

SUPPLEMENTAL MATERIAL

Gene-Environment interactions and the case of BMI and obesity: how much do they matter?

Yiyue Huangfu¹

Alberto Palloni^{1,2}

Hiram Beltrán-Sánchez³

Mary McEniry¹

¹ Center for Demography and Ecology, Center for Demography and health of Aging, University of Wisconsin-Madison

² Consejo Superior de Investigaciones Científicas (CSIC), Spain.

³ California Center for Population Research, University of California, Los Angeles

Funding: This work was supported by the National Institute on Aging (<https://www.nia.nih.gov/>), National Institute of Child Health and Development (<https://www.nichd.nih.gov/>), Fogarty International Center Global Research Training in Population and Health (<https://www.fic.nih.gov/>) and European Research Council (<https://erc.europa.eu/>) via the following project grants R01-AG016209 (AP), R03-AG015673(AP), R01-AG018016(AP), R37-AG025216 (AP), RO1-AG056608 (AP;HBS), RO1-AG052030(AP;HBS); D43-TW001586(AP), R24-HD047873(AP), P30-AG-017266(AP), R24 HD-041022(UCLA:HBS); European Union Horizon 2020 Research and Innovation Programme, Project No 788582(AP)

I. Basic models

Below we describe the data and models used to derive empirical estimates of relevant parameters.

1. Data

We use data from the Health and Retirement Survey (HRS), a nationally representative longitudinal survey of more than 37,000 individuals over age 50 in the United States. We focus on a subsample of 9,331 white individuals who provided saliva samples in the 2006 wave and for whom PGS for a selected traits could be estimated. BMI is computed with self-reports of height and weight. Genetic propensity is measured by a PGS (z-score) for BMI constructed from a total of 97 identified and imputed SNP's and weights from the GIANT GWAS (The LifeLines Cohort Study et al. 2015)

We define cohort membership by grouping individuals into the following birth year categories: before 1924, 1925-1934, 1935-1944 and 1945-1959 using three dummy variables with the oldest (born before 1925) as the excluded category¹. The model controls for principal components to neutralize stratification or population composition effects on the PGS, gender, education, and an indicator of adverse early conditions.

2. Empirical estimates of GxE effects for BMI

We estimate a growth curve model:

$$BMI_{it} = \alpha_{0i} + \alpha_1 cohort_i + \alpha_2 PGS_i + \alpha_3 cohort_i * PGS_i + \alpha_{4i} age_{it} + \alpha_{5i} age_{it}^2 + \alpha'_6 X_i + \mu_i + \varepsilon_{it}$$

where BMI_{it} is a respondent i 's BMI at year t , $cohort_i$ and PGS_i are a respondent's birth cohort and PGS for BMI, age_{it} is respondent i 's median-centered age at year t . The coefficients on age_{ij} and its squared term, β_{i4} and β_{5i} , varies across individuals. X_i is a vector of control variables including education, gender, principal components, and an indicator of adverse early conditions. μ_i is the random intercept for each respondent, and ε_{it} is the error at the individual-age level.

Descriptive statistics of the sample, alternative estimates, and sensitivity tests are in supplemental material. Supplementary Table 1 displays the descriptive statistics of the analytic sample. Column 1 in Supplementary Table 2 displays estimates of coefficients. The additive

¹ Defining the youngest group as the left out category yields the same results

genetic effect implies that a change in one standard deviation of PGS leads to a change of about 1.3 BMI units.² As expected, the additive effects of birth cohorts are large, monotonically increasing, in the expected direction (younger cohorts have higher BMI), and statistically significant. The youngest cohorts (born between 1945 and 1959) have a BMI about 3.7 units higher or 25 lbs heavier than those in the oldest cohorts (born before 1924). The contrasts between the two most recent birth cohort are milder, equivalent to a one unit of BMI or 6.4 lbs. Finally, the only statistically significant G x E we identify is associated with the youngest birth cohort. Individuals in these youngest cohorts who are 1 standard deviation above the mean of the PGS have .48 BMI units more (3.3 lbs heavier) than genetically similar individuals in the oldest cohort. These results confirm recent findings suggesting that more obesogenic environments after 1945 may have enhanced the penetrance of allelic variants associated with obesity.

