

Polygenic Scores in the National Longitudinal Study of Adolescent to Adult Health

David B. Braudt

Add Health Users Conference,

July 12th, 2022

Polygenic Scores (PGSs): An Overview

- PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

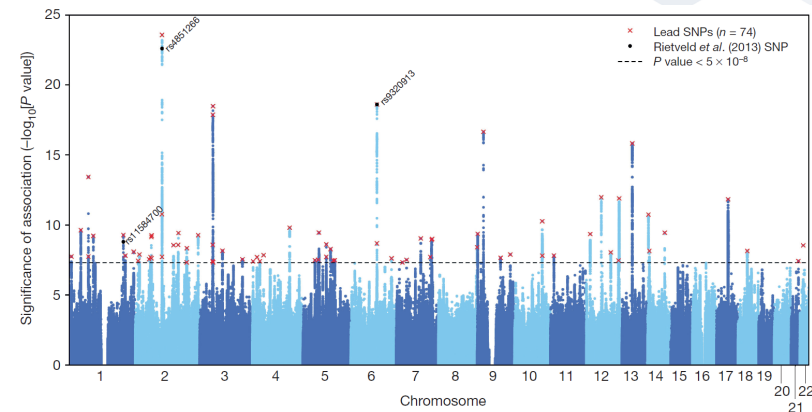
Polygenic Scores (PGSs): An Overview

- PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

ID	SNP 1	SNP 2	SNP 3	...	SNP m
1	0	2	1	...	2
2	1	1	1	...	0
3	0	1	2	...	1
n	2	0	1		1

Molecular genetic data

X



GWAS summary statistics

Polygenic Scores (PGSs): An Overview

- PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

$$PGS_i = \sum_{j=1}^k \beta_j SNP_{ij}$$

Polygenic Scores (PGSs): An Overview

- PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

$$PGS_i = \sum_{j=1}^k \beta_j SNP_{ij} \xrightarrow[\text{within ancestry groups}]{\text{standardized}} \mu_{PGS} = 0 \quad \text{and} \quad sd = 1$$

Polygenic Scores (PGSs): An Overview

- PGSs capture the combined additive genetic influence of SNPs across the entire genome on a specific trait/behavior in a single measure

$$PGS_i = \sum_{j=1}^k \beta_j SNP_{ij} \xrightarrow[\text{within ancestry groups}]{\text{standardized}} \mu_{PGS} = 0 \quad \text{and} \quad sd = 1$$

- Whole genome PGSs
 - Include genetic association from across the entire genome, but eliminate the possibility to testing hypotheses related to specific biological pathways (Belsky and Israel, 2014)

Why/How to use PGSs

- Four of the many possible uses
 - **Nature** net of **Nurture**
 - **Nurture** net of **Nature**
 - How **Nurture** modifies the effect(s) of **Nature**
 - How **Nature** modifies the effect(s) of **Nurture**

Why/How to use PGSs

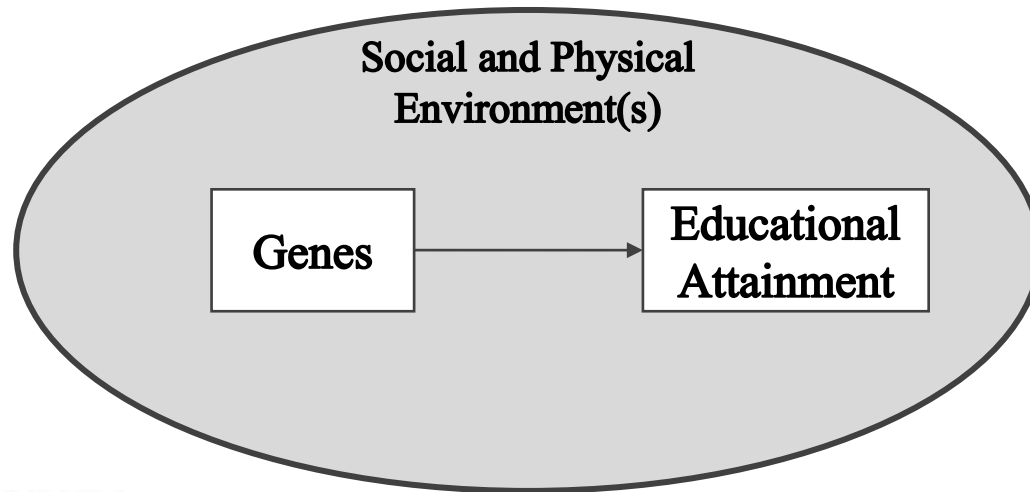
- Four of the many possible uses
 - **Nature** net of **Nurture**
 - **Nurture** net of **Nature**
 - How **Nurture** modifies the effect(s) of **Nature**
 - How **Nature** modifies the effect(s) of **Nurture**

Why/How to use PGSs

- Four of the many possible uses
 - Nature net of Nurture
 - Nurture net of Nature (Control for genetic influences... Endogeneity problem)
 - How Nurture modifies the effect(s) of Nature
 - How Nature modifies the effect(s) of Nurture

