Applications of Polygenic Scores

Colter Mitchell

2024

Outline

- Part 1: Basics of PGS Construction
- Part 2: Application of PGS
  - Reminder of Basics
  - Life course processes
  - Environment and GxE
  - Instrumental variables
  - CVFS

Past Agendas

Genomics for Social Scientists – Introduction

June 10-14, 2024

$200 course fee

Maximum 30 participants

Download the application (PDF)

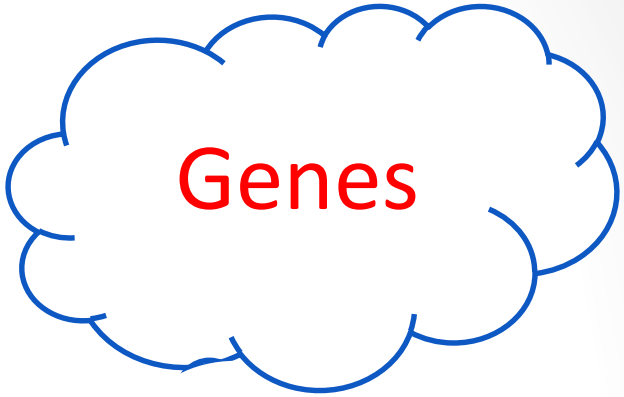
Applications are due February 19th, 2024 (extended deadline)

Travel stipends are available (letter of support is required upon application, verifying that the applicant is a student, post-doc, or early career researcher)

Researchers from the University of Michigan invite you to apply to the 8th annual Genomics for Social Scientists – Introduction workshop, held in-person June 10-14, 2024. The purpose of this NIA-sponsored workshop is to familiarize researchers with genetic data and provide hands-on training on incorporating genetic information into social science analyses. Participants will use tutorial versions of the Health and Retirement Study core survey data and genetic data files.

Potential

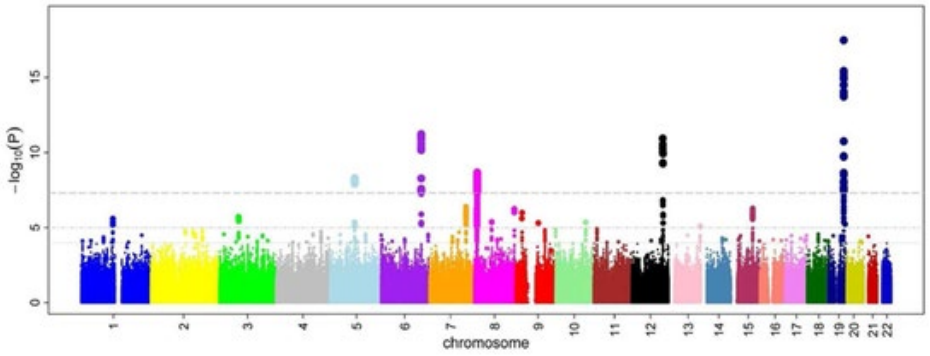
- Constant
- Predictive
- Inexpensive
- Useful for several applications



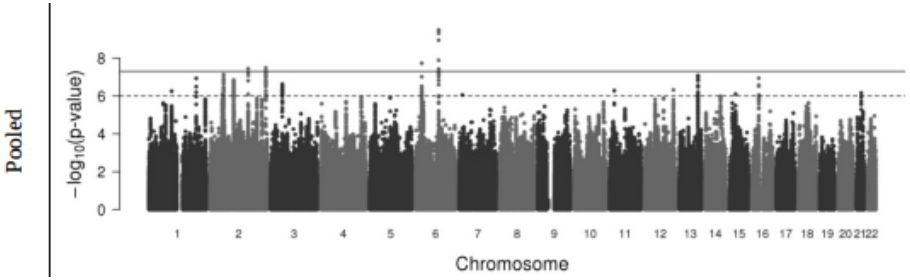
Typical approach for score construction

$$PGS_i = \sum_{j=1}^{J} W_j G_{ij}$$

IDNO	Polygenic_Score
100001	51374.52
100002	57506.1
100003	54567.35
100004	50922.69
100005	51467.5
100006	56791.58
100007	53955.28
100008	58652.57
100009	58987.74
100010	56127.94



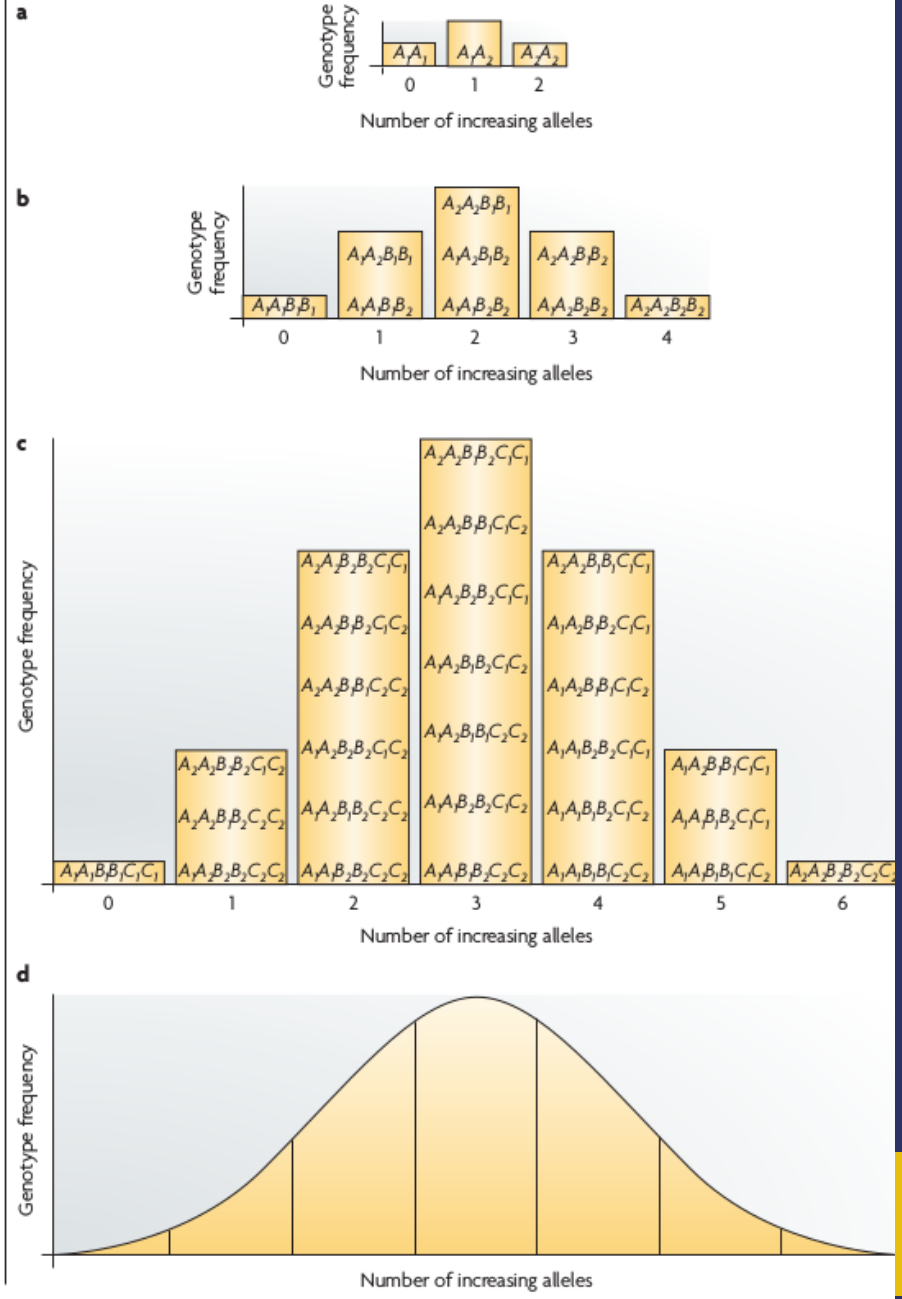
Score Construction



+

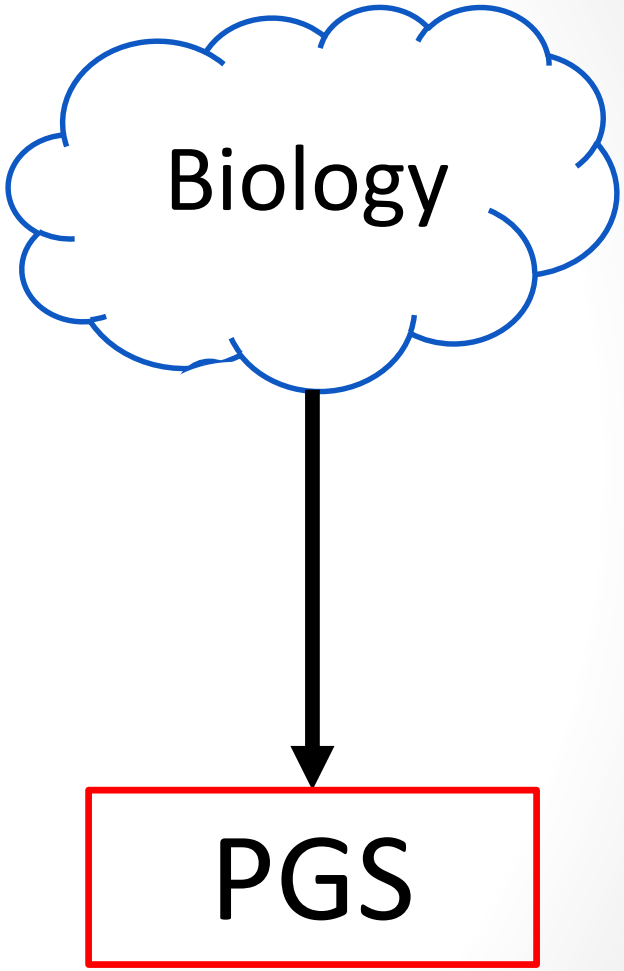
	SNP 1	SNP 2	$\cdots$	SNP 1,000,000
P1	0	1	$\cdots$	2
P2	1	0	$\cdots$	0
P3	1	2	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
P1000	2	1	$\cdots$	2

