

DISCUSSION PAPER SERIES

IZA DP No. 13671

**Association of a Genetic Risk Score with  
BMI along the Life-Cycle: Evidence from  
Several US Cohorts**

Anna Sanz-de-Galdeano  
Anastasia Terskaya  
Angie Upegui

SEPTEMBER 2020

## DISCUSSION PAPER SERIES

IZA DP No. 13671

# Association of a Genetic Risk Score with BMI along the Life-Cycle: Evidence from Several US Cohorts

**Anna Sanz-de-Galdeano**

*University of Alicante, CRES-UPF and IZA*

**Anastasia Terskaya**

*University of Navarra*

**Angie Upegui**

*University of Alicante*

SEPTEMBER 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9  
53113 Bonn, Germany

Phone: +49-228-3894-0  
Email: [publications@iza.org](mailto:publications@iza.org)

[www.iza.org](http://www.iza.org)

## ABSTRACT

---

# Association of a Genetic Risk Score with BMI along the Life-Cycle: Evidence from Several US Cohorts\*

We use data from the The National Longitudinal Study of Adolescent to Adult Health and from the Health and Retirement Study to explore how the the effect of individuals' genetic predisposition to higher BMI —measured by BMI polygenic scores— changes over the life-cycle for several cohorts. We find that the effect of BMI polygenic scores on BMI increases significantly as teenagers transition into adulthood (using the Add Health cohort, born 1974-83). However, this is not the case for individuals aged 55+ who were born in earlier HRS cohorts (1931-53), whose life-cycle pattern of genetic influence on BMI is remarkably stable as they move into old-age.

**JEL Classification:** I1, I14

**Keywords:** obesity, BMI polygenic scores, Add Health, Health and Retirement Study

**Corresponding author:**

Anna Sanz-de-Galdeano

FAE

Universidad de Alicante

Carretera de San Vicente s/n

03080 San Vicente – Alicante

Spain

E-mail: [anna.sanzdegaldeano@gmail.com](mailto:anna.sanzdegaldeano@gmail.com)

---

\* Sanz-de-Galdeano acknowledges financial support from PROMETEO/2019/037, and from the Spanish Ministry of Economy and Competitiveness Grant ECO2017-87069-P. We thank Pedro Albarrán, Dimitris Christelis, and Michael Rosholm for comments and suggestions.

# 1 Introduction

According to the World Health Organization, worldwide obesity has almost tripled since 1975, and about 39% and 13% of the world’s adult population in 2016 were overweight and obese, respectively. The prevalence of overweight and obesity among children and adolescents has risen even more dramatically from 4% in 1975 to just over 18% in 2016. The equivalent figures only for obesity among children and adolescents are just under 1% in 1975 and about 7% in 2016 (for further details see <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>).

These trends pose serious challenges to both individual and public health because raised BMI is a risk factor for noncommunicable conditions such as high cholesterol, high blood pressure, and coronary heart disease among others (see for instance [NHLBI \(2013\)](#) and the references therein), as well as some cancers ([Bhaskaran et al., 2014](#)) and mental illnesses ([Kasen et al., 2008](#); [Luppino et al., 2010](#)). Additionally, obesity has also been shown to affect health care spending and individuals’ socioeconomic outcomes (see for instance [Cawley \(2004\)](#); [Brunello and d’Hombres \(2007\)](#); [Brunello et al. \(2008\)](#); [Cawley \(2015\)](#); [Böckerman et al. \(2019\)](#)).

Obesity is a many-sided problem with multiple determinants. Hence, its analysis has not been based on a unique perspective, and scholars from several disciplines have contributed to advance knowledge in this area. Social scientists have often focused on the role played by dietary and physical activity patterns that are in turn likely affected by factors like food prices, agricultural policies, income, maternal employment and technology ([Cawley, 2011](#); [Cutler et al., 2003](#); [Cawley, 2015](#)). Importantly, BMI is also affected by genetic factors, and obesity is known to be both highly heritable and polygenic (see for instance [Maes et al. \(1997\)](#); [Yang et al. \(2007\)](#); [Speliotes et al. \(2010\)](#); [Sandholt et al. \(2012\)](#); [Visscher et al. \(2012\)](#); [Zaitlen et al. \(2013\)](#); [Jou \(2014\)](#); [Domingue et al. \(2014\)](#); [Locke et al. \(2015\)](#); [Yengo et al. \(2018\)](#)).

In this paper we study how the association between obesity-related genetic variants and BMI varies along the life-cycle or over time across several cohorts in the U.S., where obesity—which affects about 39% of adults—has increased dramatically in the past decades ([OECD, 2017](#)), and obesity-related conditions are some of the leading causes of preventable death ([NHLBI, 2013](#)). We rely on data from two longitudinal representative surveys that contain genome-wide data from respondents: the National Longitudinal Study of Adolescent Health (Add Health hereafter) and the Health and Retirement Study (HRS hereafter). Individuals’ genetic propensity to high BMI is measured using BMI polygenic scores—available in both Add Health and the HRS—constructed based on a recent large-scale genome wide association study for BMI ([Locke et al., 2015](#)). We study whether the association between BMI and BMI polygenic scores is amplified or mitigated as teenagers transition and settle into adulthood (using Add Health), and as middle-age individuals transition to old-age (using the HRS). We also test whether significantly different patterns arise by childhood socioeconomic status and gender.

Our paper is related to a growing body of research that investigates how individuals’ ge-

netic predisposition to different phenotypes interacts with the environment (Boardman et al., 2013, 2014). In regard to obesity, previous work has shown that childhood socioeconomic status (Liu and Guo, 2015), social understandings of body size (Boardman et al., 2012), and individuals' education (Barcellos et al., 2018) moderate the influence of obesity-related genetic variants on obesity-related phenotypes.

Another related strand of the literature has instead used birth cohort as an indicator for exposure to obesogenic environment. Studies for the U.S. have shown that the association between obesity-related genetic variants and BMI is larger among individuals born in later cohorts (Demerath et al., 2013; Walter et al., 2016; Liu and Guo, 2015; Rosenquist et al., 2015). Additionally, Rokholm et al. (2011b) and Rokholm et al. (2011a) have uncovered an increase in the contribution of genetic factors to variation in BMI for successive birth cohorts in Sweden and Denmark, respectively. This body of results has been interpreted as evidence that individuals' genetic risk for elevated BMI is amplified when their lives unfold in more obesogenic socio-historical contexts.

This paper focuses on a related question that has received less attention in the literature: is the association between obesity-related genetic variants and obesity-related phenotypes attenuated or strengthened as individuals from the same cohort grow older? Khera et al. (2019) have recently shown that the gap in the prevalence of severe obesity between individuals in the top and bottom polygenic score deciles widens during the transition from young adulthood to middle age in the U.S. (using data from the Framingham Offspring and Coronary Artery Risk Development in Young Adults studies), and they have also uncovered a similar pattern in children's weight from birth to 18 years of age in the UK (using data from the Avon Longitudinal Study of Parents and Children). Belsky et al. (2012) use the Dunedin Multi-disciplinary Health and Development Study, which followed individuals born in 1972-73 in Dunedin (New Zealand) from birth through 38 years, and they find that higher BMI genetic risk scores predict higher BMI growth during childhood (from ages 3 through 13 years), as well as during adulthood (from ages 13 through 38 years).

We add to the limited literature on gene-age interaction effects on BMI (Lasky-Su et al., 2008; Khera et al., 2019; Hardy et al., 2010; Belsky et al., 2012) by analysing the effect of BMI polygenic scores as individuals transition from adolescence to young adulthood, and from middle-age to old-age. Moreover, we also analyse whether life-cycle profiles of genetic influence significantly differ by individual characteristics such as gender and socioeconomic status.

We find that the effect of BMI polygenic scores on BMI significantly increases as teenagers transition into adulthood. Specifically, our results for the younger cohort (Add Health, born 1974-84) indicate that a standard deviation increase in BMI polygenic scores is associated with a 4.2% increase in BMI at ages 15-16, while the percentage increase in BMI amounts to 5.7% when individuals are about 28. For the earlier HRS Original cohort (born 1931-41), the effect of BMI polygenic scores amounts to 4.2% when respondents are about 55 years old, and it remains stable while they transition into old-age and eventually reach age 72. We uncover similarly stable life-cycle patterns when focusing on subsequent HRS cohorts (born 1942-53) of

55+ individuals. Our main result is unchanged when analyzing individuals' life-cycle profiles separately by gender and socioeconomic status: the effect of BMI polygenic scores on BMI peaks in early adulthood.

Interestingly, we also find that, in our later Add Health cohort, the effect of BMI polygenic scores is significantly stronger for individuals with lower childhood socioeconomic status than for their higher socioeconomic status counterparts. In contrast, genetic influence on BMI does not significantly vary by socioeconomic status in any of the earlier HRS cohorts we analyze. We also find that the patterns of genetic influence on BMI do not significantly vary by gender, neither in Add Health nor in the HRS.

The remainder of the paper is organized as follows. Section 2 describes the data and methods used, Sections 3 and 4 present the results, and Section 5 discusses some robustness checks. Section 6 concludes.

## 2 Materials and Methods

We use data from Add Health and from the HRS. We now describe both datasets, as well as how our working samples have been constructed. We then explain the indicator we use to measure individuals' genetic predisposition to high BMI (which is available in both datasets), and outline our empirical model.

### 2.1 The HRS Dataset

The HRS is a nationally representative longitudinal study of the U.S. public over age 50 conducted every two years since 1992. The HRS collects information on health, socioeconomic background, employment, income, wealth, and other factors relevant to aging and retirement. Genotyping was performed using DNA samples collected during enhanced face-to-face interviews conducted on half of the HRS sample each wave starting in 2006 (and in later waves for new participants). Hence, respondents must have survived at least until genotyping started (2006-08) to be part of our analysis. Detailed information on the HRS genotype data and quality control process can be found at <http://hrsonline.isr.umich.edu/modules/meta/xyear/pgs/desc/PGENSCORES3DD.pdf> and [http://hrsonline.isr.umich.edu/sitedocs/genetics/HRS2\\_qc\\_report\\_SEPT2013.pdf](http://hrsonline.isr.umich.edu/sitedocs/genetics/HRS2_qc_report_SEPT2013.pdf).

The HRS includes several birth cohorts with different entry years. In order to maximize sample size, our benchmark analysis is based on the so-called Original HRS cohort (born between 1931 and 1941) surveyed every two years from 1992 to 2012. However, we stop following this cohort in 2008 to avoid exacerbating potential biases related to mortality selection, which we discuss and address in Section 5.1. Our benchmark analysis relies on a balanced panel sample of 3,181 Original HRS cohort members of European descent who remained in the survey since 1992 until at least 2008, and for whom valid genetic data as well as information regarding their age, sex, height and weight are available. We focus on people of European descent because the BMI polygenic scores we use (described in detail in Section 2.3) were constructed using the results of a genome-wide association study that mostly relied on a sam-

ple of European-descent individuals (Locke et al., 2015). Based on self-reported height and weight information we have computed the Body Mass Index for respondents at each wave using the standard formula: weight in kilograms divided by height in meters squared ( $\text{kg}/\text{m}^2$ ). Individuals are classified as obese if their BMI is 30 or higher following the World Health Organization's recommendation regarding BMI thresholds for defining obesity in adults (World Health Organization, 2000). We use self-reports instead of measured values of weight and height because the latter are only available from 2006 onwards (see Section 5.2).

Table 1 provides basic descriptives on age, sex, BMI, and obesity prevalence for our analytic sample. Both mean BMI and obesity prevalence increase with age until individuals are almost 68, and they remain fairly stable thereafter at around 27.8 and 29%, respectively.

Additionally, we have replicated our analysis using two subsequent HRS cohorts: the War Babies cohort (born 1942-47 and followed from 1998 until 2014), and the Early Baby Boomers cohort (born 1948-55 and followed from 2004 until 2016). The sample selection criteria applied to these cohorts are analogous to those described above for the Original HRS cohort.

