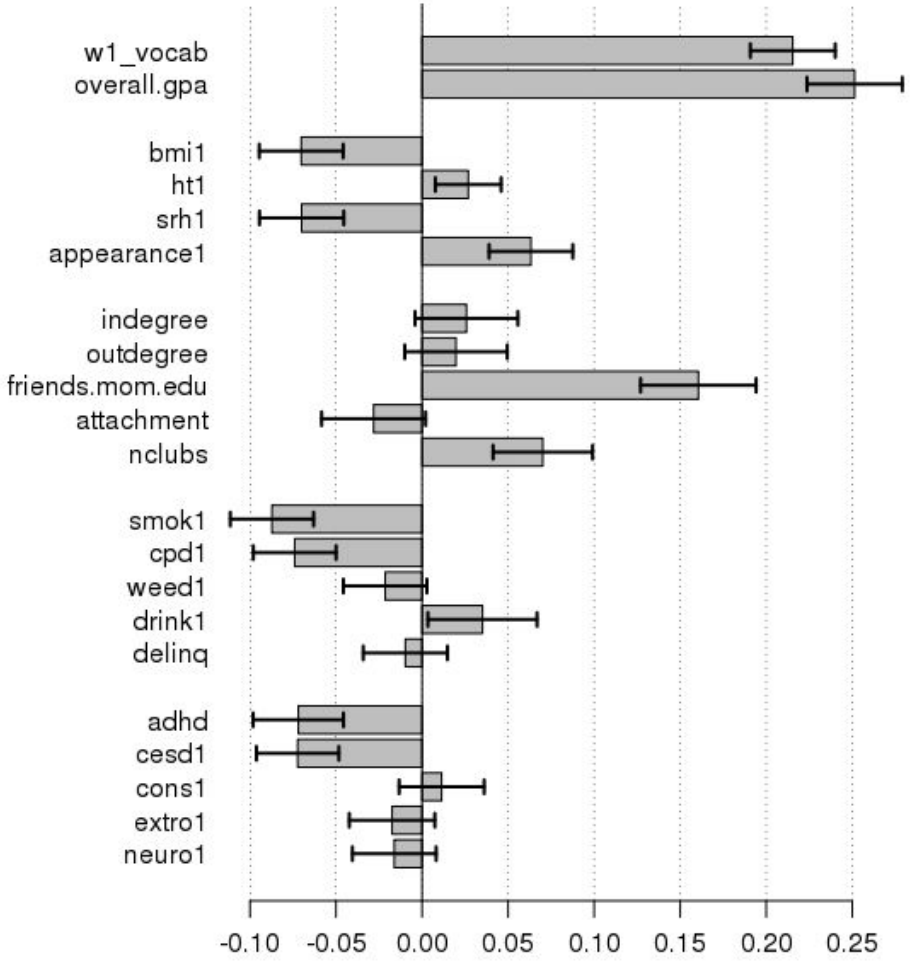
Preliminary Evidence regarding developmental pathways from Add Health

- Strong associations with GPA and verbal ability.
- Appearance!
- No connection to # of friends.



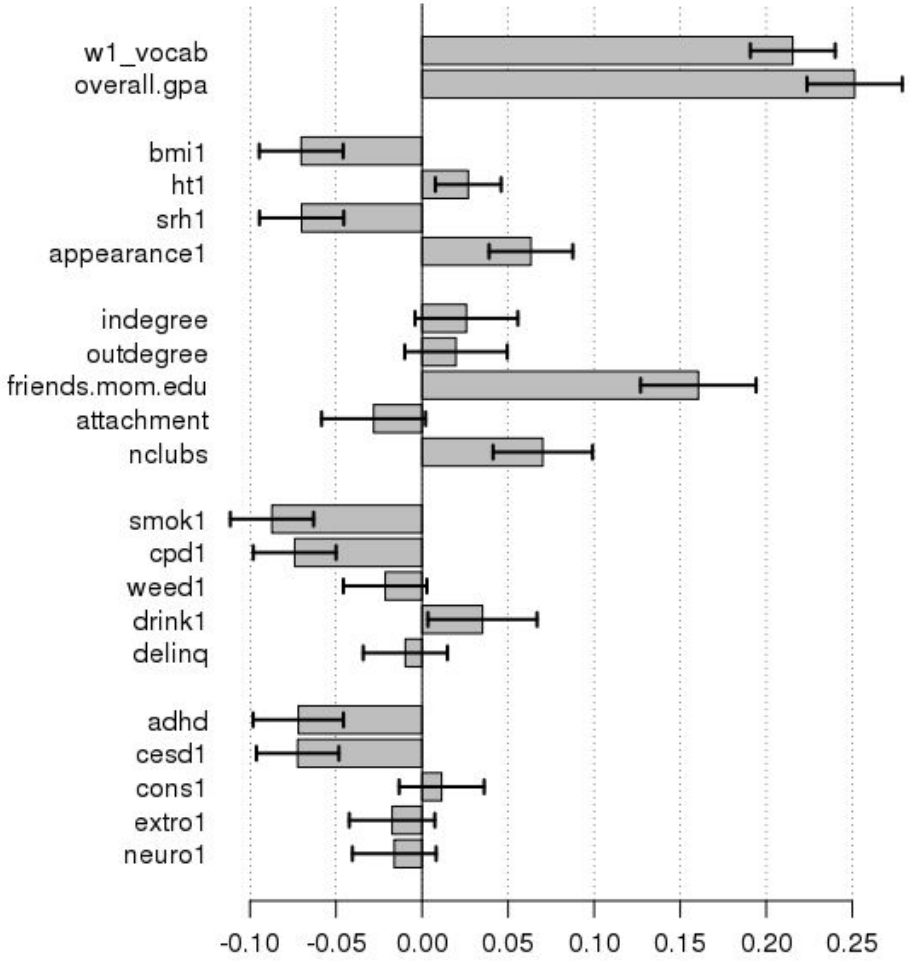
Preliminary Evidence regarding developmental pathways from Add Health

- Strong associations with GPA and verbal ability.
- Appearance!
- No connection to # of friends.
- Associated with smoking, but not other potential behavioral problems.



Preliminary Evidence regarding developmental pathways from Add Health

- Strong associations with GPA and verbal ability.
- Appearance!
- No connection to # of friends.
- Associated with smoking, but not other potential behavioral problems.
- Not associated with personality factors.



What about longer term outcomes?

[data from NZ]

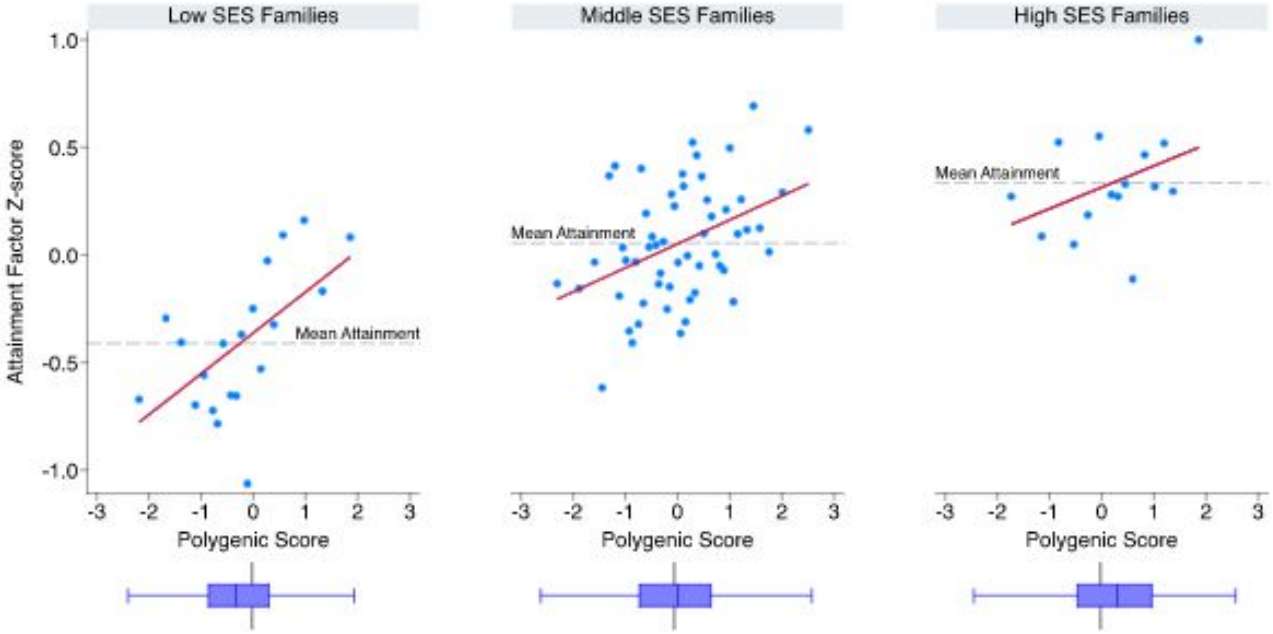
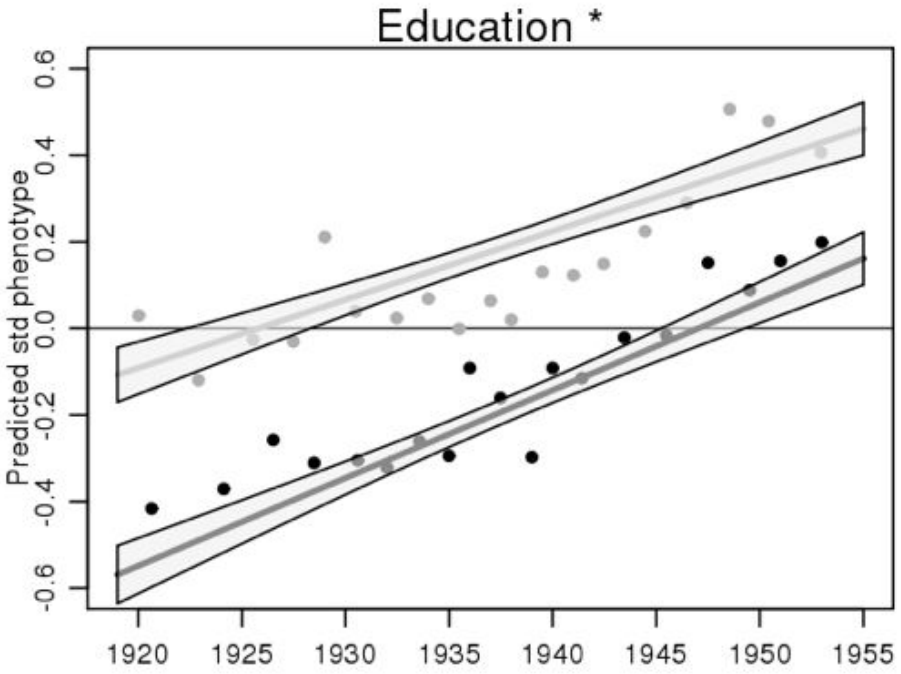