Supplementary Figure 2a plots predicted values of BMI as a function of PGS by cohort for whites only. The linear predictor of the fixed portion of the GCM shows that increases in PGS are associated with unit increases in the BMI for the entire range of observed values and that these increases are significantly larger for the youngest birth cohorts. Supplement Figure 2b plots the age trajectories of the fitted BMI values for whites by cohort. It shows the expected parabolic form and the upward shifts in younger cohorts.³

3. Empirical estimates of GxE for obesity

Because efforts to estimate a HLM for obesity comparable to that for BMI, we use two alternative approaches.

3.1. Estimation of linear probability model

We estimate a random effects linear probability model:

$$obesity_{it} = \alpha_0 + \alpha_1 cohort_i + \alpha_2 PGS_i + \alpha_3 cohort_i * PGS_i + \alpha_4 age_{it} + \alpha_5 age_{it}^2 + \alpha_6' \mathbf{X}_i + \mu_i + \varepsilon_{it}$$

where $obesity_{it}$ is a 0/1 variable for the i th respondent's obesity status ($BMI \geq 30$) at year t , $cohort_i$ and PGS_i are a respondent's birth cohort and PGS for BMI, age_{it} is respondent i 's median-centered age at year t . \mathbf{X}_i is a vector of control variables including education, gender,

² Setting height to be about 1.75 mts and 1.63 mts for males and females respectively, a 1 unit crease in BMI represents an increase of about 3kgs (6.4lbs) among males and 2.8kgs (6.2lbs) among females.

³ The results we show are based on models that use self-reported height and weight. In models estimated with measured height and weight sample sizes are much reduced, we lose power and the only statistically significant effects are those for additive PGS and for birth cohorts.

principal components, and an indicator of adverse early conditions. μ_i is the random intercept for each respondent, and ε_{it} is the error at the individual-age level

Column 2 in Supplementary Table 2 displays the estimated parameters. The additive genetic effect implies that a change in one standard deviation of PRS increases the probability of being obese by about 8 %. As expected, the additive effects of birth cohorts are quite large, monotonically increasing, in the expected direction (younger cohorts have higher probability of being obese), and statistically significant. The youngest cohorts (born between 1945 and 1959) have a probability of being obese 28% higher than those in the oldest cohorts (born before 1924). As in the case of BMI, the only statistically significant GxE we identify is associated with the youngest birth cohort. Individuals in these youngest cohorts who are one standard deviation above the mean of the PGS have obesity prevalence 4.6 % higher than genetically similar individuals in the oldest cohort.

3.2. Estimation of GxE effects consistent with those for BMI

Estimates of parameters in the linear probability model generate predicted values of the probability of obesity that do use BMI information only indirectly and they may not be consistent with each other. To impose consistence, we first compute the product of the z-score of each decile and the estimate of GxE effect in the model for BMI. We then add this quantity to the BMI of everyone in the youngest cohort in the decile and compare the fraction obese before and after the addition of the addition. The difference is a raw, decile-specific, estimate of the impact of GxE on obesity that is fully consistent with the estimated GxE effect in the BMI model. In a second step, we compute the sum of the absolute values of these differences across all deciles and divide it by 3.4, the number of standard deviations units spanned by the deciles. The resulting quantity is a global estimate of the GxE effect of cohort and PRS consistent with the GxE effects estimates using a model for BMI.

3.3. Comparison between the two procedures

In the linear probability model, the estimated GxE effects is .046. Using the second procedure we get .025 approximately. These two estimates are close but, to err on the conservative side, we choose the largest of these for all our computations. The fact that the two estimates are quite close enables us to resolve an additional problem. This is that while we were able to find alternative estimates of GxE involving BMI and birth cohort, we failed to find similar ones using obesity and cohort. To employ a range of estimates comparable to those for BMI, we computed

max and min so that they span the same fractional distance from the central estimate, .046, as do those for BMI. The max is set to be .10 or 2.4 times as large as the center of the range, the same as observed in the case of BMI. We set the minimum to about .03 or .65 of the range's central value. Again, this is erring on the conservative side as the same ration in the case of alternative estimates for BMI is .013. In any case, the min value is immaterial for our purposes and the most important alternative estimate for inferences in the max value. Strictly speaking, the lower bound should be 0.