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



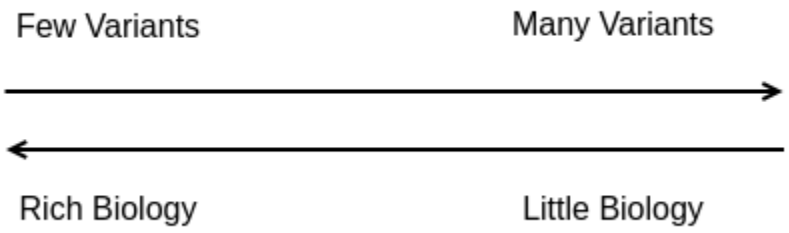
Practicalities

- GWAS of many traits
- Larger and larger effects
- Correlation within ancestry and improving representation
- Working to improve interpretation of GWAS and PGS



Genetic Predictors

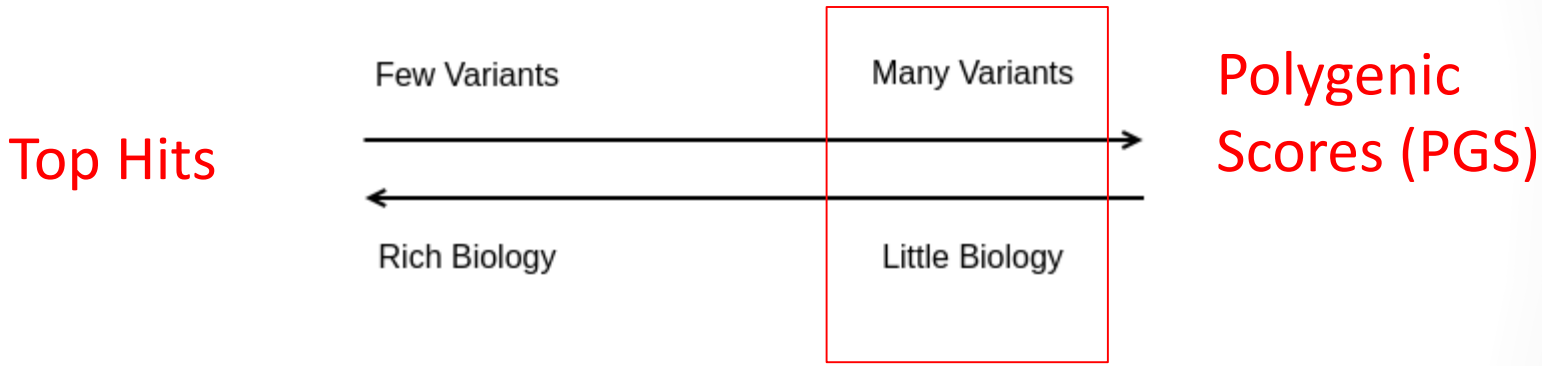
(Predictive) Power/Mechanisms trade-off



- Always important to think about the trade-off when incorporating a genetic predictor.

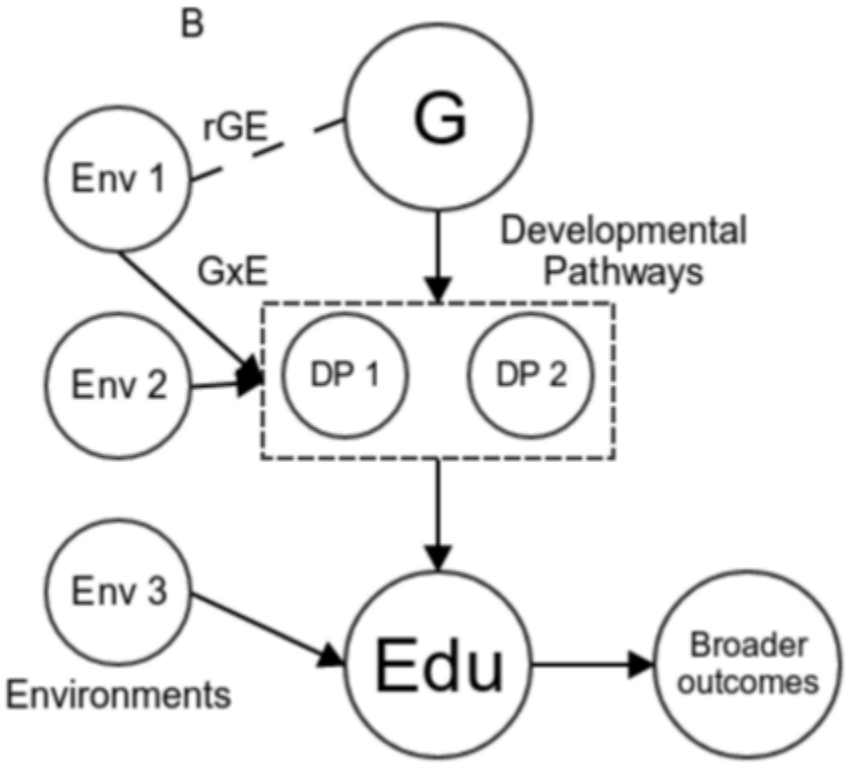
Genetic Predictors

(Predictive) Power/Mechanisms trade-off



Biologically Agnostic

- When we use more powerful genetic predictors, we generally have less of a sense for the specific biological mechanisms. E.g., pleiotropy becomes a major problem.



Recall that Genes are:

- Fixed across life course
- Not due to reverse causality

Not many other predictors have such properties

Three promising avenues for incorporation of polygenic scores

Life course

What is the process through which individual-level genetic endowments come to manifest as phenotypes?

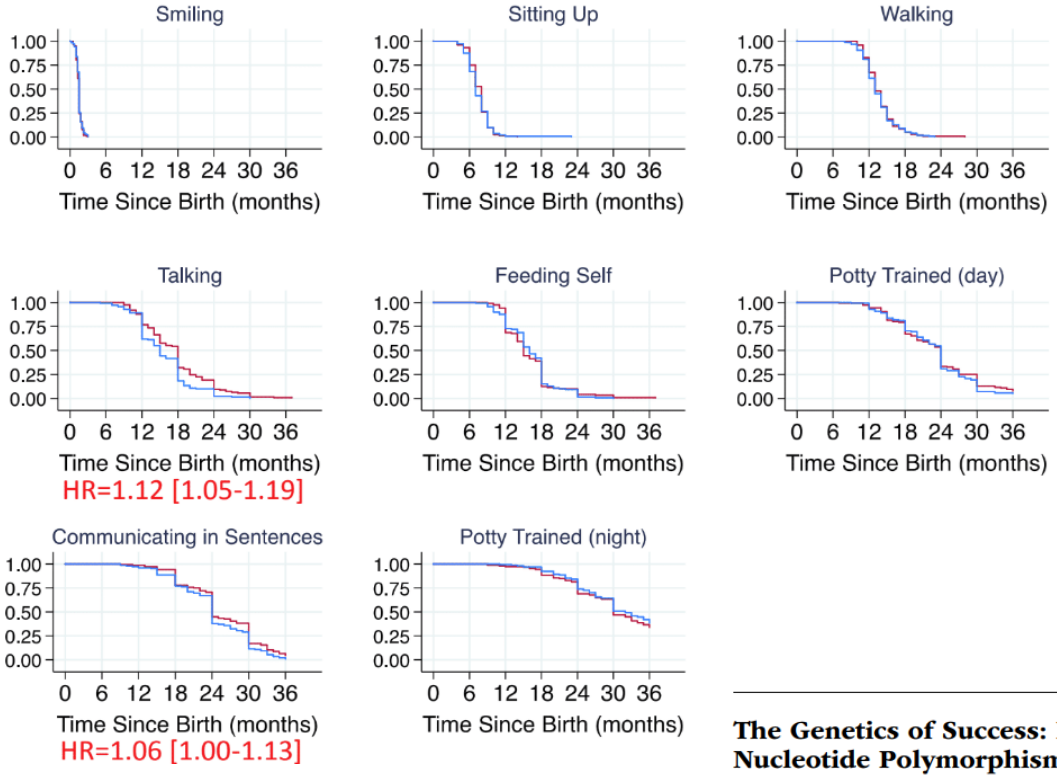
Mendelian Randomization

Using the random component of genetic inheritance as an instrumental variable.

Environment

How are genetic liabilities stratified across environmental exposures? How do genetics and environments combine to influence behavior?

Education PGS & Early Childhood



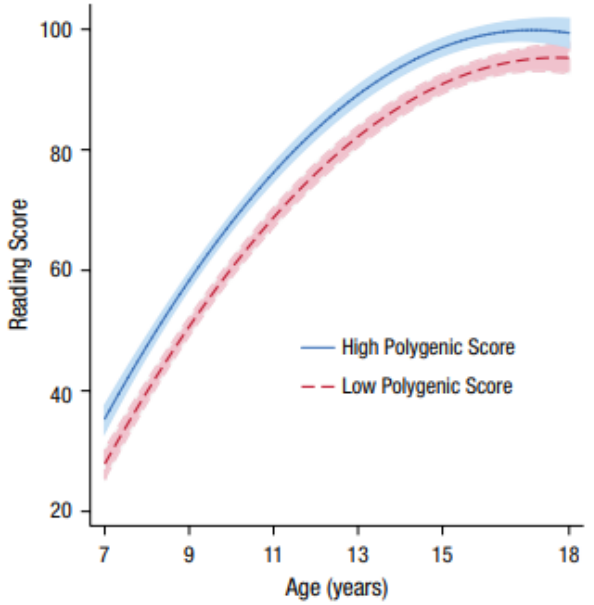


Fig. 3. Children with higher polygenic scores acquired reading skills more rapidly. Association between age and reading skill (as measured by the Burt Word Reading Test; Scottish Council for Research in Education, 1976), separately for children with high polygenic scores (≥ 1 *SD* above the mean; $n = 159$) and those with low polygenic scores (≥ 1 *SD* below the mean; $n = 147$). The shaded areas show 95% confidence intervals.