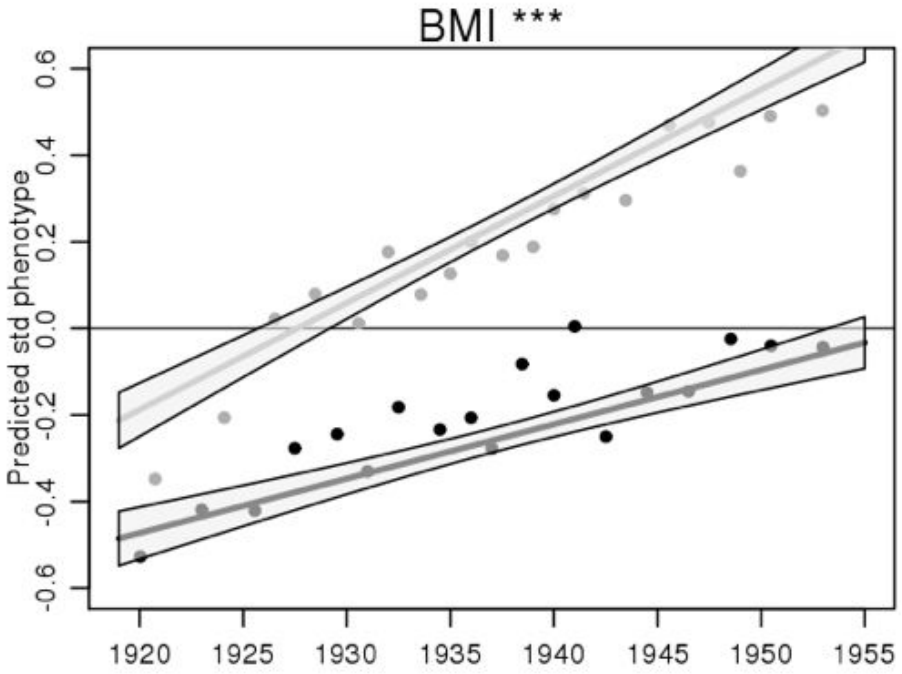
Figure 2. Polygenic scores were socially stratified, but children with higher scores were more likely to succeed no matter their social origin. The figure shows binned scatterplots of the genetic association with the Adult Attainment factor for children born in low, middle, and high socioeconomic-status (SES) families. Each plotted point represents mean X and Y coordinates for a "bin" of about 10 Study members (total n=175 for low SES families; 570 for middle SES families; 152 for high SES families). The solid red line graphs the association in the raw data. The dashed blue line shows the subgroup mean level of attainment. The distribution of polygenic scores within each subgroup is shown in the boxplots at the bottom of the figure. The black vertical line beneath the box plots shows the cohort mean polygenic score.

The Genetics of Success
How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development

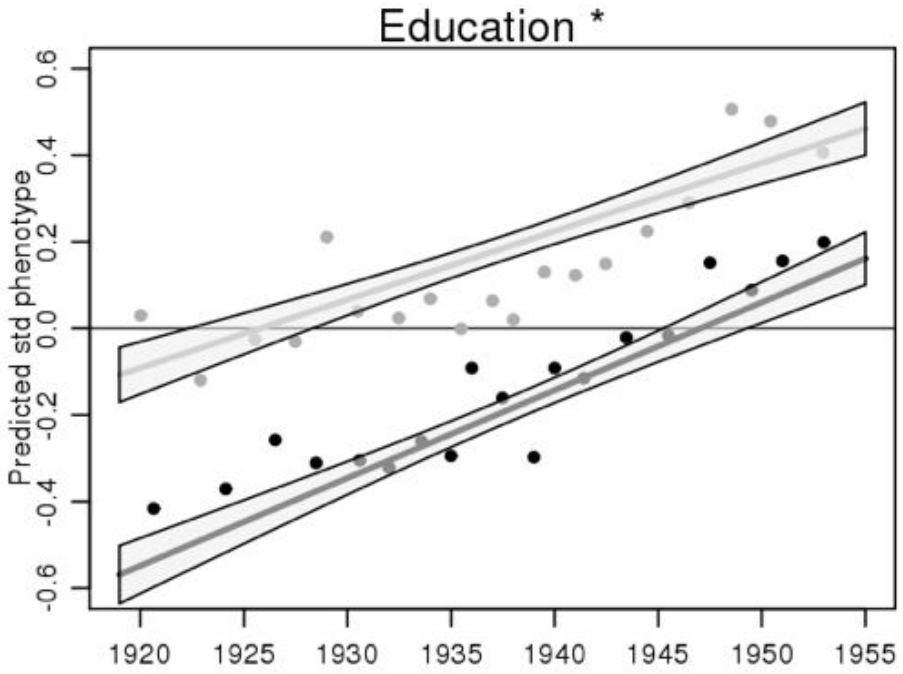
Daniel W. Belsky[1,2⇑]

How well does the polygenic score predict over time?





How well does the polygenic score predict over time?



- For an individual born in 1919, an additional standard deviation of educational genetic endowment as measured here results in more than half year of schooling more on average. By the 1955 birth cohort, that effect had been reduced by one-third.

- Largely due to changes associated with finishing high school.

- Similar trend observed in Sweden (Okbay et al., 2016, Nature).

[Forthcoming in Sociological Science]

Switching Course to BMI

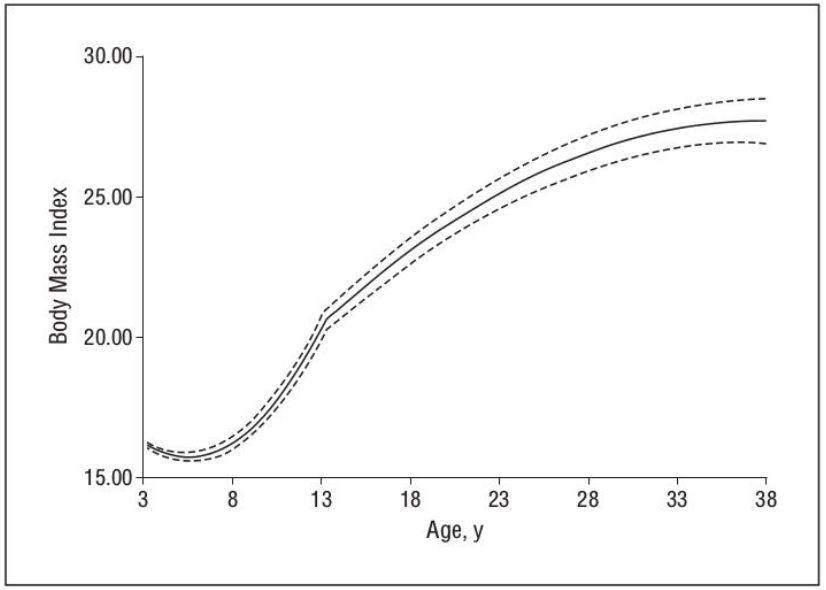


Figure 2. Life-course growth curves for children with high, low, and average genetic risk scores (GRSs). Individuals with higher-obesity GRSs were larger and grew more rapidly as children and adults. The solid line represents the population mean trajectory (average genetic risk). Dashed lines are for subgroups within 1 SD of the GRS (high and low genetic risk). Trajectories were derived from the life-course growth model (intercept fitted at 13 years of age; linear and quadratic slopes fitted during ages 3-13 years and 13-38 years), including intercept and linear slope effects for the GRS. Analyses included 856 individuals of European descent. Body mass index is calculated as weight in kilograms divided by height in meters squared.