## 2.2 The Add Health Dataset

Add Health is a school-based longitudinal study of a nationally representative cohort of adolescents in grades 7-12 in the United States during the 1994-95 school year ( $n=20,745$ , age range 12-20 at Wave 1). Add Health is based on a stratified sample of 80 high schools and 52 middle schools with probability of selection proportional to school size. Schools were stratified by region, urbanicity, school type, ethnic mix, and size. Add Health Wave I included an in-school questionnaire (administered to all the students attending the participating schools on the interview day), a more detailed in-home interview (conducted on a random sample of approximately 17 males and 17 females that were randomly selected within school and grade), and a parent questionnaire that was in general answered by the resident mothers of teenagers selected for the in-home sample. In-sample individuals have so far been followed in Waves II (1996, age range 12-21,  $n=14,738$ ), III (2000-01, age range 18-27,  $n=15,197$ ), IV (2008-09, age range 24-33,  $n=15,701$ ), and most recently in Wave V (2016-18, age range 33-43,  $n=12,300$ ).

We use data from all the waves of Add Health currently available (Waves I-V). Baseline demographic information on students and their families is obtained from Wave I, while self-reports on weight and height are used to construct BMI at each wave. We do not use objective measurements in our main analysis because they are not available in all waves of Add Health and the HRS. In Section 5.2, we replicate the main results using the objective BMI measures available in both datasets.

Saliva samples for DNA extraction were collected at Wave IV on the full sample. DNA measures were collected at Wave III for the sibling sample of Add Health (see [https://www.cpc.unc.edu/projects/addhealth/documentation/guides/PGS\\_AH1\\_UserGuide.pdf](https://www.cpc.unc.edu/projects/addhealth/documentation/guides/PGS_AH1_UserGuide.pdf) for a detailed description of genome-wide data collection and quality control protocols).

The formula used to compute BMI is the same for children and adults ( $\text{kg}/\text{m}^2$ ), but weight, height, and their relation to body fatness change along the life-cycle. All Wave I and most of Wave II respondents were still teenagers, so in those cases we followed the guidelines of the

Table 1: Summary Statistics. HRS Original Cohort Sample

Variable	Mean	Standard Deviation
BMI PGS (Normalized)	0	1.000
BMI 1992	26.951	4.562
BMI 1994	27.055	4.555
BMI 1996	27.267	4.739
BMI 1998	27.536	4.838
BMI 2000	27.689	4.935
BMI 2002	27.779	4.936
BMI 2004	27.871	5.126
BMI 2006	27.909	5.191
BMI 2008	27.825	5.247
Obese 1992	0.217	0.412
Obese 1994	0.235	0.424
Obese 1996	0.242	0.429
Obese 1998	0.263	0.440
Obese 2000	0.277	0.448
Obese 2002	0.289	0.453
Obese 2004	0.296	0.456
Obese 2006	0.292	0.455
Obese 2008	0.290	0.454
Age 1992	55.914	3.147
Age 1994	57.776	3.140
Age 1996	59.781	3.137
Age 1998	61.649	3.135
Age 2000	63.589	3.137
Age 2002	65.733	3.137
Age 2004	67.703	3.143
Age 2006	69.677	3.130
Age 2008	71.679	3.139
Female	0.553	0.497

Note: Statistics based on a balanced panel sample of 3,181 HRS Original cohort members of European descent who remained in the survey from 1992 until at least 2008, and for whom valid genetic data as well as information regarding their age, sex, height and weight are available.

U.S. Centre for Disease Control and Prevention (Kuczmarski, 2002) and classified them as obese if their BMI was equal to or greater than the 95th percentile. BMI percentiles by sex and age in the US are taken from the 2000 CDC growth charts, publicly available at [https://www.cdc.gov/growthcharts/percentile\\_data\\_files.htm](https://www.cdc.gov/growthcharts/percentile_data_files.htm). For respondents older than 20

we used instead the obesity definition applied to adults (BMI at or above 30).

Our Add Health analyses rely on a balanced panel sample of 2,730 individuals of European ancestry who remained in the survey from Wave I through Wave V with valid genetic data as well as information on age and sex, and for whom height and weight self-reports are available at all waves.

Table 2 provides basic descriptive statistics for this sample. There is a remarkable increase in both mean BMI (from 22.4 to 29.6) and obesity prevalence (which almost quadruples from 10% to 40%) as individuals transition from adolescence (average age 15.4) to young adulthood (average age 37.3).

Table 2: Summary Statistics. Add Health Sample

Variable	Mean	Standard Deviation
BMI PGS (Normalized)	-0.000	1.000
BMI 1994/95	22.379	4.442
BMI 1996	23.018	4.772
BMI 2001/02	25.668	5.920
BMI 2008/09	28.104	6.818
BMI 2016/18	29.572	7.329
Obese 1994/95	0.101	0.301
Obese 1996	0.111	0.314
Obese 2001/02	0.189	0.392
Obese 2008/09	0.316	0.465
Obese 2016/18	0.399	0.490
Age 1994/95	15.412	1.702
Age 1996	16.283	1.744
Age 2001/02	21.738	1.747
Age 2008/09	28.253	1.732
Age 2016/18	37.305	1.839
Female	0.473	0.499

Note: Statistics based on a balanced panel sample of 2,730 individuals of European ancestry who remained in the Add Health survey from Wave I through Wave V, and for whom valid genetic data as well as information regarding their age, sex, height and weight are available. Longitudinal weights are used.

### 2.3 BMI Polygenic Scores

Both Add Health and the HRS currently include BMI polygenic scores, indices that summarize individuals' genetic risk for elevated BMI (*BMIPGS* hereafter). These BMI PGS were computed based on the genome wide association (GWAS) study for BMI conducted by [Locke et al. \(2015\)](#) on a sample of 339,224 individuals. GWAS scan the entire genome in order to identify single nucleotide polymorphisms (SNPs) that are associated with a particular out-

come while using strict significance thresholds to deal with multiple hypothesis testing. SNPs are locations in the genome where there are differences across individuals that can be associated with a particular trait. [Belsky and Israel \(2014\)](#) provide further details regarding the construction of genetic risk scores from GWAS results. [Locke et al. \(2015\)](#) used conservative thresholds for statistical significance ( $P - value < 5 \times 10^{-8}$ ) and identified 97 SNPs significantly associated with BMI. *BMIPGS* are constructed for Add Health and HRS respondents by computing a weighted sum of these SNPs:

$$BMIPGS_i = \sum_{j=1}^k \hat{\beta}_j SNP_{ij} \quad (1)$$

where  $SNP_{ij} \in \{0, 1, 2\}$  is a count of the number of reference alleles for individual  $i$  at SNP  $j$ , and  $\hat{\beta}_j$  is the underlying GWAS coefficient estimated by [Locke et al. \(2015\)](#) for each SNP associated with BMI. In our Add Health working sample, BMI polygenic scores account for 4.9% (Wave I in 1994-95, mean age 15.4), 5.5% (Wave II in 1996, mean age 16.3), 5.0% (Wave III in 2001-02, mean age 21.7), 6.2% (Wave IV in 2008-09, mean age 28.3) and 5.5% (Wave V in 2016-18, mean age 37.3) of the total variation in BMI. The corresponding figures for our HRS Original cohort analytic sample are: 6.2% (Wave I in 1992, mean age 55.9), 5.9% (Wave II in 1994, mean age 57.8), 6.3% (Wave III in 1996, mean age 59.8), 5.8% (Wave IV in 1998, mean age 61.6), 5.9% (Wave V in 2000, mean age 63.6), 5.8% (Wave VI in 2002, mean age 65.7), 5.9% (Wave VII in 2004, mean age 67.8), 5.5% (Wave VIII in 2004, mean age 69.7) and 5.2% (Wave IX in 2008, mean age 71.7).

Fig 1 plot the (kernel-smoothed) densities of respondents' *BMIPGS* in our HRS and Add Health balanced panel samples, respectively. The distributions are approximately normal.

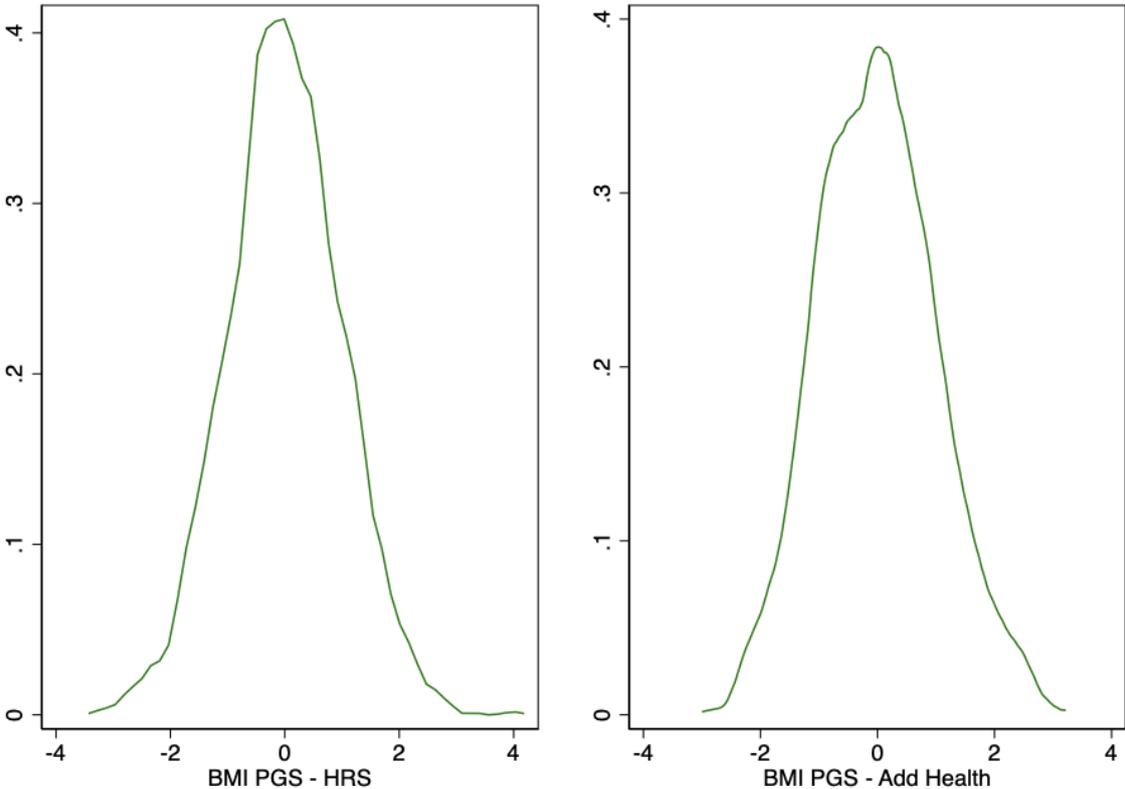
## 2.4 Empirical Model

Our baseline empirical specification is:

$$Y_{ic,t} = \beta_0 + \beta_1 BMIPGS_{ic} + X'_{ic,t} \alpha + \epsilon_{ic,t} \quad (2)$$

, where  $Y_{ic,t}$  is the log of BMI of individual  $i$  observed at time  $t$  who belongs to cohort  $c$  (Add Health or the Original HRS cohort in our main analyses).  $BMIPGS_{ic}$  denotes individuals' genetic predisposition to high BMI, which is fixed at conception.  $BMIPGS_{ic}$  is standardized to have mean 0 and standard deviation 1. The vector  $X_{ic,t}$  includes age, age squared, and a female dummy, as well as the 10 principal components of the full matrix of genetic data in order to account for population stratification ([Price et al., 2006](#); [Benjamin et al., 2012](#)). Our benchmark estimations of are based on self-reported BMI in order to avoid having different (objective vs. self-reported) BMI measurements for different ages. Add Health objective measures are available in Waves II-V (not in Wave I though, when individuals were 16.3 years old on average), while HRS objective measures are only available after 2006. In Section 5.2, we show estimation results based on objective BMI measures (whenever available), and compare them with our benchmark results based on subjective BMI measures. We estimate equation

Figure 1: BMI Polygenic Scores (Normalized) in the HRS Original Cohort and in Add Health.  
Kernel Density Estimates



Note: This figure displays the kernel-smoothed densities of HRS and Add Health respondents' BMI polygenic scores in the balanced samples described in Tables 1 and 2, respectively. Number of observations: 3,181 (HRS Original cohort) and 2,730 (Add Health).

