Impact of genetic heritability, assortative mating, and fertility differentials on population-level obesity trends

Néstor Aldea[1,2], Aitor García-Aguirre[1], Sebastian Daza[1], and Alberto Palloni[1,3]

[1]Institute of Economy, Geography and Demography, CSIC, Spain
[2]French Institute for Demographic Studies (INED), France
[3]Center for Demography and Ecology, University of Wisconsin-Madison, USA

October 20, 2023

Keywords: BMI, obesity, heritability, assortative mating, fertility, Leslie matrices, Agent Based Models

*__Corresponding author:__ Nestor Aldea: nestor.aldea@cchs.csic.es. This project received funding from the **European Research Council (ERC)** under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788582). The paper reflects only the authors' view and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

Abstract

We model the influence of selected ancillary processes that partially account for trends of obesity prevalence in modern populations. These include vertical inheritance of obesity-related genetic variants, gene-environment interactions, phenotypic assortative mating, and differential fertility. These processes operate in the background modulating the effects of more proximate determinants of obesity, including environments and individual behaviors. Modelling these processes sheds light on the spatial diffusion and intergenerational transmission of the phenotype under conditions determined by macro, micro, and mediating molecular factors affecting individuals' obesity risks. The models we use for these processes and their interactions combine Leslie matrices and Agent Base Models.

1 Introduction

The human obesity epidemic is a result of a heterogeneous set of factors that vary across time and space. There is agreement that an accounting framework of the epidemic should include distant macro (distal) determinants, such as modern food production and distribution and built-in environments, more proximate micro conditions such as individual preferences and choices responsible for caloric intake and expenditures, and mediating pathways such as molecular mechanisms that regulate energy management and tissue growth, storage, and composition[1]. This accounting framework, however, does not include background processes that operate as vehicles of the spatial diffusion and intergenerational transmission of the phenotype. In this paper we model four ancillary processes: i) vertical genetic transmission or inheritance (GH), (ii) gene-environment interactions (GxE), (iii) assortative mating (AM), and (iv) differential fertility (DF). As outcomes of interest we use BMI and WHO-defined obesity categories[2]. Mechanisms for GH and GxE drive the intergenerational transmission of the phenotype whereas those associated with AM and DF indirectly strengthen (attenuate) the impacts of the other two. Importantly, our models are anchored on input parameters whose values depend on empirical estimates generated in multiple areas of research. We are not the first to propose models that track the obesity epidemic (Dawson et al., 2013; Ejima, Thomas, & Allison, 2018; Giabbanelli, Alimadad, Dabbaghian, & Finegood, 2012; Huang, Yan, Chen, & Liu, 2016; Levy et al., 2011; Morshed et al., 2019; Shoham et al., 2015). Our paper is guided by these previous formulations and aims at enhancing them by including processes and relations not considered previously.

We address the following questions:

1. How strong is the contribution of vertical genetic inheritance to population prevalence of obesity?

2. How large can the impact of known gene-environments interactions (GxE) be on future

[1]For a recent and very thorough overview of determinants of obesity, see 'Causes of Obesity: Theories, Conjectures and Evidence', The Royal Society of London, October 17-19, 2022: https://royalsociety.org/science-events-and-lectures/2022/10/causes-obesity

[2]https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

trends of the phenotype?

3. How sensitive are population levels of obesity prevalence to changes in phenotypic and social assortative mating? How do changes in assortative mating alter vertical genetic heritability of the trait? Can increases in assortative mating be an important driver of the obesity epidemic?

4. What is the impact of differential fertility by BMI or obesity status? Is the magnitude of the effect modified by changes in assortative mating? Can reductions of assortative mating offset effects of differential fertility by obesity status?

To answer these questions we use formal models and assess first and higher order impacts of the ancillary processes. We will focus on differences in parameters of the BMI distribution and levels of obesity prevalence in the short term (5-10 generations). Our goal is not to forecast future trends but, rather, to produce an informed and empirically defensible assessment of the influence of ancillary process and their interrelations, a necessary step before undertaking any forecasting exercise.

The paper is organized as follows: Section 2 is a review of ancillary processes; Section 3 introduces two classes of models to represent them; Results are in Section 4, and in section 5 we review shortcomings and suggest future lines of investigation.

2 Ancillary processes

There is agreement that an accounting framework of the epidemic should include macro determinants, such as modern food production, distribution and built-in environments, micro conditions that shape individual preferences regulating caloric intake and expenditures, and mediating pathways consisting of molecular processes of energy management, tissue growth, storage, and composition[3]. Modern environments to which an increasing number of human populations are exposed, include conditions that lead to excess caloric intake, sedentarism, degraded nutritional quality of

[3]See footnote 1.

food staples, and calorie dense foods whose ingredients can promote fat tissue growth or disrupt metabolic functions. All of these factors promote imbalanced in caloric intake/expenditures that result in excess fat storage and are referred to as 'obesogenic environments' (Flegal, Carroll, Ogden, & Johnson, 2002; Popkin, 2011; Popkin, Corvalan, & Grummer-Strawn, 2020; Swinburn et al., 2011). They emerge rather abruptly after 1950 within a population genetic context largely shaped by ancestral adaptations evolved by natural selection that oftentimes clashes with modern conditions. The nature of the disharmony is the subject of controversy and may involve thrifty genes (Neel, 1999), thrifty phenotypes (Hales & Barker, 1992), drifty genotypes (Speakman, 2013), and crafty genotypes (Wells, 2012).

There is no doubt that characteristics of modern obesogenic environments and human evolutionary constraints are central to an understanding of modern obesity trends. But, in addition, there are distinct forces that promote the spatial diffusion and transmission of the phenotype across generations. These are ancillary processes operating in the background that could augment (diminish) the impact of obesogenic environments and mismatches with ancestral adaptations. Because they may account for an important fraction of short and long term variability within and between populations and over time, they should be considered explicitly. In particular, mechanisms responsible for GH and GxE directly account for intergenerational transmission of the phenotype whereas those associated with (ii)-(iv) have indirect effects that reinforce or attenuate the influence of the other two[4].

2.1 Vertical genetic inheritance

Human obesity can be classified into two broad categories, monogenic and polygenic. Monogenic obesity is severe and has an early onset, but is quite rare. It involves small or large chromosomal deletions or mutations in a handful of genes, particularly those controlling the leptin signaling pathway (Bouchard, 2021; Loos & Yeo, 2022). Polygenic obesity (or 'common obesity') is a result

[4]There are three additional ancillary processes we do not study here, namely, vertical cultural transmission, maternal vertical inheritance ('maternal constraints'), and cultural horizontal transmission ('intragenerational inertia').

of hundreds of polymorphisms, each of which has very small effects on storage and distribution of adipose tissue and the production of hormones and cytokines that originate in fat issue. Thus, the phenotype is partially genetically inherited. Studies based on nuclear families (including offspring-parents pairs, siblings, adopted and foster children and MZ and DZ twin pairs), estimate that heritability of BMI (and obesity) and other metrics of fat deposit distribution (waist-to-hip ratio), are in the range of .30-.90, with the majority of values concentrated around .60 (Bouchard, 2021; Yengo et al., 2018). These estimates are likely inflated by unmeasured shared environments and possibly by the contribution of gene-environment interactions.