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

Psychological Science
2016, Vol. 27(7) 957–972
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616643070
pss.sagepub.com
SAGE

Daniel W. Belsky[1,2], Terrie E. Moffitt[3,4,5,6], David L. Corcoran[5], Benjamin Domingue[7], HonaLee Harrington[3], Sean Hogan[8], Renate Houts[3], Sandhya Ramrakha[8], Karen Sugden[3], Benjamin S. Williams[3], Richie Poulton[8], and Avshalom Caspi[3,4,5,6]

[1]Department of Medicine, Duke University School of Medicine; [2]Social Science Research Institute, Duke University; [3]Department of Psychology & Neuroscience, Duke University; [4]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine; [5]Center for Genomic and Computational Biology, Duke University; [6]MRC Social, Genetic & Developmental Psychiatry Research Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London; [7]Graduate School of Education, Stanford University; and [8]Dunedin Multidisciplinary Health & Development Research Unit, Department of Psychology, University of Otago

The Education PGS & later life

Table 14: PENSIONS AND HOUSEHOLD WEALTH

Dep. Var:	Has Pension	Pension Wealth	Log Wealth	Log Wealth
	[1]	[2]	[3]	[4]
EA Score	0.012	0.004	0.120***	0.207***
	(0.008)	(0.023)	(0.021)	(0.031)
DB Pension			0.385***	0.186***
			(0.034)	(0.049)
EA Score × DB Pension				-0.169***
				(0.034)
Obs.	15660	8717	15660	15660
R^2	0.168	0.695	0.419	0.429
Standard Controls	X	X	X	X
Principal Comp.	X	X	X	X
Full Educ. Controls	X	X	X	X
Log Income				

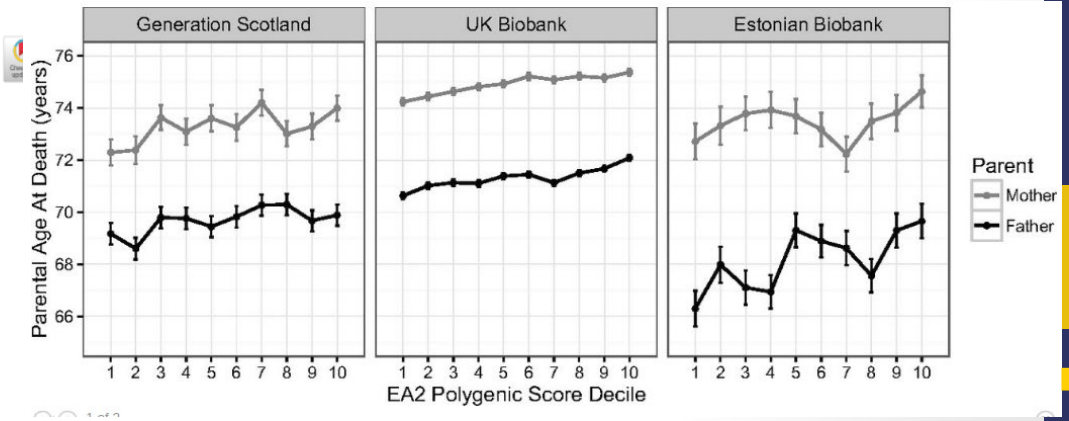
GENETIC ENDOWMENTS AND WEALTH INEQUALITY

Daniel Barth
Nicholas W. Papageorge
Kevin Thom

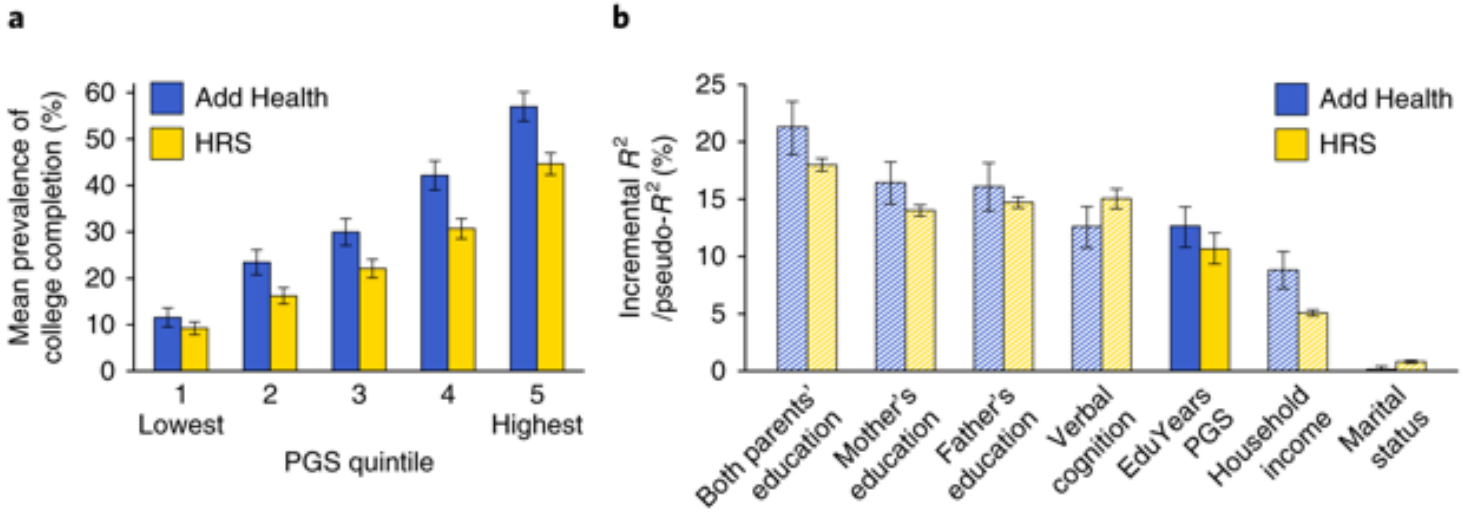
Working Paper 24642
http://www.nber.org/papers/w24642

Genetic variants linked to education predict longevity

Riccardo E. Marioni, Stuart J. Ritchie, Peter K. Joshi, Saskia P. Hagenaars, Aysu Okbay, Krista Fischer, Mark J. Adams, W. David Hill, Gail Davies, Social Science Genetic Association Consortium, Reka Nagy, Carmen Amador, Kristi Läll, Andres Metspalu, David C. Liewald, Archie Campbell, James F. Wilson, Caroline Hayward, Tõnu Esko, David J. Porteous, Catharine R. Gale, and Ian J. Deary

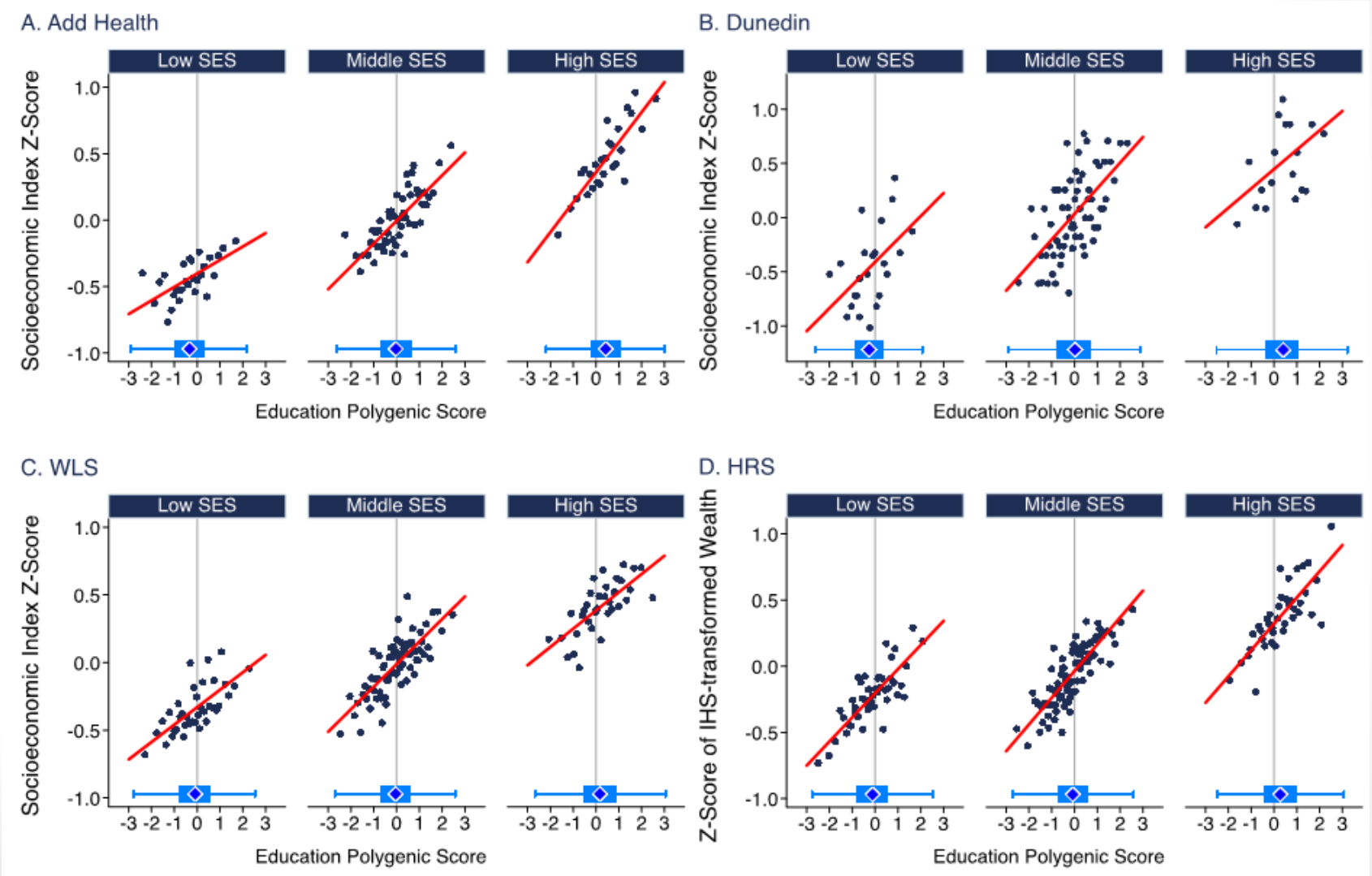


How big are these effects?



