

1. **Table of Contents**
2. **Descriptive Statistics for Confounding Variables Considered for PCA with TOB as the Collider Environment**
  - Supplemental Table I Variables Retained in PCA, TOB
  - Supplemental Table II Variables removed from PCA because both correlation 95% CI contain 0
  - Supplemental Table III Variables removed from PCA because of errors in estimating correlation
3. **Parallel Analysis Output for PCA with TOB as the Collider Environment (S4, S.Fig 1)**
  - Supplemental Table IV Eigenvalues from Parallel Analysis, TOB Environment, ALC PRS
  - Supplemental Figure 1. Scree plot of Eigenvalues for retained components, TOB PCA.
4. **Descriptive Statistics for Confounding Variables Considered for PCA with EDU as the Collider Environment**
  - Supplemental Table V Variables retained in PCA, EDU
  - Supplemental Table VI Variables removed from PCA because both correlation 95% CI contain 0
  - Supplemental Table VII Variables removed from PCA because of errors in estimating correlation
5. **Parallel Analysis Output for PCA with EDU as the Collider Environment (S10, S.Fig 2)**
  - Supplemental Table VIII Eigenvalues from Parallel Analysis, EDU Environment, ALC PRS
  - Supplemental Figure 2. Scree plot of Eigenvalues for retained components, EDU PCA.
5. **Parallel Analysis Output for PCA with EDU as the Collider Environment (S10, S.Fig 2)**
  - Supplemental Table VIII Eigenvalues from Parallel Analysis, EDU Environment, ALC PRS
  - Supplemental Figure 2. Scree plot of Eigenvalues for retained components, EDU PCA.

**Article title:**

Principal Component Analysis Reduces Collider Bias in Polygenic Risk Score Effect Size Estimation

**Journal name:**

Behavior Genetics

**Author names:**

Nathaniel S. Thomas\*<sup>1</sup>, Peter Barr<sup>5</sup>, Fazil Aliev<sup>1</sup>, Mallory Stephenson<sup>3</sup>, Sally I-Chun Kuo<sup>4</sup>, Grace Chan<sup>6,7</sup>, Daniel Salvatore<sup>4</sup>

**Affiliation:**

1 Department of Psychology, Virginia Commonwealth University, Richmond, VA

2 Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia

3 Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, Virginia

4 Department of Psychiatry, Robert Wood Johnson Medical School, Rutgers University, Piscataway, New Jersey

5 Department of Psychiatry & Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, New York

6 Department of Psychiatry, University of Connecticut School of Medicine, Farmington, Connecticut

7 Department of Psychiatry, University of Iowa, Carver College of Medicine, Iowa City, Iowa

8 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana

9 Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana

**E-mail address of the corresponding author:**  
thomasns@vcu.edu

<b>Variable</b>	<b>Total n</b>	<b>Mean / n</b>
Parents enjoy eachother*	5190	1134
Parents fight in front of offspring*	5211	1570
Parents hit each other*	5207	805
Female*	7855	4129
Baby Boomer Generation*	7842	2612
Millenial Generation*	7842	2017
Ever lived with someone or at least a year as though you w	6591	4502
Ever lived as married and ever married*	6591	1210
Cannabis Dependence Dx*	7628	1257
Cannabis Abuse Dx*	7626	1471
Major Depression Disorder Dx*	5953	1348
Conduct Disorder Dx*	6903	1077
Antisocial Personality Disorder Dx*	7361	976
Cocaine Dependence Sx	6919	0.72
Cocaine Abuse Sx	7585	0.29
Opioid Dependence Sx	6917	0.32
Opioid Abuse Sx	7584	0.13
Sedative Dependence Sx	6889	0.25
Sedative Abuse Sx	7583	0.13
Stimulant Dependence Sx	6894	0.50
Stimulant Abuse Sx	7586	0.21
Parental Relationship Quality	5209	0.37
Conflict in Household	5450	0.50
Max Drinks in 24 Hrs	7442	18.50
Age of Alcohol Initiation	6684	18.27
Age of First Alcohol Intoxication	6902	17.03
Maximum Recorded Age	7855	36.39
Maternal Age of Alcohol Initiation	3863	22.24
Maternal Age of First Alcohol Intoxication	4057	20.27
Maternal Max Drinks in 24 Hrs	4870	9.83
Paternal Max Drinks in 24 Hrs	3796	22.51
Maternal AUD Sx	4968	1.97
Paternal AUD Sx	3862	4.22