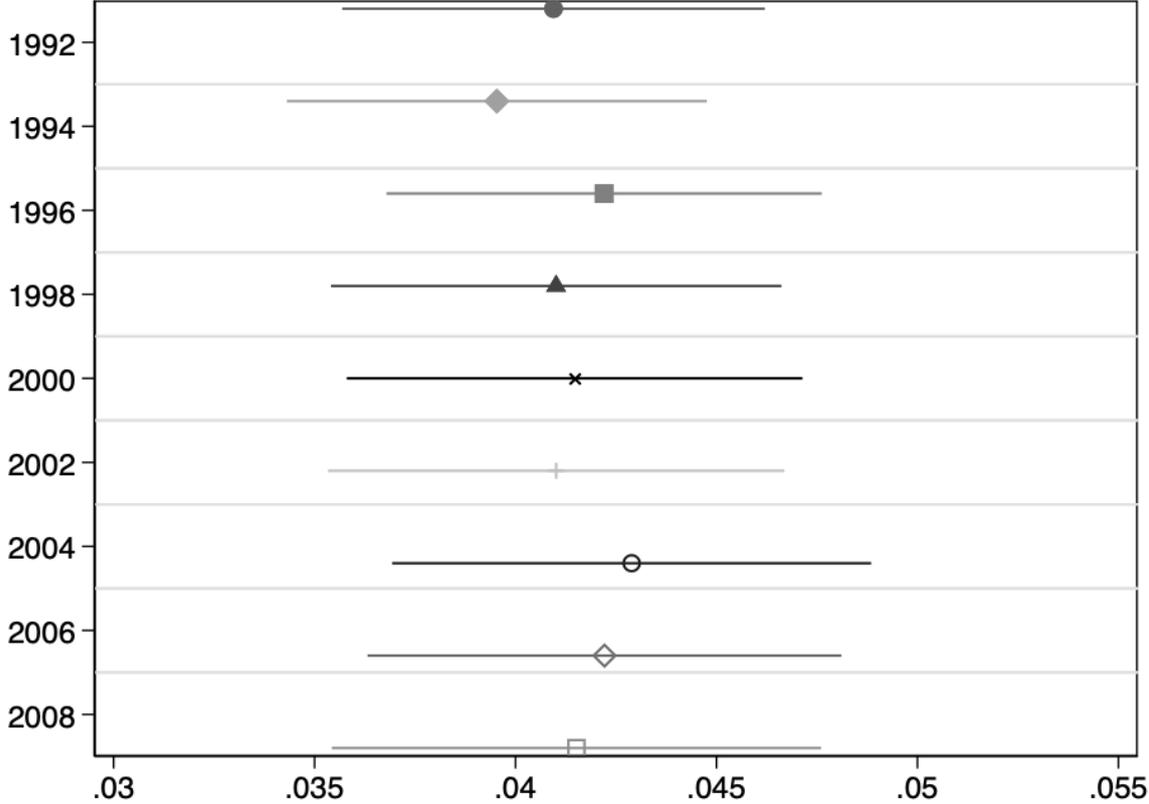
(2) for the Add Health cohort (born 1974-83) and the Original HRS cohort (born 1931-41) at different points in time: 1994-95, 1996, 2001-02, 2008-09, and 2016-18 for Add Health, and every two years since 1992 until 2008 for the Original HRS cohort. We then analyze whether genetic influence on BMI is amplified or mitigated along the life-cycle for both Add Health respondents (as they transition from adolescence to young adulthood) and Original HRS cohort members (as they transition from middle-age to old-age). Our choice of a log-level model rather than a level-level model in equation (2) is supported by AIC test results. In line with this, unconditional regression estimates (Firpo et al., 2009) indicate (see Appendix Figures A.1 and A.2) that the effect of a standard deviation increase in  $BMIPGS$  on BMI is non-linear and it is larger (in absolute terms) the higher the level of BMI.

### 3 Main Results

#### 3.1 Genetic Influence on BMI along the Life-Cycle: General Patterns

The results of estimating equation (2) on the sample of HRS Original cohort members are summarized in Fig 2, which depicts OLS coefficient estimates of  $\beta_1$  (as well as their associated 95% confidence intervals) that measure the estimated percentage increases in BMI associated with a standard deviation increase in *BMIPGS* as middle-aged adults move into old-age. OLS coefficient estimates and their corresponding standard errors (clustered at the household level) are displayed in Appendix Table A.1. The estimated life-cycle profile indicates that BMI increases associated with a standard deviation increase in *BMIPGS* remain stable around just above 4% along the life-cycle. Interestingly, a similarly flat life-cycle profile is observed in two subsequent HRS cohorts—the HRS War Babies cohort (born 1942-47) and the Early Baby Boomers cohort (born 1948-1953)—for whom a standard deviation increase in *BMIPGS* is associated with BMI increases of 5%-6% as they grow older (see Figure A.3 in Appendix A).

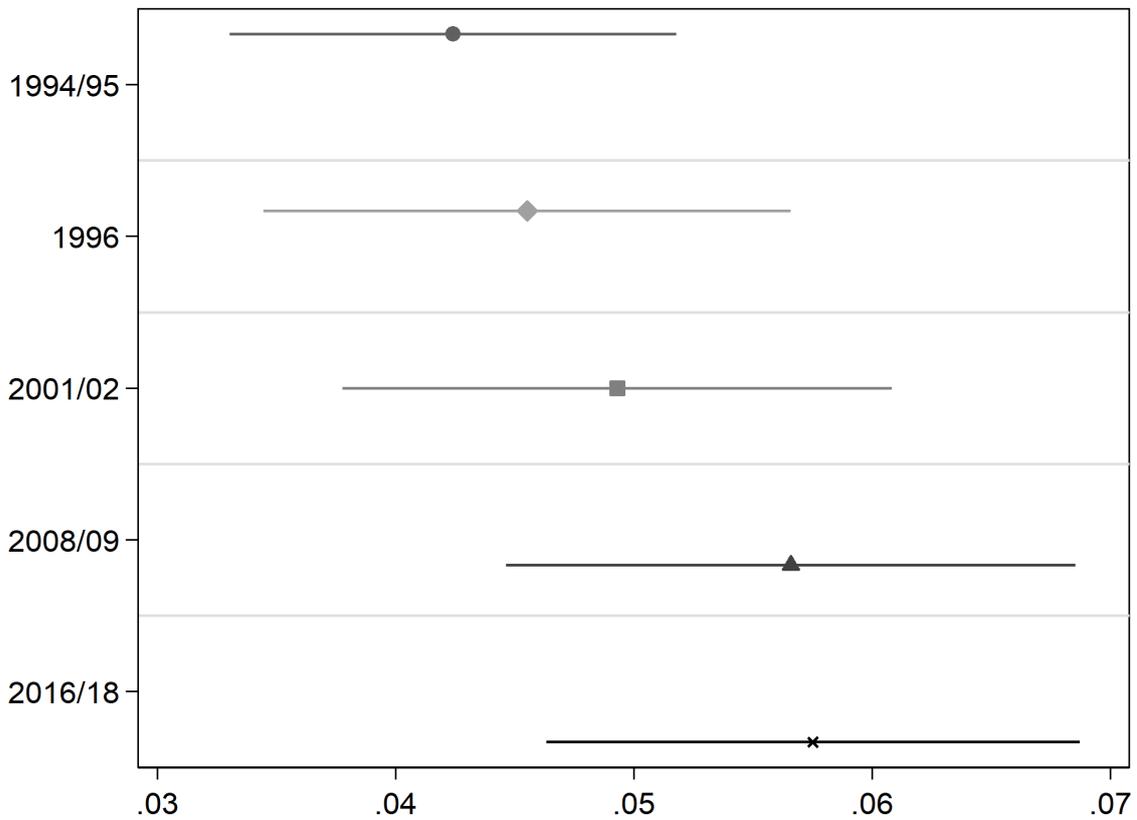
Figure 2: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle.  
HRS Original Cohort



Note: This Figure summarizes the results of estimating equation 2 on the balanced sample of 3,181 HRS Original cohort members described in Table 1. The dependent variable is Log(BMI). OLS coefficient estimates of  $\beta_1$  as well as their associated 95% confidence intervals are depicted. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. Standard errors are clustered at the household level.

Results for Add Health respondents (Fig 3, Appendix Table A.2) instead suggest that the association between *BMIPGS* and BMI increases as teenagers become adults. In particular, a standard deviation increase in *BMIPGS* increases individuals' BMI by 5.8% by the time they are about 37 in 2016-18, a significantly larger association than the one estimated (4.2%) when they were 15-16 years old (in 1994-95, one sided p-value=0.002). Interestingly, the association between *BMIPGS* and log(BMI) appears to stabilize at just above 5.5% at Wave IV (2008-09, average age 28).

Figure 3: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle.  
Add Health Cohort



Note: This Figure summarizes the results of estimating equation 2 on the balanced sample of 2,730 Add Health cohort members described in Table 2. The dependent variable is Log(BMI). OLS coefficient estimates of  $\beta_1$  as well as their associated 95% confidence intervals are depicted. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. Standard errors are clustered at the school level. Longitudinal weights are used.

Equation (2) is a reduced-form model, and disentangling the mechanisms through which individuals' *BMIPGS* may differently affect their BMI at different stages of their lives is beyond the scope of this paper. However, it is worth outlining several potential (and not mutually exclusive) determinants of the pattern of genetic influence we uncover in Add Health. First, homophily may be playing a role both at the genotypic (Domingue et al., 2018) and the phenotypic level (Crosnoe et al., 2008; De La Haye et al., 2011; Schaefer and Simpkins, 2014). In the presence of peer effects, homophily may, in turn, lead to social multiplier effects, which

would be consistent with the increasing relevance of genetic influence we observe until Add Health individuals approach age 30. These effects may, however, dissipate over time. For instance, [Brunello et al. \(2020\)](#) find that social-genetic effects on obesity induced by interactions with high school grade-mates are relevant for girls in adolescence, but they dissipate as they grow into adulthood.

Second, the effect of genes on BMI is likely to be altered by environmental factors that change during the life course ([Boardman et al., 2012](#); [Biroli, 2015](#); [Liu and Guo, 2015](#); [Barcellos et al., 2018](#)). For example, [Biroli \(2015\)](#) suggests that individuals with high genetic predisposition for obesity are more responsive to food intake than those with low genetic predisposition for obesity. To the extent that the impact of food consumption on BMI accumulates over time, the BMI gap between individuals with low and high *BMIPGS* can also grow throughout life.

Third, individuals with high *BMIPGS* may sort into more obesogenic environments. In line with this hypothesis, [Biroli \(2015\)](#) shows that individuals with a higher genetic predisposition for obesity tend to display a higher demand for food, the effect of which can also be cumulative. In contrast, [Nagata et al. \(2019\)](#) find that higher *BMIPGS* are associated with weight loss behaviors, which could reduce genetic influence throughout life.

Fourth, genetically influenced characteristics in children may evoke environmental responses that may in turn alter those characteristics, as genetic and environmental variation are not mutually exclusive ([Jencks, 1980](#)).

Finally, there may be age-related differences in genetic expression which may result in later manifestations of some genes ([Yao et al., 2014](#); [Harris et al., 2017](#)).