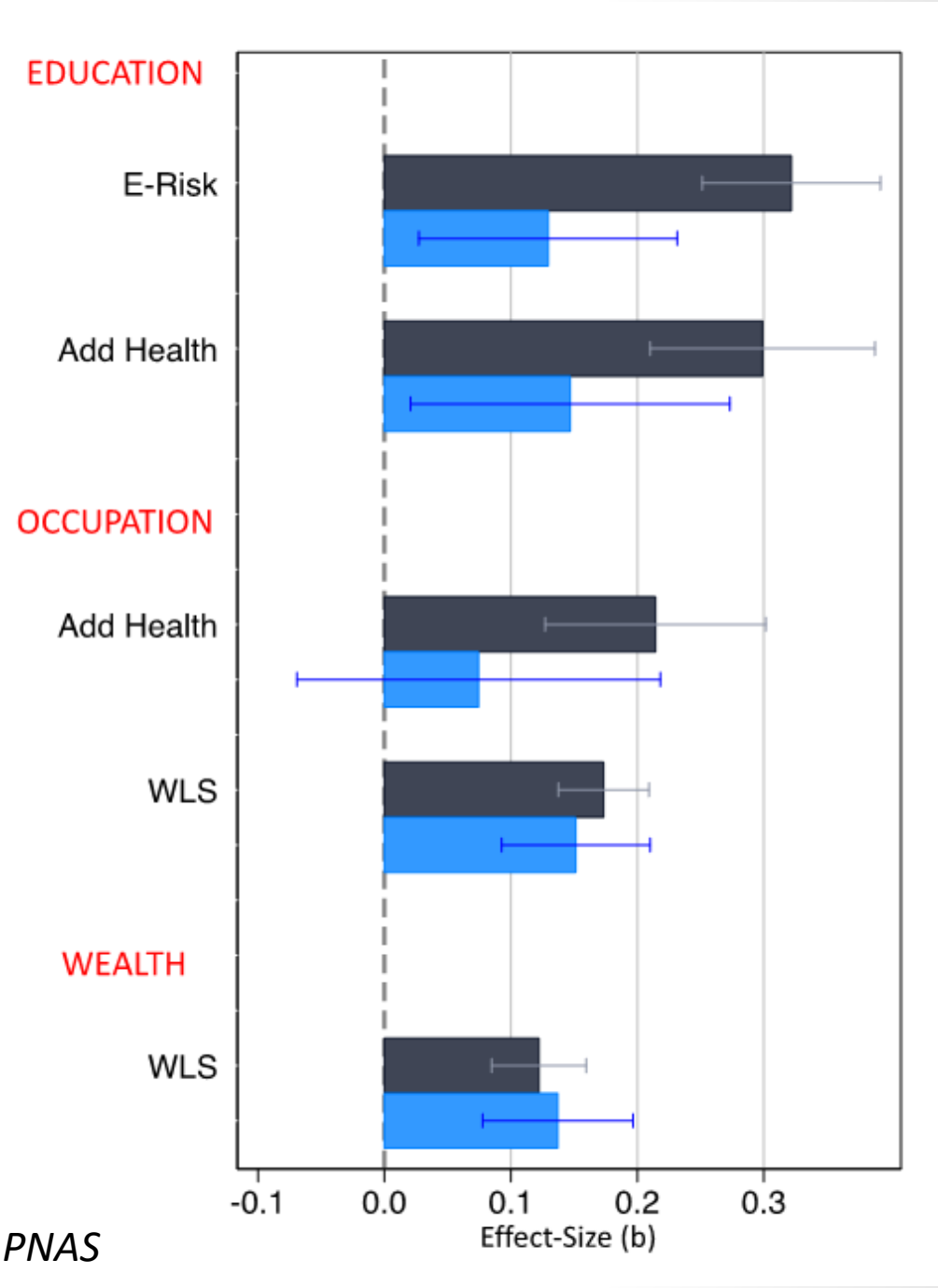
- Larger samples for GWAS will lead to higher variance explained
- Most contextual effects found in small samples are substantially smaller in large studies
- We intervene on contexts that have much smaller effects and without a causal mechanism determined

How robust is prediction?



Belsky & Domingue, et al, 2018 *PNAS*

Also Predicts within families.



Belsky & Domingue, et al, 2018 *PNAS*

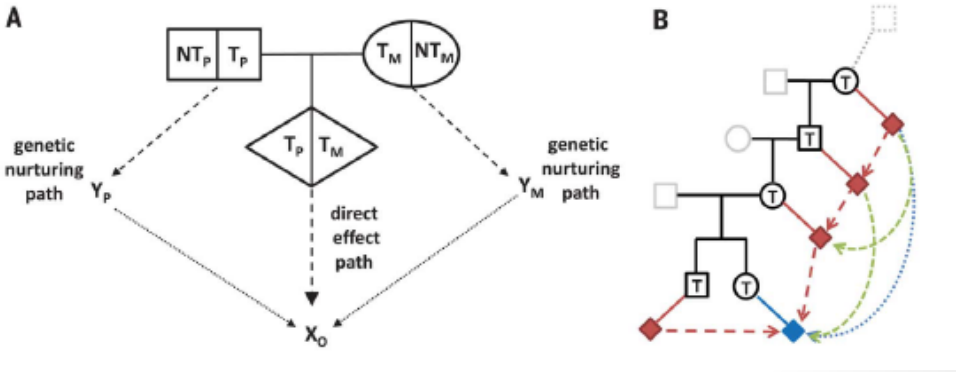
The role of environments

GWAS of socially contextualized phenotypes will presumably pick up more than just biological influences.

The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health

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

PNAS January 9, 2018. 201711803; published ahead of print January 9, 2018. https://doi.org/10.1073/pnas.1711803115



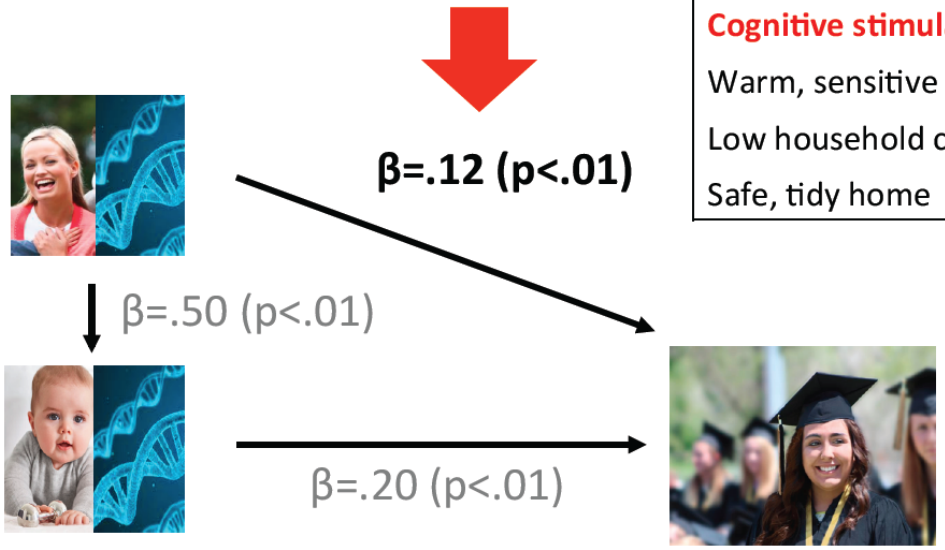
HUMAN GENOMICS

The nature of nurture: Effects of parental genotypes

Augustine Kong,[1,2,3*] Gudmar Thorleifsson,[1] Michael L. Frigge,[1] Bjarni J. Vilhjalmsson,[4,5] Alexander I. Young,[1,2,6] Thorgeir E. Thorgeirsson,[1] Stefania Benonisdottir,[1] Asmundur Oddsson,[1] Bjarni V. Halldorsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1,3] Agnar Helgason,[1,7] Gyda Bjornsdottir,[1] Unnur Thorsteinsdottir,[1,8] Kari Stefansson[1,8*]

The role of environments

Evidence for genetic nurture

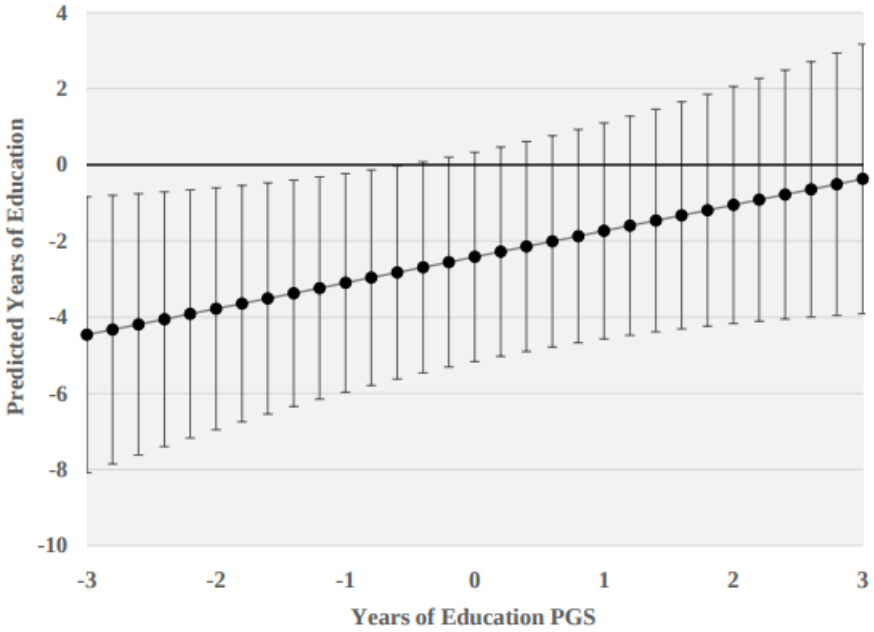


MEDIATION BY MOMS PARENTING	% mediated	P-val.
Cognitive stimulation	**75%**	**<.01**
Warm, sensitive parenting	17%	ns
Low household chaos	42%	<.01
Safe, tidy home	25%	<.01

Wertz et al 2019 *Dev Psych*

The role of environments

Figure 2. Difference in Predicted Years of Education: Veterans versus Non-Veterans