<b>Variable</b>	<b>Total n</b>	<b>Mean / n</b>
Parental Divorce*	5888	794
Paternal Age of Alcohol Initiation	3603	19.12
Paternal Age of First Alcohol Intoxication	3615	17.52

<b>Variable</b>	<b>Total n</b>	<b>Suppleme Mean / n</b>
Silent Generation*	7842	1210
Generation X*	7842	1804

**Supplemental Table I Variables Retained**

<b>SD / Proportion</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Correlation with AUD</b>
0.22			0.16
0.30			0.18
0.15			0.22
0.53			-0.31
0.33			0.26
0.26			-0.20
0.68			-0.34
0.18			0.28
0.16			0.49
0.19			0.32
0.23			0.15
0.16			0.39
0.13			0.52
1.88	0	7	0.49
0.83	0	4	0.48
1.28	0	7	0.30
0.56	0	4	0.31
1.03	0	7	0.35
0.53	0	4	0.36
1.53	0	7	0.40
0.69	0	4	0.39
0.33	0	1	0.18
0.33	0	1	0.18
16.25	0	100	0.64
5.54	10	79	-0.25
4.64	10	68	-0.33
14.95	7	91	0.05
8.95	10	79	0.13
7.80	10	74	0.14
9.97	0	100	-0.01
16.75	1	100	0.06
2.98	0	11	0.03
3.69	0	11	0.13

**Supplemental Table II Variables removed from PCA because bo**

<b>SD / Proportion</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Correlation with AUD</b>
0.16			0.01
6.04	10	63	0.01
4.75	10	68	0.02

**Appendix Table III Variables removed from PCA because of errors in estimating correlation**

<b>SD / Proportion</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Correlation with AUD</b>
0.15			-0.13
0.23			-0.02

**1 in PCA, TOB**

<b>AUD lower 95%</b>	<b>AUD upper 95%</b>	<b>Correlation with TOB</b>
0.13	0.20	0.17
0.14	0.21	0.16
0.19	0.26	0.25
-0.34	-0.29	-0.15
0.24	0.29	0.22
-0.23	-0.17	-0.26
-0.36	-0.31	-0.38
0.25	0.31	0.29
0.47	0.51	0.50
0.29	0.35	0.36
0.12	0.18	0.15
0.36	0.42	0.41
0.49	0.54	0.48
0.47	0.50	0.37
0.46	0.49	0.35
0.28	0.33	0.34
0.29	0.33	0.32
0.33	0.37	0.29
0.34	0.38	0.28
0.37	0.42	0.33
0.37	0.41	0.33
0.15	0.20	0.18
0.15	0.21	0.16
0.63	0.66	0.46
-0.27	-0.23	-0.09
-0.35	-0.30	-0.18
0.02	0.07	0.26
0.10	0.16	0.13
0.11	0.17	0.15
-0.04	0.02	0.06
0.03	0.09	0.17
0.00	0.06	0.08
0.10	0.16	0.21

**th correlation 95% CI contain 0, TOB**

<b>AUD lower 95%</b>	<b>AUD upper 95%</b>	<b>Correlation with TOB</b>
-0.04	0.05	0.05
-0.02	0.04	0.00
-0.02	0.05	0.01

**Correlation matrix of all confounders for PCA after Knn imputation, TOB**

<b>AUD lower 95%</b>	<b>AUD upper 95%</b>	<b>Correlation with TOB</b>
-0.16	-0.09	0.22
-0.05	0.01	-0.08