Recent GWAS studies identify hundreds of SNPs that collectively explain no more than 6% - 10% of the phenotype (obesity or BMI), thus leaving a large unexplained gap as *missing heritability* (Abadi et al., 2017; Brandkvist et al., 2019; Fall & Ingelsson, 2014; Goodarzi, 2018; Khera et al., 2019; Qasim et al., 2018; Rohde et al., 2019; Wang et al., 2011; Yengo et al., 2018).[5] These studies find closely related genomic regions that regulate fat deposition and storage and code for proteins expressed in brain tissues. Some genes regulate addiction-reward brain pathways and are involved in traits associated with obesity (reproduction, immune function, tissue growth and repair). Also, there is no evidence that these genes (and those associated with them by Gamete Phase Disequilibrium (GPD)) have been subjected to recent selection pressures, a finding that supports the so-called two-intervention point conjecture about the origin of the modern obesity epidemic (Speakman, 2013, 2018; Speakman, Djafarian, Stewart, & Jackson, 2007).

The above suggests two inferences. First, it is highly unlikely that the sudden global increase in obesity is a result of changes in the gene pool because the time span involved is too short, the genetic diversity at the base of the trait is too large, and footprints of selection pressure of gene variants associated with obesity are missing (Speakman, 2018). We explore the possibility that increased assortative mating plays a role in recent trends by modifying genotypes distributions and, therefore, explain part of the heterogeneity within and between populations. Second, and related to our goal, the fact that the phenotype of interest is highly polygenic suggests that the

[5]For a very accessible description of the nature of GWAS studies and their foundation on DNA sequencing, see (Mills, Barban, & Tropf, 2020).

vertical genetic component of an obesity model should be represented by a synthetic indicator of genetic penetrance, such as a Polygenic Risk Score (PRS) for BMI (obesity), instead of including effects of one or a handful of identifiable loci[6].

2.2 Gene-Environments interactions, GxE

We would expect high levels of GxE interaction effects in populations that experience sharp contrasts between ancestral and modern environments. Mismatching is more likely to happen among individuals with high genetic predisposition to high body weight who live in obesogenic settings that differ sharply from ancestral ones under which they may not have been significantly affected by their genetic propensity. In fact, it has been shown that when placed in more obesogenic environments, some individuals experience larger increases in BMI than expected given their genotypes (Huangfu, Palloni, Beltrán-Sánchez, & McEniry, 2023; Walter et al., 2016). Although GxE interactions effects will influence aggregate obesity trends and can alter the composition by genotypes of some population subgroups, they cannot directly change genotypic frequencies of subsequent generations.

2.3 Assortative mating

An important mechanism of the evolution of polygenic traits is the nature of the population mating regime. To the extent that mating is random relative to the phenotype (or genotype), genotype frequencies involving alleles associated with the phenotype will remain constant across generations, and the trait's evolution will be a function of other forces (drift, gene flow, and selection). However, when individuals mate non-randomly, genotype frequencies (though not allelic frequencies) will change and deviate from Hardy-Weinberg equilibrium, potentially contributing to the evolutionary trajectory of the phenotype.

[6]It is possible that epigenetic changes may contribute to the phenotype as a result either of fetal exposures or vertical maternal (paternal) inheritance (Bateson & Gluckman, 2011; Fall, 2011; Godfrey et al., 2011; Herrera, Keildson, & Lindgren, 2011; Kempf et al., 2022; Ling & Rönn, 2019; Sharp et al., 2017; Vickers, 2014; Waterland & Jirtle, 2003). In what follows we ignore the role of epigenetic modifications.

Assortative mating by BMI or obesity can be either phenotypically driven or the result of social homogamy. Preference for partners' physical appearance is culturally and socially defined and can result in deviations from random mating. It is known, for example, that there are societies in which mates choose among the most obese as the trait is a sign of wealth, health, or good luck (Macchi, 2022). Departures from random mating can also be the result of social and cultural stratification, that is, when the set of suitable partners is constrained by membership in social classes, residential location, ethnicity or religious affiliations. This is important in the case of obesity since, at least in high income countries, recent increases in obesity are disproportionately concentrated among individuals of low socioeconomic status (SES). Because mating in these populations is non-random relative to SES, it will also lead to assortative mating by the phenotype.[7]

Because BMI is partially genetically determined, couples that resemble each other are more likely to be genetically similar (Abdellaoui, Verweij, & Zietsch, 2014; Domingue, Fletcher, Conley, & Boardman, 2014) and thus influence their offspring genotype and phenotype. In addition, however, pairings of individuals with similar body shapes can have non-genetic, cultural impacts on their offspring phenotypes. Both types of obesity assortative mating entail the formation of parental household settings that could strengthen (discourage) the creation of obesogenic 'local' environments associated with social class, residential location, religion or ethnic-specific practices. This opens a gateway through which parental risks are passed on to offspring not just through genetic variants but through shared environments (vertical cultural inheritance[8]). The creation of local, household-based, obesogenic environments may also have a feedback effect as it can reinforce assortative mating in subsequent generations either via imprinting or through parental influences on offspring mate choices (Kong et al., 2018; Zietsch, Verweij, Heath, & Martin, 2011).

Assortative mating increases the frequency of homozygosity and its ultimate effects on the phenotype distribution is a function of the number and penetrance of loci associated with the phenotype. The smaller is the set of genes responsible for the trait and stronger is their penetrance, the

[7] A similar argument may apply to place of residence, religion, ethnicity, etc.

[8] There has been a rapid growth of research on these non-genetic parental influences (Boyd & Richerson, 1988; Cavalli-Sforza' & Feldman, 1973; Cavalli-Sforza & Feldman, 1981; Costa-Font, Jofre-Bonet, & Le Grand, 2020; Feldman & Ramachandran, 2018; Uchiyama, Spicer, & Muthukrishna, 2021; Wells, 2011; Willyard, 2014)

larger will be the impact of assortative mating on the trait's evolution[9] and through it augment the genetic variance of the trait. Also, when assortative mating is strong, it can increase gametic linkage disequilibrium (GLD) which is itself a factor that increases additive genetic variance (Hedrick, 2016).