THE EFFECT OF VIETNAM-ERA CONSCRIPTION AND GENETIC POTENTIAL FOR EDUCATIONAL ATTAINMENT ON SCHOOLING OUTCOMES

Lauren L. Schmitz
Dalton Conley

Working Paper 22393
http://www.nber.org/papers/w22393

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

NATURE HUMAN BEHAVIOUR LETTERS

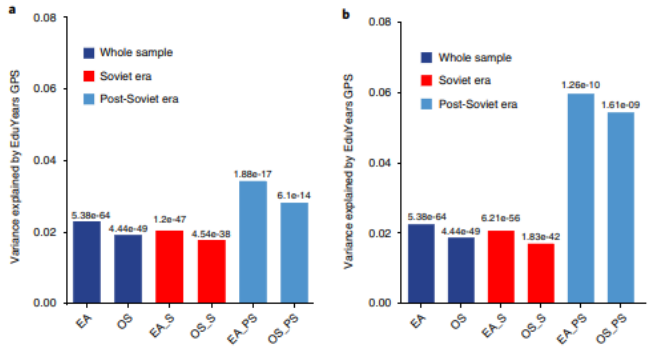


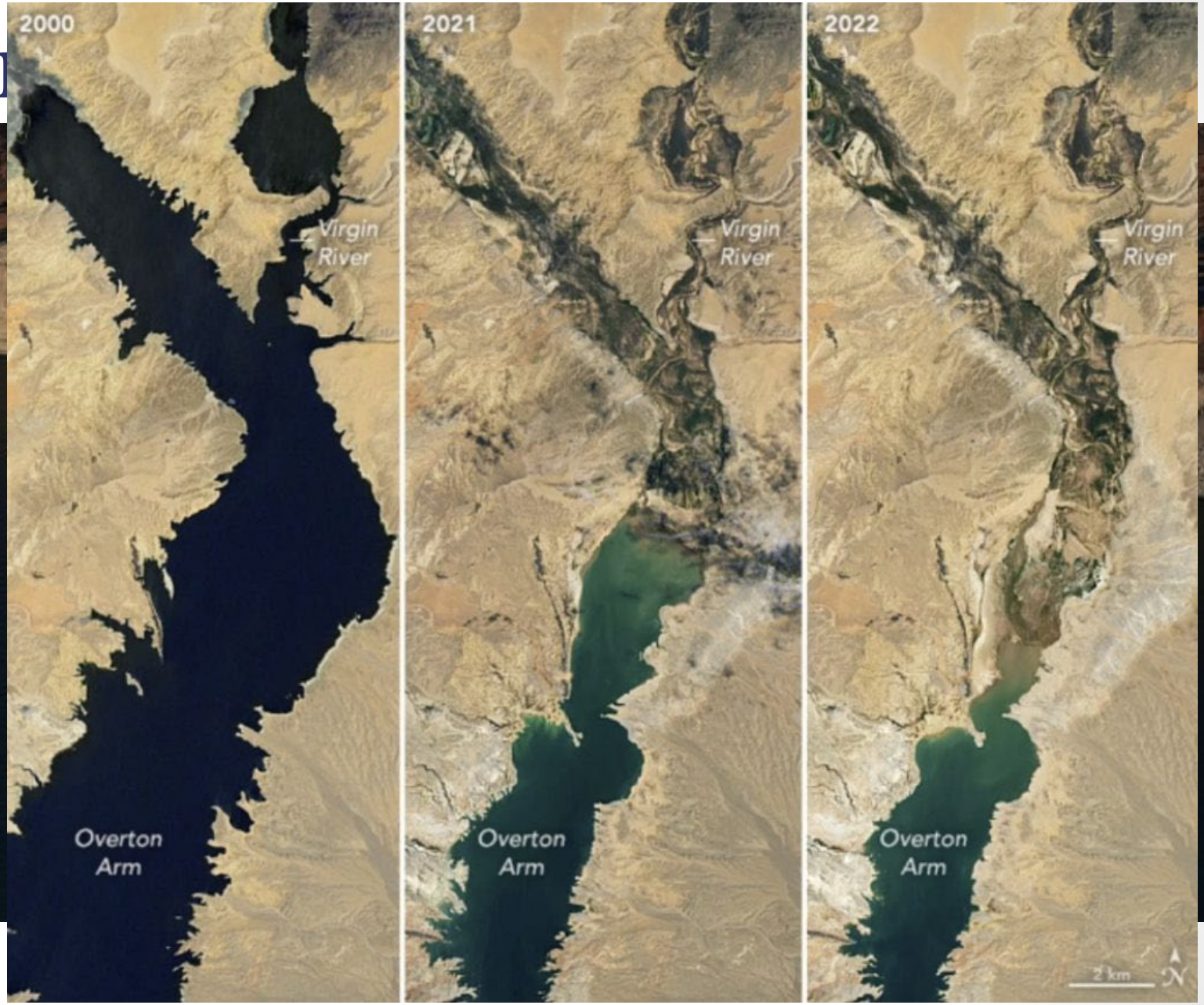
Fig. 1 | Variance explained by EduYears GPS in the post-Soviet and Soviet groups. a,**b**, The GPS was calculated using a 0.1 GWA study *P* value threshold for educational attainment (EA) and occupational status (OS) for the whole EGCUT sample (N(EA) = 12,483; N(OS) = 11,419) and when divided into historical eras using two cutoffs: the post-Soviet (PS) group included participants 15 years or younger when independence was regained, and the Soviet (S) group included the rest of the participants (N(EA_S) = 10,381; N(OS_S) = 9,417; N(EA_PS) = 2,102; N(OS_PS) = 2,002) (**a**); the post-Soviet (PS) group included participants 10 years or younger when independence was regained and the Soviet (S) group included the rest of the participants (N(EA_S) = 11,808; N(OS_S) = 10,767; N(EA_PS) = 675; N(OS_PS) = 652) (**b**).

nature human behaviour
LETTERS
https://doi.org/10.1038/s41562-018-0332-5

Genetic influence on social outcomes during and after the Soviet era in Estonia

Kaili Rimfeld[1*], Eva Krapohl[1], Maciej Trzaskowski[2], Jonathan R. I. Coleman[1,3], Saskia Selzam[1], Philip S. Dale[4], Tonu Esko[5], Andres Metspalu[5] and Robert Plomin[1]

E


2000
Virgin River
Overton Arm
2021
Virgin River
Overton Arm
2022
Virgin River
Overton Arm
2 km
N

GxE Reviews

American Journal of PUBLIC HEALTH

Special Issue: Integration of Behavioral, Social Science and Genetics Research,

Vol. 103, S1 (October 2013)

Genetic Differential Sensitivity to Social Environments: Implications for Research

Colter Mitchell, PhD, Sara McLanahan, PhD, Jeanne Brooks-Gunn, PhD, Irwin Garfinkel, PhD, John Hobcraft, BSc, and Daniel Notterman, MD

Defining the Environment in Gene–Environment Research: Lessons From Social Epidemiology

Jason D. Boardman, PhD, Jonathan Daw, PhD, and Jeremy Freese, PhD

Annual Review of Psychology
2014. 65:41–70

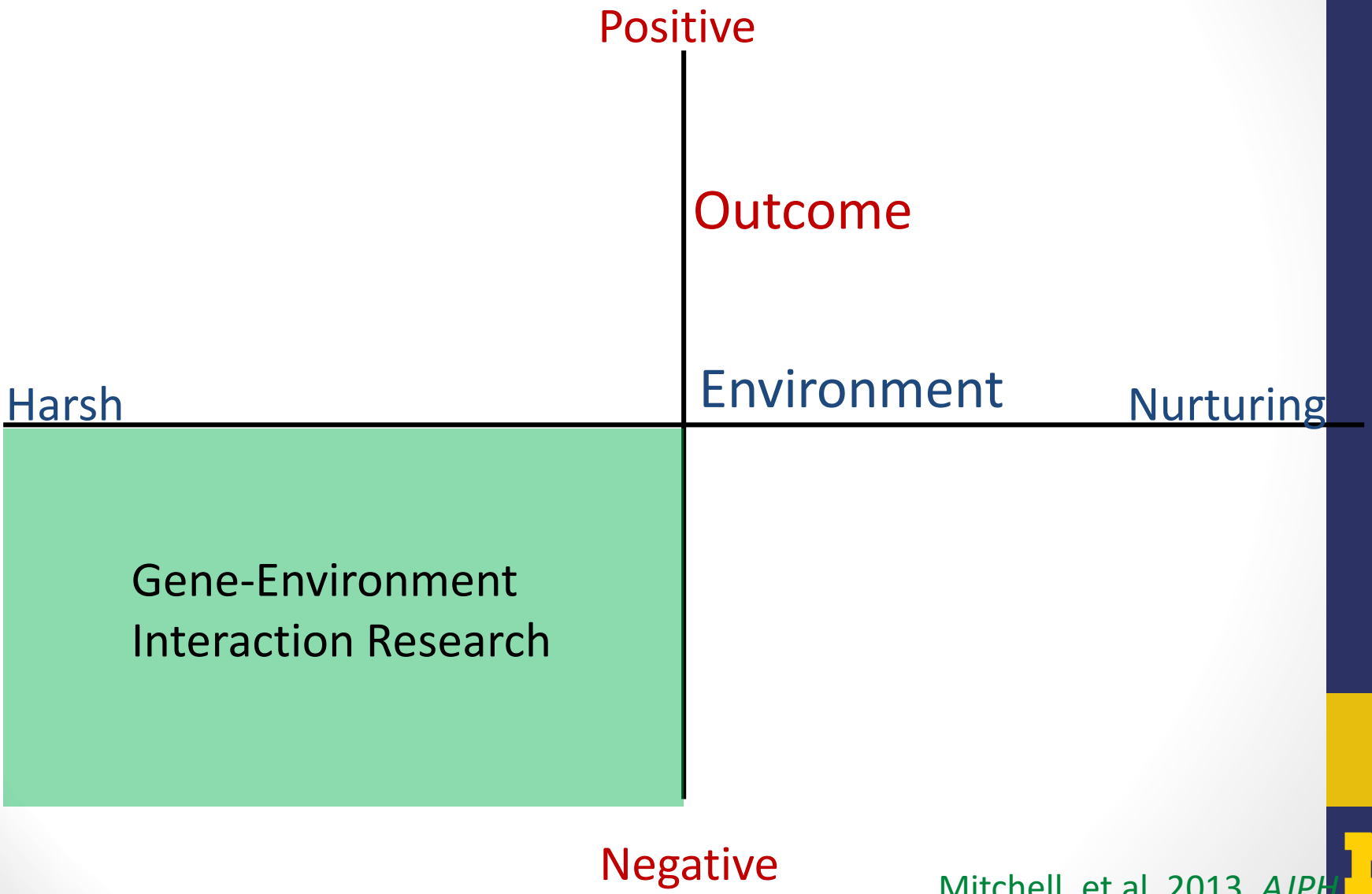
Gene-Environment Interaction

Stephen B. Manuck[1] and Jeanne M. McCaffery[2]