<b>TOB lower 95%</b>	<b>TOB upper 95%</b>
0.13	0.22
0.12	0.21
0.20	0.30
-0.18	-0.11
0.19	0.26
-0.30	-0.23
-0.42	-0.35
0.25	0.34
0.47	0.54
0.32	0.40
0.11	0.19
0.37	0.45
0.44	0.52
0.34	0.40
0.32	0.38
0.30	0.38
0.28	0.36
0.25	0.33
0.25	0.32
0.30	0.37
0.30	0.36
0.15	0.21
0.13	0.20
0.44	0.49
-0.12	-0.06
-0.21	-0.15
0.23	0.28
0.09	0.17
0.11	0.19
0.03	0.10
0.13	0.21
0.05	0.12
0.17	0.24

<b>TOB lower 95%</b>	<b>TOB upper 95%</b>
-0.01	0.10
-0.04	0.04
-0.03	0.05

<b>TOB lower 95%</b>	<b>TOB upper 95%</b>
0.18	0.26
-0.12	-0.04

**Supplemental Table IV Eigenvalues from Para**

Component	Adjusted Eigenvalue
1	7.04
2	4.38
3	3.26
4	2.25
5	1.82
6	1.38
7	1.19
8	1.14

**Allele Analysis, TOB Environment, ALC PRS**

Unadjusted Eigenvalue

---

7.16

4.48

3.36

2.33

1.90

1.45

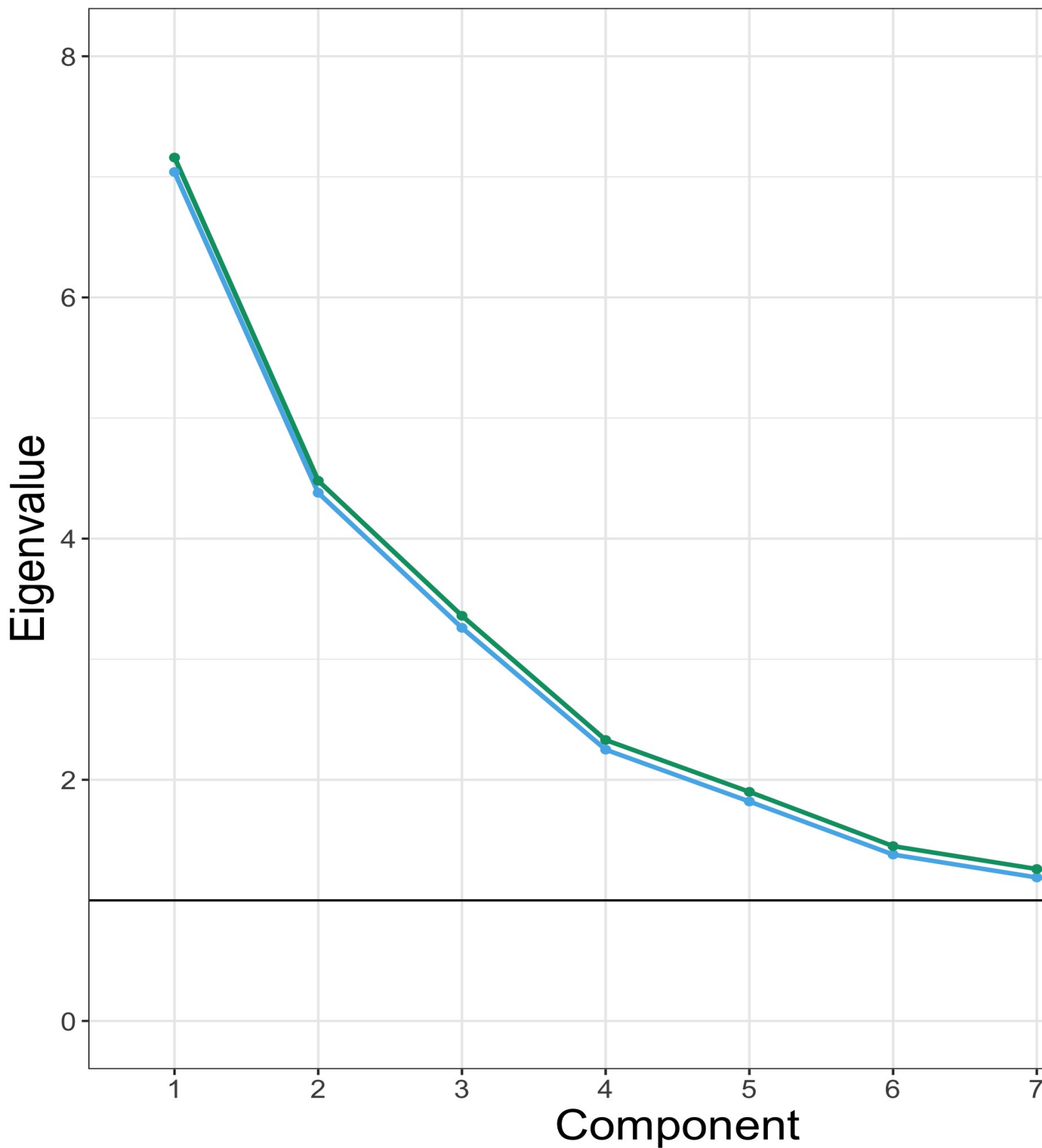
1.26


1.20

---

Supplemental Figure 1. Scree plot of Eigenvalues for retained components, TOB PCA.

### TOB PCA Scree Plot, Retained Components



Type  Adjusted.Eigenvalue  Unadjusted.Eigenvalue

<b>Variable</b>	<b>Total n</b>	<b>Mean / n*</b>
Parents enjoy eachother*	5190	1134
Parents fight in front of offspring*	5211	1570
Parents hit each other*	5207	805
Female*	7855	4129
Parental Divorce*	5888	794
Baby Boomer Generation*	7842	2612
Millenial Generation*	7842	2017
Ever lived with someone or at least a year as though you we	6591	4502
Ever lived as married and ever married*	6591	1210
Cannabis Dependence Dx*	7628	1257
Cannabis Abuse Dx*	7626	1471
Major Depression Disorder Dx*	5953	1348
Conduct Disorder Dx*	6903	1077
Antisocial Personality Disorder Dx*	7361	976
Cocaine Dependence Sx	6919	0.72
Cocaine Abuse Sx	7585	0.29
Opioid Dependence Sx	6917	0.32
Opioid Abuse Sx	7584	0.13
Sedative Dependence Sx	6889	0.25