There is empirical evidence that assortative mating by body weight is non-trivial and that its strength has increased over time, hand in hand with the prevalence of the phenotype (Ajslev, 2012; Hebebrand et al., 2000; Power et al., 2011; Speakman, Djafarian, Stewart, & Jackson, 2007). Furthermore, some data suggest that the strength of assortative mating may be higher in subpopulations located at the upper tail of the BMI distribution (Ajslev, 2012; Bouchard, 2021; Jacobson, Torgerson, Sjostrom, & Bouchard, 2006). If this finding is confirmed more generally, prevalence of obesity could increase even under constant genetic penetrance, thus enhancing the influence that assortative mating has on increases of BMI and obesity prevalence.

2.4 Fertility differentials

An important determinant of the increase of obesity prevalence over time is the differential rate of growth of the obese and non-obese subpopulations. In the absence of significant pre-reproductive mortality between these two subgroups, differences in rates of growth must be accounted for by gross reproduction rates (GRR) (possibly weighted by parent-offspring vertical cultural heritability). Fertility differentials will also influence the timing of trajectories of BMI and obesity. From first demographic principles, we know that even if there is a successful intervention that reduces the net reproduction rate of the obese population to 1, the rate of growth of obesity prevalence will continue to increase for some time afterward, and so will any existing obesity differentials by SES (Keyfitz, 1971).

Human GRR depends on male-female fecundity, on the one hand, and on lifetime fertility conditional on fecundity, on the other. In humans, obesity influences both fertility and fecundity but in ways that offset each other. There is strong empirical evidence showing that both male and

[9] Assortative mating will increase the association between distant loci (Yengo et al., 2018).

female obesity impair fecundity (Craig, Jenkins, Carrell, & Hotaling, 2017; Sermondade, 2012; Silvestris, de Pergola, Rosania, & Loverro, 2018)[10].

However, obesity (and BMI) is also associated with GRR as a result of factors ostensibly unrelated to fecundity. Thus, empirical evidence from high-income countries (HIC), suggests that GRR is positively related to BMI, that the relation is strong and increases over the life cycle of individuals (Swan & Colino, 2021). It is likely that this positive association is due to common influences rooted in socioeconomic conditions.

In some, not all, low-income countries, the association between fertility and obesity is negative (Pampel, Denney, & Krueger, 2012; Vazquez & Cubbin, 2020) though the reverse may be true in high SES subgroups[11]. The existence of both positive and negative relations is consistent with the fact that, during the first stages of the epidemic, the association in HIC is weakly positive or negative. It is quite possible that LIC are in the early stages of the process and that a positive relation will also prevail as they move through future stages.

3 Models

We aim to assess the effects of different assortative mating regimes under varying conditions of vertical genetic inheritance, GxE interactions, and fertility differentials by body weight. Although there is a voluminous literature on each of these, there are few studies that attempt to integrate them into a coherent model to account for obesity trends (Dawson et al., 2013; Daza, S. & Palloni, A, 2022; Ejima, Thomas, & Allison, 2018). The modeling approach we propose is simplified because it ignores three ancillary processes (identified in footnote 4). Despite this and with some caveats we will state later, the model is complex enough to approximate the relative contribution of the processes included in it.

We use two modelling approaches. The backbone of the first is based on Leslie matrices whose

[10] In contrast, empirical evidence from animal studies shows that the genetic propensity to higher body weight is associated with higher fecundity and gross reproduction rates (Allison et al., 1996; McAllister et al., 2009). This is likely due to a direct impact of excess adipose tissue on reproductive capacity.

[11] See Supplementary Material for empirical evidence of this relation.

properties shed light on the nature of relations between the various components. The second approach consists of an Agent Based Model (ABM). Each approach has its own strengths and weaknesses but they complement each other and, despite their differences, lead to similar inferences.

3.1 Leslie matrices

We follow a modification of a stable population approach first employed by Preston and Campbell to study IQ trajectories (Preston & Campbell, 1993). Instead of dealing with the population distribution by age, as is done in standard stable population applications, we focus on generation-specific distributions of the population in four BMI classes[12]. We employ the WHO classification and define four categories: underweight (BMI $\leq$ 18.5), normal weight (18.5 $<$ BMI $\leq$ 25), overweight (25 $<$ BMI $\leq$ 30), and obese (BMI $>$ 30). To represent intergenerational changes in the vector of obesity categories, we define a matrix of transition probabilities that depends on mating rules, net fertility, and vertical inheritance probabilities.

Let the distribution of the first generation by obesity status be given by a 1x4 row vector $BMI(0)$ and that of the subsequent generation by $BMI(1)$. These two vectors are related as follows:

$$BMI(1) = BMI(0) \times N \times R \times H = BMI(0) \times \Sigma \tag{1}$$

where N is a 4x16 matrix of mating probabilities of pairing between obesity categories that produces 16 classes of couples, R is a 16x16 matrix of fertility rates for the 16 classes of couples and, finally, H is a 16x4 matrix of genetic heritability, containing the probabilities of offspring allocation by obesity status conditional on the obesity status of each parent[13]. The mating (N) and fertility (R) matrices are defined to accommodate different combinations of mating rules and differential fertility by obesity status. These two matrices are fixed and deterministic, e.g. their entries are not influenced by random variation and remain constant throughout. In contrast, entries

[12]Throughout we will ignore the role of age and assume that couple formation and childbearing occur at the mean age of childbearing.

[13]Full description of these matrices is in Section 1 of Supplemental Materials.

of the heritability matrix, H, can be either fixed or changing across generations and, furthermore, may or may not contain random components.

When H is fixed and deterministic, the second generation's distribution by categories will be

$$BMI(2) = BMI(1) \times N \times R \times H = BMI(0) \times \Sigma^2 \tag{2}$$

where Σ^2 is the square of the matrix product $N \times R \times H$.

More generally,

$$BMI(t+1) = BMI(t) \times N \times R \times H = BMI(0) \times \Sigma^t \tag{3}$$

When H changes across generations, we will have

$$BMI(t+1) = BMI(0) \times \Pi^t \times \Sigma(t) \tag{4}$$

where Π^t is the t^{th} power of the product of matrices N and R, and $\Sigma(t)$ is the product of $H(0), H(1)..., H(t)$.

In the section that follows we define rules for assortative mating, fertility differentials, and heritability[14].

3.1.1 Rules of assortative mating

To specify assortative mating rules, we define a one-parameter continuous function, $\omega \in [0,1]$. When its value is 0 the mating regime is completely random and when it is 1 the mating regime is completely endogamous, e.g. individuals only mate within their obesity category. Other ω values within the closed interval [0,1] define mixed regimes. Let π_i, be the i^{th} entry of the row vector $BMI(t)$ corresponding to the fraction of individuals in obesity category i. The probability that an individual belonging to category i mates with a partner in the same category is:

[14] A full empirical justification of the values of the input parameters that define these rules is in Section 3 of Supplementary Materials.

$$p_{ii} = \pi_i + (1 - \pi_i) \cdot \omega \tag{5}$$

Because the sum of over all j of p_{ij} must add up to 1[15], the probability of mating with individuals from a different BMI category ($i \neq j$) is

$$p_{ij} = (1 - \omega) \cdot \pi_j \tag{6}$$

The probabilities from equations (5) and (6) fully specify the entries of matrix N (see (S1) of Supplemental Material).