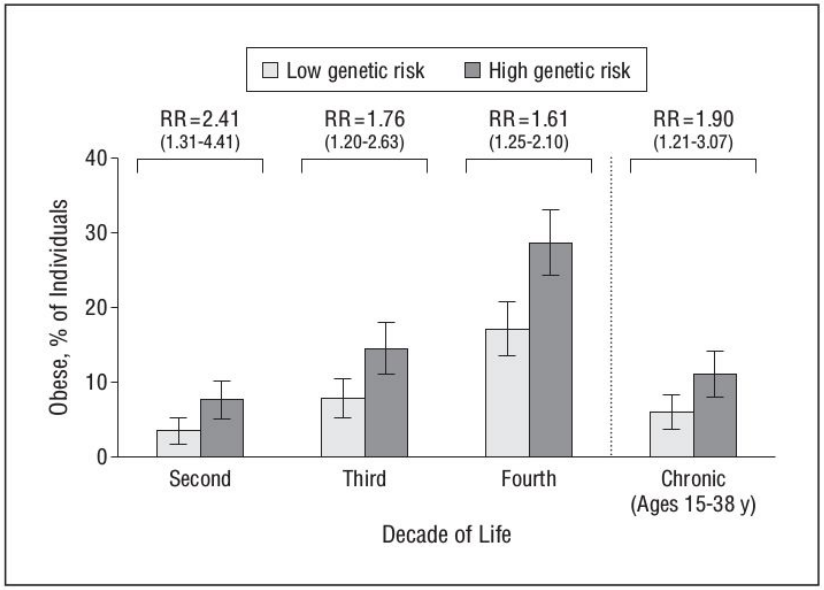
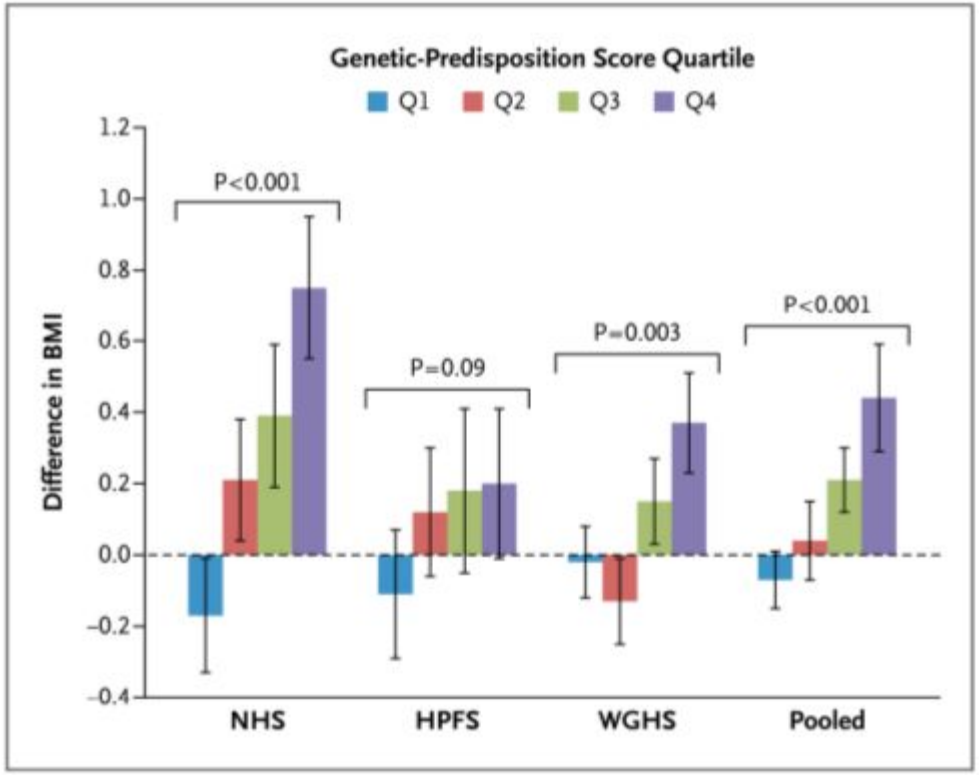


Figure 3. Obesity prevalence among low and high genetic risk cohort members in their second, third, and fourth decades of life and chronically across ages 15 to 38 years. Individuals with higher genetic risk scores (GRSs) were more likely to be obese across 2 decades of adult follow-up. Error bars and numbers in parentheses reflect 95% CIs. The GRS was dichotomized at the sample mean to create low and high genetic risk categories. Relative risks (RRs) (95% CIs) are reported from Poisson regression models adjusted for sex that included the 856 individuals of European descent in the analysis sample.

Polygenic Risk, Rapid Childhood Growth, and the Development of Obesity

Evidence From a 4-Decade Longitudinal Study

Daniel W. Belsky, PhD; Terrie E. Moffitt, PhD; Renate Houts, PhD; Gary G. Bennett, PhD; Andrea K. Biddle, PhD; James A. Blumenthal, PhD; James P. Evans, MD, PhD; HonaLee Harrington, BA; Karen Sugden, PhD; Benjamin Williams, BS; Richie Poulton, PhD; Avshalom Caspi, PhD



Sugar-Sweetened Beverages and Genetic Risk of Obesity

Qibin Qi, Ph.D., Audrey Y. Chu, Ph.D., Jae H. Kang, Sc.D., Majken K. Jensen, Ph.D., Gary C. Curhan, M.D., Sc.D., Louis R. Pasquale, M.D., Paul M. Ridker, M.D., M.P.H., David J. Hunter, M.B., B.S., Sc.D., Walter C. Willett, M.D., Dr.P.H., Eric B. Rimm, Sc.D., Daniel I. Chasman, Ph.D., Frank B. Hu, M.D., Ph.D., and **Lu Qi, M.D., Ph.D.**
Departments of Nutrition (Q.Q., M.K.J., D.J.H., W.C.W., E.B.R., F.B.H., L.Q.) and Epidemiology (G.C.C., D.J.H., W.C.W., E.B.R., F.B.H.), Harvard School of Public Health; and the Divisions of Preventive Medicine (A.Y.C., P.M.R., D.I.C.), Cardiovascular Disease (P.M.R.), and Genetics (D.I.C.), and the Channing Division of Network Medicine (J.H.K., G.C.C., L.R.P., D.J.H., W.C.W., E.B.R., F.B.H., L.Q.), Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; and the Department of Ophthalmology (L.R.P.), Massachusetts Eye and Ear Infirmary, Harvard Medical School — all in Boston

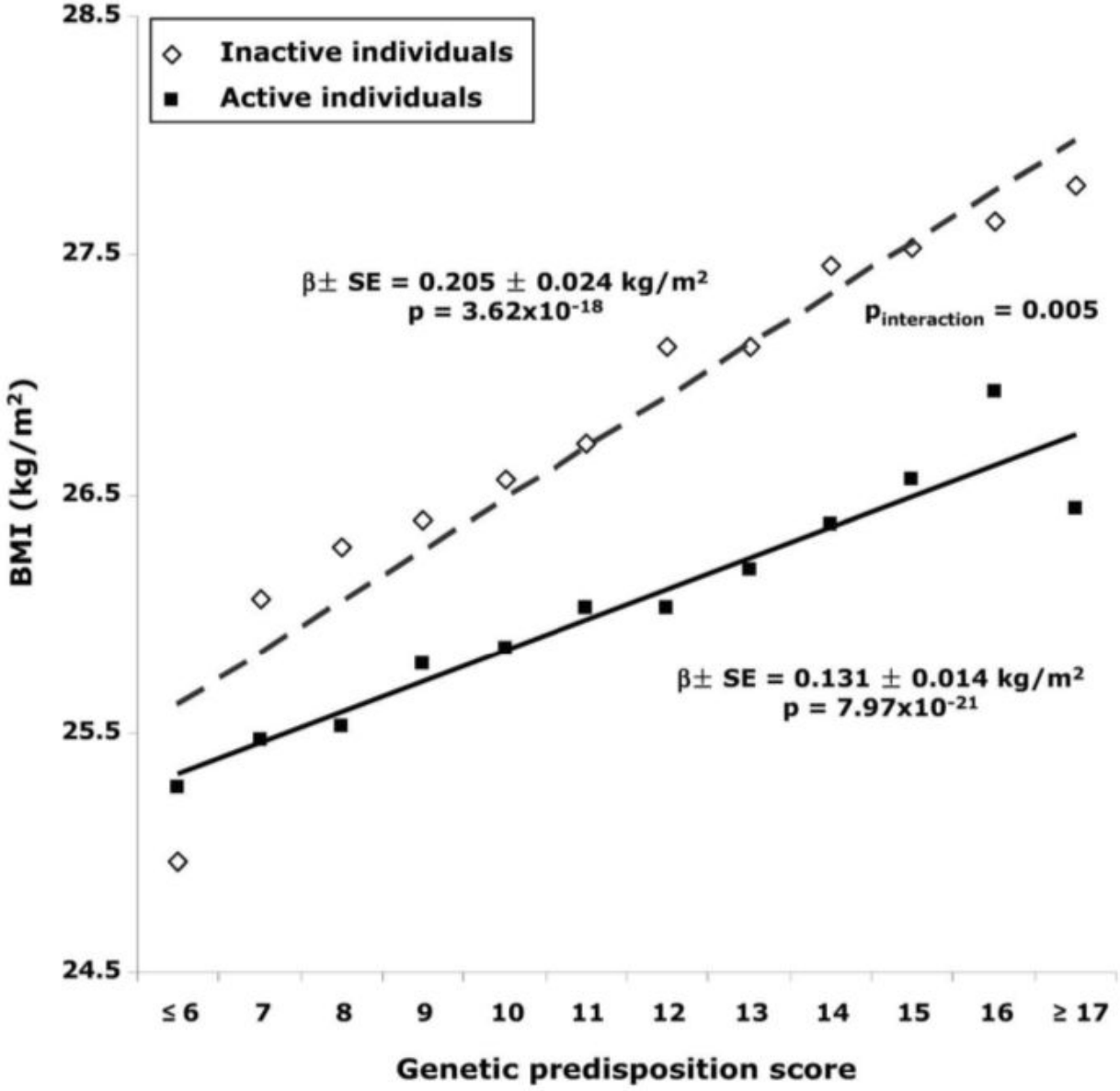
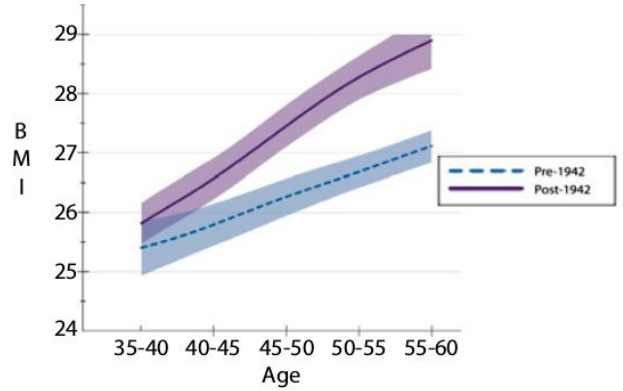