Supplementary Figure 1 plots changes in fraction of obese using six alternative estimates: three grounded in the first procedure (linear probability model) and the construction of a range of values and three based on the second procedure described above. Examining each of the pairs (HRS, max and min) we conclude that differences are minor.

II. Computation of second order effects

In this section we describe in a step-by-step fashion the calculations estimate second order effects for GxE effects of obesity on T2D, disability and mortality. With a few modifications, the approach we followed to do computations using BMI instead of obesity, are the same.

To assess impacts of GxE effects on the probabilities of T2D and disability, we use the HRS sample and estimate separate hazard models for each outcome. WE estimate two versions of these models, one including BMI as predictor and a second one including the logit of the probability of obesity. These models are of the following form:

$$\mu_{ik}(t; BMI_{it}, Z_i) = \mu_{ok}(t) \exp(\gamma BMI_{it} + \tau Z_i)$$

and

$$\mu_{ik}(t; LO_{it}, Z_i) = \mu_{ok}(t) \exp(\kappa LO_{it} + \tau Z_i)$$

where $\mu_{ik}(t; BMI_{it}, Z_i)$ and $\mu_{ik}(t; LO_{it}, Z_i)$ are hazards for outcome $k = 1$ (T2D) and $k = 2$ (disability) for the i th individual as functions of BMI_{it} , or of the logit of the probability of obesity, LO_{it} . \mathbf{Z} is a vector of controls including age at baseline survey, t is duration of follow-up, and μ_{ok} is a baseline hazard (Gompertz for $k=2$ and log-logistic for $k=1$).

Parameter estimates of the hazard model for mortality, T2D and disability are in Supplementary Tables 3-5.

III. Alternative estimates of GxE effects

To get a sense of the range of the impact of GxE effects when using different data and model specifications, we conducted a literature search to identify alternative empirical estimates of GxE effects and BMI. We used 3 selection criteria: (1) polygenic scores are used as a measure of genetic propensity; (2) birth cohort is used as an indicator for environment; (3) BMI is the phenotype. We identified 4 estimates of GxE effects that met these criteria and Supplementary Table 6 shows relevant information. Assessment of first and second order effects in this paper are based on our own HRS based estimates and these four alternative ones. In all cases, we focus on the effect of GxE for cohort born after 1945 relative to cohort born before 1924.

As mentioned before, our search for alternative estimates for obesity that are analogous to those found for BMI was not successful. In its place, we resorted to a roundabout procedure described in section II above.

IV. Model for intergenerational transmission of BMI and obesity risks

We propose a simple model to assess potentially intergenerational transmission of obesity risks. We invoke four assumptions:

- i. Parental effects are equivalent to sociocultural heritability, as defined previously, but exclude mechanisms of ideational influence that emanate from outside the individuals' household. In addition, we assume parental effects are independent of other environments and uncorrelated with genetic BMI-obesity propensities.
- ii. Any environmental change that takes place in one generation is preserved in the next generation. An obesogenic environment established at time t does not disappear overnight or is cancelled before the next generation is influenced by it.
- iii. All effects are invariant across generations
- iv. Offspring' phenotypes and genotypes are averages of parental phenotypes and genotypes.

Let O_t and O_{t+1} be the probabilities of becoming obese in generation t and $t+1$. O_t can be expressed as

$$\begin{aligned} O_t &= \alpha_1 E_t + (\alpha_2 + \alpha_3 E_t) * G_t + \theta O_{t-1} + \varepsilon_t \\ &= O_t^b + \theta O_{t-1} + \varepsilon_t \end{aligned} \quad (1)$$

where E_t and G_t stand for environment and genetic propensity in generation t, α' s are effects, $0 \leq \theta \leq 1$ is a coefficient of sociocultural heritability⁴, ε_t is Gaussian error, and O_t^b is the local, current contribution of E and G. The offspring generation probability of obesity is:

$$O_{t+1} = O_t^b(1 + \theta) + \theta^2 O_{t-1} + \varepsilon_{t+1} \quad (2)$$

an equality holds under the assumption that $G_t \approx G_{t+1}$.

