An example of **GxE with real data** and a recap of **main biases**

Polygenic index workshop

Canazei Winter School on Inequality and Social Welfare Theory

Gaia Ghirardi * & Rosa Cheesman **

*Department Statistical Sciences “Paolo Fortunati”, University of Bologna, Italy
**Department of Psychology, University of Oslo

7th January 2025

1) Contextualising the practical with real data:

The Development of Body Mass Index from Adolescence to Adulthood: A Genotype-Family Socioeconomic Status Interaction Study

Gaia Ghirardi

Research Questions

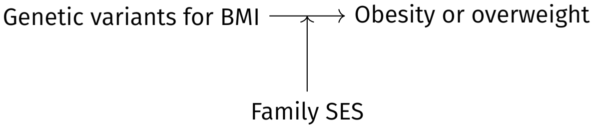
1) Does the genetic propensity for high BMI matter more in explaining overweight or obesity among individuals with high or low SES family of origin?

2) If so, does this differ according to the age at which we observe BMI?

Research Questions

1) Does the genetic propensity for high BMI matter more in explaining overweight or obesity among individuals with high or low SES family of origin?

2) If so, does this differ according to the age at which we observe BMI?



Data and Variables

Data

The National Longitudinal Study of Adolescent to Adult Heath study (Add Health): Nationally representative panel study of **U.S. adolescents** enrolled in grades 7-12 and born between **1974 and 1983** ◂ Data

Predictors

1 **BMI PGI**: a summary indicator of individuals' genetic propensity to have a high BMI
2 **Family SES**: Principal component loadings for four social origin measures
   1. Parental education
   2. Parental occupation
   3. Household income
   4. Household receipt of public assistance

Outcomes

↪ Mass (kg) divided by squared height (m) (kg/m2)
1 **BMI adolescence** (∼ 16 years old)
2 **BMI early adulthood** (∼ 22 years old)
3 **BMI adulthood** (∼ 28 years old)
4 **BMI later adulthood** (∼ 37 years old)

Controls

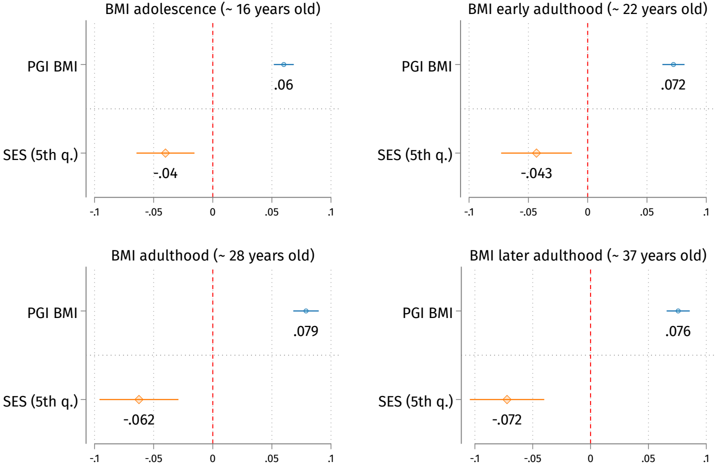
Gender, years of birth, first 10 principal components

◂ Methods

Results: Prediction family SES and PGI BMI

OLS regression models:

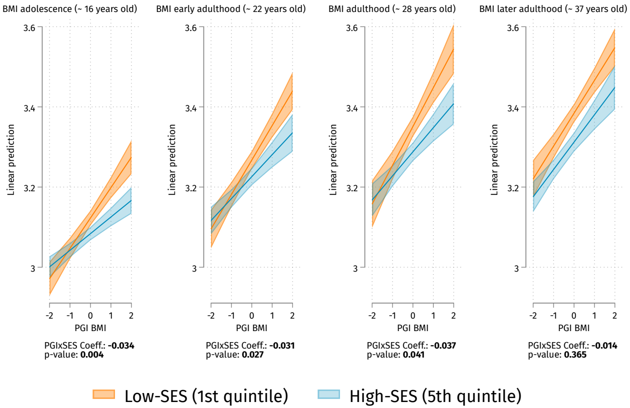
$$log(BMI)_i = \alpha + \beta_1 PGI\ BMI + \beta_2 Family\ SES_i + Z_i + \epsilon_i$$



Results: Interaction family SES and PGI BMI

OLS regression models:

$$log(BMI)_i = \alpha + \beta_1 PGI\ BMI_i + \beta_2 Family\ SES_i + \beta_3 Family\ SES_i \times PGI\ BMI_i$$
$$+ \beta_\gamma Family\ SES_i \times Z_i + \beta_\delta PGI\ BMI_i \times Z_i + \beta_\lambda Z_i + \epsilon_i$$