Why/How to use PGSs

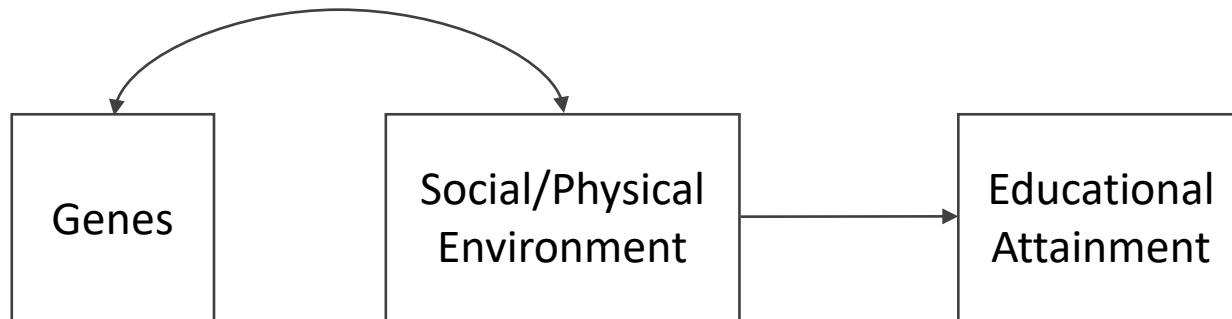
- Four of the many possible uses
 - Nature net of Nurture
 - Nurture net of Nature
 - How Nurture/Nature modifies the effect(s) of Nature/Nurture
 - How Nature modifies the effect(s) of Nurture



Gene-Environment Interactions

Why/How to use PGSs

- Four possible uses
 - Nature net of Nurture
 - Nurture net of Nature
 - How Nurture/Nature modifies the effect(s) of Nature/Nurture
 - How Nature/Nurture modifies the effect(s) of Nurture/Nature



Gene-Environment Correlations

Why/How to use PGSs

Research Article



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

Daniel W. Belsky^{1,2}, Terrie E. Moffitt^{3,4,5,6}, David L. Corcoran⁵, Benjamin Domingue⁷, HonaLee Harrington³, Sean Hogan⁸, Renate Houts³, Sandhya Ramrakha⁸, Karen Sugden³, Benjamin S. Williams³, Richie Poulton⁸, and Avshalom Caspi^{3,4,5,6}

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

Why/How to use PGSs

Research Article



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

Daniel W. Belsky^{1,2}, Terrie E. Moffitt³, Benjamin Domingue⁷, HonaLee Harrington⁴, Renate Houts⁵, Sandhya Ramrakha⁸, Benjamin S. Williams³, Richie Poulton⁹, Avshalom Caspi^{3,4,5,6}

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

Abstract

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study ($N = 918$). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, genes and environments were correlated: Children with higher polygenic scores were born into better-off homes. Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social-mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores. Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.

Nature net of Nurture

Why/How to use PGSs

Research Article



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

Daniel W. Belsky^{1,2}, Terrie E. Moffitt³, Benjamin Domingue⁷, HonaLee Harrist⁴, Renate Houts⁵, Sandhya Ramrakha⁸, Benjamin S. Williams³, Richie Poulton¹, Avshalom Caspi^{3,4,5,6}

Psychological Science
2016, Vol. 27(7) 957–972
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616666666
ps.sagepub.com

Gene-Environment Interactions

Abstract

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study ($N = 918$). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, genes and environments were correlated: Children with higher polygenic scores were born into better-off homes. Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social-mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores. Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.

Why/How to use PGSs

Research Article

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

Daniel W. Belsky^{1,2}, Terrie E. Moffitt^{3,4}, Benjamin Domingue⁷, HonaLee Harrington⁵, Renate Houts⁵, Sandhya Ramrakha⁸, K. Benjamin S. Williams³, Richie Poulton⁸, Avshalom Caspi^{3,4,5,6}



Psychological Science
2016, Vol. 27(7) 957–972
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616666666