[1]Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260; email: manuck@pitt.edu

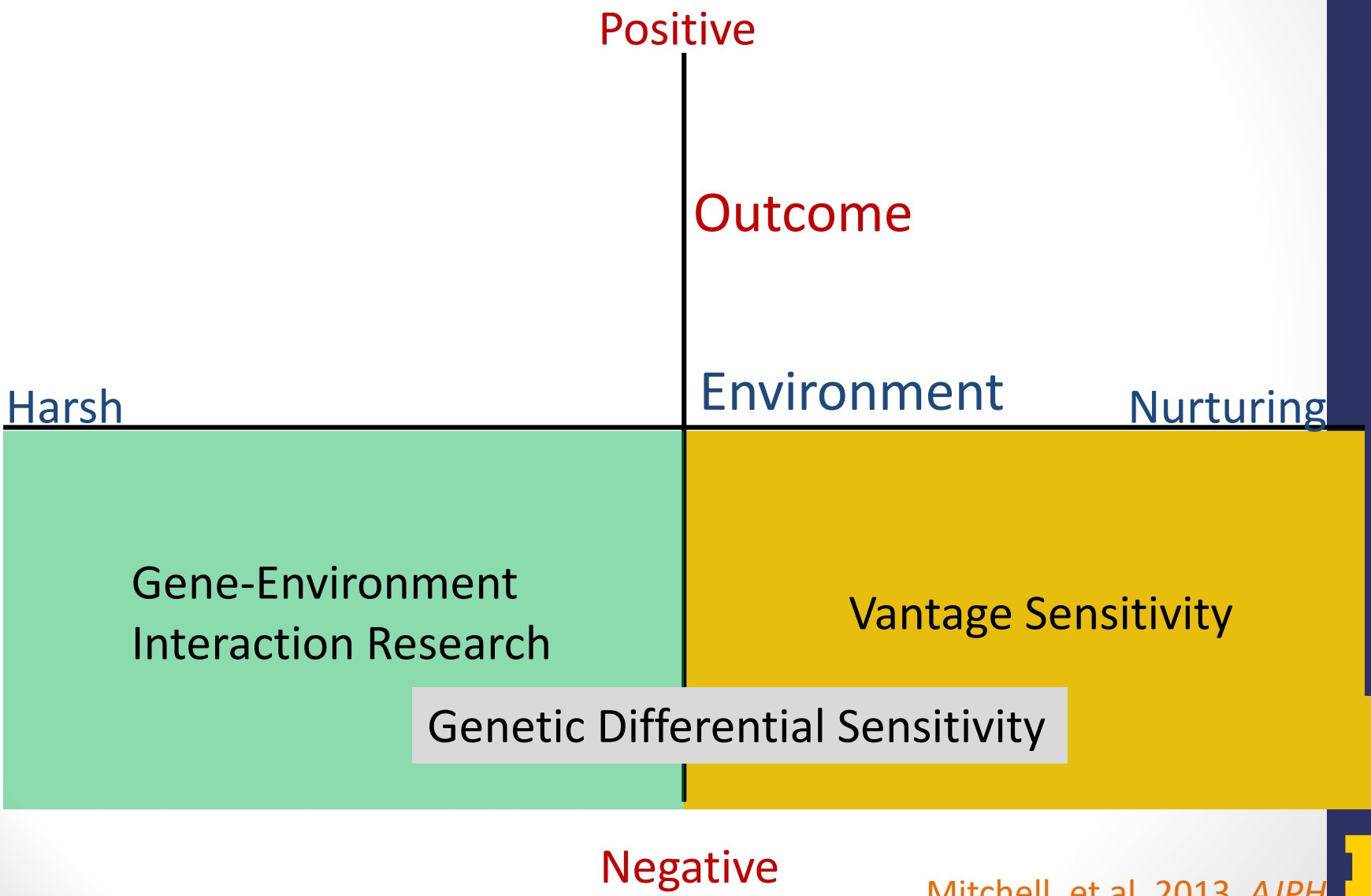
[2]Department of Psychiatry and Human Behavior, The Miriam Hospital, and Warren Alpert School of Medicine at Brown University, Providence, Rhode Island 02903; email: jeanne_mccaffery@brown.edu

Recent Past of GxE Work



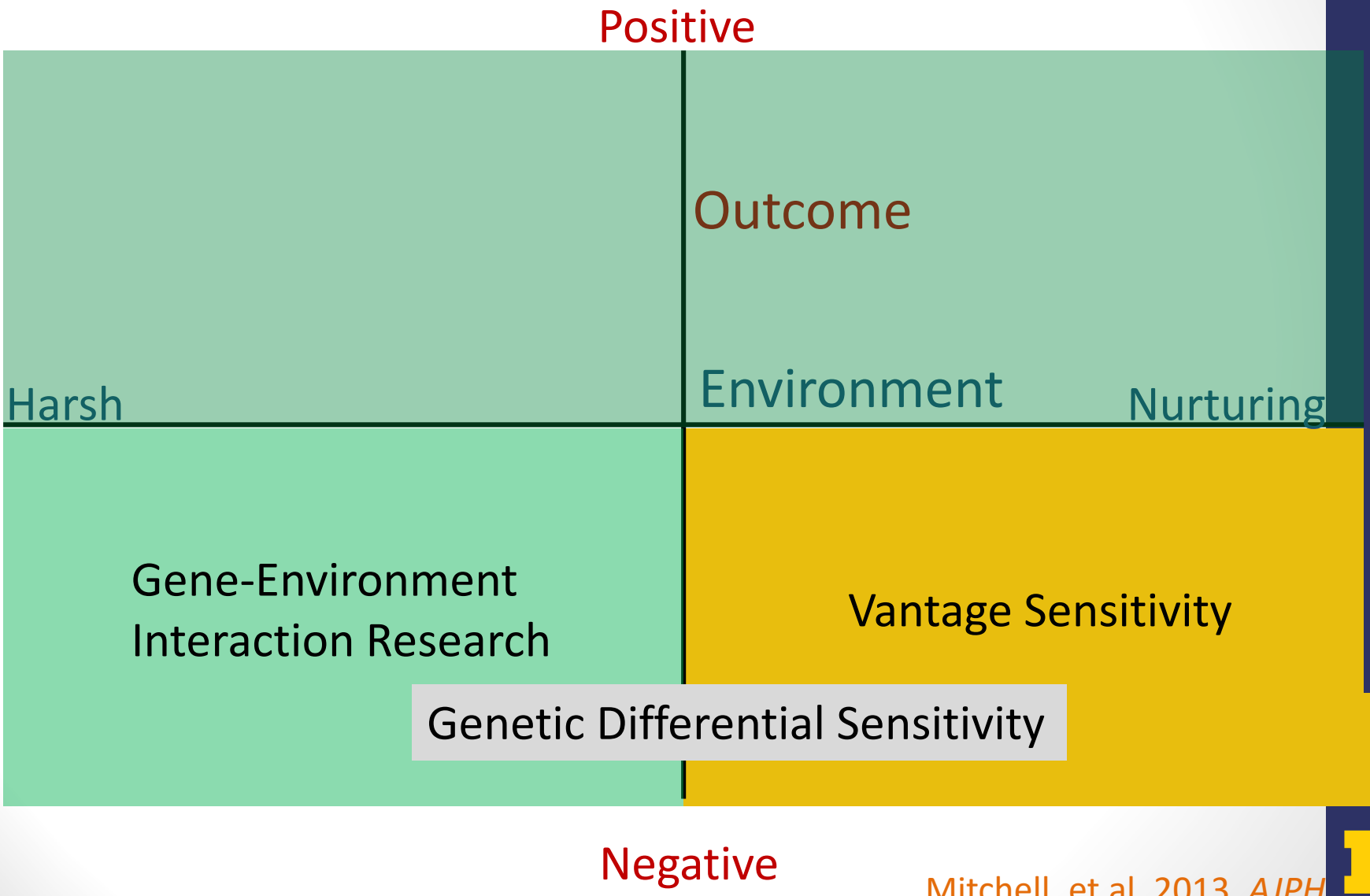
Mitchell, et al. 2013. *AJPH*

Measurement of E



Mitchell, et al. 2013. *AJPH*

Measuring Outcome



Mitchell, et al. 2013. *AJPH*

Formal Tests of Models

- Original work tested for an interaction and then visually examined (cross-over, vantage, etc).
- Larger push to be able to distinguish these models
  - Belsky, Pluess, & Widaman, 2013; Lee, Lei, & Brody, 2015; Roisman et al., 2012; Widaman et al., 2012; M. Del. Giudice 2017
- Regions of Significance (RoS)- values of the environmental variable for which the moderator is significantly associated with the outcome