Sedative Abuse Sx	7583	0.13
Stimulant Dependence Sx	6894	0.50
Stimulant Abuse Sx	7586	0.21
Parental Relationship Quality	5209	0.37
Conflict in Household	5450	0.50
Max Drinks in 24 Hrs	7442	18.50
Age of Alcohol Initiation	6684	18.27
Age of First Alcohol Intoxication	6902	17.03
Maximum Recorded Age	7855	36.39
Maternal Age of Alcohol Initiation	3863	22.24
Paternal Age of Alcohol Initiation	3603	19.12
Maternal Age of First Alcohol Intoxication	4057	20.27
Paternal Age of First Alcohol Intoxication	3615	17.52
Maternal Max Drinks in 24 Hrs	4870	9.83
Paternal Max Drinks in 24 Hrs	3796	22.51
Maternal AUD Sx	4968	1.97
Paternal AUD Sx	3862	4.22

**Sup**

<b>Variable</b>	<b>Total n</b>	<b>Mean / n</b>
Generation X*	7842	1804

**Supplemental Table VII Variabl**

<b>Variable</b>	<b>Total n</b>	<b>Mean / n</b>
Silent Generation*	7842	1210



**Supplemental Table V Variables retained in PCA, EDU**

<b>SD / Proportion*</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Correlation with AUD</b>	<b>AUD lower 95%</b>
0.22			0.16	0.13
0.30			0.18	0.14
0.15			0.22	0.19
0.53			-0.31	-0.34
0.13			0.01	-0.04
0.33			0.26	0.24
0.26			-0.20	-0.23
0.68			-0.34	-0.36
0.18			0.28	0.25
0.16			0.49	0.47
0.19			0.32	0.29
0.23			0.15	0.12
0.16			0.39	0.36
0.13			0.52	0.49
1.88	0	7	0.49	0.47
0.83	0	4	0.48	0.46
1.28	0	7	0.30	0.28
0.56	0	4	0.31	0.29
1.03	0	7	0.35	0.33
0.53	0	4	0.36	0.34
1.53	0	7	0.40	0.37
0.69	0	4	0.39	0.37
0.33	0	1	0.18	0.15
0.33	0	1	0.18	0.15
16.25	0	100	0.64	0.63
5.54	10	79	-0.25	-0.27
4.64	10	68	-0.33	-0.35
14.95	7	91	0.05	0.02
8.95	10	79	0.13	0.10
6.04	10	63	0.01	-0.02
7.80	10	74	0.14	0.11
4.75	10	68	0.02	-0.02
9.97	0	100	-0.01	-0.04
16.75	1	100	0.06	0.03
2.98	0	11	0.03	0.00
3.69	0	11	0.13	0.10