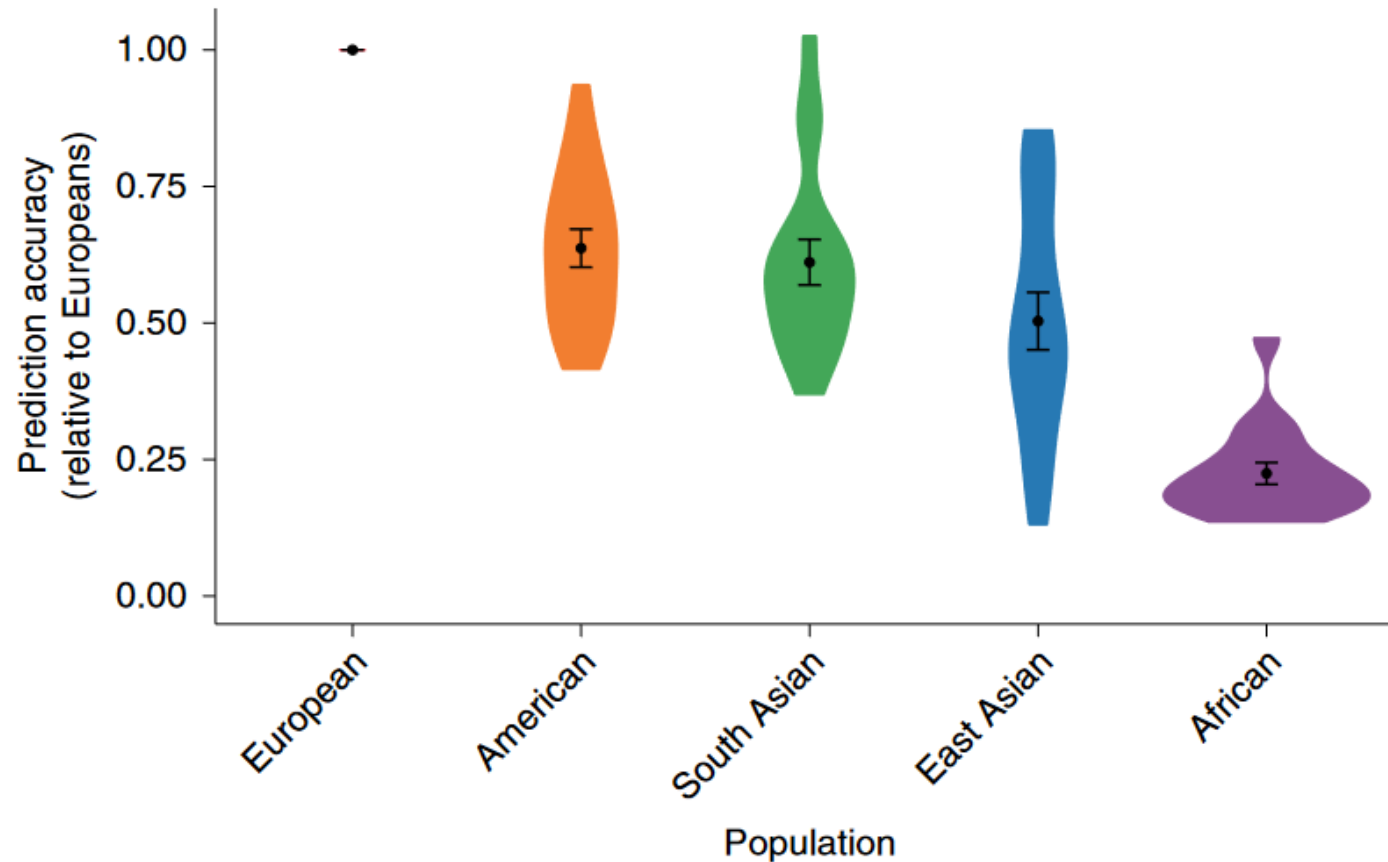
Gene-Environment Correlations

Abstract

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study ($N = 918$). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, **genes and environments were correlated: Children with higher polygenic scores were born into better-off homes.** Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social-mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores. Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.

PGSs and GWAS Genetic Ancestry

PGSs and GWAS Genetic Ancestry

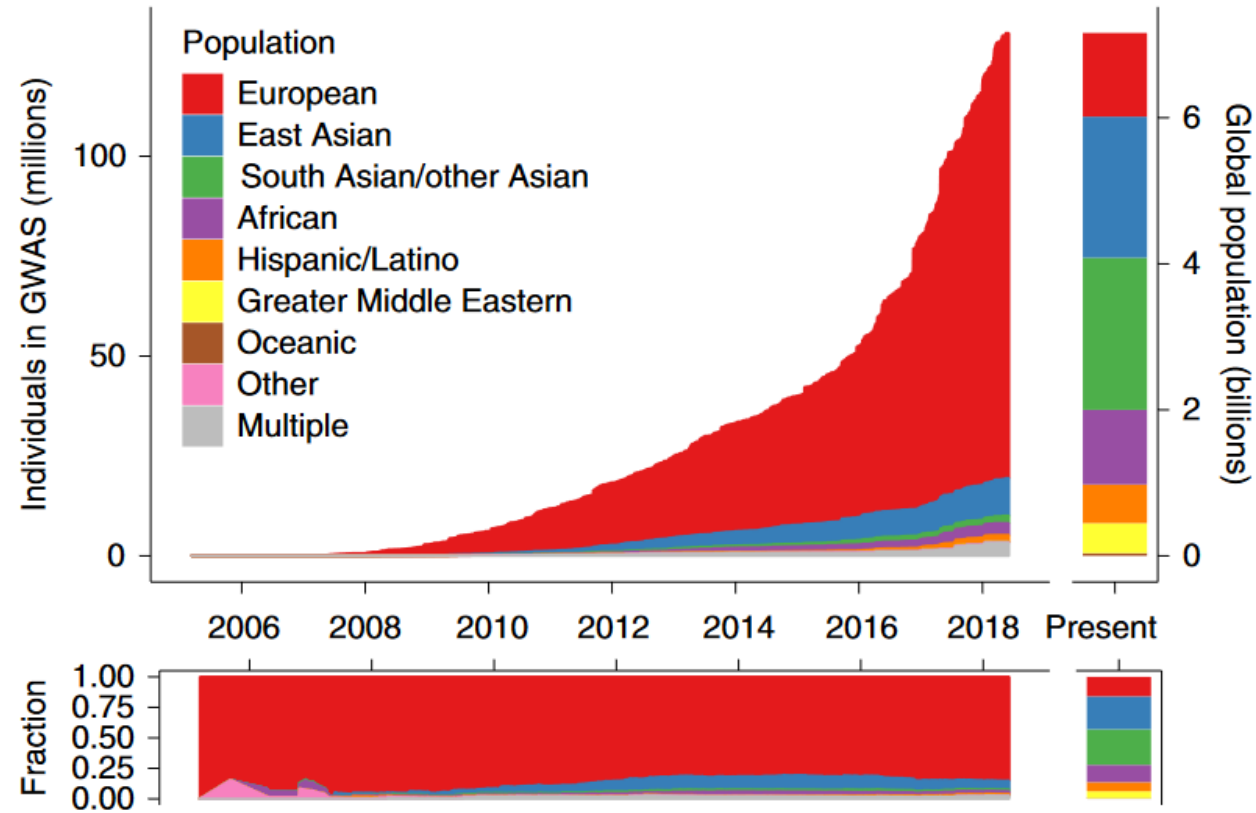


PGSs for individuals not of the same genetic ancestry group(s) as the GWAS sample from which summary statistics are retrieved are less predictive (Martin et al. 2019)

- Violin plots show distributions of relative prediction accuracies
- Points represent mean values
- Error bars represent standard errors

Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51(4):584–91. doi: .