3.1.2 Size and direction of fertility differentials

Regimes of differential fertility are specified in a manner analogous to assortative mating. We assume a strictly positive association between fertility and BMI and define a continuous, one-parameter function, $\varphi \in [0, 1]$. A couple whose members are in obesity categories i and j will produce a number of offspring given by

$$f_{ij} = 2 + (i + j - 6) \cdot \frac{\varphi}{2} \tag{7}$$

The magnitude of the differential is maximized when couples formed by two individuals in the categories ($i, j = 4$), have 3 children whereas couples consisting of two underweight individuals ($i, j = 1$), have none. When there are no fertility differentials, all couples have two offspring. The values f_{ij} from equation (7) fully specify the entries of the diagonal fertility matrix R defined in the Supplemental Material.

3.1.3 Vertical genetic inheritance

We formulate two variants of the heritability matrix H. The first has deterministic and generation-invariant entries. The second matrix's entries include a random component and change across

[15] In this version of the model, all individuals find a partner.

generations.

1. *A one biallelic locus for obesity (constant H)*:

We assume a one biallelic locus responsible for the phenotype obesity and that the obesity allele, A, is dominant over the non-obese allele, a[16]. We then employ standard Mendelian segregation rules to define H and determine the offspring phenotypic distribution. Because this matrix remains invariant throughout, the population distribution by obesity categories at time $t+1$ will be given by expression 4. Also, because Σ is irreducible and primitive, the system is ergodic and the population distribution by categories converges to a stable distribution that depends only on Σ, not on the initial distribution $BMI(0)$[17]. The entries of H are defined by simple Mendelian rules applied to a one biallelic locus. The probabilities of an obese offspring according to combinations of parental phenotype are:

Table 1: Probabilities for offspring phenotypes as a function of parents'

Parents	**2 obese**	**1 obese, 1 non-obese**	**2 non-obese**
Child obese	8/9	2/3	0
Child non-obese	1/9	1/3	1

Children with a non-obese phenotype are then distributed in the three lowest BMI categories as a function of their parents' BMI categories, as described in Supplemental Materials. These probabilities, $p(k|ij)$ will be the entries of the matrix H.

2. *Multiple loci for BMI (time dependent H or $H(t)$)*:

The second variant of the matrix approach assumes multiple loci with variable penetrance, each with small effects on the phenotype of interest (BMI), summarized by a BMI polygenic risk score (PRS). Because the offspring' PRS depends on her parents's PRS and a random component[18], the entries of the matrix $H(t)$ ($\forall t$) are random and, importantly, depend on t.

[16]None of the inferences regarding the role of the ancillary processes change when the obese allele is the recessive one or when the heterozygotes' phenotype is determined as a function of a dominance parameter.

[17]The mathematics of transformations of an initial population distribution by age groups using an invariant transition matrix is well-known in demography and ecology (Caswell, 2000; Keyfitz,N., 1977).

[18]See below and Section 1 of Supplementary Material.

To define numerical values for $H(t)$'s entries we start out with observed distributions of PRS and BMI in 1,568 intact couples (3,136 individuals) with DNA information in the 2006 wave of the Health & Retirement Study[19].These individuals play the role of generation 0, *G0*, and we will use their PRS as well as the observed relation between their PRS and BMI to determine their offspring's (generation G1) values of PRS and BMI (and corresponding obesity category). This is done in four steps:

- Formation of couples in the first generation, G0: we use the set of 1,568 intact couples in the 2006 HRS' wave including their information on BMI and PRS. We then assign to each couple a random number of children following a Poisson distribution with mean 2. Thus, at least initially, there is no association between realized fertility and BMI (obesity). The set of offspring defines generation G1.
- Allocation of offspring' PRS in G1: after fusion of gametes, the zygote genome is composed of approximately half of maternal and paternal genes. To reflect this, we compute G1's PRS as the arithmetic mean of their G0 parents' plus a random term to represent inherent stochasticity as in S8. The random component is scaled so that the offspring PRS distribution is normal (0,1), as is that of their parents'.

$$PRS_{off} = \frac{PRS_{par1} + PRS_{par2}}{2} + \eta(\mu = 0, \sigma = 0.7) \tag{8}$$

- Assignment of offspring' BMI in G1: This is done using estimated parameters for a simple linear relation between BMI and PRS observed in the G0 sample. The estimated relation is

$$BMI = 27.796 + 1.071 \times PRS + \varepsilon(\mu = 0, \sigma = 4.55) \tag{9}$$

[19]Health and Retirement Study, RAND HRS Longitudinal File 2023, public use dataset. Produced by the RAND Center for the Study of Aging (Santa Monica, CA.) with funding from the National Institute on Aging and the Social Security Administration and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

The value of σ is scaled so that the variance of the BMI distribution is identical to the observed one in G0. We then assign obesity categories to members of G1 according to their predicted BMI. This last step defines matrix $H(0)$. [20]

- Couple formation in G1: After classifying members of G1's couples by obesity categories, we pair them up according to the mating rules, that is, a fraction $1-\omega$ of the total target couples $(N/2)$ pair up ramdomly and then match the remaining individuals within their BMI category (endogamously).
- Assignment of offspring in G1: we assign to each G1 couple a number of offspring (G2) according to differential fertility rules[21].

Subsequent generations are computed by repeatedly applying the steps described above. To reflect the presence of GxE in well-defined environments and subpopulations, we will also allow increases of the parameter for penetrance in equation (9), $\alpha = 1.0715$. To do so, we will define the penetrance parameter as :

$$\alpha = 1.071 \cdot (1+p) \tag{10}$$

with p attaining values 0, 0.1, and 0.2.

Because the values of PRS and BMI in each generation after G0 are functions of random normal variates (σ and ε), we assign values of PRS and BMI to each individual in a generation by (independently) randomly draws from normal distributions, $N(0,\sigma)$ and $N(0,\varepsilon)$. This is done for all generations we simulate. The collection of distributions by obesity categories defines a single realization of the sequence of matrices $(H(0),H(1),H(2),...H(T))$ associated with one combination of parameters or set of rules. We repeat this process 20 times gathering 20 different realizations associated with those rules. To summarize results of these

[20] The correlation of BMI and the PRS for BMI in the sample of 3,136 individuals is $\rho = 0.23$.

[21] The expression for f_{ij} from equation (7) may result in a fractional number of children, between n and $n+1$, say $n+v$, where $0 \leq v \leq 1$. In order to convert this figure to an integer we randomly assign integer n with probability $1-v$ and $n+1$ with probability v.

20 realizations, we compute means and standard deviations of the outcomes of interest (BMI and PRS distributions and obesity prevalence)[22].