**Supplemental Table VI Variables removed from PCA because both correlation 95% CI contain 0**

<b>SD / Proportion</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Correlation with AUD</b>	<b>AUD lower 95%</b>
0.23			-0.02	-0.05

**les removed from PCA because of errors in estimating correlation matrix of all confounder**

<b>SD / Proportion</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Correlation with AUD</b>	<b>AUD lower 95%</b>
0.15			-0.13	-0.16

<b>AUD upper 95%</b>	<b>Correlation with EDU</b>	<b>EDU lower 95%</b>	<b>EDU upper 95%</b>
0.20	-0.16	-0.20	-0.13
0.21	-0.18	-0.21	-0.15
0.26	-0.25	-0.28	-0.21
-0.29	0.06	0.03	0.09
0.05	-0.23	-0.26	-0.19
0.29	0.09	0.06	0.11
-0.17	-0.02	-0.05	0.01
-0.31	0.16	0.13	0.19
0.31	-0.10	-0.13	-0.06
0.51	-0.17	-0.20	-0.13
0.35	-0.05	-0.09	-0.02
0.18	-0.02	-0.05	0.02
0.42	-0.21	-0.25	-0.18
0.54	-0.26	-0.29	-0.23
0.50	-0.12	-0.14	-0.09
0.49	-0.12	-0.14	-0.10
0.33	-0.09	-0.11	-0.06
0.33	-0.08	-0.11	-0.06
0.37	-0.07	-0.10	-0.05
0.38	-0.08	-0.10	-0.06
0.42	-0.12	-0.14	-0.09
0.41	-0.11	-0.14	-0.09
0.20	-0.14	-0.17	-0.11
0.21	-0.11	-0.14	-0.09
0.66	-0.15	-0.17	-0.13
-0.23	0.06	0.04	0.09
-0.30	0.09	0.06	0.11
0.07	0.11	0.09	0.13
0.16	0.06	0.03	0.09
0.04	0.09	0.06	0.12
0.17	0.05	0.02	0.08
0.05	0.11	0.08	0.14
0.02	-0.18	-0.20	-0.15
0.09	-0.21	-0.24	-0.18
0.06	-0.15	-0.18	-0.13
0.16	-0.19	-0.22	-0.16

**in 0, EDU**

<b>AUD upper 95%</b>	<b>Correlation with EDU</b>	<b>EDU lower 95%</b>	<b>EDU upper 95%</b>
0.01	0.02	-0.01	0.05

**'s for PCA after Knn imputation, EDU**

<b>AUD upper 95%</b>	<b>Correlation with EDU</b>	<b>EDU lower 95%</b>	<b>EDU upper 95%</b>
-0.09	-0.04	-0.08	-0.01

**Supplemental Table VIII Eigenvalues from P:**

Component	Adjusted Eigenvalue
1	7.41
2	4.97
3	3.36
4	2.24
5	1.83
6	1.70
7	1.18
8	1.14
9	1.02