### 3.2 Genetic Influence on BMI along the Life-Cycle: Patterns by Gender and Socioeconomic Background

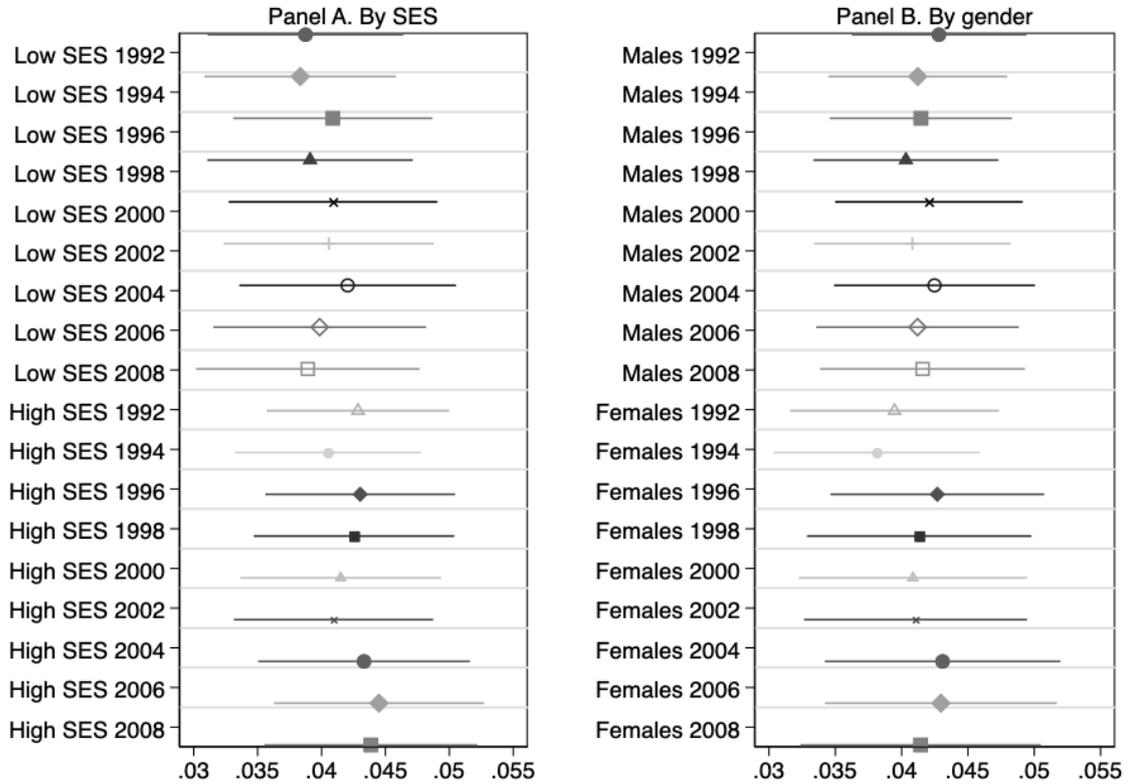
We now explore whether the life-cycle patterns we have uncovered so far significantly vary by gender and by childhood socioeconomic status (SES). We use parental background information from both Add Health and the HRS in order to construct summary indices of childhood SES. The construction of these summary indices is detailed in Appendix B. Individuals are classified as High SES and Low SES if the value of their childhood SES index is above and below the median, respectively.

Fig 4 shows how genetic influence on BMI varies by gender and by socio-economic status in the sample of HRS Original cohort members as they age. The life-cycle profile of genetic influence is stable for all subgroups. Additionally, the association between *BMIPGS* and  $\log(\text{BMI})$  does not significantly differ neither by gender nor by SES at any point in time.

The results for the Add Health cohort are depicted in Figure 5. There is a remarkable SES gradient in the influence of *BMIPGS*: the effect of BMI polygenic scores is significantly stronger for individuals with lower family socioeconomic status than for those with higher socioeconomic status. In contrast, there are no significant differences by gender.

Regarding life-cycle patterns of genetic influence in Add Health, the conclusions are the same for all subgroups: the association between *BMIPGS* and  $\log(\text{BMI})$  significantly increases as adolescents transition into adulthood.

Figure 4: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle: Patterns by Gender and Socioeconomic Background. HRS Original Cohort



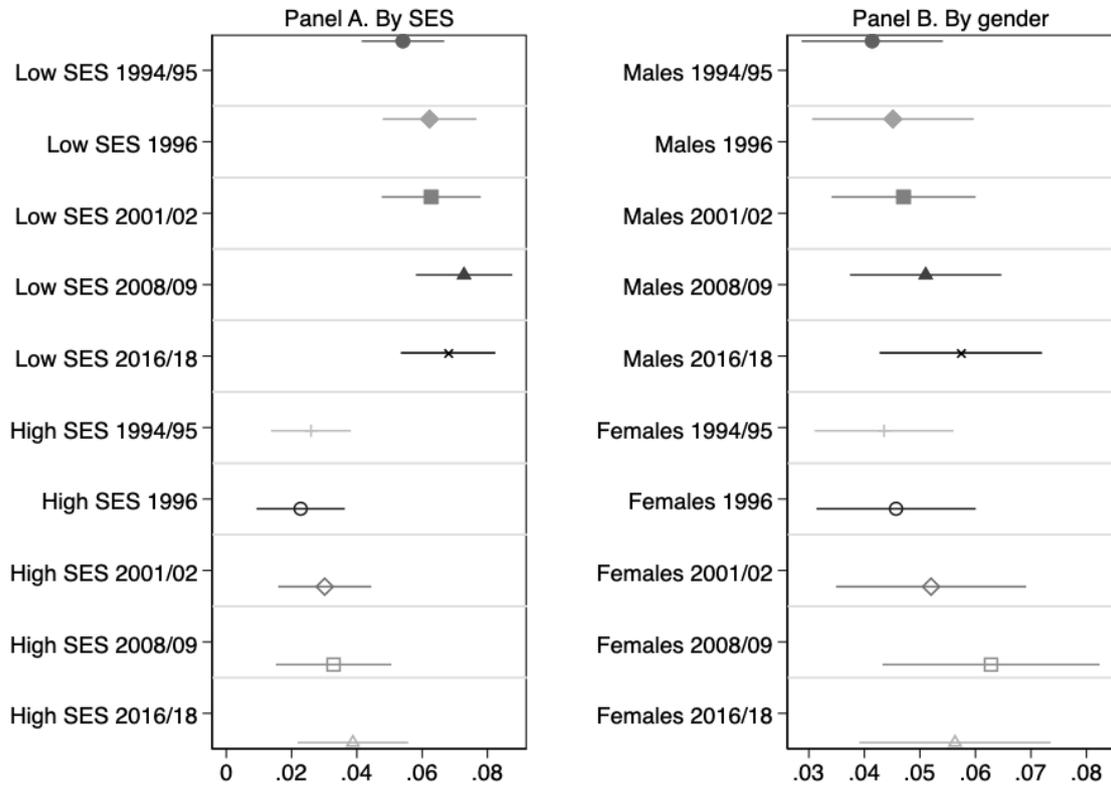
Note: This Figure summarizes the results of estimating equation 2 on the balanced sample of 3,181 HRS Original cohort members described in Table 1 by parental socioeconomic status (SES, in Panel A) and by gender (Panel B). The dependent variable is Log(BMI). OLS coefficient estimates of  $\beta_1$  as well as their associated 95% confidence intervals are depicted. All regressions include age, age squared, and the first 10 principal components of the full matrix of genetic data. Regressions by SES (in Panel A) also include a female dummy as a covariate. Low and High SES individuals are those whose parental socioeconomic status is below and above the median, respectively. Standard errors are clustered at the household level.

## 4 Additional Results

### 4.1 Pubertal Stage and the Association of BMI PGS with BMI

Puberty and BMI are likely related (Ong et al. (2012); Solorzano and McCartney (2010), among others), and pubertal timing differs across individuals. Therefore, part of BMI variation during adolescence may be due to pubertal stage differences across teenage respondents. Hence, the variance of the error in equation (2) is likely larger for adolescents than for older individuals. Moreover, there is evidence that pubertal timing and BMI have a common genetic component and therefore part of the effect of genes on BMI might be explained by the effect of genes on pubertal timing (Elks et al., 2010; Day et al., 2017). To study whether our previous results are affected by these factors, we replicate our baseline analyses including gender-specific information on the stage of development of adolescents that Add Health collected in Waves I

Figure 5: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle: Patterns by Gender and Socioeconomic Background. Add Health Cohort



Note: This Figure summarizes the results of estimating equation 2 on the balanced sample of 2,730 Add Health cohort members described in Table 2 by parental socioeconomic status (SES, in Panel A) and by gender (Panel B). The dependent variable is Log(BMI). OLS coefficient estimates of  $\beta_1$  as well as their associated 95% confidence intervals are depicted. All regressions include age, age squared, and the first 10 principal components of the full matrix of genetic data. Regressions by SES (in Panel A) also include a female dummy as a covariate. Low and High SES individuals are those whose parental socioeconomic status is below and above the median, respectively. Standard errors are clustered at the school level. Longitudinal weights are used.

and II, as by Wave III individuals were already between 18 and 26 years old (21.7 years old on average in our analytic sample).

In particular, we use the following questions that were asked to boys in Waves I and II: i) “How much hair is under your arms now? 1 I have no hair at all, 2 I have a little hair, 3 I have some hair, but not a lot; it has spread out since it first started, 4 I have a lot of hair that is thick, 5 I have a whole lot of hair that is very thick, as much hair as a grown man”; ii) “How thick is the hair on your face? 1 I have a few scattered hairs, but the growth is not thick, 2 The hair is somewhat thick, but you can still see a lot of skin under it, 3 The hair is thick; you can’t see much skin under it, 4 The hair is very thick, like a grown man’s facial hair”; iii) “Is your voice lower now than it was when you were in grade school? 1 No, it is about the same as when you were in grade school, 2 Yes, it is a little lower than when you were in grade school, 3 Yes, it is somewhat lower than when you were in grade school, 4 Yes, it is a lot lower than when

you were in grade school, 5 Yes, it is a whole lot lower than when you were in grade school; it is as low as an adult man’s voice”; and iv) “How advanced is your physical development compared to other boys your age? 1 I look younger than most, 2 I look younger than some, 3 I look about average, 4 I look older than some, 5 I look older than most”.

As for girls, we use the following questions that were asked in Waves I and II: i) “As a girl grows up her breasts develop and get bigger. Which sentence best describes you? 1 My breasts are about the same size as when I was in grade school, 2 My breasts are a little bigger than when I was in grade school, 3 My breasts are somewhat bigger than when I was in grade school, 4 My breasts are a lot bigger than when I was in grade school, 5 My breasts are a whole lot bigger than when I was in grade school, they are as developed as a grown woman’s breasts”; ii) “As a girl grows up her body becomes more curved. Which sentence best describes you? 1 My body is about as curvy as when I was in grade school, 2 My body is a little more curvy than when I was in grade school, 3 My body is somewhat more curvy than when I was in grade school, 4 My body is a lot more curvy than when I was in grade school, 5 My body is a whole lot more curvy than when I was in grade school”; iii) “Have you ever had a menstrual period (menstruated)? 0 No, 1 Yes”; and iv) “How advanced is your physical development compared to other girls your age? 1 I look younger than most, 2 I look younger than some, 3 I look about average, 4 I look older than some, 5 I look older than most”. We construct binary indicators for all the possible answers to these questions and we add them as controls to our estimations of equation (2) for Waves I and II. The results of this analysis, reported in Appendix Table A.3, indicate that the effect of *BMIPGS* on  $\log(\text{BMI})$  is lower after the inclusion of puberty stage controls. This is consistent with the fact that pubertal timing and BMI have a common genetic component. As a consequence, the estimated association between *BMIPGS* and  $\log(\text{BMI})$  increases more markedly as individuals transition from adolescence into adulthood when we control for pubertal stage indicators than when we do not (see Figure 3 and/or Appendix Table A.2). While it is reassuring that our conclusion is robust to the addition of pubertal stage indicators, our preferred specification excludes this set of controls in order to avoid reverse causality bias, as there is evidence that childhood obesity increases the risk of premature puberty for girls and boys (Solorzano and McCartney, 2010). Moreover, we re-estimate our benchmark model including pubertal timing as an additional regressor in Appendix Table A.4. Females’ puberty onset is classified as early vs. delayed if age of menarche was lower 13 (which is the median in our sample) vs. 13+. Establishing males’ puberty onset is more complex. We do so following the recommendations from Mendle et al. (2019). In particular, we regress a pubertal status index on age, and we then save the residuals. The pubertal status index has been constructed using principal component analysis on the variables related to pubertal stage for boys previously described and measured in Wave I, as they display more variation in Wave I than in Wave II. Males’ puberty onset is subsequently classified as early vs. delayed if the regression’s residuals are below vs. above the median. As the comparison between Columns 1 and 2 of Appendix Table A.4 reveals, the inclusion of pubertal timing as a control barely alters the estimated coefficients of *BMIPGS*. In summary, this evidence indicates that the increasing pattern of association between *BMIPGS*

and  $\log(\text{BMI})$  we find for Add Health adolescents as they transition into adulthood is robust to the inclusion of controls for pubertal stage and the timing of puberty onset.