PGSs and GWAS Genetic Ancestry



Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51(4):584–91.

PGSs and GWAS Genetic Ancestry

<https://gwasdiversitymonitor.com/>



Melinda Mills

Nuffield Professor of Sociology, Director Leverhulme Centre for Demographic Science



Department of Sociology

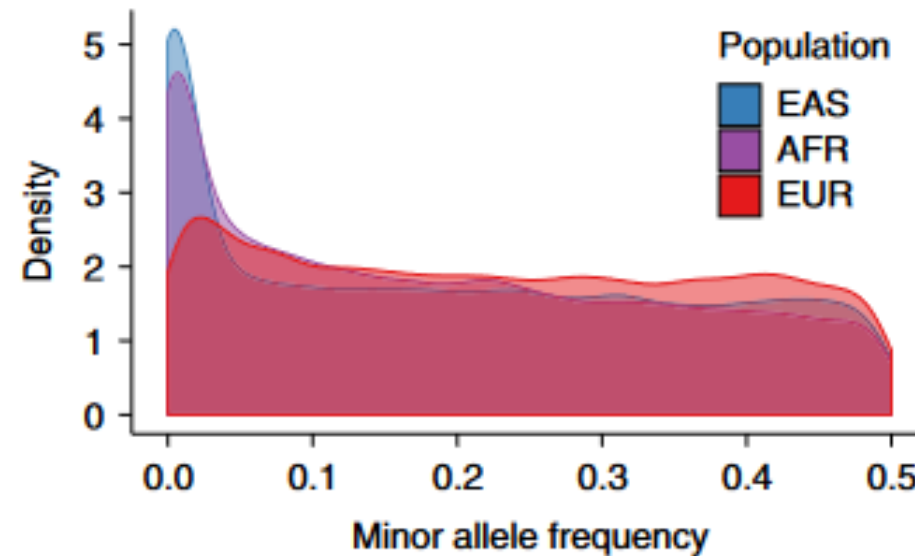
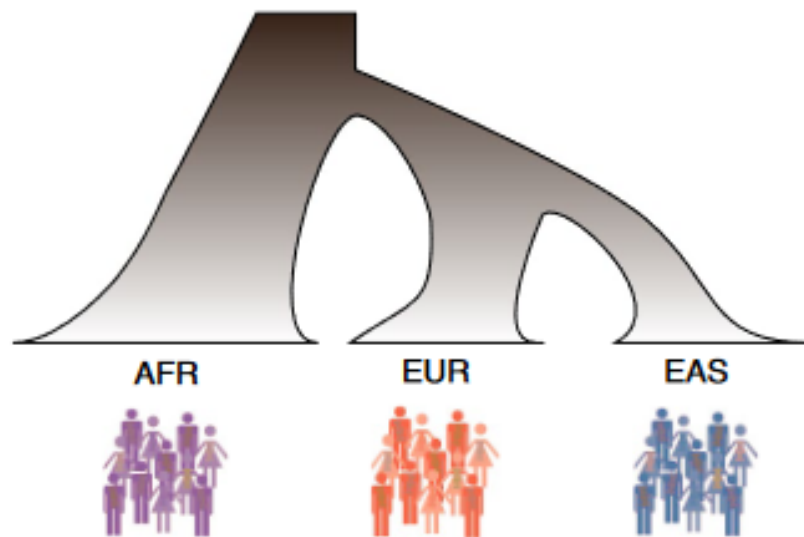
PGSs and GWAS Genetic Ancestry

Phenotype	GWAS Ancestry Group(s)
Coronary Artery Disease	European
Myocardial Infarction	European, South Asian, East Asian
Plasma Cortisol	European
Low-density Lipoprotein Cholesterol	European
High-density Lipoprotein Cholesterol	European
Total Cholesterol	European
Triglycerides	European
Type II Diabetes (2012)	European
Type II Diabetes (2018)	European, East Asian, South Asian, Mexican/Mexican-American
BMI	European
Waist Circumference	European
Waist-to-Hip Ratio	European
Height	European
Age at Menarche	European
Age at Menopause	European
Number of Children Ever Born	European
Age at First Birth	European
Ever/Current Smoker	European
Number of Cigarettes per day	European
Extraversion	European
Attention-deficit/hyperactivity Disorder (2010)	European
Attention-deficit/hyperactivity Disorder (2017)	European, Chinese
Bipolar Disorder	European
Major Depressive Disorder (2013)	European
Major Depressive Disorder (2018)	European
Schizophrenia	European, East Asian
Mental Health Cross Disorder	European
Alzheimer's Disease	European
Educational Attainment (2016)	European
Educational Attainment (2018)	European

PGSs for individuals not of the same genetic ancestry group(s) as the GWAS sample from which summary statistics are retrieved are less predictive (Martin et al. 2019)