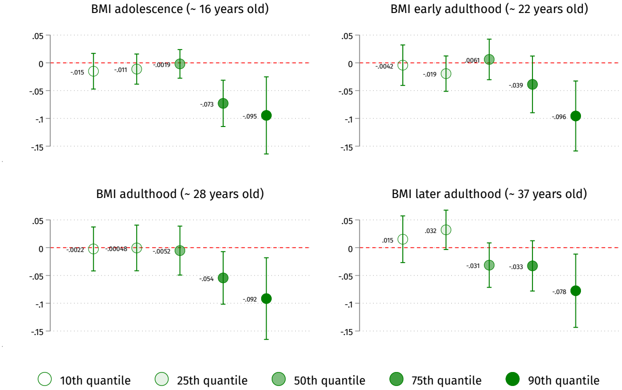


→ The negative interaction between family SES and PGI BMI is present at early ages until 28 years old (PGI BMI more predictive among low-SES individuals), and the interaction disappears in later adulthood (at age 37)

Results: Interaction family SES and PGI BMI

Unconditional quantile regression models

$$RIF(Y_i; q_t, F_Y) = \alpha + \beta_1 PGI\ BMI_i + \beta_2 Family\ SES_i + \beta_3 Family\ SES_i \times PGI\ BMI_i$$
$$+ \beta_\gamma Family\ SES_i \times Z_i + \beta_\delta PGI\ BMI_i \times Z_i + \beta_\lambda Z_i + \epsilon_i$$



→ PGI x SES interaction plays a crucial role in determining higher levels of BMI

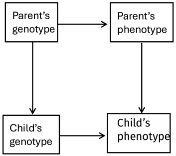
- influencing the prevalence of overweight and obesity among low-SES individuals
- while it remains inconsequential at lower and medium BMI levels

2) Recapping the biases in GxE studies

Using PGI in GxE studies: some limitations & partial solutions

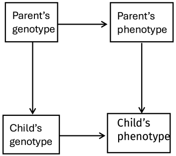
Using PGI in GxE studies: some limitations & partial solutions

1. **Gene-environment correlation (rGE)**
   - Definition: Genetic characteristics of the parents can indirectly affect the children's outcomes through the environment (Passive rGE)
   - Partial solution: Controlling for **parents' genetic characteristics** (using trios, imputing parents' information, performing family FEs or use PGIs from family GWAS)

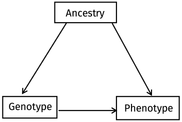


Using PGI in GxE studies: some limitations & partial solutions

1. **Gene-environment correlation (rGE)**
   - Definition: Genetic characteristics of the parents can indirectly affect the children's outcomes through the environment (Passive rGE)
   - Partial solution: Controlling for **parents' genetic characteristics** (using trios, imputing parents' information, performing family FEs or use PGIs from family GWAS)



2. **Population stratification**
   - Definition: Individuals with shared genetic ancestry exhibit similar genotypes. If PGIs are associated with ancestry—which itself aligns with ethnic and regional backgrounds—PGIs may capture cultural or environmental influences linked to these demographic factors
   - Partial solution: controlling for **5-20 PCs** i.e., variables representing "ancestry" on the basis of PC analysis (PCA). However, not perfect solution (PCs are inadequate for correcting for population substructure)



Using PGI in GxE studies: some limitations & partial solutions

3 **Low portability of the PGI between ancestry groups**

- Definition: Up to now, PGIs are more predictive for European ancestry individuals because in GWAS participate mainly European ancestry individuals
- Partial solution: **Recognize this selection bias** during the interpretation of the results + including only **European ancestry individuals**

Using PGI in GxE studies: some limitations & partial solutions

3 **Low portability of the PGI between ancestry groups**

- Definition: Up to now, PGIs are more predictive for European ancestry individuals because in GWAS participate mainly European ancestry individuals
- Partial solution: **Recognize this selection bias** during the interpretation of the results + including only **European ancestry individuals**

4 **Low portability of the PGI within same ancestry group**

- Definition: Selection bias in GWAS participation $\Rightarrow$ PGIs are more predictive of healthy, female, older and high-SES individuals
- Partial solution: **Recognize this selection bias** during the interpretation of the results + construct and include **sample weights**

Using PGI in GxE studies: some limitations & partial solutions

3 **Low portability of the PGI between ancestry groups**
  - Definition: Up to now, PGIs are more predictive for European ancestry individuals because in GWAS participate mainly European ancestry individuals
  - Partial solution: **Recognize this selection bias** during the interpretation of the results + including only **European ancestry individuals**

4 **Low portability of the PGI within same ancestry group**
  - Definition: Selection bias in GWAS participation $\Rightarrow$ PGIs are more predictive of healthy, female, older and high-SES individuals
  - Partial solution: **Recognize this selection bias** during the interpretation of the results + construct and include **sample weights**