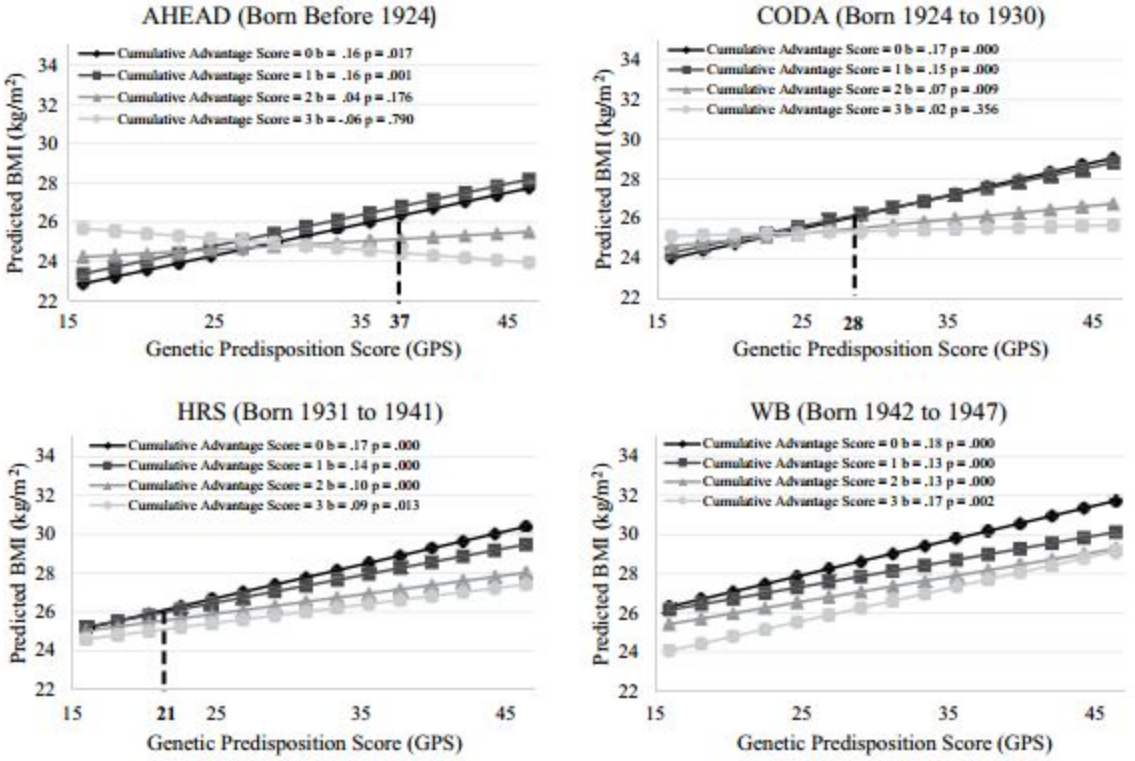
Figure 1. BMI with different genetic predisposition scores in inactive versus active individuals.
doi:10.1371/journal.pmed.1000332.g001



Cohort of birth modifies the association between FTO genotype and BMI

James Niels Rosenquist[a,1], Steven F. Lehrer[b,c], A. James O'Malley[d], Alan M. Zaslavsky[e], Jordan W. Smoller[f], and Nicholas A. Christakis[g,h,i,j]

Fig. 3. BMI over the ages of 35–60 by birth cohort for AT/TA–FTO genotype.



Lifetime Socioeconomic Status, Historical Context, and Genetic Inheritance in Shaping Body Mass in Middle and Late Adulthood

Hexuan Liu[a] and Guang Guo[a]

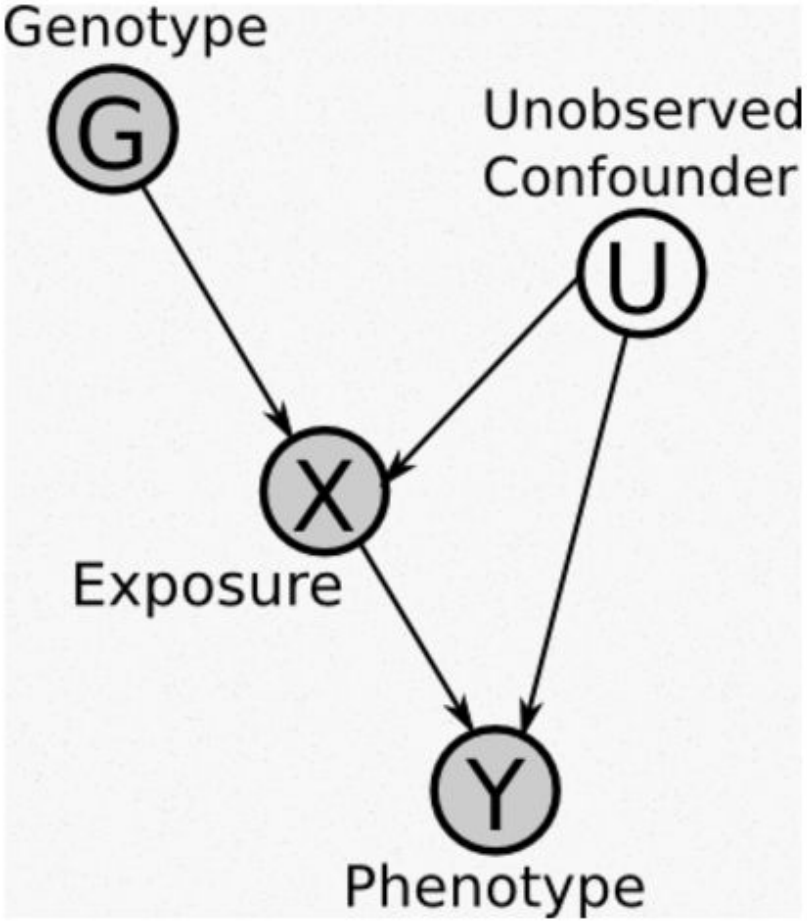
Polygenic Scores

- Exciting area of research.
- Limited by available GWAS
  - These have been performed on insufficiently diverse samples.

- Our knowledge of how individual genetic variants influence traits can be used in other manners as well.