In the nth generation we have

$$O_{t+n} = O_t^b(1 + \theta + \theta^2 \dots) + \theta^n O_{t-1} + \varepsilon_{t+n} \quad (3)$$

After a change in E equivalent to ΔE_t in the parental generation the probability of obesity is:

$$O_t^* = O_b + \Delta E_t * (\alpha_1 + \alpha_3 * G_t) + \theta O_{t-1} + \varepsilon_t$$

$$O_t^* = O_t^{b*} + \theta O_{t-1} + \varepsilon_t$$

with $O_t^{b*} = O_t^b + \Delta O_t^b$. The quantity $\Delta O_t^b = \Delta E_t * (\alpha_1 + \alpha_3 * G_t)$ is the *excess* probability of obesity in generation t due to the change in E. Note that ΔO_t^b is composed of one part associated with the direct impact of the change in E, $\Delta E_t * \alpha_1$, and one associated with an increase in the probability of obesity induced by the interaction of environmental change and genotype, $\alpha_3 * \Delta E_t * G_t$.

After the environmental change, the offspring generation experiences a probability of obesity given by

$$\begin{aligned} O_{t+1}^* &= O_t^{b*} + \theta O_t^* + \varepsilon_{t+1} \\ &= (O_t^b + \Delta O_t^b) * (1 + \theta) + \theta^2 O_{t-1} + \varepsilon_{t+1} \end{aligned}$$

In the nth generation we have

$$O_{t+n}^* = (O_t^b + \Delta O_t^b) * (1 + \theta + \theta^2 \dots) + \theta^n O_{t-1} + \varepsilon_{t+n}$$

Three important quantities follow:

i. Offspring excess probability of obesity after the environmental change:

$$\delta_{t+1} = O_{t+1}^* - O_{t+1} = (1 + \theta) * \Delta O_t^b$$

The first component of this is ΔO_t^b or the excess obesity in offspring as a direct result of the GxE effect that applies to them as much as to their parents. The quantity $\Delta O_t^b * \theta$ is the secondary increase that results from sociocultural heritability of the parental excess due to the GxE effect

And after n generations (n large):

⁴ The slope of a regression line of offspring obesity risk on the parental obesity risk

$$\delta_{t+n} = O_{t+n}^* - O_{t+n} \sim (1/(1 - \theta)) * \Delta O_t^b$$

ii. Differences between offspring and parental generation before and after the environmental change

$$(O_{t+1}^* - O_t^*) - (O_{t+1} - O_t) = \theta * O_t^b$$

and after n generations.

$$\sim \left(\frac{\theta^2}{1 - \theta}\right) * O_t^b$$

iii. Finally, in each case, the fraction of differences in probabilities of obesity induced by the change in E that is attributable to the GxE interaction effect is the quantity

$$\gamma = (\alpha_3 * G_t) / (\alpha_1 + \alpha_3 * G_t)$$

The quantity δ_{t+1} is an estimate of the increase in the probability of obesity in the offspring generation due to the GxE. There are two contributors only: the persistent effects of E and GxE includes in ΔO_b (a function of parameters α_1 and α_3) and the strength of intergenerational heritability, θ . Having estimates of these parameters is all one needs to address the question of the magnitude of the impact of the original GxE on the probability of obesity in successive generations.

IV. Estimates of θ

To estimate the impact of GxE on the offspring generation following the simple model outlined above, we searched for estimates of θ in studies correlating parental and child BMI/obesity. We focus on studies that explicitly show the correlation between parental and child BMI/obesity, or the regression coefficients of offspring on parental phenotype. Supplementary Table 7 displays the estimates. We chose two values of θ , 0.2 and 0.37, as the minimal and maximal correlation between parental and child BMI/obesity in our analysis.

Supplementary Table 1. Sample Characteristics by Cohort, white HRS respondents.