**Parallel Analysis, EDU Environment, ALC PRS**

Unadjusted Eigenvalue

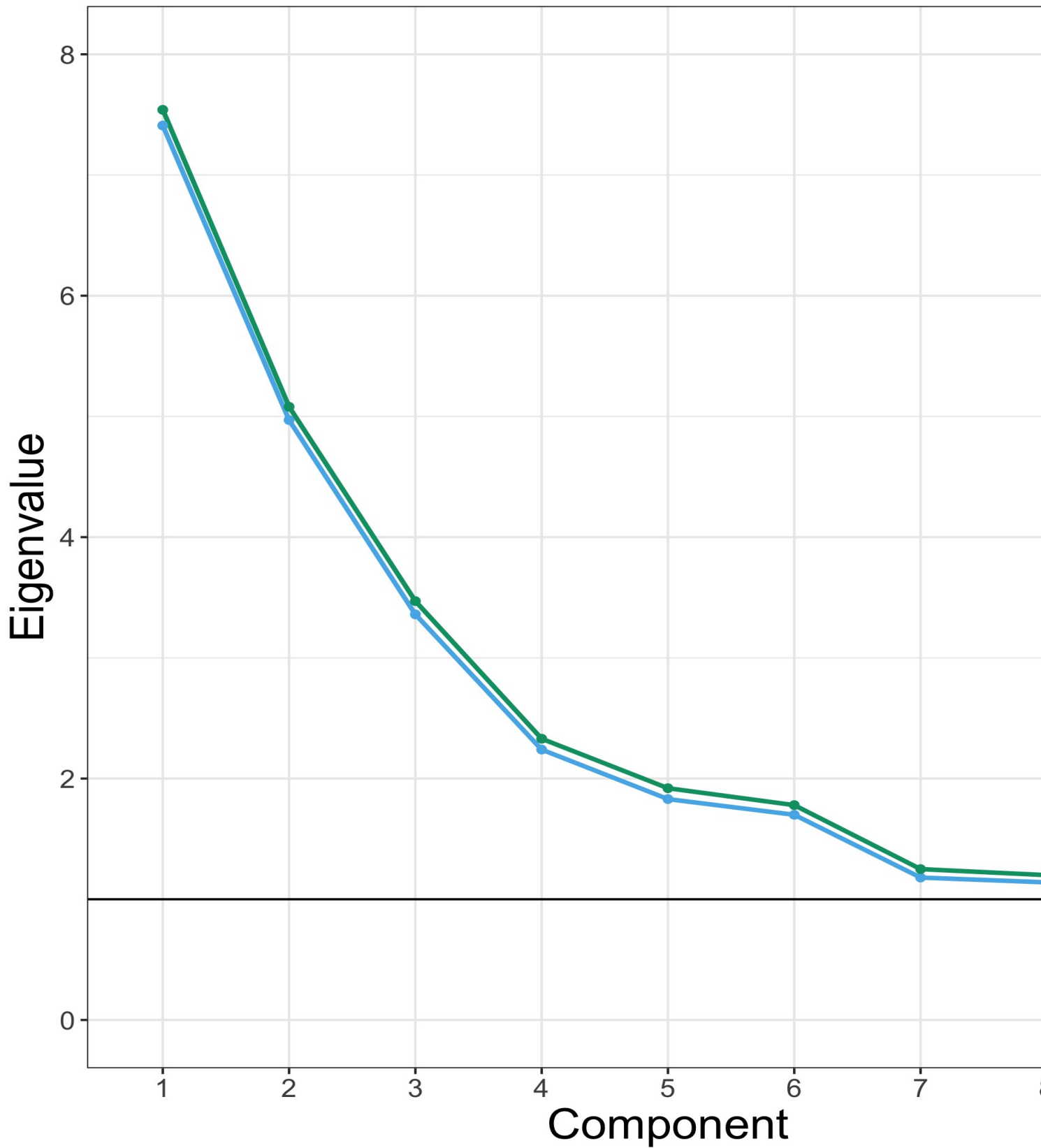
---

7.54
5.08
3.47
2.33
1.92
1.78
1.25
1.20
1.08

---

Supplemental Figure 2. Scree plot of Eigenvalues for retained components, EDU PCA.