## 4.2 Morbidity and the Association of BMI PGS with BMI

Chronic diseases are more prevalent among the elderly, and they may in turn lead to wasting (BMI loss). We investigate whether our previous results for HRS Original cohort members are affected by the prevalence of the following conditions: heart disease, cancer, diabetes, lung disease, and arthritis. First, we study how the prevalence of these conditions correlates with both BMI and with *BMIPGS* in our analytic sample. The prevalence of heart disease, diabetes, and arthritis is positive and significantly correlated with BMI, while the prevalence of cancer, lung disease, and BMI are not significantly correlated. This pattern is the same for all sample years, that is, since individuals are on average 55.9 years old (in 1992) until they reach 71.7 years of age on average (in 2008). Hence, we find no evidence of BMI reductions being linked to higher prevalence of chronic diseases in our sample. The correlation between *BMIPGS* and chronic diseases is positive and significant for heart disease, diabetes, and arthritis, while it is generally insignificant for cancer and lung disease.

Next, we replicate our baseline analyses including the prevalence of these five chronic conditions as additional controls in all our sample years. The results of this analysis, reported in Appendix Table A.5 reveal that the inclusion of this set of controls slightly attenuates the estimated association between *BMIPGS* and  $\log(\text{BMI})$ . This is consistent with our previous finding that *BMIPGS* are positively and significantly correlated with several chronic diseases. Importantly, the life-cycle association between *BMIPGS* and  $\log(\text{BMI})$  remains stable as individuals transition from middle-age to old-age once these additional controls are included in our benchmark model (2). However, we do not include them in our preferred specification because their relationship with BMI is likely bidirectional.

## 5 Robustness checks

### 5.1 Attrition

The longitudinal nature of our analyses implies that there is attrition in both our Add Health and HRS samples. This could be problematic if attrition is systematically related to *BMIPGS*. We cannot directly test whether this is the case because individuals were not genotyped in the first wave we observe them neither in Add Health (genotyping took place in Wave IV) nor in the HRS (genotyping took place in 2006-08).

We can, however, investigate whether attrition is related to obesity and BMI measured the first time individuals were interviewed. We do so by regressing a binary variable identifying missing individuals due to attrition between the first and the last waves analyzed on initial BMI and obesity. We find that attrition is not significantly related to initial BMI or obesity status neither in Add Health nor in the HRS.

Concerns about attrition due to selective mortality may remain in the HRS because mem-

bers of the HRS Original cohort (55.9 years old on average the first time we observe them in 1992) may have died by the time genetic data were collected (Zajacova and Burgard, 2013; Domingue et al., 2017), and elevated BMI is known to have adverse health consequences. Actually, if we regress a dummy identifying attrited individuals due to death between the first (1992) and the last wave (2012) analyzed (instead of a dummy identifying overall attrition) on BMI and obesity measured in 1992, the estimated coefficients are positive and significant. Hence, we adjust our benchmark results for the HRS Original cohort by using inverse probability weighting methods as in Domingue et al. (2017). Fitted values from a logit survival regression are used to obtain probability weights which are used as inverse probability weights to adjust estimates for selective mortality. In particular, our inverse probability weights are based on fitted values obtained from estimating a logit model of the probability of survival (until genotyping took place) as a function of respondents' educational attainment, year of birth, and several health indicators (the means of individuals' BMI, CES depression scale, and self-reported health over all available years, indicators of whether respondents ever reported smoking, having diabetes, and having heart disease, and respondents' maximum height over all available waves).

The results of this adjustment, presented in Table A.6 in Appendix, suggest that our results are robust to selective mortality because they are extremely similar to those obtained in our benchmark analysis.

## 5.2 Objective Measurements versus Self-Reports of Weight and Height

Objective measurements of height and weight are only available in some waves of Add Health (Waves II, III, IV, and V) and the HRS (2006 and 2008). We use this information to investigate whether it is likely that using self-reports may affect our results, and our findings are reassuring. We show estimation results based on objective BMI measures (whenever available), and compare them with our benchmark results based on subjective BMI measures in Table A.7 in Appendix. Panel A of Table A.7 in Appendix displays the estimated associations between *BMIPGS* and objective (Column 1) and self-reported (Column 2)  $\log(\text{BMI})$  for the HRS Original cohort for years 2006 and 2008 (our sample years with available objective BMI measures). The comparison of Columns 1 and 2 reveals that the estimated associations between *BMIPGS* and objective and self-reported  $\log(\text{BMI})$  barely differ. Therefore, our conclusion that the link between *BMIPGS* and  $\log(\text{BMI})$  is stable over as middle-age individuals transition to old-age remains when using objective BMI measures. Panel B of Table A.7 in Appendix does the same comparative analysis for the Add Health cohort. The estimated coefficients of *BMIPGS* do not significantly differ (at the 5% level) across columns for all waves. Importantly, our finding that the association between *BMIPGS* and  $\log(\text{BMI})$  increases as adolescents transition into adulthood prevails when using objective BMI measures.

### 5.3 Socioeconomic Status and the Association of BMI PGS with BMI

We now replicate our benchmark analyses including childhood SES among the set of control variables. This allows us to explore further whether the observed life-cycle associations between *BMIPGS* and  $\log(\text{BMI})$  reflect similar patterns as association between SES and  $\log(\text{BMI})$  as individuals grow older. The results of these analyses are shown in Tables A.8 and A.9 in Appendix.

The association between SES and  $\log(\text{BMI})$  for Add Health cohort members is negative, significant, and it increases (in absolute terms) as they transition from adolescence into adulthood (Table A.8 in Appendix, Column 2). However, the inclusion of SES among the control set barely changes the estimated coefficients *BMIPGS* (Table A.8 in Appendix, comparison of Columns 1 and 3). This indicates that SES effects across the life course cannot explain the observed increasing association between *BMIPGS* and  $\log(\text{BMI})$  between adolescence and early adulthood, which remains basically unaltered when SES is held constant.

The association between SES and  $\log(\text{BMI})$  for HRS Original cohort members is negative and significant, and it does not significantly change as individuals get older (Table A.9 in Appendix, Column 2). The inclusion of the childhood SES index among the set of control variables hardly modifies the estimated coefficients of *BMIPGS* (Table A.9 in Appendix, comparison of Columns 1 and 3).

In summary, this evidence indicates that SES cannot account for the life-cycle patterns of association between *BMIPGS* and  $\log(\text{BMI})$  we have uncovered so far, neither for Add Health nor for HRS Original cohort members.

## 6 Discussion

In this paper we find that the effect of BMI polygenic scores on  $\log(\text{BMI})$  increases significantly as teenagers transition into adulthood (using the Add Health cohort, born 1974-83). However, this is not the case for individuals aged 55+ who were born in earlier cohorts (HRS Original cohort born 1931-41, War Babies cohort born 1942-47, and Early Baby Boomers cohort born 1948-53), whose life-cycle pattern of genetic influence on BMI is remarkably stable. We uncover similar life-cycle patterns for all the cohorts we study when we separately analyse males and females, and low and high socioeconomic status groups.

One possible explanation for our results is that the effect of BMI polygenic scores on BMI increases until people reach a certain age, and remains stable thereafter. This hypothesis is consistent with Hardy et al. (2010), who find that the association between genes and BMI peaks in early adulthood.

Interestingly, we also find that the association between BMI polygenic scores and BMI significantly differs by childhood socioeconomic status in the Add Health cohort, while this is not the case in earlier HRS cohorts. In particular, childhood socioeconomic status significantly moderates the effect of BMI polygenic scores for Add Health cohort members. In contrast, the effect of BMI polygenic scores does not significantly differ by gender in any of the cohorts analysed.

Last but not least, our findings also indicate that the effect of BMI polygenic scores on BMI is likely to be non-linear. In fact, the AIC test rejects the linear model in favor of a log-linear model. This simple transformation might be considered when conducting future GWAS in order to improve the predictive power of polygenic scores.

## 6.1 Strengths and limitations

In this paper, we use two longitudinal surveys to provide new evidence on gene-age interaction effects on BMI for several cohorts. In particular, we study teenagers from the Add Health cohort (born 1974-1983) as they transition into adulthood as well as individuals aged 55+ who were born in earlier HRS cohorts (1931-53) as they move into old-age. The use of panel data is crucial in this context because it allows one to disentangle age/time associations from cohort effects. In contrast, as argued by [Lasky-Su et al. \(2008\)](#), cross-sectional studies may fail to detect age-varying associations as they cannot disentangle age/time from cohort effects. Our analyses are based on different cohorts observed at different stages of the life cycle. Hence, our contrasting findings for Add Health and the HRS may reflect differing patterns of genetic influence along the life cycle, but they could also stem from systematic differences across cohorts in their life-cycle patterns of genetic influence.

Note also that in this paper we estimate a reduced-form model without digging into the mechanisms behind gene-age interactions because of data limitations. Our results therefore could be explained by changes in the biology of BMI across the life course as well as by environmental changes that may reinforce or mitigate the effect of genes on BMI ([Sanz-de Galdeano and Terskaya, 2019](#); [Liu and Guo, 2015](#); [Barcellos et al., 2018](#)). Understanding the mechanisms behind the patterns we uncover is worth further investigation.

Another limitation of our analyses is that the genome-wide association study employed to compute the *BMIPGS* used mostly relies on European-descent individuals ([Locke et al., 2015](#)). Therefore, our results cannot be generalized to individuals from different ancestries. The availability of GWAS for other ancestries would allow to overcome this limitation.

Finally, another potential limitation stems from the fact that the strength of genotype-phenotype associations may vary by age. Hence, GWAS results may not replicate in samples where the age distribution differs from that of the GWAS sample ([Lasky-Su et al., 2008](#)). The *BMIPGS* we use rely on the GWAS conducted by [Locke et al. \(2015\)](#), which is in turn mostly based on a sample of midlife individuals. Hence, their predictive power may be lower for younger individuals. A similar argument may apply to other demographic characteristics like childhood socioeconomic status (as our Add Health results by socioeconomic status suggest). While the strongest *BMIPGS*-*BMI* association we uncover is for young adults (Waves 4 and 5 of Add Health), this warrants further investigation.