Regions of Significance

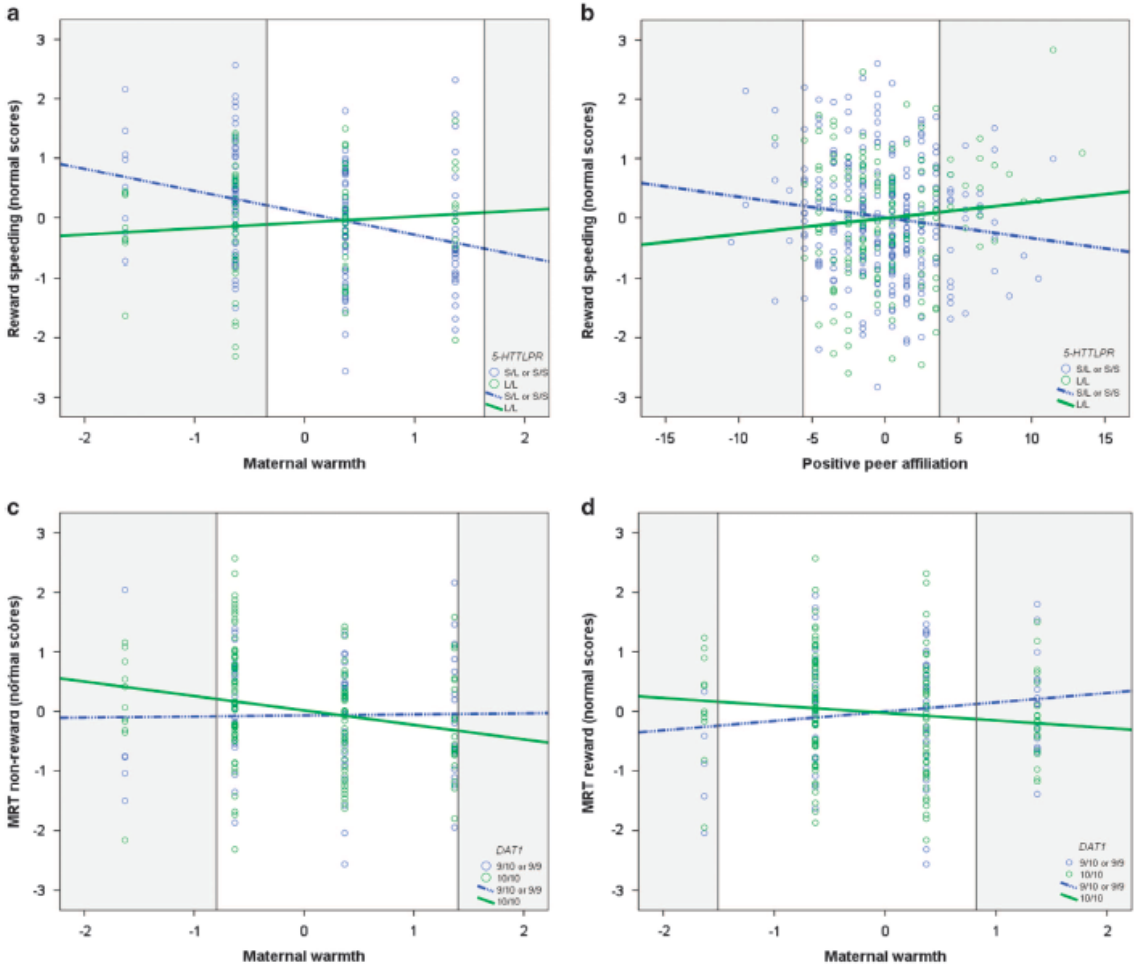


Figure 1. (**a**) Interaction between *5-HTT* and maternal warmth on reward speeding ($B = -0.45$, $P = 0.005$; normal score (0) = 27.71 ms). The shaded areas indicate the regions of significance (RoS), lower threshold $X = -0.34$; upper threshold $X = 1.64$. (**b**) Interaction between *5-HTT* and positive peer affiliation on reward speeding ($B = -0.07$, $P = 0.012$; normal score (0) = 25.52 ms). The shaded areas indicate the RoS, lower threshold $X = -5.61$; upper threshold $X = 3.71$. (**c**) Interaction between *DAT1* and maternal warmth on the mean reaction time during non-reward ($B = 0.40$, $P = 0.012$; normal score (0) = 324.90 ms). The shaded areas indicate the RoS, lower threshold $X = -0.80$; upper threshold $X = 1.40$. (**d**) Interaction between *DAT1* and maternal warmth on the mean reaction time during reward ($B = 0.41$, $P = 0.013$; normal score (0) = 296.31 ms). The shaded areas indicate the RoS, lower threshold $X = -1.51$; upper threshold $X = 0.82$. Values in the RoS are significant. MRT, mean reaction time.

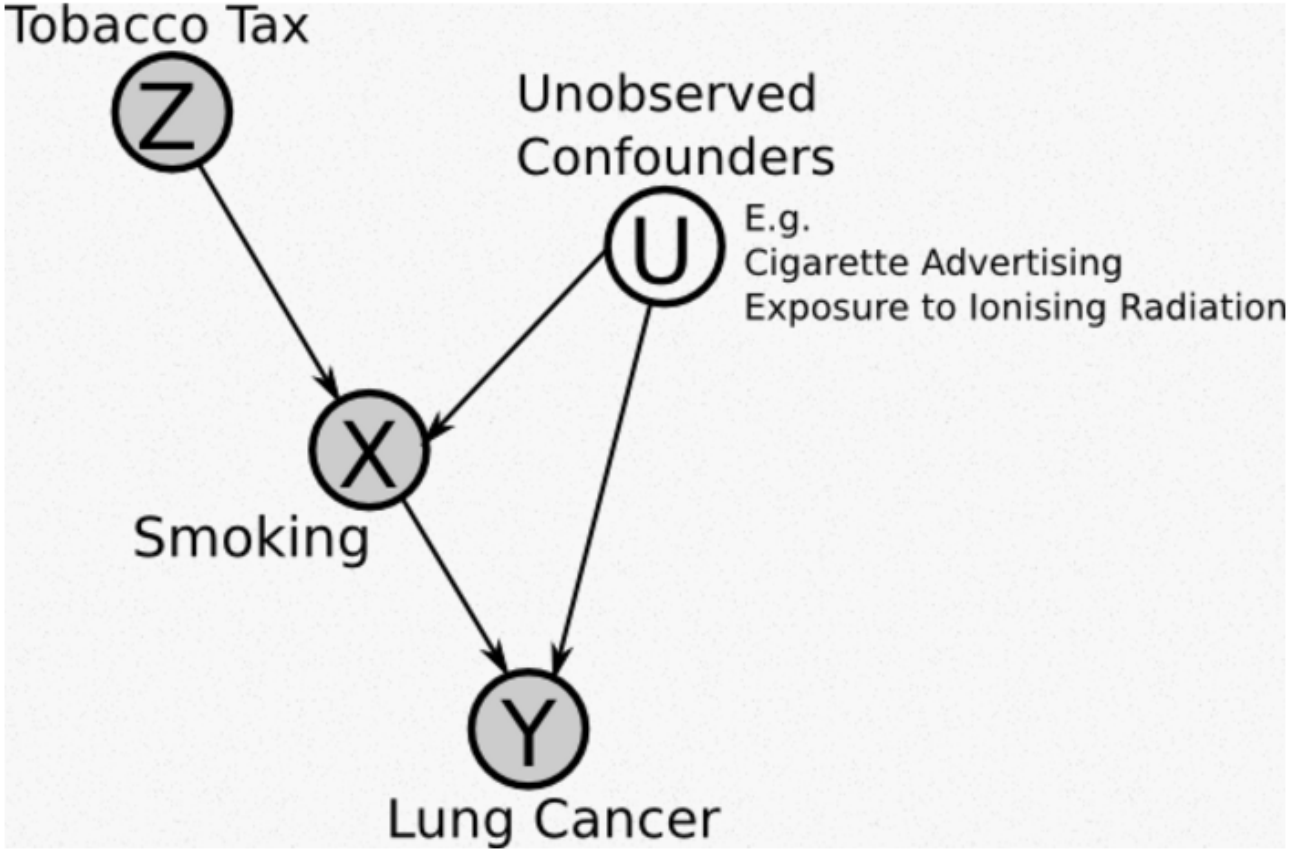
Richards et al *Translational Psych*, 2016

Genetic Predictors

(Predictive) Power/Mechanisms trade-off

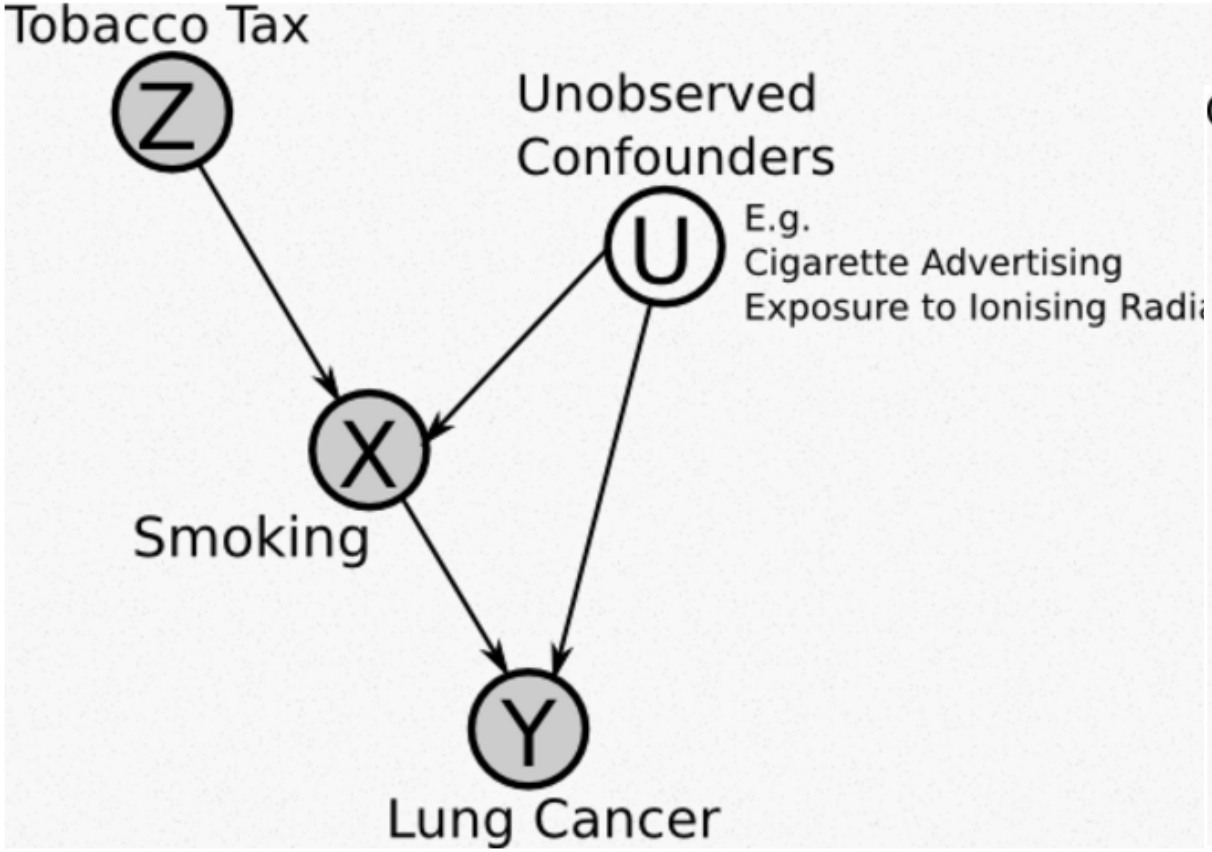
Mendelian Randomization (MR)