3.2 ABM model

The matrix approach has an important advantage in that the time trajectory of the population distribution by obesity categories is associated with matrix properties that can be completely specified. The disadvantage is that it cannot easily represent individual behavior(s) and, as a consequence, is not flexible enough to integrate features additional to those considered in the paper. For example, inclusion of horizontal heredity via peers or residential location renders the matrix treatment cumbersome or outright intractable.

An ABM approach is useful when the dynamic of a system cannot be transparently translated into a formal representation (such as Leslie matrices). It is flexible enough to integrate multiple processes, interactions between them, and complex feedback mechanisms. An ABM model can be easily defined to integrate peers' influence and other potential sources of life course changes, vertical maternal inheritance, and characteristics associated with individuals' residential location. An example of a model with some of these properties was developed by Giabanelli and colleagues and validated using data from the National Longitudinal Study of Youth (NLSY) (Giabbanelli, Alimadad, Dabbaghian, & Finegood, 2012). Because our final goal is to represent all relevant ancillary process, including those we ignore in this paper, and because this cannot be accomplished by extensions of the matrix approach, we must verify that an ABM following rules exactly identical to those in the matrix approach generates the same (or approximately the same) results and inferences.

The design of the ABM follows the same rules for ancillary processes used in the matrix approach. The only difference is that these rules apply to single (simulated) individuals who follow them as agents rather than to categories or groups. Individual agents select mates, choose number

[22]We assign 6 different values to ω and 3 each to φ and p thus producing 54 combinations of rules. For each of these we collect 20 realizations.

of offspring and offspring' inherit parental genetic variants and are thereby assigned a BMI and to an obesity category[23].

4 Results

4.1 Results from Leslie matrices

4.1.1 Obesity as a monogenic phenotype under dominance

Figure 1: Distribution of the population by BMI category and generation, in four scenarios of assortative mating and differential fertility (Leslie matrices)

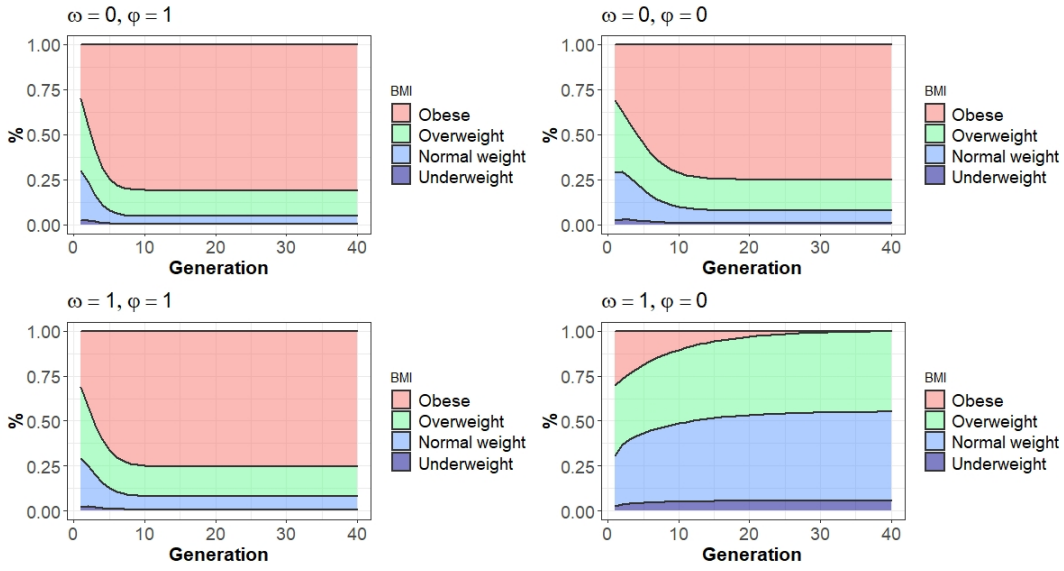


Figure 1 displays the distribution by obesity categories in four different regimes of FD and AM. To avoid cluttering, we only show results for two AM regimes ($\omega = 1$ and 0) and two FD regimes ($\varphi = 0$ and 1). Under maximum fertility differential favoring the obese subpopulation ($\varphi = 1$), the steady state distributions by obesity categories in a regime with random mating (top left) and an endogamous one (bottom left), are virtually identical: in both cases the prevalence of obesity attains values around .75-.81. However, a comparison of the two assortative mating regimes under conditions of no fertility differentials, ($\varphi = 0$), reveals stark differences: when there

[23]The full description of the ABM model is in Supplementary Materials.

is random mating, the steady state distribution continues to favor the obese category (top right) and the prevalence of obesity attains a value of about .75. Instead, in the case of endogamous mating, the population in the obese category vanishes[24]. This occurs because the obesity allele is dominant and under an endogamous regime couples whose members are obese may produce offspring who are not, but the reverse cannot happen, e.g. when neither member of the couple is obese, they can only produce non-obese offspring. Thus, in every generation the obese category sheds population to the other categories and these losses are not offset by higher fertility (as they are in the case of differential fertility). The main inference we draw from these results is that when there are fertility differentials that favor those with higher BMI, assortative mating plays a marginal or no role at all.[25]

4.1.2 BMI as a polygenic phenotype

When the distribution by obesity categories is a function of individuals' BMI whose heritability depends on a PRS, the transition matrix $H(t)$ will have random entries and change across generations. Except in one case (no fertility differentials), there will be no steady state distributions. Because of this, the assessment of effects of AM, FD and PE, requires that we choose outcomes in one or more generations. In what follows we will choose the distributions attained in the 10th generation[26].