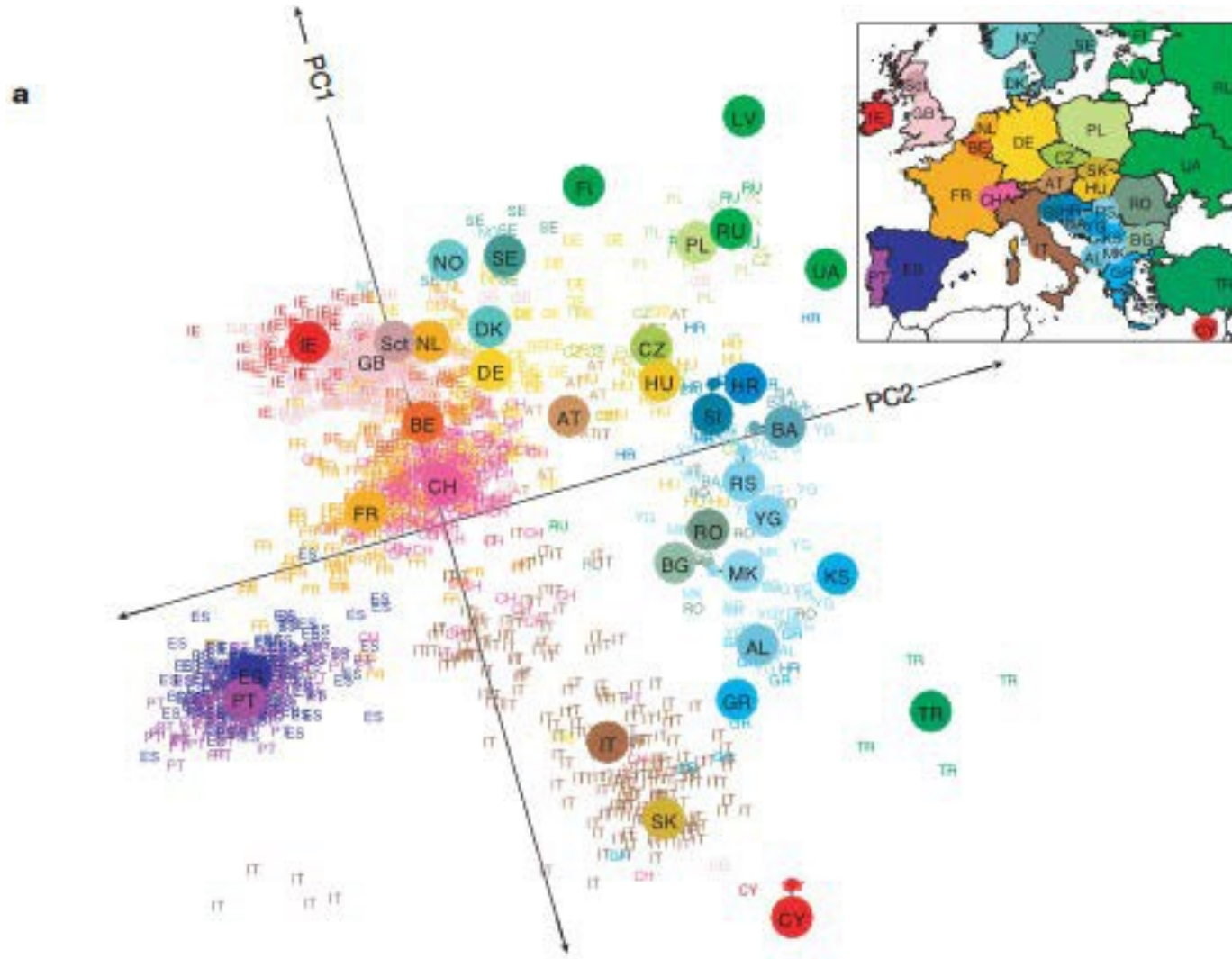
Population Structure and Genetics

Why / How this difference arises - The genetic explanation

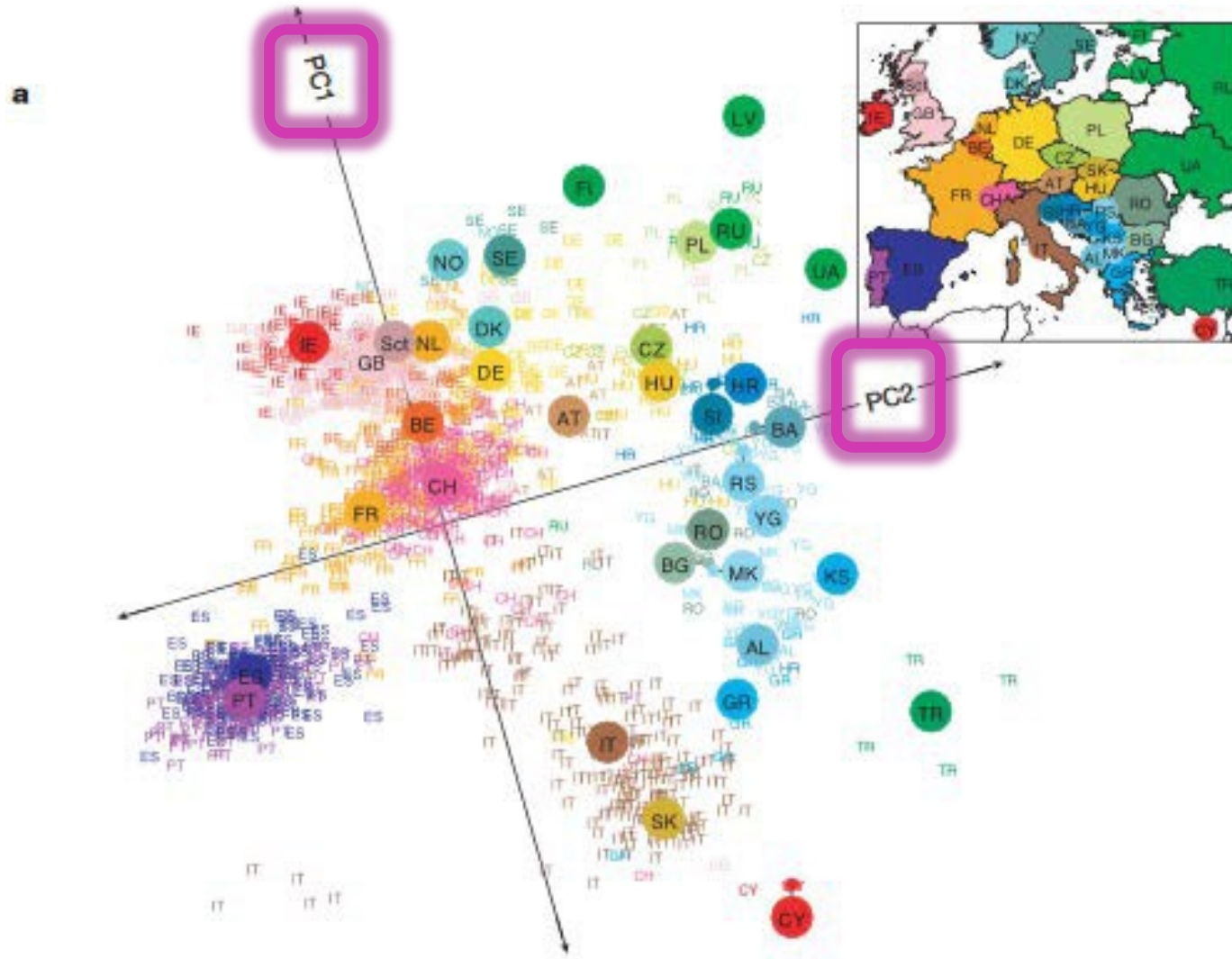


Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51(4):584–91. doi: .