5 **Not accurate and biased GxE coefficients**
  - Definition: Potential confounders are not properly controlled for in the statistical models used to test GxE effects since confounders are incorrectly entered as covariates in linear model
  - Partial solution: Including **covariate-by-gene & covariate-by-environment interactions** in the model in which the GxE term is estimated (Keller, 2014)

Using PGI in GxE studies: some limitations & partial solutions

3. **Low portability of the PGI between ancestry groups**
   - Definition: Up to now, PGIs are more predictive for European ancestry individuals because in GWAS participate mainly European ancestry individuals
   - Partial solution: **Recognize this selection bias** during the interpretation of the results + including only **European ancestry individuals**
4. **Low portability of the PGI within same ancestry group**
   - Definition: Selection bias in GWAS participation $\Rightarrow$ PGIs are more predictive of healthy, female, older and high-SES individuals
   - Partial solution: **Recognize this selection bias** during the interpretation of the results + construct and include **sample weights**
5. **Not accurate and biased GxE coefficients**
   - Definition: Potential confounders are not properly controlled for in the statistical models used to test GxE effects since confounders are incorrectly entered as covariates in linear model
   - Partial solution: Including **covariate-by-gene & covariate-by-environment interactions** in the model in which the GxE term is estimated (Keller, 2014)
6. **Endogeneity in the GxE because E is endogenous**
   - Definition: If there is something that affects both the environmental measure and the outcome, the GxE effect might be biased (e.g., Akimova et al., 2021)
   - Partial solution: Find credible **exogenous sources of variation in the environmental measure** (Schmitz & Conley, 2017; Lahtinen et al., 2024) or **conduct sensitivity analyses** (Akimova et al., 2021) ◂ Others: e.g., Assortative Mating

Thank you!

✉
gaia.ghirardi@eui.eu

Data

Add Health

- The National Longitudinal Study of Adolescent to Adult Heath study
- Nationally representative panel study of **U.S. adolescents** enrolled in grades 7-12 and born between **1974 and 1983**
- Iia use info from Waves I, II and IV. Wave I of the study took place during the 1994–1995 school year when the subjects were age 12–19. Wave II surveyed the same adolescents a year later, in 1996, Wave IV followed in 2008–2009, when the subjects were aged 25–34 years
- Sample: **3,977**

Back

Methods

Empirical model OLS and unconditional quantile regression

y =BMI or obesity

i =Child

Z =gender + birth year + 20 genetic principal components

$$log(BMI)_i = \alpha + \beta_1 PGI\ BMI + \beta_2 Family\ SES_i + Z_i + \epsilon_i$$

$$\begin{aligned} log(BMI)_i =& \alpha + \beta_1 PGI\ BMI_i + \beta_2 Family\ SES_i + \beta_3 Family\ SES_i \times PGI\ BMI_i \\ &+ \beta_\gamma Family\ SES_i \times Z_i + \beta_\delta PGI\ BMI_i \times Z_i + \beta_\lambda Z_i + \epsilon_i \end{aligned}$$

$$\begin{aligned} RIF(Y_i; q_t, F_Y) =& \alpha + \beta_1 PGI\ BMI_i + \beta_2 Family\ SES_i + \beta_3 Family\ SES_i \times PGI\ BMI_i \\ &+ \beta_\gamma Family\ SES_i \times Z_i + \beta_\delta PGI\ BMI_i \times Z_i + \beta_\lambda Z_i + \epsilon_i \end{aligned}$$

Additional information

- I include
  1. *survey weights* to correct for **nonresponse and attrition**
  2. *IPW weights* to correct for **participation in the genetic sample**
- I exclude **non-European ancestry** individuals
- I add **covariate-gene interactions** and **covariate-environment interactions** to provide more statistically consistent estimates of GxE (Keller, 2014, Domingue et al., 2020)

Back

Assortative Mating

- Assortative mating occurs when individuals are more likely to choose partners with similar characteristics than random individuals.
- The individual's decision to mate with a person with similar characteristics may not only induce a correlation among phenotypes but also a correlation among genes associated with these phenotypes.
- As a result, the offspring of these individuals will have alleles inherited from their parents that have an effect on the phenotype of interest and are non-independent.
- For example, if both parents have a high level of education and this is influenced by their genetic variants for different traits (i.e., variant 1 and variant 2 in Panel C of Figure 1), they will pass to their child both these variants, that will be correlated among each other (dotted line in Panel C of Figure 1) and that they will both affect the phenotype, such as educational attainment. Therefore, children who inherit genetic variants associated with higher educational attainment from one of their parents (e.g., genetic variant 1 in Panel C of Figure 1) are also more likely than average to inherit genetic variants (e.g., genetic variant 2 in Panel C of Figure 1) associated with higher educational attainment from their other parent.