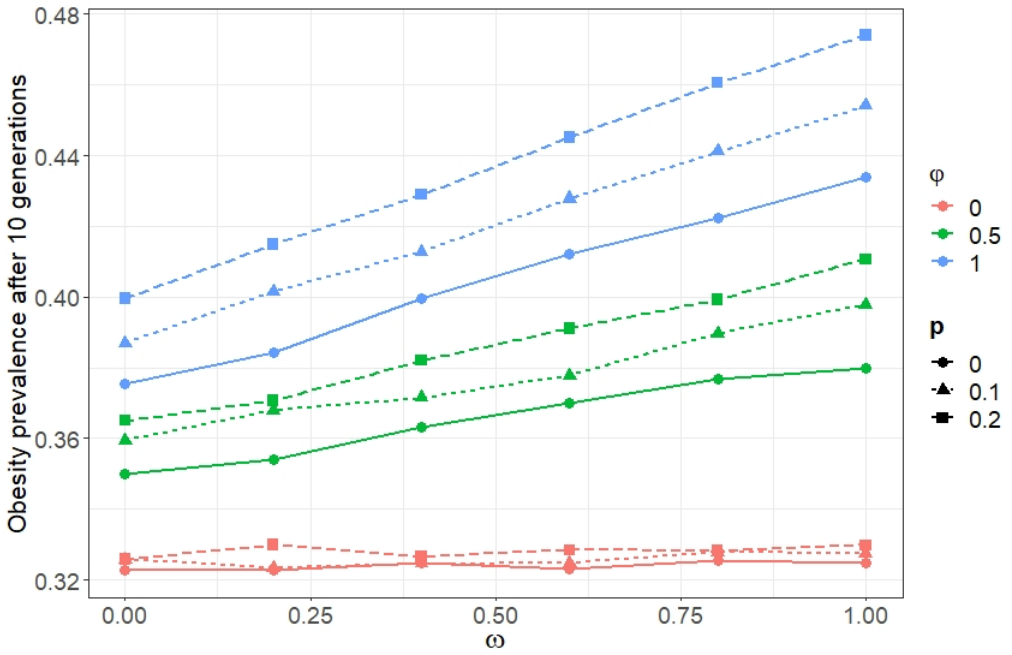
Figure 2 displays values of obesity prevalence in the 10th generation as a function of the AM parameter, ω, for multiple combinations of PE and FD. When the parameter for FD is greater than 0 (green and blue lines), prevalence levels increase linearly as one moves from a regime of random mating to one that is completely endogamous. Changes in AM induce the largest changes in prevalence. In fact, the maximum range of variation (with higher FD) is approximately .16

[24]The relatively high levels of steady state prevalence of obesity in 3 of the scenarios in the figure are due to the assumption of a single locus with a dominant allele for obesity. Had we replaced the dominance assumption for a milder version (half of those possessing a biallelic genotype become obese and the other half do not), the levels of steady state prevalence would have been reduced.

[25]A feature in Figure 1 is that the steady states in three of the scenarios are attained relatively rapidly (10 generations or less) whereas in the fourth case it requires about 15 to 20 generations.

[26]Because the obesity prevalence across generations are monotonically increasing (with FD) or relatively steady (without FD), inferences are the same irrespective of the generation we choose to focus on.

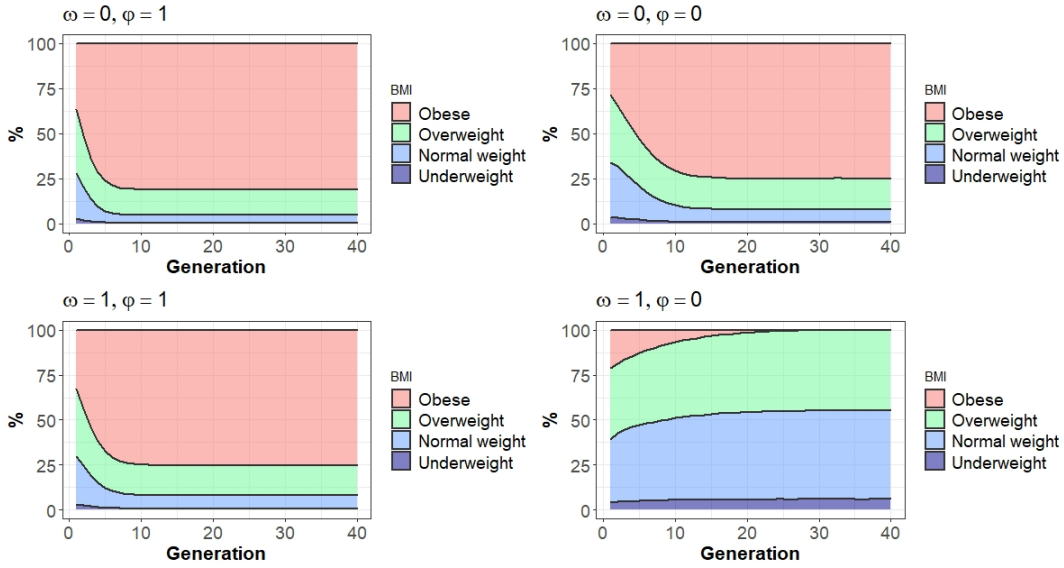
Figure 2: Prevalence of obesity after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.



$(.48 - .32)$ or about 42 percent of the mean of the range. Second, when there are no FD (red lines), AM plays no role and neither does PE. Third, differences associated with increases in PE (solid and dashed lines) grow in absolute value as FD increase and do so irrespective of ω. Note that, when in a regime with no FD, obesity prevalence does not converge to 0 as it did in the simple Mendelian segregation case.

In summary, the largest impact is exerted by GH and FD. AM plays no *direct* role, e.g. its additive effects are negligible and the impact of PE is indirect. This is an important finding for it casts doubts on the possibility that AM could have had discernible influence on the global obesity epidemic.

Figure 3: Distribution of the population by BMI category and generation, in four scenarios of assortative mating and differential fertility (ABM)



4.2 Results from ABM

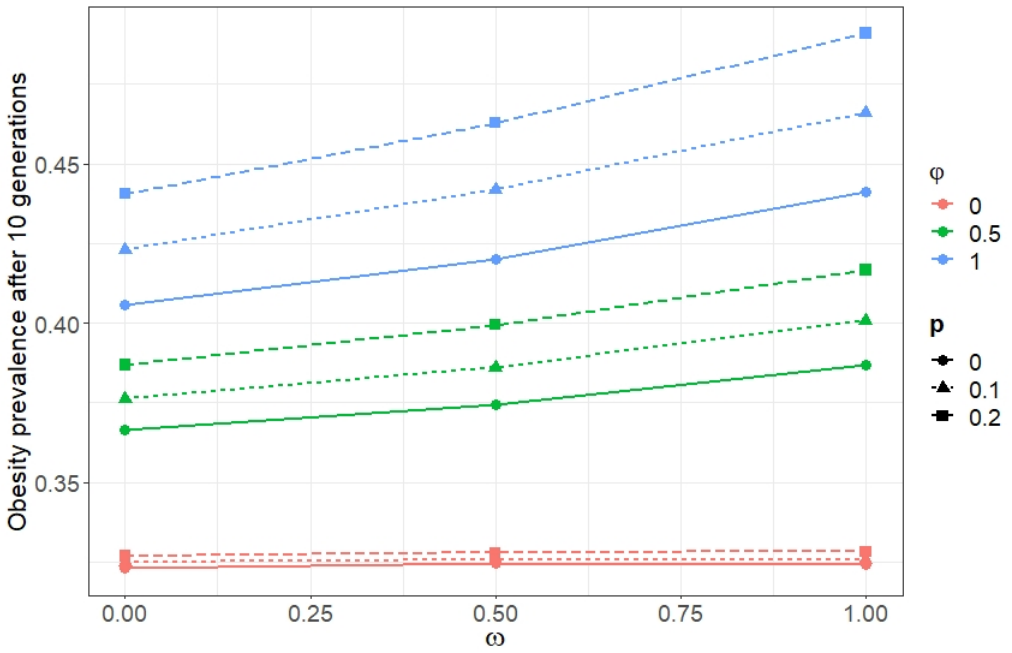
4.2.1 Obesity as a monogenic phenotype under dominance

Figure 3 shows the distributions by obesity category in four scenarios of AM and FD defined using identical rules to those in the case of Leslie matrices. As in results from matrix approach, in scenarios with FD favoring the obese subpopulation, the contrasts between a regime of full random mating and one fully endogamous are negligible. Further, both approaches estimate obesity prevalence in the range .75-.82. Similarly, and as in Figure 1, in the absence of fertility differentials a fully endogamous regime leads to the disappearance of the obese population.