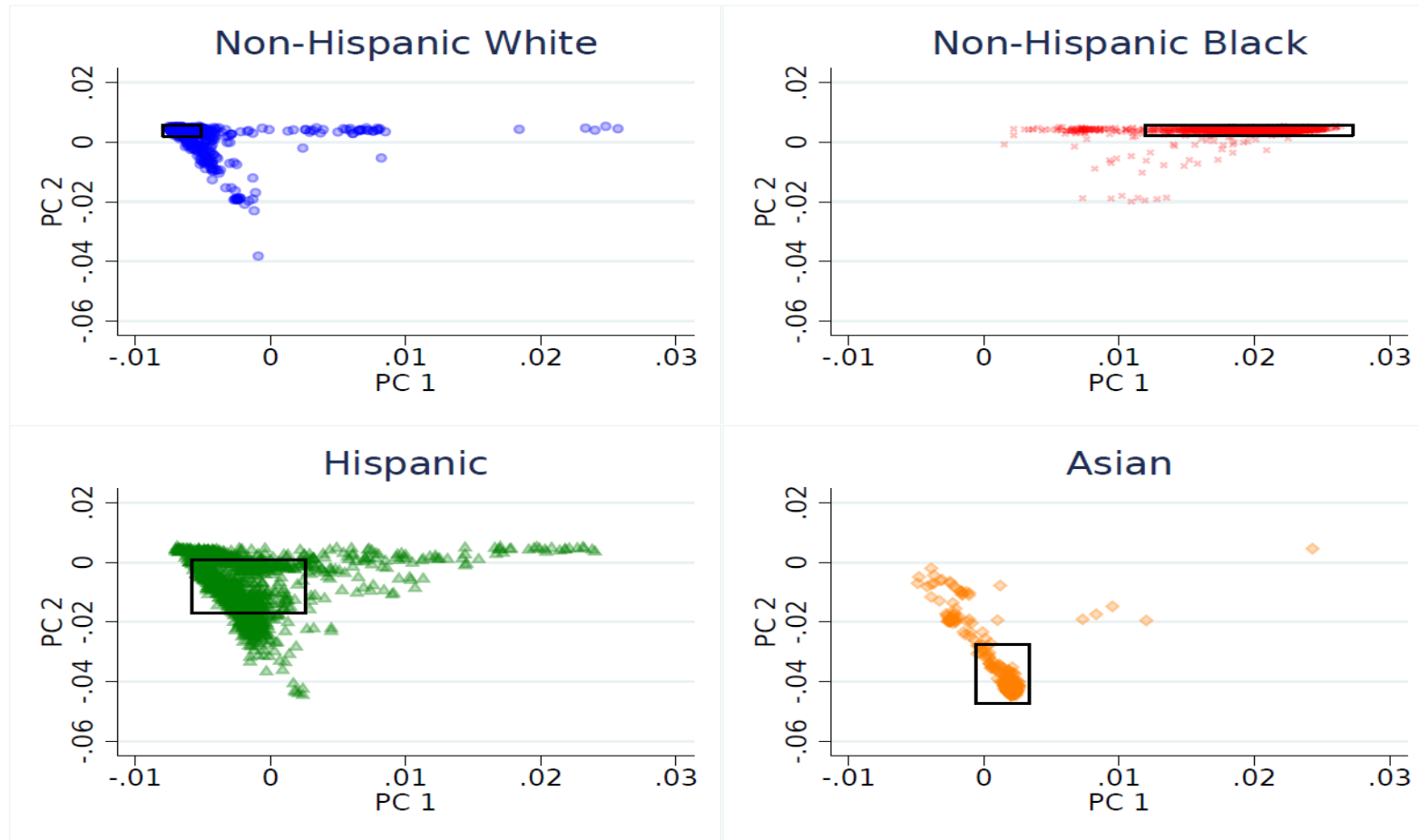
Population Structure and Genetics



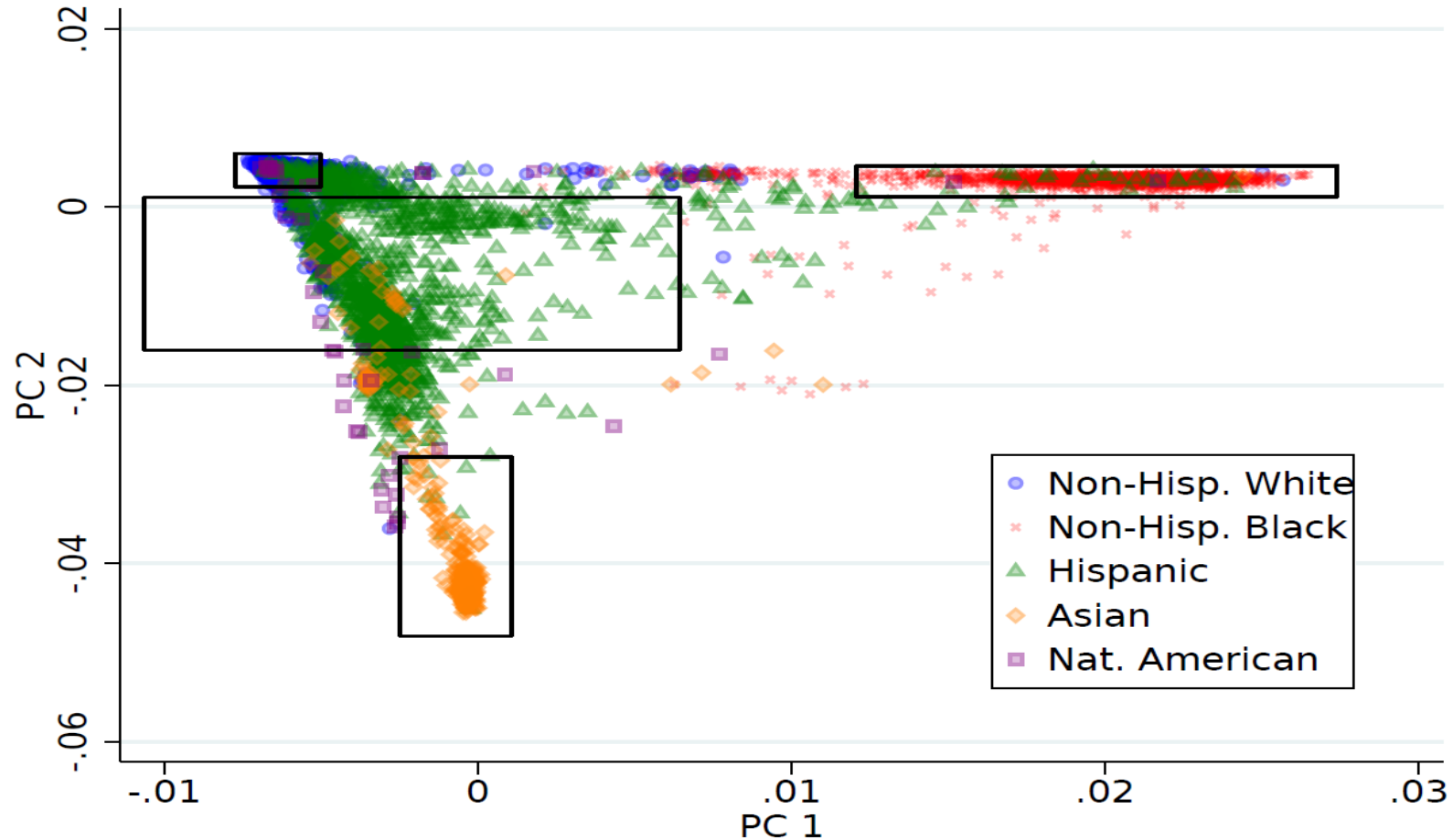
Population Structure and Genetics



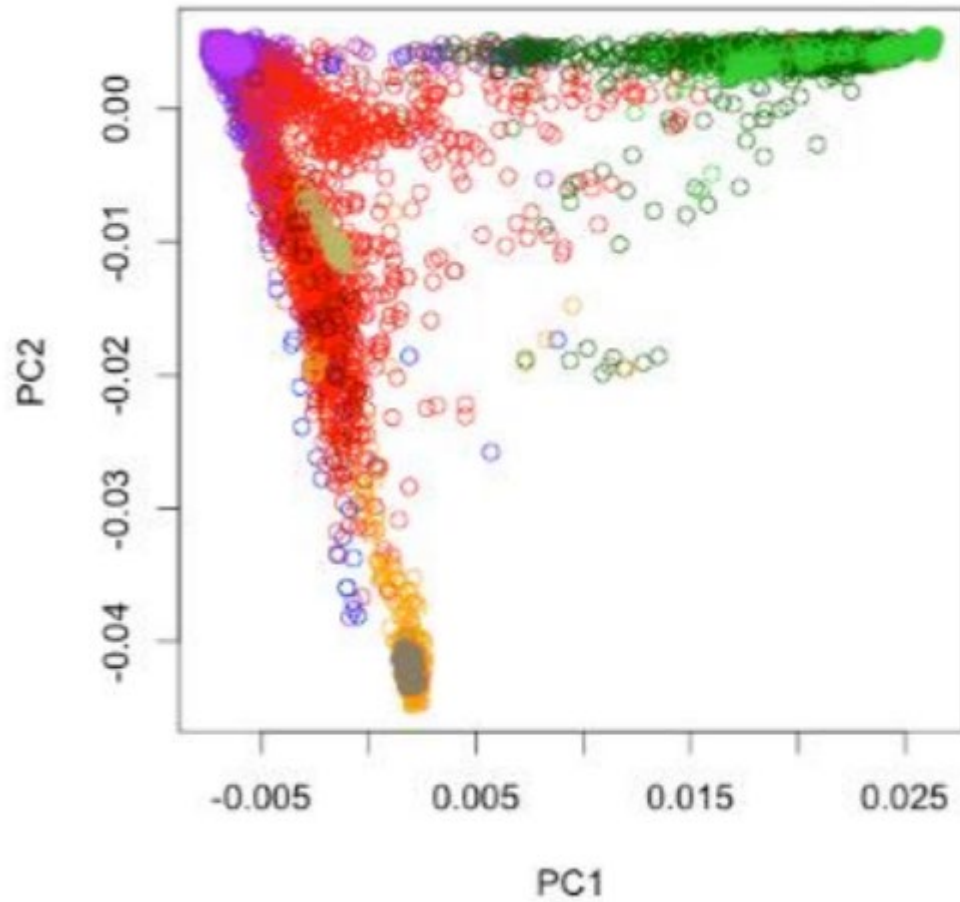
Add Health Genetic Ancestry Groups



Add Health Genetic Ancestry Groups



Add Health Genetic Ancestry Groups



- White
- African American
- Hispanic
- Native American
- Asian American
- CEU TSI
- ASW LWK MKK YRI
- MEX
- GHI
- CHB CHD JPT

Add Health Self-identified Race/Ethnicity

Western Europe & Italy
African ancestry (South West USA), Kenya, Nigeria
Mexican ancestry (Los Angeles, CA)
Gujarati Indians in Houston, TX
Han Chinese (Beijing), Chinese (Den, CO), Japanese (Tokyo)

Add Health Genetic Ancestry Groups

Self-Identified Race/Ethnicity	Genetic Ancestry				Total
	European	African	East Asian	Hispanic	
Non-Hispanic White	5,645	5	0	105	5,755
Non-Hispanic Black	0	1,938	0	1	1,939
Native American	14	2	0	6	22
Asian	0	1	422	26	449
Hispanic	72	27	15	849	963
Missing	0	2	0	0	0
Total Sample Size	5,731	1,975	437	987	9,130

Race/Ethnicity VS Genetic Ancestry

While the two constructs are highly correlated, race is socially defined while genetic-ancestry is biologically defined; and **conflating the two does harm to both scientific knowledge and efforts to reduce the continuing harms caused by systemic racism** (Phelan and Link 2013).

Those of us who pursue the expansion of knowledge in this area of science must be careful in our use of language generally as well as avoid applying terms used to, and/or created for the purpose of defining racial/ethnic categories while describing the current geo-political region(s) of the world that, on average, most closely correlate with a person's own genetic variation.

Add Health Polygenic Scores – Release 2

[Link to Documentation](https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/WaveIVPGSRelease2UserGuide.pdf)