## References

- BARCELLOS, S. H., CARVALHO, L. S. AND TURLEY, P. (2018): "Education can reduce health differences related to genetic risk of obesity," *Proceedings of the National Academy of Sciences*, 115, E9765–E9772.
- BELSKY, D. W., DOMINGUE, B. W., WEDOW, R. ET AL. (2018): "Genetic analysis of social-class mobility in five longitudinal studies," *Proceedings of the National Academy of Sciences*, 115, E7275–E7284.
- BELSKY, D. W. AND ISRAEL, S. (2014): "Integrating genetics and social science: Genetic risk scores," *Biodemography and Social Biology*, 60, 137–155.
- BELSKY, D. W., MOFFITT, T. E., HOUTS, R. ET AL. (2012): "Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study," *Archives of pediatrics & adolescent medicine*, 166, 515–521.
- BENJAMIN, D. J., CESARINI, D., CHABRIS, C. F. ET AL. (2012): "The promises and pitfalls of geneconomics," *Annu. Rev. Econ.*, 4, 627–662.
- BHASKARAN, K., DOUGLAS, I., FORBES, H. ET AL. (2014): "Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5· 24 million UK adults," *The Lancet*, 384, 755–765.
- BIROLI, P. (2015): "Genetic and economic interaction in the formation of human capital: the case of obesity," Tech. rep., Mimeo. University of Zurich.
- BOARDMAN, J. D., DAW, J. AND FREESE, J. (2013): "Defining the environment in gene–environment research: Lessons from social epidemiology," *American journal of public health*, 103, S64–S72.
- BOARDMAN, J. D., DOMINGUE, B. W., BLALOCK, C. L. ET AL. (2014): "Is the gene-environment interaction paradigm relevant to genome-wide studies? The case of education and body mass index," *Demography*, 51, 119–139.
- BOARDMAN, J. D., ROETTGER, M. E., DOMINGUE, B. W. ET AL. (2012): "Gene-environment interactions related to body mass: School policies and social context as environmental moderators," *Journal of Theoretical Politics*, 24, 370–388.
- BÖCKERMAN, P., CAWLEY, J., VIINIKAINEN, J. ET AL. (2019): "The effect of weight on labor market outcomes: An application of genetic instrumental variables," *Health economics*, 28, 65–77.
- BRUNELLO, G. AND D’HOMBRES, B. (2007): "Does body weight affect wages?: Evidence from Europe," *Economics & Human Biology*, 5, 1–19.
- BRUNELLO, G., MICHAUD, P.-C. AND SANZ-DE GALDEANO, A. (2008): "The rise in obesity across the Atlantic: An economic perspective," .

- BRUNELLO, G., SANZ-DE GALDEANO, A. AND TERSKAYA, A. (2020): "Not Only in My Genes: The Effects of Peers' Genotype on Obesity," *Journal of Health Economics*, 102349.
- CAWLEY, J. (2004): "The impact of obesity on wages," *Journal of Human resources*, 39, 451–474.
- (2011): *The Oxford handbook of the social science of obesity*, Oxford University Press.
- (2015): "An economy of scales: A selective review of obesity's economic causes, consequences, and solutions," *Journal of health economics*, 43, 244–268.
- CROSNOE, R., FRANK, K. AND MUELLER, A. S. (2008): "Gender, body size and social relations in American high schools," *Social Forces*, 86, 1189–1216.
- CUTLER, D. M., GLAESER, E. L. AND SHAPIRO, J. M. (2003): "Why have Americans become more obese?" *Journal of Economic perspectives*, 17, 93–118.
- DAY, F. R., THOMPSON, D. J., HELGASON, H. ET AL. (2017): "Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk," *Nature genetics*, 49, 834–841.
- DE LA HAYE, K., ROBINS, G., MOHR, P. ET AL. (2011): "Homophily and contagion as explanations for weight similarities among adolescent friends," *Journal of Adolescent Health*, 49, 421–427.
- DEMERATH, E. W., CHOH, A. C., JOHNSON, W. ET AL. (2013): "The positive association of obesity variants with adulthood adiposity strengthens over an 80-year period: a gene-by-birth year interaction," *Human heredity*, 75, 175–185.
- DOMINGUE, B. W., BELSKY, D. W., FLETCHER, J. M. ET AL. (2018): "The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health," *Proceedings of the National Academy of Sciences*, 115, 702–707.
- DOMINGUE, B. W., BELSKY, D. W., HARRATI, A. ET AL. (2017): "Mortality selection in a genetic sample and implications for association studies." *International Journal of Epidemiology*, 46, 1285–1294.
- DOMINGUE, B. W., BELSKY, D. W., HARRIS, K. M. ET AL. (2014): "Polygenic risk predicts obesity in both white and black young adults," *PloS one*, 9, e101596.
- ELKS, C. E., PERRY, J. R., SULEM, P. ET AL. (2010): "Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies," *Nature genetics*, 42, 1077.
- FIRPO, S., FORTIN, N. M. AND LEMIEUX, T. (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.
- HARDY, R., WILLS, A. K., WONG, A. ET AL. (2010): "Life course variations in the associations between FTO and MC4R gene variants and body size," *Human molecular genetics*, 19, 545–552.

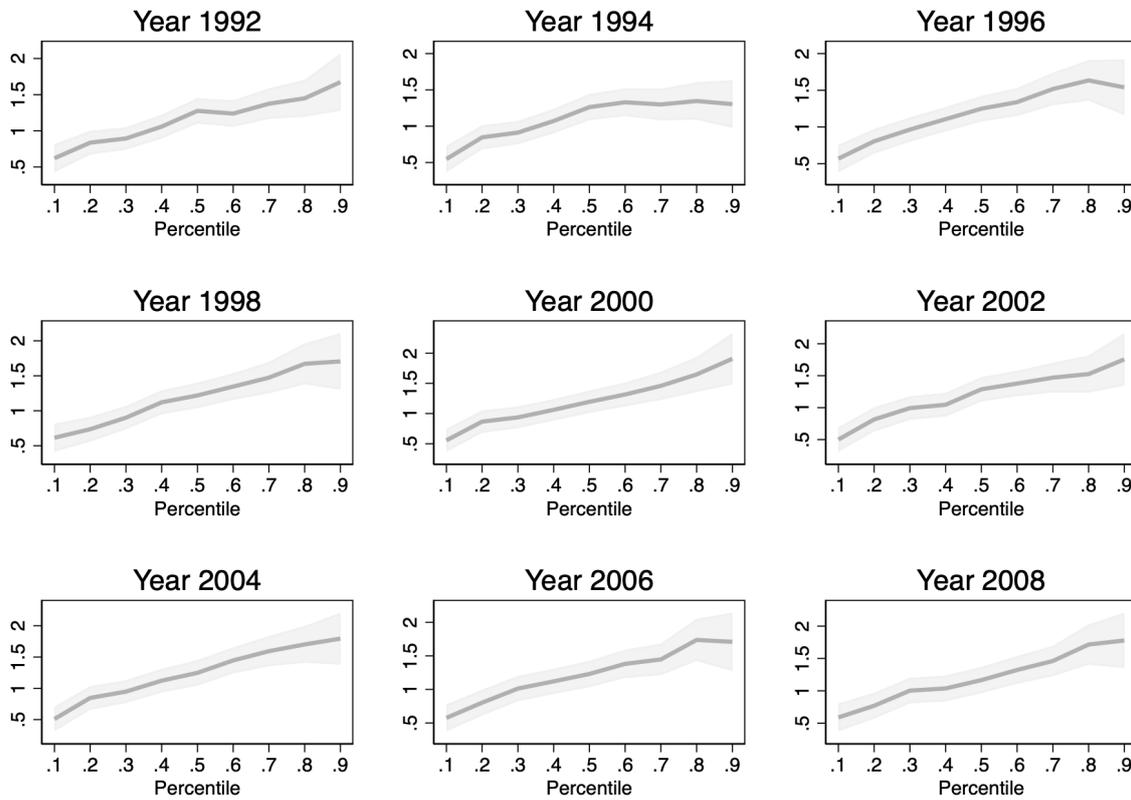
- HARRIS, S. E., RIGGIO, V., EVENDEN, L. ET AL. (2017): "Age-related gene expression changes, and transcriptome wide association study of physical and cognitive aging traits, in the Lothian Birth Cohort 1936," *Aging (Albany NY)*, 9, 2489.
- JENCKS, C. (1980): "Heredity, environment, and public policy reconsidered," *American Sociological Review*, 723–736.
- JOU, C. (2014): "The biology and genetics of obesity—a century of inquiries," *New England Journal of Medicine*, 370, 1874–1877.
- KASEN, S., COHEN, P., CHEN, H. ET AL. (2008): "Obesity and psychopathology in women: a three decade prospective study," *International Journal of Obesity*, 32, 558.
- KHERA, A. V., CHAFFIN, M., WADE, K. H. ET AL. (2019): "Polygenic prediction of weight and obesity trajectories from birth to adulthood," *Cell*, 177, 587–596.
- KLING, J. R., LIEBMAN, J. B. AND KATZ, L. F. (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75, 83–119.
- KUCZMARSKI, R. J. (2002): "2000 CDC growth charts for the United States; methods and development," .
- LASKY-SU, J., LYON, H. N., EMILSSON, V. ET AL. (2008): "On the replication of genetic associations: timing can be everything!" *The American Journal of Human Genetics*, 82, 849–858.
- LIU, H. AND GUO, G. (2015): "Lifetime socioeconomic status, historical context, and genetic inheritance in shaping body mass in middle and late adulthood," *American sociological review*, 80, 705–737.
- LOCKE, A. E., KAHALI, B., BERNDT, S. I. ET AL. (2015): "Genetic studies of body mass index yield new insights for obesity biology," *Nature*, 518, 197.
- LUPPINO, F. S., DE WIT, L. M., BOUVY, P. F. ET AL. (2010): "Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies," *Archives of general psychiatry*, 67, 220–229.
- MAES, H. H., NEALE, M. C. AND EAVES, L. J. (1997): "Genetic and environmental factors in relative body weight and human adiposity," *Behavior genetics*, 27, 325–351.
- MENDLE, J., BELTZ, A. M., CARTER, R. ET AL. (2019): "Understanding puberty and its measurement: ideas for research in a new generation," *Journal of Research on Adolescence*, 29, 82–95.
- NAGATA, J. M., BRAUDT, D. B., DOMINGUE, B. W. ET AL. (2019): "Genetic risk, body mass index, and weight control behaviors: Unlocking the triad," *International Journal of Eating Disorders*, 52, 825–833.

- NHLBI (2013): *Managing Overweight and Obesity in Adults: Systematic Evidence Review from the Obesity Expert Panel*, National Institutes of Health, U.S. Department of Health and Human Services.
- OECD (2017): *Obesity Update*, OECD Publishing, <https://www.oecd.org/els/health-systems/Obesity-Update-2017.pdf> (accessed 18.01.2020).
- ONG, K. K., BANN, D., WILLS, A. K. ET AL. (2012): "Timing of voice breaking in males associated with growth and weight gain across the life course," *The Journal of Clinical Endocrinology & Metabolism*, 97, 2844–2852.
- PRICE, A. L., PATTERSON, N. J., PLENGE, R. M. ET AL. (2006): "Principal components analysis corrects for stratification in genome-wide association studies," *Nature genetics*, 38, 904.
- ROKHOLM, B., SILVENTOINEN, K., ÄNGQUIST, L. ET AL. (2011a): "Increased genetic variance of BMI with a higher prevalence of obesity," *PloS one*, 6, e20816.
- ROKHOLM, B., SILVENTOINEN, K., TYNELIUS, P. ET AL. (2011b): "Increasing genetic variance of body mass index during the Swedish obesity epidemic," *PloS one*, 6, e27135.
- ROSENQUIST, J. N., LEHRER, S. F., O'MALLEY, A. J. ET AL. (2015): "Cohort of birth modifies the association between FTO genotype and BMI," *Proceedings of the National Academy of Sciences*, 112, 354–359.
- SANDHOLT, C., HANSEN, T. AND PEDERSEN, O. (2012): "Beyond the fourth wave of genome-wide obesity association studies," *Nutrition & diabetes*, 2, e37.
- SANZ-DE GALDEANO, A. AND TERSKAYA, A. (2019): "Sibling Differences in Educational Polygenic Scores: How Do Parents React?" Tech. rep., IZA Discussion Paper No. 12375.
- SCHAEFER, D. R. AND SIMPKINS, S. D. (2014): "Using social network analysis to clarify the role of obesity in selection of adolescent friends," *American journal of public health*, 104, 1223–1229.
- SOLORZANO, C. M. B. AND MCCARTNEY, C. R. (2010): "Obesity and the pubertal transition in girls and boys," *Reproduction (Cambridge, England)*, 140, 399.
- SPELIOTES, E. K., WILLER, C. J., BERNDT, S. I. ET AL. (2010): "Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index," *Nature genetics*, 42, 937.
- VISSCHER, P. M., BROWN, M. A., MCCARTHY, M. I. ET AL. (2012): "Five years of GWAS discovery," *The American Journal of Human Genetics*, 90, 7–24.
- WALTER, S., MEJÍA-GUEVARA, I., ESTRADA, K. ET AL. (2016): "Association of a genetic risk score with body mass index across different birth cohorts," *Jama*, 316, 63–69.
- WORLD HEALTH ORGANIZATION (2000): *Obesity: preventing and managing the global epidemic*, 894, World Health Organization.