Cohort	before 1924	1925-1934	1935-1944	1945-1960
BMI ^a	25.76	26.68	26.80	28.17
(SD)	(3.93)	(4.40)	(4.82)	(5.86)
No. of BMI report	9.44	10.45	11.38	6.44
(SD)	(1.89)	(2.35)	(2.18)	(2.89)
PGS-BMI	-0.07	0.00	0.00	0.00
(SD)	(1.00)	(0.98)	(0.99)	(1.00)
Age	74.46	64.39	54.07	52.00
(SD)	(3.58)	(5.38)	(3.43)	(4.14)
% male	0.40	0.46	0.44	0.44
Education (%)				
< HS	0.19	0.15	0.11	0.04
HS	0.36	0.39	0.40	0.30
>HS	0.44	0.46	0.49	0.66
Early life condition ^b				
Adverse	0.83	0.56	0.48	0.28
Median	0.12	0.19	0.21	0.20
Good	0.05	0.25	0.31	0.52
N	881	2,058	2,866	3,526

Notes:

a. BMI calculated with self-reported weight and height;

b. tertiles of an early life condition index summarizing family financial situation, health, father's occupation, parental education, mother's effort and attention, relationship with mother, stressful events during childhood.

Supplementary Table 2. Impact of GxE on BMI and Obesity, white HRS respondents. Growth curve model for BMI; Linear mixed effect model for obesity.

VARIABLES	BMI		Obesity	
	b	(SE)	b	(SE)
BMI PGS	1.32***	(0.18)	0.08***	(0.01)
1925-1934	1.59***	(0.35)	0.10***	(0.03)
1935-1944	2.55***	(0.34)	0.19***	(0.03)
1945-1959	3.67***	(0.34)	0.28***	(0.03)
Age (median centered)	0.09***	(0.02)	0.00	(0.00)
Age (median centered) square	-0.01***	(0.00)	-0.00	(0.00)
Median early life condition	-0.12	(0.14)	-0.01	(0.01)
Good early life condition	-0.43**	(0.13)	-0.03***	(0.01)
1925-1934 X BMI PGS	-0.14	(0.21)	0.01	(0.02)
1935-1944 X BMI PGS	0.15	(0.20)	0.02	(0.02)
1945-1959 X BMI PGS	0.48*	(0.19)	0.05***	(0.02)
1925-1934 X Age	-0.02	(0.02)	0.01**	(0.00)
1935-1944 X Age	-0.05*	(0.02)	0.00	(0.00)
1945-1959 X Age	-0.10***	(0.02)	0.00	(0.00)
1925-1934 X Age square	-0.00	(0.00)	-0.00***	(0.00)
1935-1944 X Age square	0.00	(0.00)	-0.00***	(0.00)
1945-1959 X Age square	0.00	(0.00)	-0.00**	(0.00)
BMI PGS X Age	-0.00	(0.00)	-0.00	(0.00)
Middle school	-0.11	(0.19)	-0.01	(0.02)
High school or higher	-0.72***	(0.19)	-0.04***	(0.02)
Female	-0.90*	(0.36)	0.02	(0.03)
Female X age	0.02**	(0.00)	0.00	(0.00)
Cohort 1925-1934 X Female	-0.18	(0.42)	-0.04	(0.03)
Cohort 1935-1944 X Female	0.16	(0.40)	-0.04	-0.03
Cohort 1945-1959 X female	0.57	(0.40)	-0.01	-0.03
Principle components	yes		yes	
Constant	26.43***	(0.34)	-5.60***	(0.49)
Random effects				
Age	0.03	(0.00)	0.00	(6.11e-06)
Age square	0.00008	(2.88e-06)	4.21E-07	(2.71e-08)
Intercept	23.70	(0.37)	0.14	(0.00)
Observations			85,136	
Number of individuals			9,331	

Standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Supplementary Table 3. Mortality hazard as a function of ADL and T2D

VARIABLES	Hazard of mortality	
	b	(SE)
Age	0.11***	(0.00)
Female	0.44***	(0.04)
Less than high school (ref. high school)	0.05	(0.05)
Above high school (ref. high school)	-0.15***	(0.04)
ADL	0.38***	(0.01)
T2D	0.43***	(0.04)
Constant	-12.57***	(0.18)
gamma	0.11***	(0.00)
Observations	85,587	
Number of individuals	8,753	
Number of failures	2,672	

Standard errors in parentheses

*** p<0.001, ** p<0.05, * p<0.1

Supplementary Table 4. T2D hazard as a function of obesity

VARIABLES	Hazard of T2D	
	b	(SE)
Age	0.01***	(0.00)
Female	-0.33***	(0.05)
Less than high school (ref. high school)	-0.25***	(0.07)
Above high school (ref. high school)	0.19***	(0.05)
Logit of obesity	0.05***	(0.01)
Constant	3.70***	(0.17)
Gamma	0.86***	(0.02)
Observations	62,022	
Number of individuals	6,940	
Number of failures	1,991	