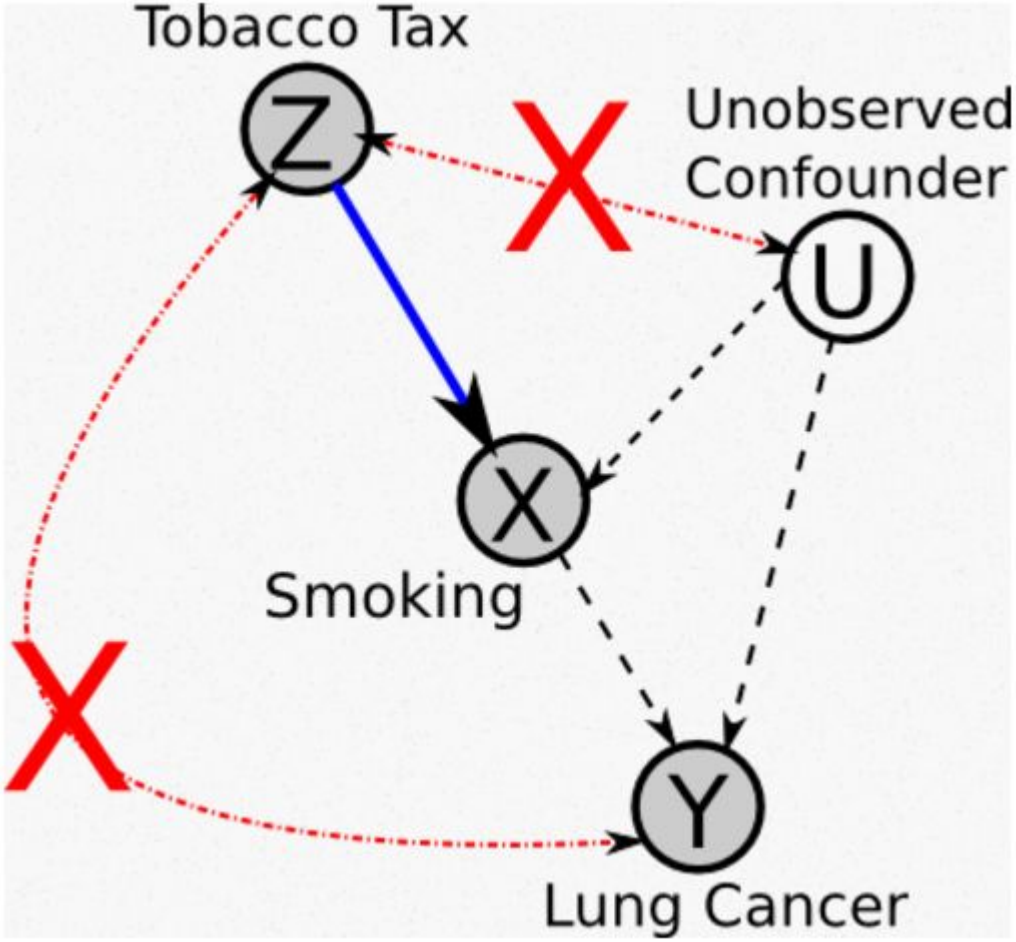
3. Mendelian Randomization

- Genes as IVs

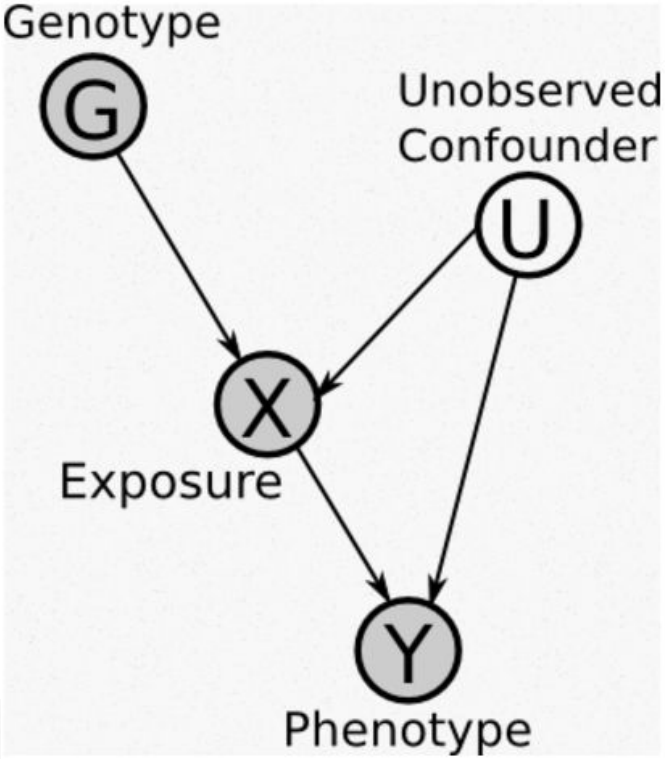


http://jamesmcm.github.io/blog/2014/08/17/mendelian/

Assumptions



- Z → X (not X → Z)
- No Z→Y
- Z not associated with U

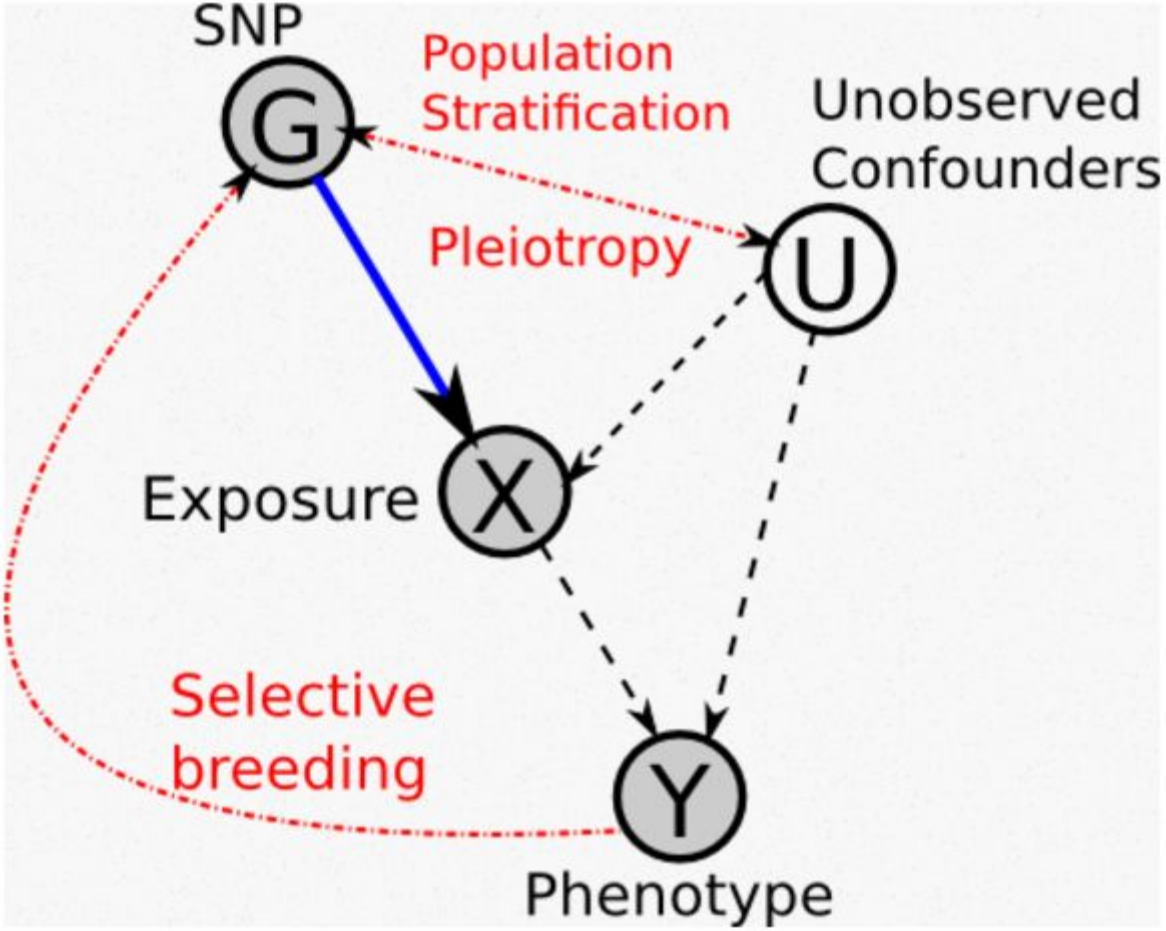


No Pleiotropy

Pleiotropy is the phenomena whereby one gene can affect many (even seemingly unrelated) phenotypes. Mendelian Randomisation makes the assumption of **no pleiotropy**.

In this case, this means that we assume the genotype is only influencing the phenotype via the considered exposure. I.e. ApoE2 **only** affects serum cholesterol levels, and cannot affect cancer risk by other, unobserved means.

This is a big assumption, and **prior knowledge** is necessary. If possible, using multiple, independent SNPs which act through the same path, can help to alleviate this issue, because, if they are all consistent then it is unlikely that they would all have other pathways causing the same change in phenotype. But note that they must be independent, and so cannot be in Linkage Disequilibrium.


SNP
G
Population
Stratification
Unobserved
Confounders
Pleiotropy
U
Exposure
X
Selective
breeding
Y
Phenotype

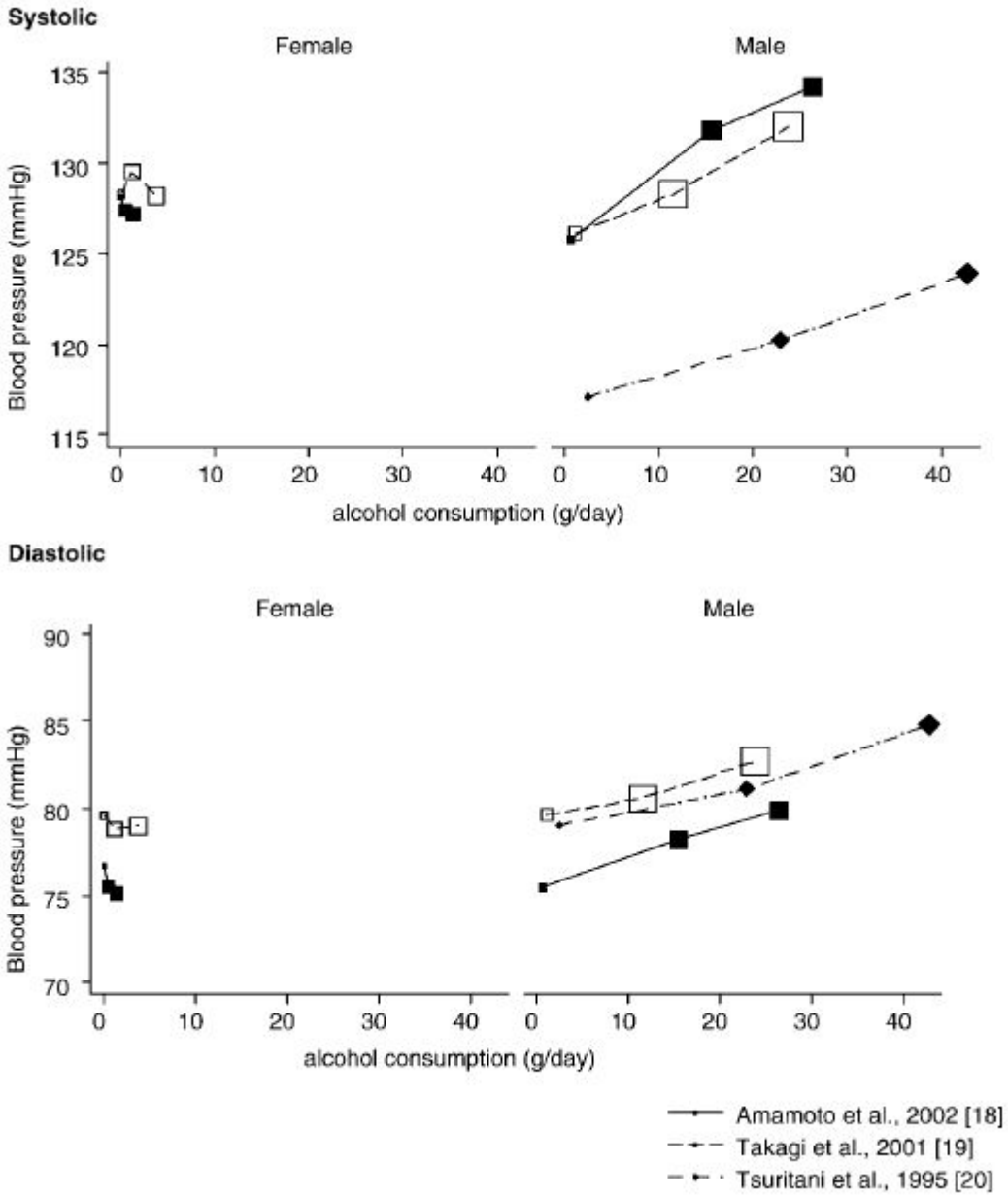
Alcohol Intake and Blood Pressure: A Systematic Review Implementing a Mendelian Randomization Approach

Lina Chen, George Davey Smith, Roger M Harbord, Sarah J Lewis

Published: March 4, 2008 • http://dx.doi.org/10.1371/journal.pmed.0050052

Background

Alcohol has been reported to be a common and modifiable risk factor for hypertension. However, observational studies are subject to confounding by other behavioural and sociodemographic factors, while clinical trials are difficult to implement and have limited follow-up time. Mendelian randomization can provide robust evidence on the nature of this associatio by use of a common polymorphism in *aldehyde dehydrogenase 2* (*ALDH2*) as a surrogate for measuring alcohol consumption. *ALDH2* encodes a major enzyme involved in alcohol metabolism. Individuals homozygous for the null variant (**2*2*) experience adverse symptoms when drinking alcohol and consequently drink considerably less alcohol than wild-type homozygotes (**1*1*) or heterozygotes. We hypothesise that this polymorphism may influence the risk of hypertension by affecting alcohol drinking behaviour.