https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/WaveIVPGSRelease2UserGuide.pdf

Using Add Health PGSs

- Accounting for ancestry
 - Separate analyses by ancestry
 - Ancestry-specific principal components
- Ancestry-specific principal components
 - Randomized in sets of five
 - 1-5, 6-10, 11-15, 16-20

Data Access/Contracts

[About](#) ▾[Documentation](#) ▾[Data](#) ▾[Publications](#)[News](#)[Events](#)[Participants](#)

Data

The Add Health study has collected data of interest to investigators from many disciplines in the social and behavioral sciences and from many theoretical traditions, making it a global data resource for over 50,000 researchers. The Add Health data is available in two forms, public-use data and restricted-use data, and offer endless options in the types of analysis, choices of data, and presentation.

Restricted-use Data Sets

The extensive restricted-use data, available by contractual agreement, contains data for all study participants and will be distributed only to certified researchers who commit themselves to maintaining limited access. To be eligible to enter into a contract, researchers must have an IRB-approval letter, security plan for handling and storing sensitive data, and sign a data-use contract agreeing to keep the data confidential.

To learn more about the restricted-use data and how to apply for restricted-use data, please access the [CPC Data Portal](#).

Data Access/Contracts



Add Health

Initiated in 1994 and supported by five program project grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) with co-funding from 23 other federal agencies and foundations, Add Health is the largest, most comprehensive, nationally-representative longitudinal survey of adolescents ever undertaken. Beginning with