Standard errors in parentheses

*** p<0.001, ** p<0.05, * p<0.1

Supplementary Table 5. Disability hazard as a function of obesity

VARIABLES	Hazard of T2D	
	b	(SE)
Age	0.03***	(0.00)
Female	-0.02	0.04)
Less than high school (ref. high school)	0.36***	(0.05)
Above high school (ref. high school)	-0.16***	(0.04)
Logit of obesity	0.06*	(0.03)
Constant	-5.67***	(0.12)
Gamma	0.06***	(0.00)
Observations	62,022	
Number of individuals	6,940	
Number of failures	1,991	

Standard errors in parentheses

*** p<0.001, ** p<0.05, * p<0.1

Supplementary Table 6. Alternative Estimates of GxE on BMI

	Polygenic Score for BMI	Environment	Effect size of GxE ^a
Demereth et al, (2013)	PGS is constructed using 32 SNPs from Speliotes et al., (2010).	Linear specification of birth year.	0.06 ^b
Liu and Guo, (2015)	PGS is constructed using 32 SNPs from Speliotes et al., (2010)	Linear specification of cohort (1-5). AHEAD: before 1924 = 1; CODA: 1924-1930 = 2; HRS: 1931-1941 = 3; WB: 1942-1947 = 4; EBB: 1948-1953 = 5.	0.155
Walter et al., (2016)	PGS is constructed using 32 SNPs from Speliotes et al., (2010)	Dummy specification of cohort. Cohort 1: before 1924; Cohort 2: 1924-1933; Cohort 3: 1934-1943; Cohort 4: 1944-1958.	1.196
Conley et al, (2016)	PGS is constructed from multiple GWAS results.	Linear specification of birth year.	0.18 ^c

Notes.

a. Effect of GxE for cohort born after 1945 relative to cohort born before 1924, as shown or implied by the estimates.

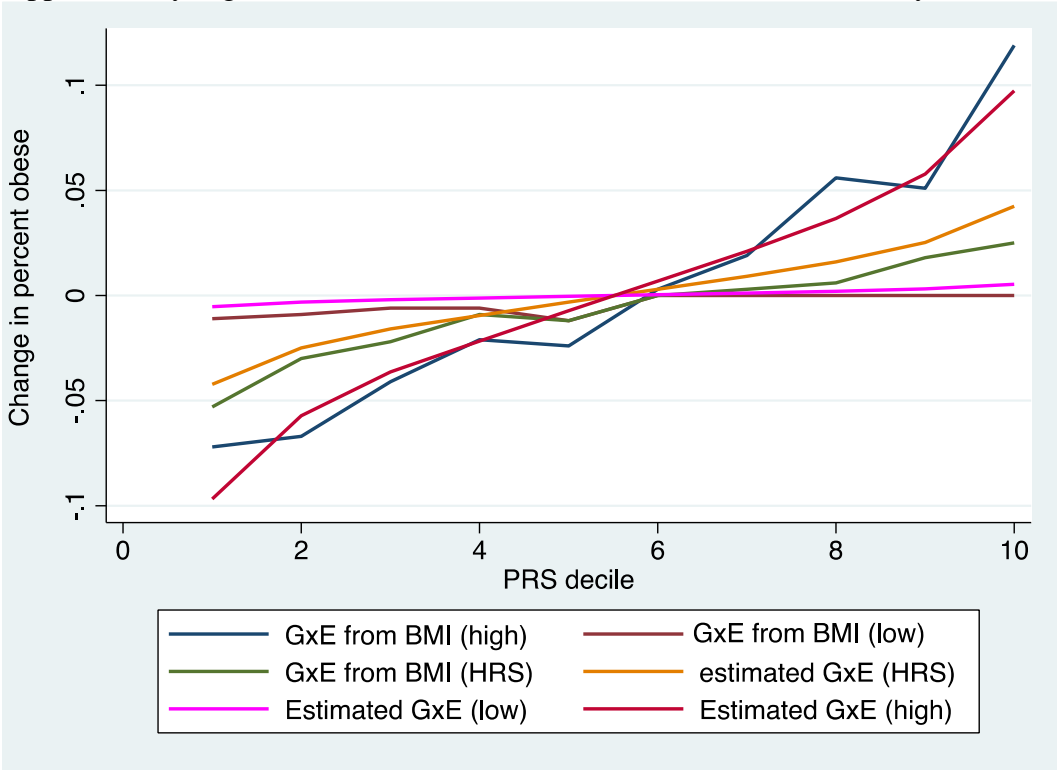
b. In our analysis, the difference between the birth year of individuals in cohort 1945-1959 and those in cohort born before 1924 is about 30 years on average. For analyses using single-year birth cohort, the coefficient as reported in the article is multiplied by 30.