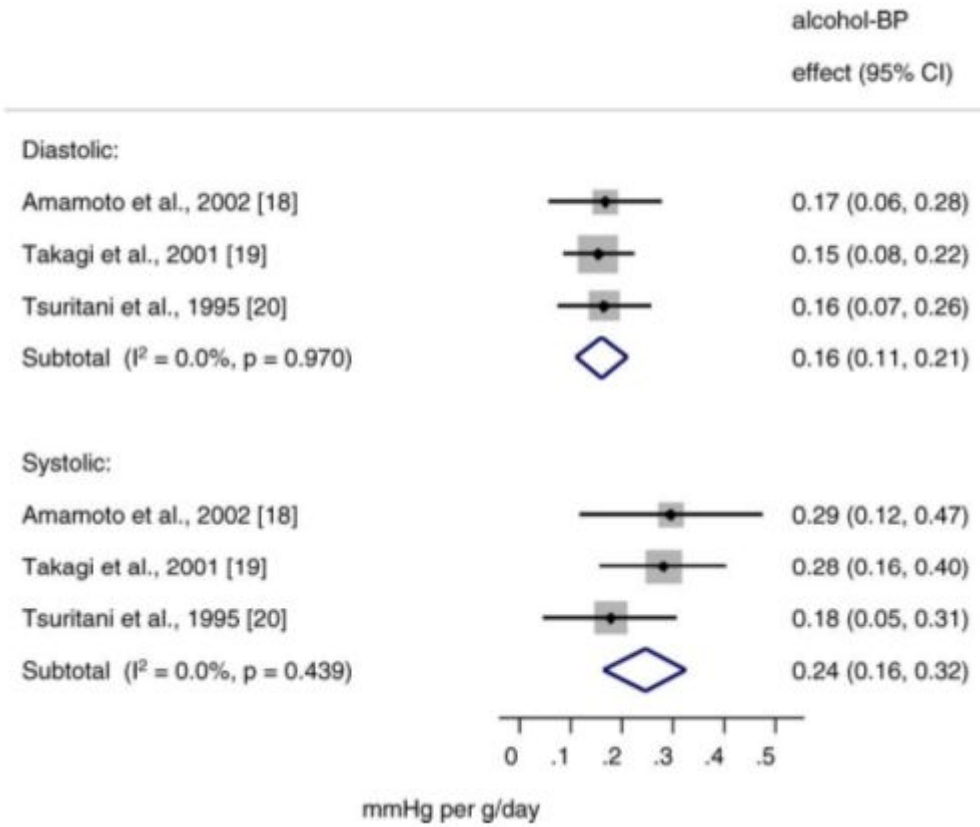


How seriously do we take this relationship?

Data Type	Reference	No. of Participants	Alcohol Exposure Groups, Grams of Ethanol per Day	Sex	Alcohol Status by Genotype[a]		
					*1*1	*1*2	*2*2
Categorical data	Saito et al., 2003 [28]	335	<18.6	Male	39	87	21
			18.6–37		54	33	0
			≥37.1		81	15	0
			Ex-drinker		3	1	0
	Yamada et al., 2002 [29]	828	Nondrinker	Male	21	56	28
			<2.26		75	86	16
			2.27–47.4		114	75	2
			47.5–71.1		210	64	2
			>71.1		63	16	0
Continuous data	Amamoto et al., 2002 [18]	2,035	N/D	Male	26.3	15.5	0.66
			N/D	Female	1.4	0.5	0.03
	Hashimoto et al., 2002 [13]	133	N/D	Male	54.9 ± 13.1	51 ± 9.8	—
	Mackenzie et al., 2005 [12]	28	N/D	Mixed	23.7 ± 39.5	31.6 ± 39.5	—
	Okayama et al., 1994 [14]	159	N/D	Male	26.8 ± 19.1	17.6 ± 17.9	—
	Takagi et al., 2001 [19]	4,057	N/D	Male	23.7 ± 16.7	11.5 ± 15.8	1.1 ± 16.9
			N/D	Female	3.8 ± 6.0	1.2 ± 6.8	0.08 ± 6.5
	Tsuritani et al., 1995 [20]	403	N/D	Male	42.7 ± 28.8	22.9 ± 25.1	2.4 ± 3.6

Table 3. Distribution of Potential Confounding Factors by Genotype among Studies Included in the Meta-analysis

Author	No. of Participants	Covariates	Category	Covariates by Genotypes[a]			Published *p*-Values[b]
				*1*1	*1*2	*2*2	
Amamoto et al., 2002 [18]	2,035	Sex	Male	335 (44.7)	351 (46·9)	63 (8.4)	—
			Female	644 (50.1)	555 (43.2)	87 (6.8)	—
		Age	Male	56.8 ± 15.7	58.1 ± 15.6	52.8 ± 17.7	0.05
			Female	54.8 ± 16.0	55.1 ± 15.4	58.5 ± 14.3	0.12
		BMI	Male	22.9 ± 3.0	22.3 ± 2.8	21.9 ± 2.7	0.004
			Female	22.5 ± 3.1	22.4 ± 3.3	22.4 ± 3.1	0.89
		Current smoker	Male	176 (52.5)	173 (49.3)	36 (56.5)	0.49
			Female	39 (6.1)	48 (8.6)	9 (10.5)	0.16
		Cigarette smoking, cigarettes per day	Male	10.4	10.8	11.5	0.78
			Female	0.69	1.04	1.02	0.17
Hashimoto et al., 2002 [13]	133	Age	—	46.0 ± 7.1	46.2 ± 5.8	—	>0.05
		BMI	—	22.5 ± 2.0	22.2 ± 1.6	—	>0.05
		Smoking, cigarettes per day	—	14.2 ± 16.3	12.7 ± 13.8	—	>0.05
		Exercise, times per month	—	2.6 ± 3.1	2.6 ± 2.7	—	>0.05
Iwai et al., 2004 [31]	1,849	Sex	Male	443 (51.6)	348 (40.6)	67 (7·8)	—
			Female	520 (52.3)	381 (38.3)	93 (9·4)	—
Mackenzie et al., 2005 [12]	28	Sex	Male	8 (47.1)	9 (52.9)	—	—
		Age	Female	9 (81.8)	2 (18.2)	—	—
			—	30 ± 7 (12); 29 ± 5 (5)	31 ± 10	—	—
		BMI	—	23.8 ± 3.1(12); 22.0 ± 3.4 (5)	23.0 ± 3.1	—	—
Nishimura et al., 2002 [11]	36	Age	—	22.1 ± 3.3	23.1 ± 3.4	—	—
Okayama et al., 1994 [14]	159	Age	—	45.7 ± 10.2	48.0 ± 9.4	—	—
		BMI	—	22.9 ± 2.8	21.8 ± 2.7	—	—
Saito et al., 2003 [28]	335	Age	—	53.3 ± 8.6	53.4 ± 8.8	52.4 ± 8.4	0.89
		BMI	—	23.5 ± 3.0	23.3 ± 3.1	23.0 ± 2.5	0.71
		Smoking	Current smokers	100 (56.3)	77 (55.9)	14 (66..7)	0.12
			Never-smokers	49 (27.8)	25 (18.4)	4 (19.1)	
			Ex-smokers	28 (15.9)	35 (25.7)	3 (14.3)	
		Physical activity	Inactive	66 (37.5)	52 (38.2)	3 (14.3)	0.32
			Moderate	50 (28.4)	41 (30.2)	8 (38.1)	
			Active	60 (34.1)	43 (31.6)	10 (47.6)	
Takagi et al., 2001 [19]	4,057	Sex	Male	924 (48.2)	825 (43.0)	170 (8.8)	—
			Female	1,112 (52.0)	838 (39.2)	188 (8.8)	—
		Age	Male	60.2 ± 12.2	61.6 ± 11.5	61.3 ± 11.7	0.05
			Female	58.5 ± 13.3	59.6 ± 11.6	58.·4 ± 12·3	0.09
		BMI	Male	23.2 ± 3.0	22.8 ± 2.9	23.1 ± 2.6	0.02
			Female	22.4 ± 3.3	22.1 ± 2.9	22.3 ± 2.7	0.11
		Current smoker	Male	367 (39.7)	326 (39.5)	61 (35.9)	n.s.
			Female	93 (8.4)	68 (8.2)	13 (7.5)	n.s.
Tsuritani et al., 1995 [20]	403	Age	—	46.4 ± 5.8	45.7 ± 5.0	46.4 ± 4.4	n.s.
		BMI	—	24.7 ± 2.8	24.3 ± 3.0	24.9 ± 2.4	n.s.