an in-school questionnaire administered to a nationally representative sample of students in grades 7-12, the study followed up with a series of in-home interviews conducted in 1995, 1996, 2001-02, 2008, and 2016-18. Add Health participants are now full-fledged adults, aged 33-44, and will soon be moving into midlife. Over the years, Add Health has added a substantial amount of additional data for users, including contextual data on the communities and states in which participants reside, genomic data and a range of biological health markers of participants, and parental survey data.

For further information about the study, please see the Add Health Website.

Public-Use Data consists of one-half of the core sample, and one-half of the oversample of African-American adolescents with a parent who has a college degree, chosen at random. Public-Use Data is available for Waves I-V.

Restricted-Use Data will be distributed only to certified researchers (this includes researchers that are located outside of the US) who commit themselves to maintaining limited access. To be eligible to enter into a contract, researchers must complete Contract Application which includes:

- Security plan
- IRB approval letter
- \$1000 payment by check (NEW contract only)

If you are applying for data in addition to the Core Files, you will need to submit an additional justification description for each data file requested.

For the Romantic Pairs Data, researchers must complete a separate Contract Application for Romantic Pairs which includes:

- Security plan for standalone computer
- IRB approval letter
- \$1000 payment by check (NEW contract only)
- Additional justification description for each data file requested

[Getting Started](#)

[Website](#)

[Public-Use Data](#)

[Codebooks](#)

Thank you!

david.braudt@hci.utah.edu