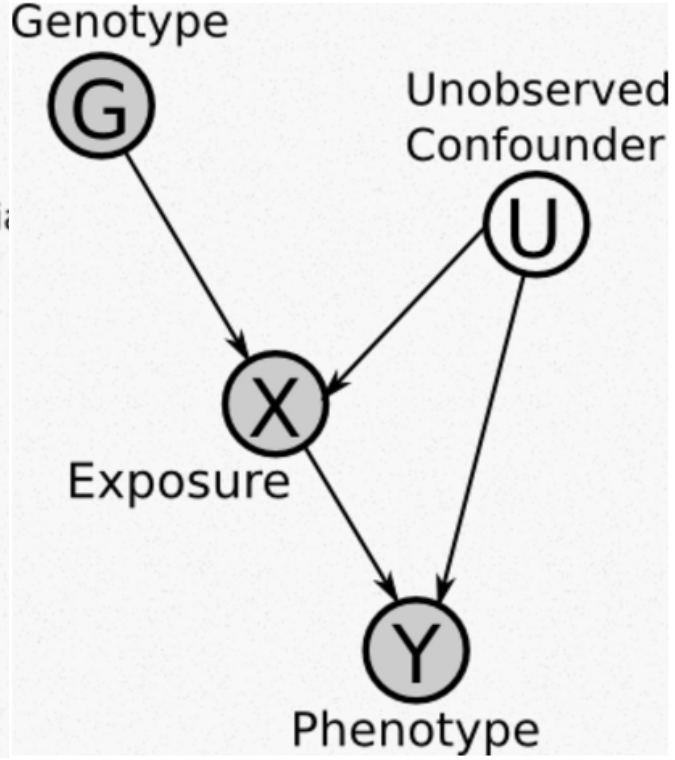
MR: Genes as Instruments



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

MR: Genes as Instruments





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

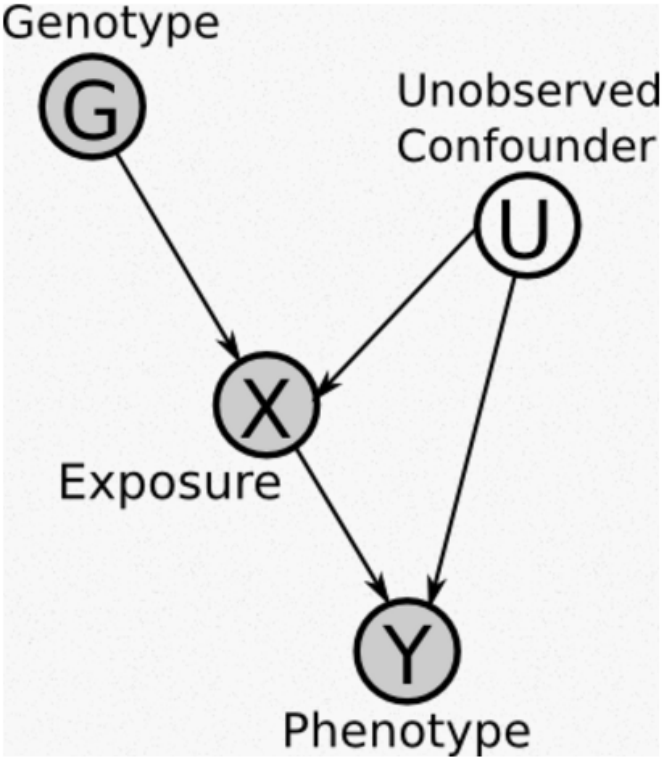
Alcohol Intake and Blood Pressure: A Systematic Review Implementing a Mendelian Randomization Approach

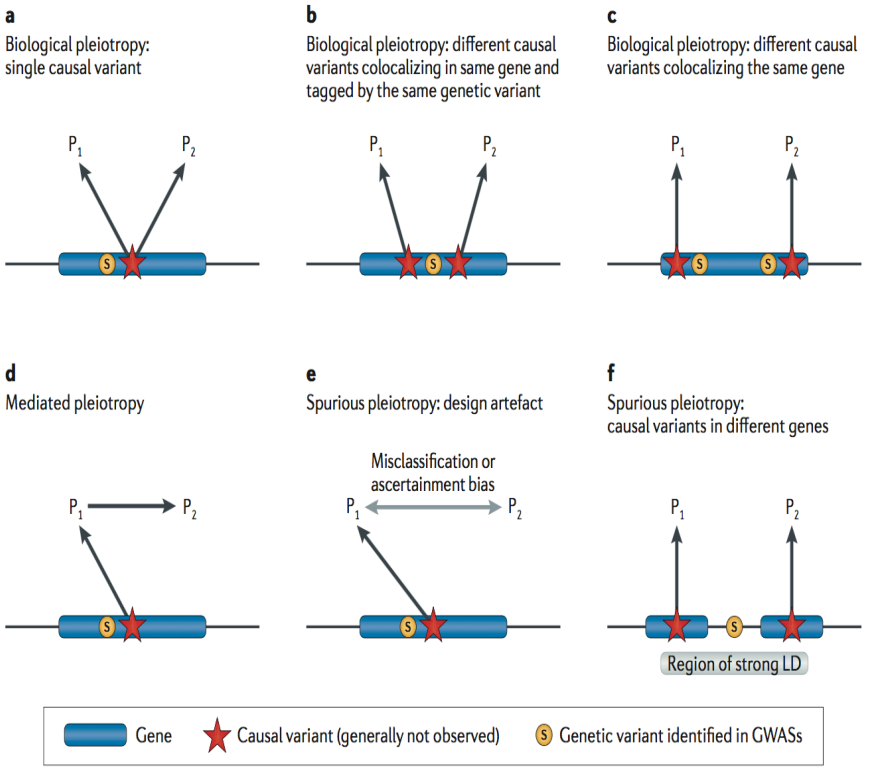
Lina Chen, George Davey Smith, Roger M Harbord, Sarah J Lewis

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

Beware exclusion restriction

Pleiotropy is a problem





New techniques

- Clinical Prediction
- PGS as IV
- Rapid development here, so keep your eyes open

PNAS

Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data

Thomas A. DiPrete[a,1,2], Casper A. P. Burik[b,1], and Philipp D. Koellinger[b,1,2]

[a]Department of Sociology, Columbia University, New York, NY 10027; and [b]Department of Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved March 21, 2018 (received for review May 3, 2017)

CVFS CIDI Data

Table 1. Sample description

	CVFS sample[a] 2016–2018 (*N* = 10 714)		Chitwan district 2011 (*N* = 579 984)		Nepal census 2011 (*N* = 26 494 504)	
Gender	Number	%	Number	%	Number	%
Male	4923	46.0 (45.0, 46.9)	279 087	48.1	12 849 041	48.5
Female	5791	54.1 (53.1, 55.0)	300 897	51.9	13 645 463	51.5
Age						
15–24	3935	36.7 (35.8, 37,6)	127 870	35.6	5 290 051	35.1
25–34	3008	28.1 (27.2, 28.9)	90 545	25.2	3 814 659	25.3
35–44	2080	19.4 (18.7, 20.2)	70 718	19.7	2 990 440	19.8
45–59	1691	15.8 (15.1, 16.5)	69 919	19.5	2 996 698	19.9
Ethnicity						
Brahmin/Chhetri	4634	43.3 (42.3, 44.2)	239 466	41.3	8 499 061	32.1
Hill Janajati	2106	19.7 (18.9, 20.4)	176 875	30.5	5 886 260	22.2
Dalit	1301	12.1 (11.5, 12.8)	50 655	8.7	3 474 767	13.1
Newar	640	6.0 (5.5, 6,4)	30 256	5.2	1 321 933	5.0
Terai Janajati	1942	18.1 (17.4, 18.9)	63 592	11.0	2 257 951	8.5
Others	91	0.9 (0.7, 1.0)	19 140	3.3	5 054 532	19.1
S.L.C. or more						
Yes	4098	38.3 (37.3, 39.2)	3 288 783	12.4	102 483	17.7
No	6615	61.8 (60.8, 62.7)	23 205 721	87.6	477 501	82.3

SLC = school leaving certificate.
[a]The age range in the CVFS sample is between 15 years and 59 years. For comparison, the calculated age distribution for the Chitwan district and Nepal census excludes people younger than 15 years and older than 59 years.

K. M. Scott et al. *Psychological Medicine*(2020)

CVFS Genetic Data

- Saliva-easy to collect, room temperature storage for months
- 96% of CIDI participants
- Separate consents for DNA collection and sharing Genetic Data

Expected CVFS Genetic Data Family Relatedness

Family Relatedness	Number of Individuals
Trios (Mother, Father, Child)	5,398*
One parent, >1 sib-pair	1,435
Parent-child duo	618
No parents, >1 sib-pair	939
Unrelated	1,918
Total	**10,308**

*Families can have more than one set of trios if multiple children with both parents participated

Likely close to 2,000 trios (very large from this type of study)
Over 5,100 mother-child pairs
Over 4,500 father-child pairs
Over 2,000 sibling pairs

Applications to CVFS

- MDD and AUD PGS
- Other PGS-Cardiovascular Disease, BMI, Educational Attainment, Alzheimer's Disease, Menarche, Fertility, etc
- Tracing it through life course process
  - Genetic Nurture

- Selection into different environments
- Interact with different environments

- Causal tests:
  - Family models
  - Instrumental Variables