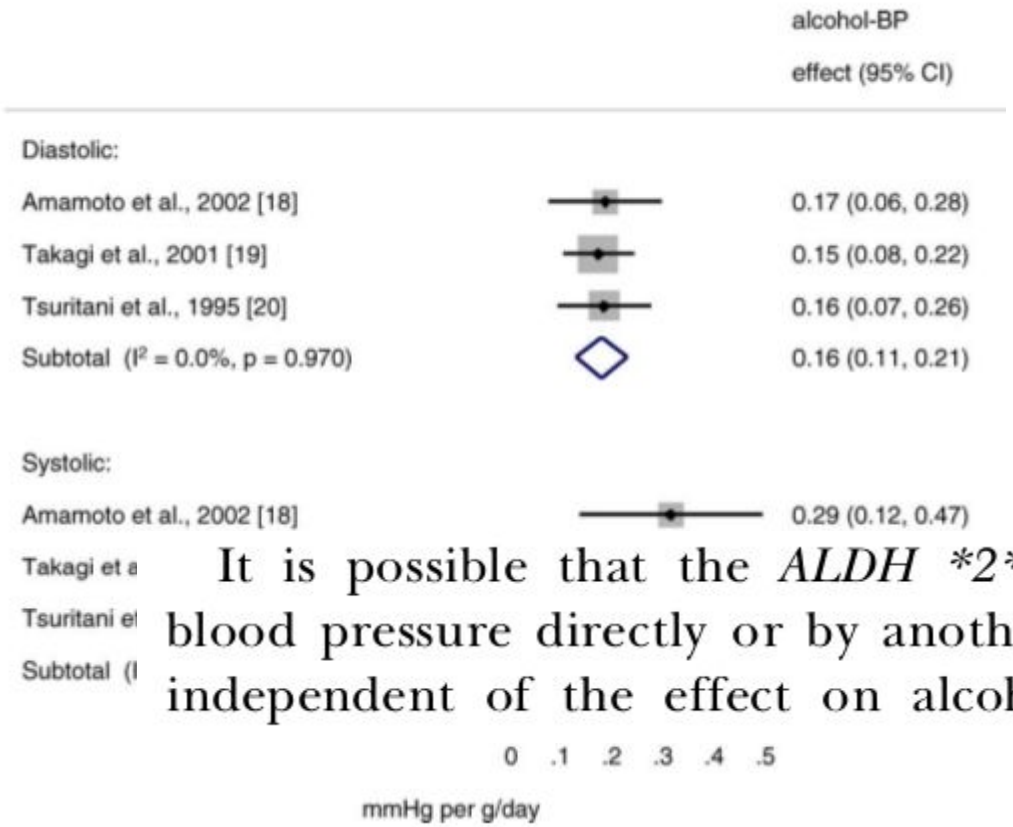


Methods and Findings

We carried out fixed effect meta-analyses of the *ALDH2* genotype with blood pressure (five studies, $n = 7{,}658$) and hypertension (three studies, $n = 4{,}219$) using studies identified via systematic review. In males, we obtained an overall odds ratio of 2.42 (95% confidence interval [CI] 1.66–3.55, $p = 4.8 \times 10^{-6}$) for hypertension comparing **1*1* with **2*2* homozygotes and an odds ratio of 1.72 (95% CI 1.17–2.52, $p = 0.006$) comparing heterozygotes (surrogate for moderate drinkers) with **2*2* homozygotes. Systolic blood pressure was 7.44 mmHg (95% CI 5.39–9.49, $p = 1.1 \times 10^{-12}$) greater among **1*1* than among **2*2* homozygotes, and 4.24 mmHg (95% CI 2.18–6.31, $p = 0.00005$) greater among heterozygotes than among **2*2* homozygotes.

Conclusions

These findings support the hypothesis that alcohol intake has a marked effect on blood pressure and the risk of hypertension.



It is possible that the *ALDH *2*2* genotype influenced blood pressure directly or by another mechanism that was independent of the effect on alcohol intake (pleiotropy).

Methods and Findings

We carried out fixed effect meta-analyses of the *ALDH2* genotype with blood pressure (five studies, $n = 7{,}658$) and hypertension (three studies, $n = 4{,}219$) using studies identified via systematic review. In males, we obtained an overall odds ratio of 2.42 (95% confidence interval [CI] 1.66–3.55, $p = 4.8 \times 10^{-6}$) for hypertension comparing **1*1* with **2*2* homozygotes and an odds ratio of 1.72 (95% CI 1.17–2.52, $p = 0.006$) comparing heterozygotes (surrogate for moderate drinkers) with **2*2* homozygotes. Systolic blood pressure was 7.44 mmHg (95% CI 5.39–9.49, $p = 1.1 \times 10^{-12}$) greater among **1*1* than among **2*2* homozygotes, and 4.24 mmHg (95% CI 2.18–6.31, $p = 0.00005$) greater among heterozygotes than among **2*2* homozygotes.

Conclusions

These findings support the hypothesis that alcohol intake has a marked effect on blood pressure and the risk of hypertension.

OPEN ACCESS Freely available online

PLOS | MEDICINE

Causal Relationship between Obesity and Vitamin D Status: Bi-Directional Mendelian Randomization Analysis of Multiple Cohorts

Abstract

Background: Obesity is associated with vitamin D deficiency, and both are areas of active public health concern. We explored the causality and direction of the relationship between body mass index (BMI) and 25-hydroxyvitamin D [25(OH)D] using genetic markers as instrumental variables (IVs) in bi-directional Mendelian randomization (MR) analysis.

Genotyping

We selected 12 established BMI-related SNPs (fat mass and obesity-associated, [*FTO*]- rs9939609, melanocortin 4 receptor [*MC4R*]- rs17782313, transmembrane protein 18 [*TMEM18*]- rs2867125, SH2B adaptor protein 1 [*SH2B1*]- rs7498665, brain-derived neurotrophic factor [*BDNF*]- rs4074134, potassium channel tetramerisation domain containing 15 [*KCTD15*]- rs29941, ets variant 5 [*ETV5*]- rs7647305, SEC16 homolog B [*SEC16B*]- rs10913469, Fas apoptotic inhibitory molecule 2 [*FAIM2*]- rs7138803, neuronal growth regulator 1 [*NEGR1*]- rs3101336, mitochondrial carrier 2 [*MTCH2*]- rs10838738, and glucosamine-6-phosphate deaminase 2 [*GNPDA2*]- rs10938397) for our analysis based on the study by Li et al. [24] and previously published genome-wide association studies for obesity-related traits [23,25,26]. The four vitamin D-related SNPs (*DHCR7*- rs12785878, *CYP2R1*- rs10741657, *GC*- rs2282679, and *CYP24A1*- rs6013897) were chosen on the basis of the recent genome-wide association study on 25(OH)D [27]. The studies that did not have