- YANG, W., KELLY, T. AND HE, J. (2007): "Genetic epidemiology of obesity," *Epidemiologic reviews*, 29, 49–61.
- YAO, C., JOEHANES, R., JOHNSON, A. D. ET AL. (2014): "Sex-and age-interacting eQTLs in human complex diseases," *Human molecular genetics*, 23, 1947–1956.
- YENGO, L., SIDORENKO, J., KEMPER, K. E. ET AL. (2018): "Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry," *Human molecular genetics*, 27, 3641–3649.
- ZAITLEN, N., KRAFT, P., PATTERSON, N. ET AL. (2013): "Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits," *PLoS genetics*, 9, e1003520.
- ZAJACOVA, A. AND BURGARD, S. A. (2013): "Healthier, wealthier, and wiser: a demonstration of compositional changes in aging cohorts due to selective mortality," *Population research and policy review*, 32, 311–324.

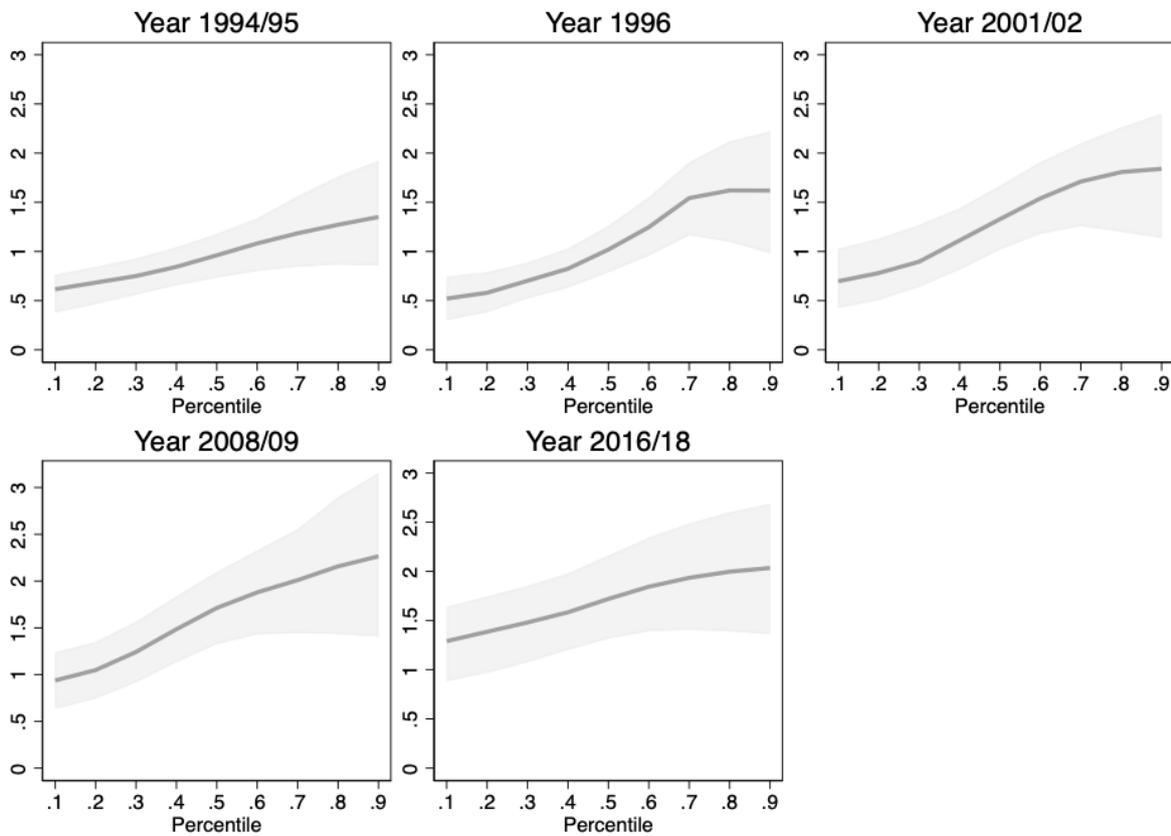
## A Figures and Tables

Figure A.1: Association between BMI Polygenic Scores and BMI Along the Life-Cycle and Along the BMI Distribution. HRS Original Cohort



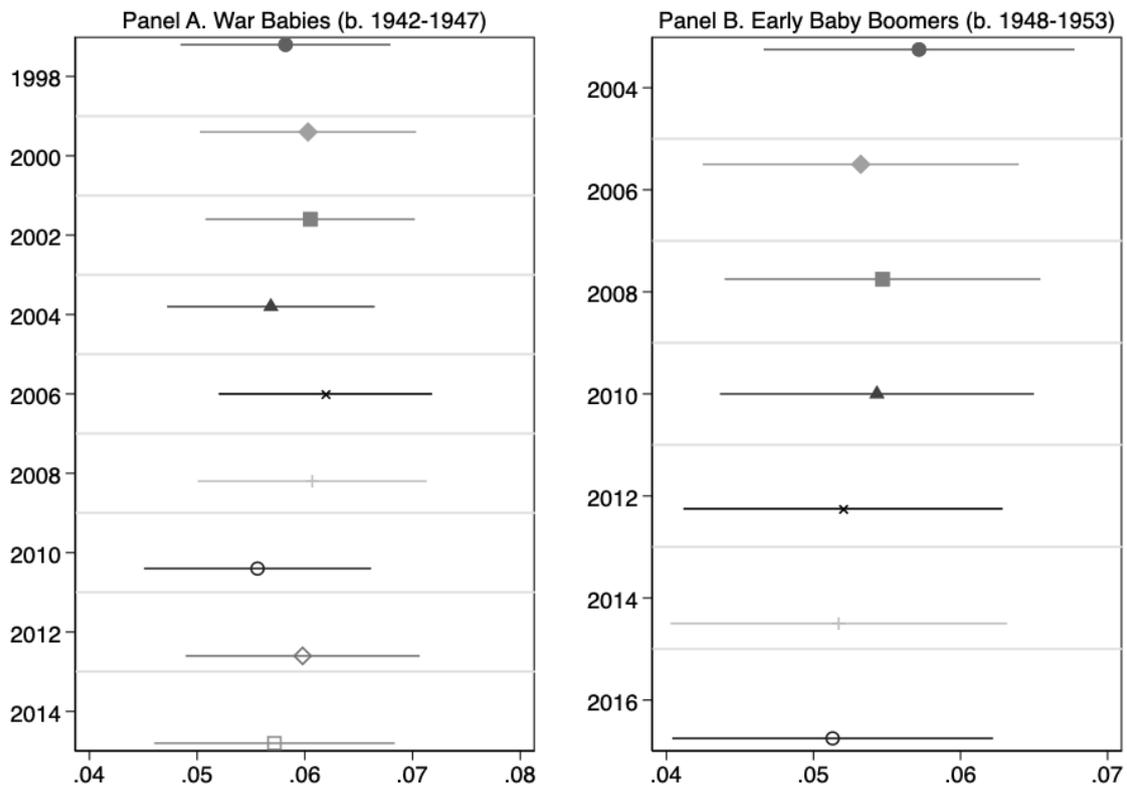
Note: Results based on our benchmark balanced panel sample of HRS Original cohort members described in Table 1 of the main text. The dependent variable is BMI. The figure shows unconditional quantile regression coefficient estimates of *BMIPGS* (normalized to have mean 0 and standard deviation 1) in equation 2, and their associated 95% confidence intervals. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data as controls. Standard errors are clustered at the household level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Figure A.2: Association between BMI Polygenic Scores and BMI Along the Life-Cycle and Along the BMI Distribution. Add Health Cohort



Note: Results based on our benchmark balanced panel sample of HRS Original cohort members described in Table 2 of the main text. The dependent variable is BMI. The figure shows unconditional quantile regression coefficient estimates of *BMIPGS* (normalized to have mean 0 and standard deviation 1) in equation 2, and their associated 95% confidence intervals. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data as controls. Standard errors are clustered at the school level. Longitudinal weights are used. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Figure A.3: Association between BMI Polygenic Scores and log(BMI) Along the Life-Cycle.  
HRS War Babies and Early Baby Boomers Cohorts



Note: This Figure summarizes the results of estimating equation 2 on two balanced samples of 1,201 HRS War Babies cohort members (Panel A) and 1,191 Early Baby Boomers cohort members (Panel B). The dependent variable is Log(BMI). OLS coefficient estimates of  $\beta_1$  as well as their associated 95% confidence intervals are depicted. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. Standard errors are clustered at the household level.

Table A.1: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle.  
HRS Original Cohort

	(1)	(2)	(3)	(4)	(5)
	All	Low SES	High SES	Males	Females
1992: Age 55.9	0.041*** (0.003)	0.039*** (0.004)	0.043*** (0.004)	0.043*** (0.003)	0.039*** (0.004)
1994: Age 57.8	0.040*** (0.003)	0.038*** (0.004)	0.040*** (0.004)	0.041*** (0.003)	0.038*** (0.004)
1996: Age 59.8	0.042*** (0.003)	0.041*** (0.004)	0.043*** (0.004)	0.041*** (0.004)	0.043*** (0.004)
1998: Age 61.6	0.041*** (0.003)	0.039*** (0.004)	0.043*** (0.004)	0.040*** (0.004)	0.041*** (0.004)
2000: Age 63.6	0.041*** (0.003)	0.041*** (0.004)	0.041*** (0.004)	0.042*** (0.004)	0.041*** (0.004)
2002: Age 65.7	0.041*** (0.003)	0.041*** (0.004)	0.041*** (0.004)	0.041*** (0.004)	0.041*** (0.004)
2004: Age 67.8	0.043*** (0.003)	0.042*** (0.004)	0.043*** (0.004)	0.042*** (0.004)	0.043*** (0.005)
2006: Age 69.7	0.042*** (0.003)	0.040*** (0.004)	0.044*** (0.004)	0.041*** (0.004)	0.043*** (0.004)
2008: Age 71.7	0.042*** (0.003)	0.039*** (0.004)	0.044*** (0.004)	0.042*** (0.004)	0.041*** (0.005)
Observations	3,181	1,567	1,614	1,423	1,758

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of *BMIPGS* (normalized to have mean 0 and standard deviation 1) in equation 2. The benchmark sample used for these analyses is described in Table 1 of the main text. All regressions include age, age squared, and the first 10 principal components of the full matrix of genetic data as controls. In Columns 1-3 a female dummy is also included as a regressor. The samples used in Columns 2 and 3 include individuals whose parental socioeconomic status is below and above the median, respectively. Standard errors (in parentheses) are clustered at the household level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.2: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle.  
Add Health Cohort

	(1)	(2)	(3)	(4)	(5)
	All	Low SES	High SES	Males	Females
1994/95: Age 15.4	0.042*** (0.005)	0.054*** (0.006)	0.026*** (0.006)	0.041*** (0.006)	0.044*** (0.006)
1996: Age 16.3	0.046*** (0.006)	0.062*** (0.007)	0.023*** (0.007)	0.045*** (0.007)	0.046*** (0.007)
2001/02: Age 21.7	0.049*** (0.006)	0.063*** (0.008)	0.030*** (0.007)	0.047*** (0.007)	0.052*** (0.009)
2008/09: Age 28.3	0.057*** (0.006)	0.073*** (0.007)	0.033*** (0.009)	0.051*** (0.007)	0.063*** (0.010)
2016/18: Age 37.3	0.058*** (0.006)	0.068*** (0.007)	0.039*** (0.009)	0.057*** (0.007)	0.056*** (0.009)
Observations	2,730	1,361	1,369	1,204	1,526