4.2.2 BMI as a polygenic phenotype

Figure 4 displays estimated prevalence of obesity as a function of the parameter ω for 6 combinations of ancillary processes FD and PE. These plots are very similar to those in Figure 2 and lead to the same inferences. In particular, AM does not have a significant direct influence on trajectories of obesity prevalence but does appear to have a modest interaction effect with FD and PE (see below).

Figure 4: Prevalence of obesity after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), ABM.



4.3 Estimation of models for the effects of the ancillary processes

A useful way to summarize the above results is by modeling outcomes of interest (prevalence of obesity and parameters of the distributions of BMI and PRS) as a function of the input parameters. Because results from both the matrix and ABM approach depend on microsimulation (with 20 replicas, in the case of Leslie matrices and about 300 in the case of ABM), we can estimate the average relation between selected outcomes of interest and input parameters. In particular, we estimate the following models

$$ln(O_{ij}) = \alpha_i + \Sigma_k(\beta_{ik} \times Z_{ik}) \tag{11}$$

where $ln(O_{ij})$ is the the log of the outcome i, namely, mean BMI and standard deviations of BMI and PRS for the j^{th} simulation, Z_{ik} is the value of the k^{th} ancillary parameter (ω for AM, φ for FD and p for PE) used in the simulation i, α_i is an outcome-specific constant, and β_{ik}'s are

effects of the input parameter k on outcome i[27].

Table 2: Coefficients of regression of three outcomes for ABM and Leslie matrices

	log(prev.)		**log(BMI)**		**Mean PRS**	
	ABM	**LM**	**ABM**	**LM**	**ABM**	**LM**
Constant	-1.177***	-1.129***	3.325***	3.325***	.009***	.008
$\beta(\omega)$	**.006*** **	**.011***	**.001** **	.001	**.015*** **	.005
$\beta(\varphi)$	**.225*** **	**.152*** **	**.037*** **	**.024*** **	**.978*** **	**.635*** **
$\beta(p)$	**.068*** **	**.064** **	.001	.002	.013	-.029
$\beta(\omega \times \varphi)$	**.082*** **	**.140*** **	**.015*** **	**.024*** **	**.410*** **	**.674*** **
$\beta(\omega \times p)$	.019	.034	.002	.003	.032	.136
$\beta(\varphi \times p)$	**.360*** **	**.262*** **	**.072*** **	**.047*** **	**.799*** **	**.576*** **
$\beta(\omega \times \varphi \times p)$	**.113*** **	**.118***	**.030*** **	**.037*** **	**.342*** **	.244
N	7,944	1,080	7,944	1,080	7,944	1,080
R^2	.982	.963	.989	.976	.993	.988
Adj. R^2	.982	.962	.989	.976	.993	.988
RSE	.019	.022	.0026	.003	.046	.051
df	7,936	1,072	7,936	1,072	7,936	1,072

The N for LM models corresponds to 54 combinations of parameter values and 20 replicas for each of them. The N for ABM models corresponds to 27 combinations of parameter values and about 300 replicas for each.

Table 2 displays estimates of parameters and measures of fit for linear models for (the logs of) prevalence of obesity, mean BMI, and mean PRS as of the 10th generation (results for the standard deviations of BMI and PRS are displayed in Suppl. Materials, Table S5).

All estimates have the same signs, attain comparable statistical significance levels, and their magnitude is similar. The strongest direct effect is associated with FD: a change from a regime with no fertility differential to the maximum implies a relative change of obesity prevalence between 15 and 23 percent (first two columns, second row) and of 2 to 4 percent of the mean BMI. Changes in PE exert a weaker (prevalence) or no effect (BMI and PRS) whereas the direct effect of AM is small in all cases (albeit statistically significant in the ABM simulations). As surmised by Figures 2 to 4, there are multiple and strong interaction effects. Thus, as expected, the effects of fertility differentials become larger with increases in assortative mating and penetrance. For example, the proportional increase in prevalence associated with maximum fertility differentials grows from .23

[27]Because the PRS is in standard units and can attain negative values we do not use its logarithmic value.

(ABM) or .15 (matrix) to about .31 (ABM) or .29 (matrix) under a fully endogamous regime. These are large changes and suggest that non-linearities and higher order effects are likely to be an important part of observed time trends of BMI and obesity.

5 Summary and discussion

The results of the simulations lead to a handful of potentially useful inferences. First, and unlike empirical results from the IQ phenotype (Preston & Campbell, 1993), we find that fertility differentials by obesity categories are central to the story. Furthermore, we find that the other two ancillary processes, but particularly assortative mating, do not exert a significant *direct* influence. Instead, they operate as modifiers of the influence of fertility differentials and vertical genetic heritability. This undermines the conjecture that assortative mating on its own, even though apparently on the rise, could have driven past or could drive future trends of the obesity epidemic, both in populations that are in the midst of it and in those experiencing the initial stages (see also Supplementary Materials for added reasons to suspect a weak role of AM)[28]. There is one caveat of importance though: the direct effect of assortative mating must be stronger in the presence vertical cultural heritability (eg. parental and household influences) of the phenotype.

Second, even though we use rather low values of excess genetic penetrability (p) to reflect the impact of GxE interaction, its role is not insignificant. In fact, it augments the impact of fertility differentials and through it, reinforces the power of potential interventions designed to eliminate or reduce exposure of some subgroups to obesogenic environments (see Section 4 of Supplementary Materials). In addition, its influence could be felt in areas we did not explore. For example, a GxE interaction that emerges in generation G0 may have important effects on genotypic composition, couples' distribution by obesity category, and genetic heritability of the phenotype, that will be subsequently expressed in G1. Finally, we ignored the implications of a GxE interaction that

[28]Hedrick's formal representation shows that AM can contributes little to aggregate genetic variance via LD. Our simulation suggest that its total contribution via increased homozygosity is minor and a relatively relatively minor player in the growth of global obesity.

emerges in a social context with strong vertical and horizontal cultural heritability.

Third, we confirmed that, at least within the bounds of the ancillary processes considered, inferences from a matrix and ABM approach are quite similar. This is an important result for the matrix approach rests on massive simplifications that the ABM approach avoids. The fact that they can be used interchangeably is strong evidence supporting the use of the ABM model to include ancillary process that a matrix approaches cannot represent.