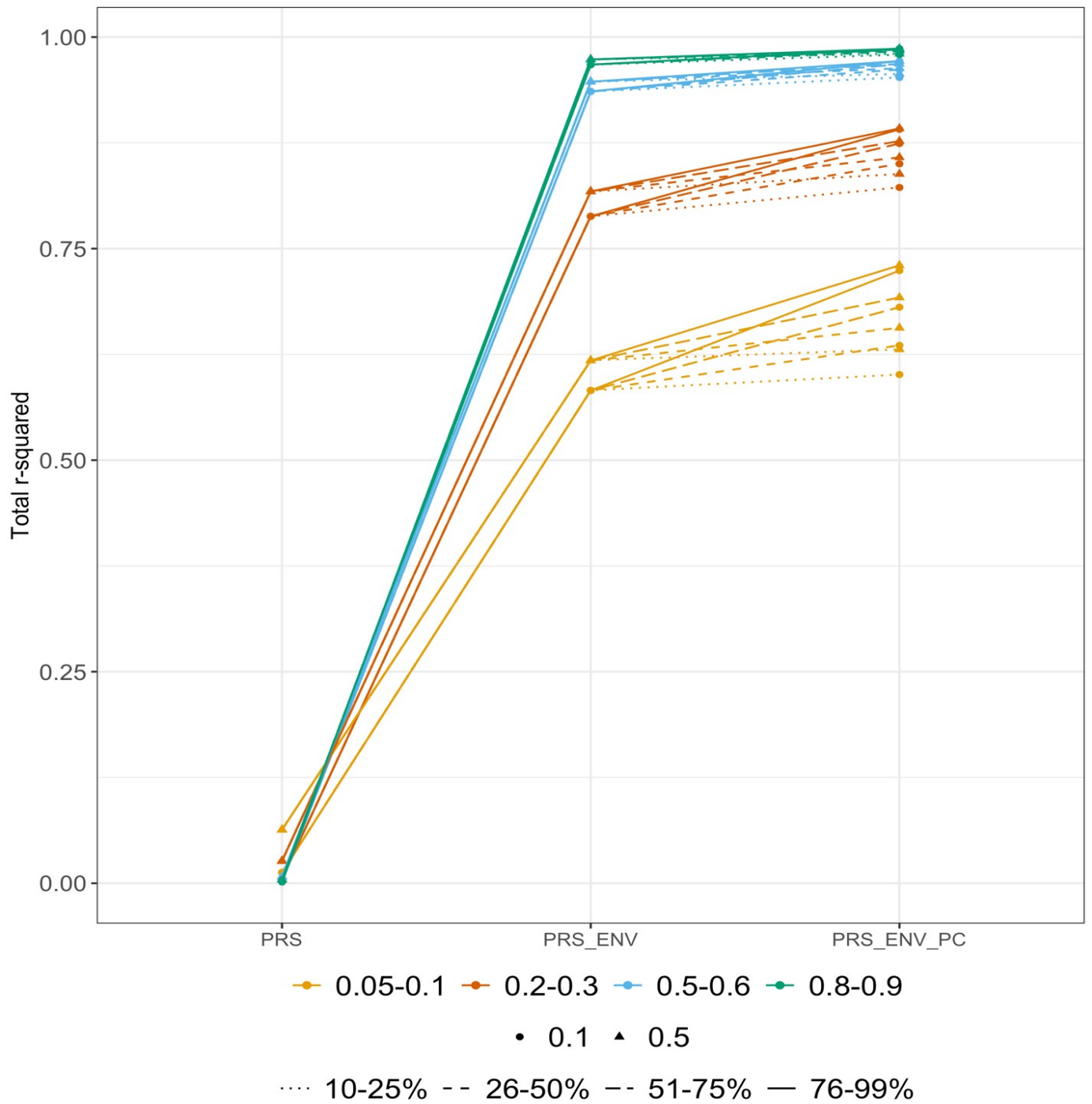
### EDU PCA Scree Plot, Retained Components



Type  Adjusted.Eigenvalue  Unadjusted.Eigenvalue



**Supplemental Figure 3. Simulation total model R-square by magnitude of confounder correlation (color), magnitude of rGE (dot shape), and proportion of data included in the PC correction (linetype)**