c. as the comparison between cohort EBB and cohort AHEAD

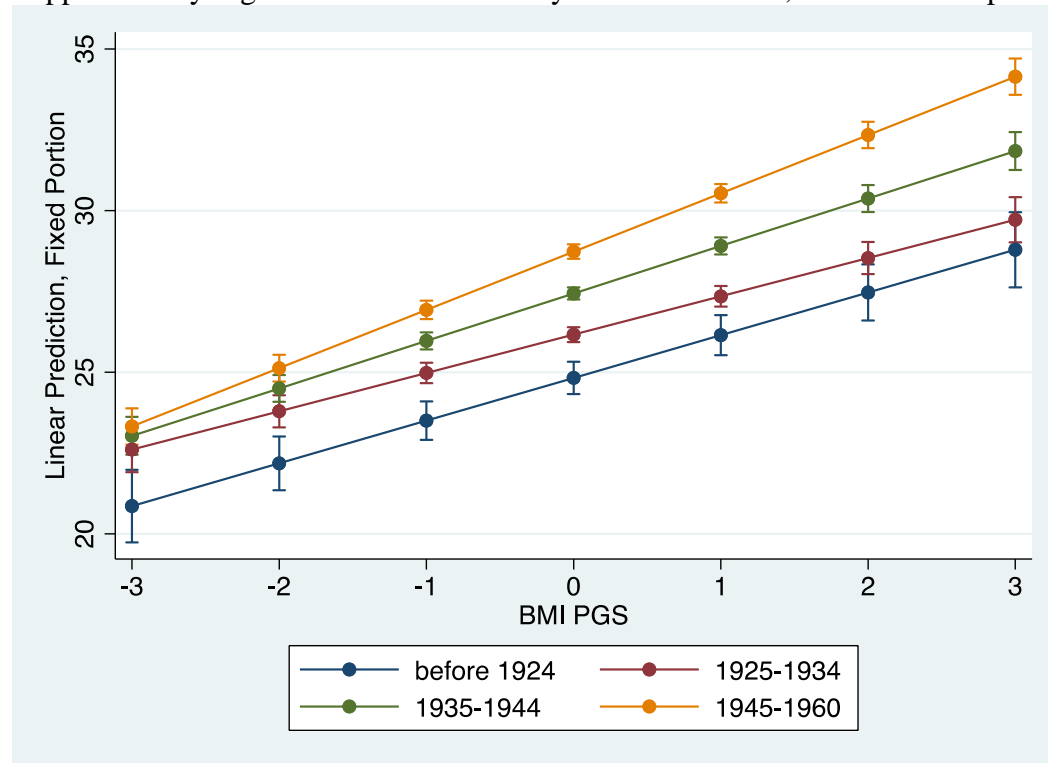
Supplementary Table 7. Estimates of the Correlation between Parental and Child BMI/Obesity

Fogelhom et al., (1999)	The unadjusted parent \pm child correlations of weight status ranges from 0.20 - 0.37.
Bralic et al., (2005)	The association between parental BMI and children's BMI is about 0.2 (0.19 for mother-child and 0.22 for father-child).
Svensson et al., (2011)	The association between parental BMI and children being severely obese is about 0.22-0.36 at age 15, 0.07-0.17 at age 7.
Wang et al., (2017)	The pooled OR from 32 publications reviewing the parental-child obesity or overweight is 2.22

Supplementary Figure 1. Alternative estimates of GxE effects for obesity.



Supplementary Figure 2. Predicted BMI by PGS and Cohort, white HRS respondents.



Supplementary Figure 2b. Age Trajectories of BMI by Cohort, white HRS respondents.