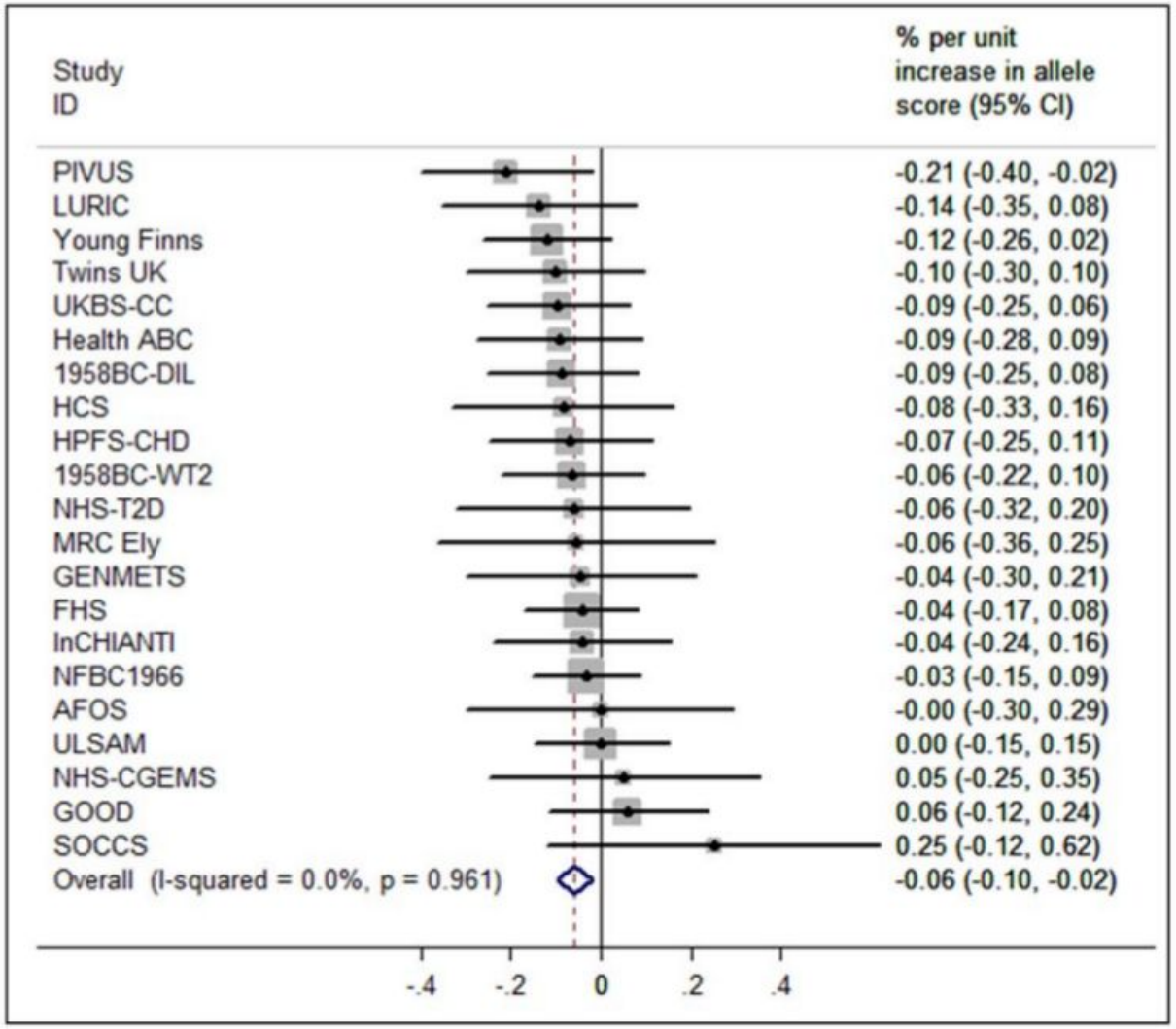
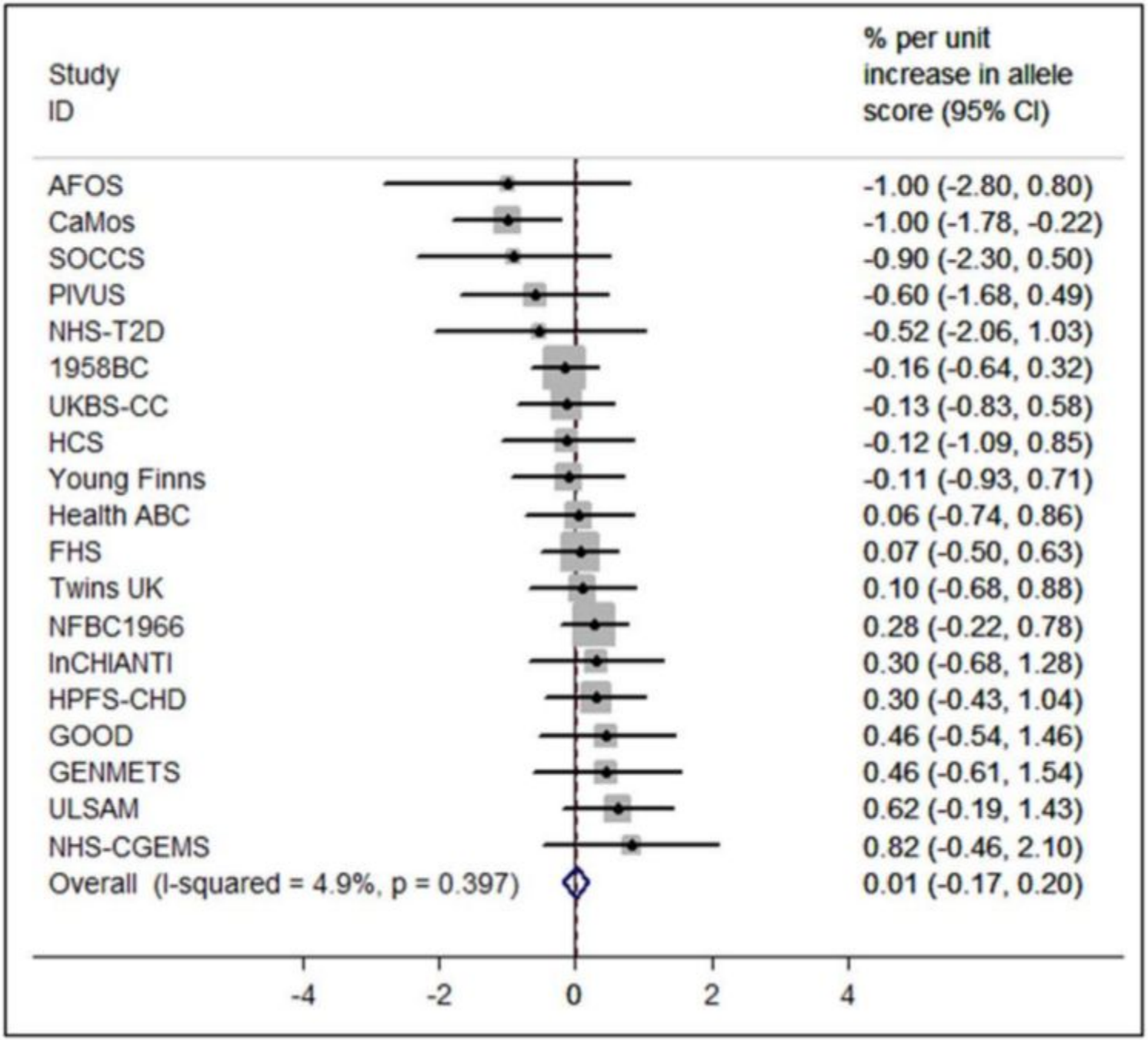
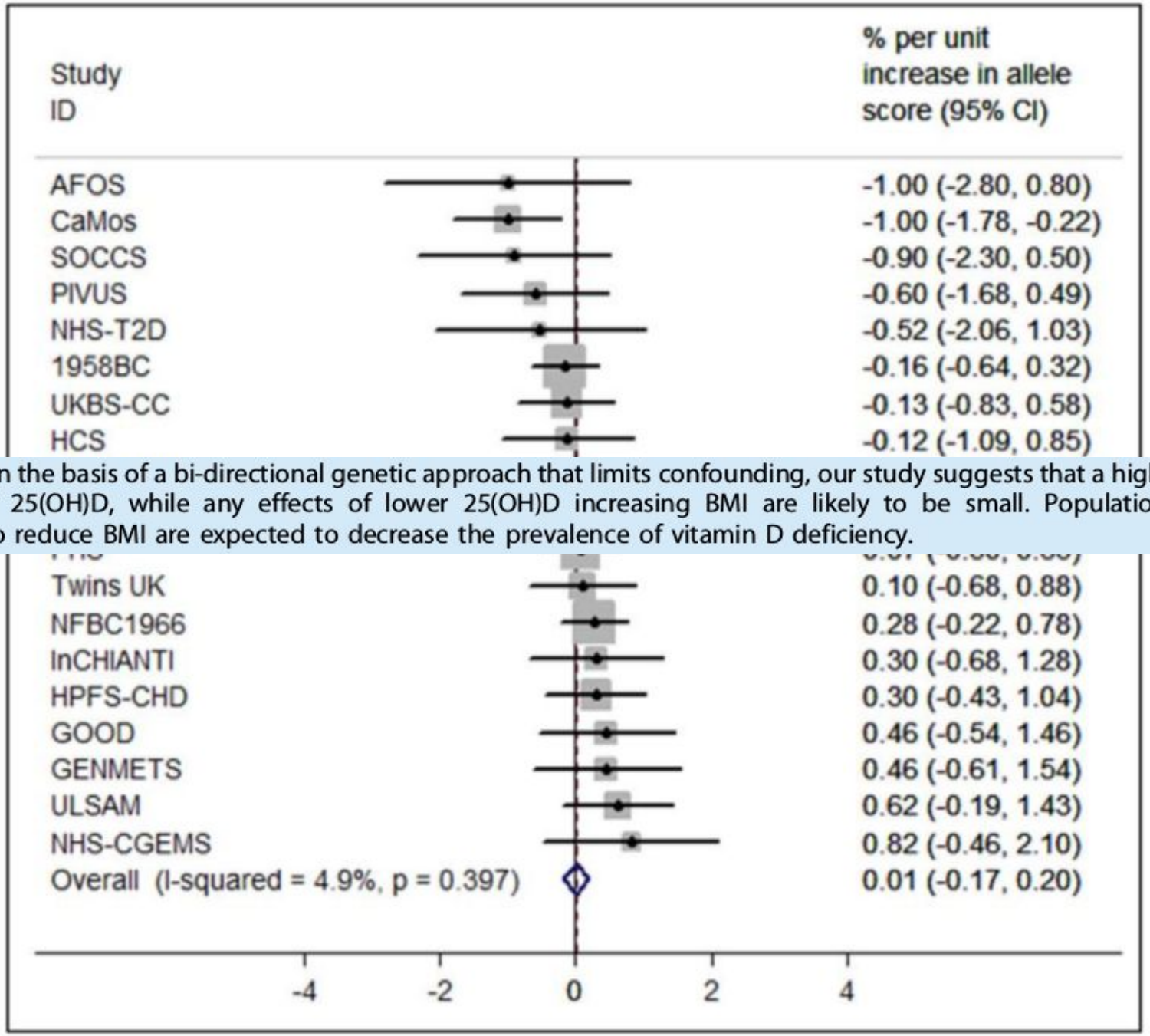


Figure 3. Meta-analysis of the BMI allele score association with 25(OH)D (*n*=31,120). 95% confidence intervals given by error bars.
doi:10.1371/journal.pmed.1001383.g003

A


Study ID
% per unit increase in allele score (95% CI)
AFOS -1.00 (-2.80, 0.80)
CaMos -1.00 (-1.78, -0.22)
SOCCS -0.90 (-2.30, 0.50)
PIVUS -0.60 (-1.68, 0.49)
NHS-T2D -0.52 (-2.06, 1.03)
1958BC -0.16 (-0.64, 0.32)
UKBS-CC -0.13 (-0.83, 0.58)
HCS -0.12 (-1.09, 0.85)
Young Finns -0.11 (-0.93, 0.71)
Health ABC 0.06 (-0.74, 0.86)
FHS 0.07 (-0.50, 0.63)
Twins UK 0.10 (-0.68, 0.88)
NFBC1966 0.28 (-0.22, 0.78)
InCHIANTI 0.30 (-0.68, 1.28)
HPFS-CHD 0.30 (-0.43, 1.04)
GOOD 0.46 (-0.54, 1.46)
GENMETS 0.46 (-0.61, 1.54)
ULSAM 0.62 (-0.19, 1.43)
NHS-CGEMS 0.82 (-0.46, 2.10)
Overall (I-squared = 4.9%, p = 0.397) 0.01 (-0.17, 0.20)
-4
-2
0
2
4

A



Conclusions: On the basis of a bi-directional genetic approach that limits confounding, our study suggests that a higher BMI leads to lower 25(OH)D, while any effects of lower 25(OH)D increasing BMI are likely to be small. Population level interventions to reduce BMI are expected to decrease the prevalence of vitamin D deficiency.

Mendelian Randomization

- With small number of variants:
  - Weak instrument
- With large numbers of variants
  - Pleiotropy is likely to be a concern.
  - Issues regarding population stratification will become more difficult to address.

Quick Summary

- GCTA
  - Chip heritability can be a useful concept,
    especially when exploring issues above & beyond simple univariate heritability.

Quick Summary

- Polygenic Scores
  - How do biological differences, in form of genetics, (1) manifest as individual differences and (2) interplay with environments.
    - Due to mechanics of GWAS, PGS offer conservative test of GxE.
- Mendelian Randomization
  - How can we leverage biological differences to better understand other associations?

Thanks!

bdomingue@stanford.edu