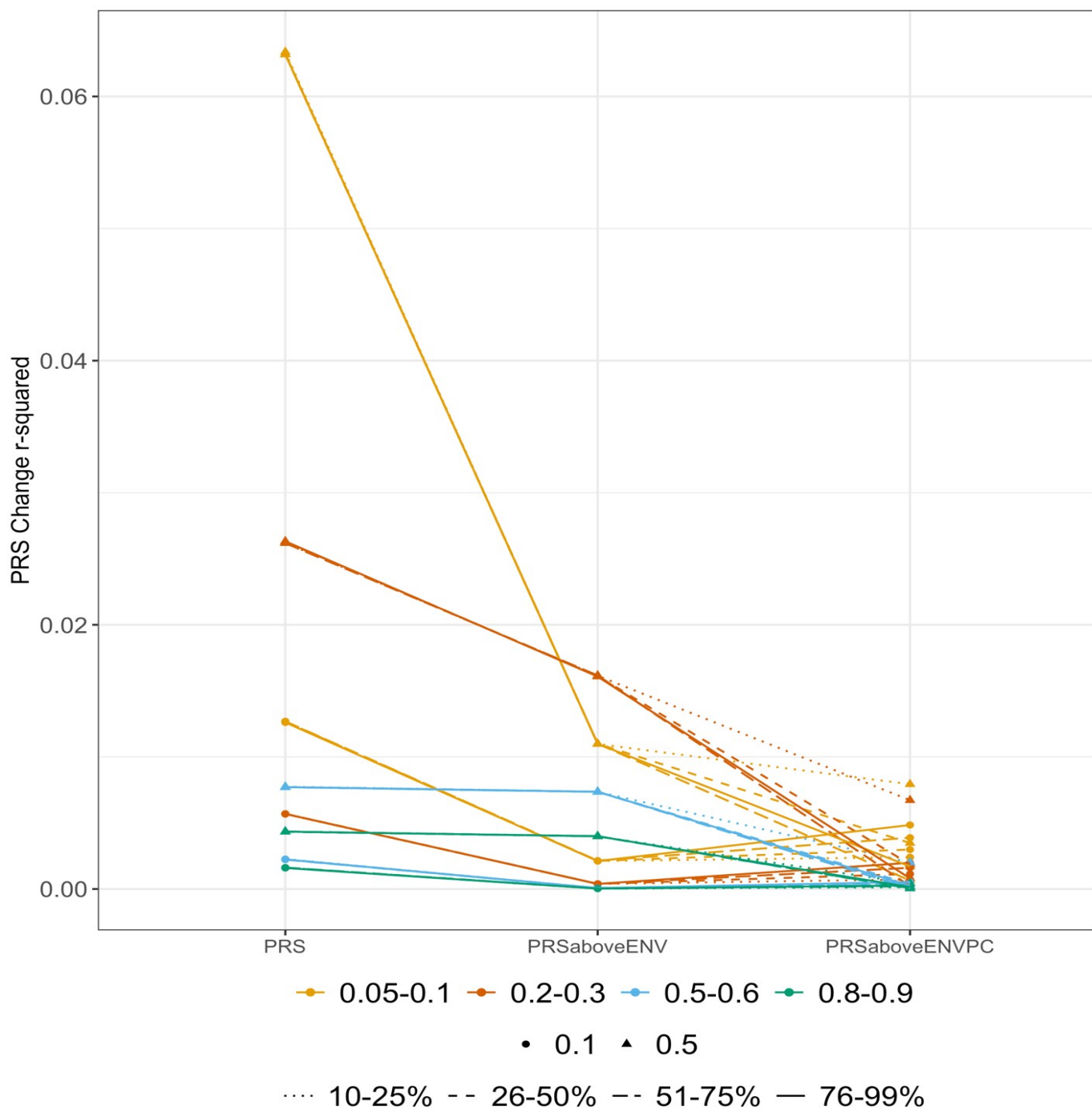


PRS : Just PRS

PRS\_ENV : PRS + Environment

PRS\_ENV\_PC : PRS + Environment + PC

**Supplemental Figure 4. Simulation PRS change R-square by magnitude of confounder correlation (color), magnitude of rGE (dot shape), and proportion of data included in the PC correction (linetype)**



PRS : Just PRS

PRSaboveENV : Change R2 when PRS is added to model with only Environment

PRSaboveENVPC : Change R2 when PRS is added to model with Environment and PC