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of *BMIPGS* (normalized to have mean 0 and standard deviation 1) in equation 2. The benchmark sample used for these analyses is described in Table 2 of the main text. All regressions include age, age squared, and the first 10 principal components of the full matrix of genetic data as controls. In Columns 1-3 a female dummy is also included as a regressor. The samples used in Columns 2 and 3 include individuals whose parental socioeconomic status is below and above the median, respectively. Standard errors (in parentheses) are clustered at the school level. Longitudinal weights are used. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.3: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle and Pubertal Stage. Add Health Cohort

	(1) Benchmark	(2) With Controls for Pubertal Stage
1994/95: Age 15.4	0.042*** (0.005)	0.039*** (0.004)
1996: Age 16.3	0.046*** (0.006)	0.038*** (0.006)
2001/02: Age 21.7	0.049*** (0.006)	0.049*** (0.006)
2008/09: Age 28.3	0.057*** (0.006)	0.057*** (0.006)
2016/18: Age 37.3	0.058*** (0.006)	0.058*** (0.006)
Observations	2,730	2,730

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of *BMIPGS* (normalized to have mean 0 and standard deviation 1) in equation 2. All specifications include the following covariates: a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. The specifications for Waves I (1994/95) and II (1996) in Column (2) also include gender and wave specific controls for pubertal stage. Standard errors (in parentheses) are clustered at the school level. Longitudinal weights are used. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.4: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle and Pubertal Timing. Add Health Cohort

	(1)	(2)
	Benchmark	Puberty Onset Timing Among Controls
1994/95: Age 15.4	0.042*** (0.005)	0.041*** (0.005)
1996: Age 16.3	0.046*** (0.006)	0.044*** (0.006)
2001/02: Age 21.7	0.049*** (0.006)	0.048*** (0.006)
2008/09: Age 28.3	0.057*** (0.006)	0.055*** (0.006)
2016/18: Age 37.3	0.058*** (0.006)	0.056*** (0.006)
Observations	2,730	2,708

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of BMIPGS (normalized to have mean 0 and standard deviation 1) in equation 2. All specifications include the following covariates: a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. An indicator for early vs. delayed puberty onset is added in Column 2. Standard errors (in parentheses) are clustered at the school level. Longitudinal weights are used. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.5: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle and the Prevalence of Chronic Diseases. HRS Original Cohort

	(1) Benchmark	(2) With Controls for Chronic Diseases
1992: Age 55.9	0.041*** (0.003)	0.039*** (0.003)
1994: Age 57.8	0.040*** (0.003)	0.038*** (0.003)
1996: Age 59.8	0.042*** (0.003)	0.039*** (0.003)
1998: Age 61.6	0.041*** (0.003)	0.038*** (0.003)
2000: Age 63.6	0.041*** (0.003)	0.038*** (0.003)
2002: Age 65.7	0.041*** (0.003)	0.038*** (0.003)
2004: Age 67.8	0.043*** (0.003)	0.039*** (0.003)
2006: Age 69.7	0.042*** (0.003)	0.037*** (0.003)
2008: Age 71.7	0.042*** (0.003)	0.036*** (0.003)
Observations	3,181	3,161

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of BMIPGS (normalized to have mean 0 and standard deviation 1) in equation 2. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. The specification in Column (2) adds period specific indicators for the prevalence of the following diseases: cancer, lung disease, heart disease, diabetes, and arthritis. Standard errors (in parentheses) are clustered at the household. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.6: Association between BMI Polygenic Scores and Obesity Along the Life-Cycle. HRS Original Cohort. Correction for Mortality Selection Using Inverse Probability Weights

	(1)	(2)	(3)	(4)	(5)
	All	Low SES	High SES	Males	Females
1992: Age 55.9	0.041*** (0.003)	0.040*** (0.004)	0.043*** (0.004)	0.043*** (0.003)	0.040*** (0.004)
1994: Age 57.8	0.040*** (0.003)	0.039*** (0.004)	0.040*** (0.004)	0.041*** (0.003)	0.039*** (0.004)
1996: Age 59.8	0.043*** (0.003)	0.042*** (0.004)	0.043*** (0.004)	0.041*** (0.004)	0.044*** (0.004)
1998: Age 61.6	0.041*** (0.003)	0.040*** (0.004)	0.042*** (0.004)	0.040*** (0.004)	0.042*** (0.004)
2000: Age 63.6	0.042*** (0.003)	0.042*** (0.004)	0.041*** (0.004)	0.042*** (0.004)	0.042*** (0.004)
2002: Age 65.7	0.041*** (0.003)	0.042*** (0.004)	0.041*** (0.004)	0.040*** (0.004)	0.042*** (0.004)
2004: Age 67.8	0.043*** (0.003)	0.043*** (0.004)	0.043*** (0.004)	0.042*** (0.004)	0.044*** (0.005)
2006: Age 69.7	0.042*** (0.003)	0.041*** (0.004)	0.044*** (0.004)	0.040*** (0.004)	0.044*** (0.004)
2008: Age 71.7	0.042*** (0.003)	0.040*** (0.005)	0.044*** (0.004)	0.040*** (0.004)	0.043*** (0.005)
Observations	3,181	1,567	1,614	1,423	1,758

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of *BMIPGS* (normalized to have mean 0 and standard deviation 1) in equation 2. The benchmark sample used for these analyses is described in Table 1 of the main text. All regressions include age, age squared, and the first 10 principal components of the full matrix of genetic data as controls. In Columns 1-3 a female dummy is also included as a regressor. The samples used in Columns 2 and 3 include individuals whose parental socioeconomic status is below and above the median, respectively. In these estimations we have corrected for selective mortality using inverse probability weights as described in Section 5.1. Standard errors (in parentheses) are clustered at the household level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.7: Association between BMI Polygenic Scores and Log(BMI) Along the Life-Cycle.  
 BMI based on Objective Measurements vs. Self-Reports

	(1) Objective	(2) Self-Reported
Panel A: HRS Original Cohort		
2006: Age 69.7	0.044*** (0.005)	0.042*** (0.003)
2008: Age 71.7	0.044*** (0.005)	0.042*** (0.003)
Panel B: Add Health Cohort		
1996: Age 16.3	0.046*** (0.005)	0.046*** (0.006)
2001/02: Age 21.7	0.053*** (0.006)	0.049*** (0.006)
2008/09: Age 28.3	0.059*** (0.006)	0.057*** (0.006)
2016/18: Age 37.3	0.069*** (0.009)	0.058*** (0.006)

Note: The dependent variables are Log(BMI) based on objective measurements (Column 1) and self reports (Column 2), respectively. The Table displays OLS coefficient estimates of BMIPGS (normalized to have mean 0 and standard deviation 1) in equation 2. All regressions include a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. Standard errors (in parentheses) are clustered at the household (Panel A) and school (Panel B) level, respectively. Longitudinal weights are used in Panel A. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table A.8: Association between BMI Polygenic Scores, Childhood Socioeconomic Status, and Log(BMI) Along the Life-Cycle. Add Health Cohort

	(1)	(2)	(3)
	With SES Included as Control		Benchmark
	BMI PGS	SES	BMI PGS
1994/95: Age 15.4	0.041*** (0.005)	-0.017*** (0.006)	0.042*** (0.005)
1996: Age 16.3	0.044*** (0.005)	-0.019** (0.007)	0.046*** (0.006)
2001/02: Age 21.7	0.048*** (0.006)	-0.020*** (0.007)	0.049*** (0.006)
2008/09: Age 28.3	0.055*** (0.006)	-0.027*** (0.008)	0.057*** (0.006)
2016/18: Age 37.3	0.055*** (0.006)	-0.034*** (0.007)	0.058*** (0.006)
Observations		2,730	2,730

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of BMIPGS and childhood SES (both normalized to have mean 0 and standard deviation 1). All specifications include the following covariates: a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. The specification used in Columns 1 and 2 adds childhood SES as an additional covariate. Standard errors (in parentheses) are clustered at the school level. Longitudinal weights are used. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A.9: Association between BMI Polygenic Scores, Childhood Socioeconomic Status, and Log(BMI) Along the Life-Cycle. HRS Original Cohort

	(1)	(2)	(3)
	With SES Included as Control		Benchmark
	BMI PGS	SES	BMI PGS
1992: Age 55.9	0.041*** (0.003)	-0.010* (0.004)	0.041*** (0.003)
1994: Age 57.8	0.039*** (0.003)	-0.010* (0.004)	0.040*** (0.003)
1996: Age 59.8	0.042*** (0.003)	-0.012** (0.004)	0.042*** (0.003)
1998: Age 61.6	0.041*** (0.003)	-0.010* (0.005)	0.041*** (0.003)
2000: Age 63.6	0.041*** (0.003)	-0.013** (0.005)	0.041*** (0.003)
2002: Age 65.7	0.041*** (0.003)	-0.012** (0.005)	0.041*** (0.003)
2004: Age 67.8	0.042*** (0.003)	-0.012** (0.005)	0.043*** (0.003)
2006: Age 69.7	0.042*** (0.003)	-0.010* (0.005)	0.042*** (0.003)
2008: Age 71.7	0.041*** (0.003)	-0.011* (0.005)	0.042*** (0.003)
Observations		3,181	3,181

Note: The dependent variable is Log(BMI). The Table displays OLS coefficient estimates of BMIPGS and childhood SES (both normalized to have mean 0 and standard deviation 1). All specifications include the following covariates: a female dummy, age, age squared, and the first 10 principal components of the full matrix of genetic data. The specification used in Columns 1 and 2 adds childhood SES as an additional covariate. Standard errors (in parentheses) are clustered at the household level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## B Childhood Socioeconomic Status Indices Construction

### B.1 HRS

The HRS parental socioeconomic status index is based on several retrospective questions asked to survey participants that allow one to create the following variables:

1. Paternal education: years of schooling completed by respondents' fathers.
2. Family well off: an indicator variable identifying respondents who declare that their families were pretty well off or about average financially from birth to age 16.
3. Family never moved or asked for help for financial reasons: an indicator variable identifying respondents whose families ever had to move residence, or who never had to ask relatives for help due to financial reasons.
4. Father's employment: an indicator variable identifying respondents whose father never spent several months or more unemployed.

Next, we compute a summary index as in [Kling et al. \(2007\)](#) that is equal to the unweighted average of the previous standardized variables:

$$SES^* = \frac{\sum_k SES_k^*}{K}, \text{ where } SES_k^* = \frac{SES_k - \mu_k}{\sigma_k}$$

, where  $SES_k$  is the  $k$ th component of the index,  $\mu_k$  denotes its mean and  $\sigma_k$  its standard deviation. Since all the components are associated with higher socioeconomic status, higher values of the summary index  $SES^*$  are associated with higher socioeconomic status.

### B.2 Add Health

To measure childhood SES in Add Health we construct an index based on parental education, parental occupation prestige, household income, and household receipt of public assistance following [Belsky et al. \(2018\)](#) and [Sanz-de Galdeano and Terskaya \(2019\)](#). We have constructed an occupational prestige indicator using occupational prestige scores from the National Opinion Research Center (NORC) occupational classification. See <http://ibgwww.colorado.edu/~agross/NNSD/prestige%20scores.html>. The information we use was gathered at Wave I. Some of the variables used are based on questions included in the parental questionnaire (household income), while others rely on adolescents' responses (parental occupation and household receipt of public assistance). In some cases (parental education) we have complemented information based on questions addressed to parents with information based on adolescents' answers.

Finally, we conducted principal components analysis of parental education, parental occupational attainment, family income, and household receipt of public assistance to produce a factor score. The first principal component explained 49.2% of the variance. We used loadings on this component to compute a SES index, and then we standardized it to have mean 0 and standard deviation 1.