Fourth, the manner in which we represent vertical genetic heritability is consequential. The differences in results between the simple Mendelian mechanism and the one embedded in the use of a PRS are an indication of this. The assumption of a single dominant allele and Mendelian segregation rules leads to a scenario (no differential fertility) in which the obese population disappears. Using the PRS leads to a stationary scenario in which the levels of obesity remain close to those that existed in G0. Substituting a monogenic scenario by a polygenic one will generate more complex relations between the number of gene variants involved, their direct effects, and the steady states. It also introduces complications in the computation of probabilities of genotypes and requires information on the net effect of each of the alleles. Both these difficulties are circumvented by using a PRS.

The paper has a number of shortcomings. First, it excludes ancillary processes of importance. These may not only have important direct impacts but, more importantly, could modify the impact of the other ancillary processes. Future versions of the model will include vertical cultural heritability, specially parental influences through households ”niches”, horizontal heritability, such as peer's and residential influences and, finally, maternal effects *in utero* and during the first few years of life, both known to influence adult obesity risks.

Second, the model only represents situations in which the relation between fertility and obesity is positive, e.g the observed pattern in populations that are in advanced stages of the obesity epidemic. A more realistic model should include both regimes simultaneously, one in which the initial stages are characterized by an inverse relation that is reversed once the population attains certain levels of obesity prevalence. This may turn out to be a powerful feedback mechanism that

only an ABM model can handle efficiently.

Lastly, one could ask if our model fits *observed* data well. The only avenue to address this is to compare some of the outcomes we compute (aggregate obesity prevalence, mean BMI) with the *observed* population totals. A number of strategies could be used to accomplish this, including simple numerical methods, standard Maximum Likelihood methods, or Approximate Bayesian Computation. However, the results of the exercise will be inconclusive as the omission of ancillary processes will operate as an omitted variable does in standard econometric models and contaminate estimates of ancillary processes that were included. To validate the model thoroughly, we must be able to simulate richer observed outcomes, including age-specific prevalence by gender, spatial distribution of the phenotype, and patterns of diffusion over time.

Despite these limitations,there is value in our contribution. In particular, because we relied on empirically derived, not guessed, input parameters and because there are no better ways to represent the four ancillary processes included in the model, the goal of producing an informed and empirically defensible assessment of the influence of ancillary process and their interrelations, has been met.

References

Abadi, A., Alyass, A., Robiou du Pont, S., Bolker, B., & et al. (2017). Penetrance of Polygenic Obesity Susceptibility Loci across the Body Mass Index Distribution. *The American Journal of Human Genetics*, *101*(6), 925–938.

Abdellaoui, A., Verweij, K. J. H., & Zietsch, B. P. (2014). No evidence for genetic assortative mating beyond that due to population stratification. *Proceedings of the National Academy of Sciences*, *111*(40).

Ajslev, T. A. (2012). Assortative marriages by body mass index have increased simultaneously with the obesity epidemic. *Frontiers in Genetics*, *3*.

Allison, D. B., Neale, M. C., Kezis, M. I., Alfonso, V. C., & et al. (1996). Assortative mating for relative weight: Genetic implications. *Behavior Genetics*, *26*(2), 103–111.

Bateson, P., & Gluckman, P. (2011). *Plasticity, Robustness, Development and Evolution*. Cambridge University Press.

Bouchard, C. (2021). Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity*, *29*(5), 802–820.

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process*. University of Chicago Press.

Brandkvist, M., Bjørngaard, J. H., Ødegård, R. A., Åsvold, B. O., & et al. (2019). Quantifying the impact of genes on body mass index during the obesity epidemic: Longitudinal findings from the HUNT Study. *BMJ*, *366*, l4067.

Caswell, H. (2000). *Matrix Population Models: Construction, Analysis, and Interpretation* (2nd edition). Sinauer Associates Inc.

Cavalli-Sforza', L. L., & Feldman, M. W. (1973). Cultural versus Biological Inheritance: Phenotypic Transmission from Parents to Children (A Theory of the Effect of Parental Phenotypes on Children's Phenotypes), 20.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution*.

Costa-Font, J., Jofre-Bonet, M., & Le Grand, J. (2020). Vertical transmission of overweight: Evidence from a sample of English adoptees. *Food Policy*, *97*, 101972.

Craig, J. R., Jenkins, T. G., Carrell, D. T., & Hotaling, J. M. (2017). Obesity, male infertility, and the sperm epigenome. *Fertility and Sterility*, *107*(4), 848–859.

Dawson, J. A., Dhurandhar, E. J., Vazquez, A. I., Peng, B., & Allison, D. B. (2013). Propagation of Obesity Across Generations: The Roles of Differential Realized Fertility and Assortative Mating by Body Mass Index. *Human heredity*, *75*(0), 204–212.

Daza, S. & Palloni, A. (2022). Modeling the impact of heritability, assortative mating and fertility on population-level obesity trends.

Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences*, *111*(22), 7996–8000.

Ejima, K., Thomas, D., & Allison, D. B. (2018). A Mathematical Model for Predicting Obesity Transmission With Both Genetic and Nongenetic Heredity. *Obesity (Silver Spring, Md.)*, *26*(5), 927–933.

Fall, C. (2011). Evidence for the intra-uterine programming of adiposity in later life. *Annals of Human Biology*, *38*(4).

Fall, T., & Ingelsson, E. (2014). Genome-wide association studies of obesity and metabolic syndrome. *Molecular and Cellular Endocrinology*, *382*(1), 740–757.

Feldman, M. W., & Ramachandran, S. (2018). Missing compared to what? Revisiting heritability, genes and culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170064.

Flegal, K., Carroll, M., Ogden, C., & Johnson, C. (2002). Prevalence and Trends in Obesity Among US Adults, 1999-2000. *JAMA*, *288*(14), 1723–1727.

Giabbanelli, P. J., Alimadad, A., Dabbaghian, V., & Finegood, D. F. (2012). Modeling the influence of social networks and environment on energy balance and obesity. *Journal of Computational Science*, *3*(1-2), 17–27.

Godfrey, K. M., Sheppard, A., Gluckman, P. D., Lillycrop, K. A., & et al. (2011). Epigenetic Gene Promoter Methylation at Birth Is Associated With Child's Later Adiposity. *Diabetes*, *60*(5), 1528–1534.

Goodarzi, M. O. (2018). Genetics of obesity: What genetic association studies have taught us about the biology of obesity and its complications. *The Lancet Diabetes & Endocrinology*, *6*(3), 223–236.

Hales, C., & Barker, D. (1992). Type 2 (non-insulin-dependent) diabetes mellitus: The thrifty phenotype hypothesis. *Diabetologia*, *35*(7), 595–601.

Hebebrand, J., Wulftange, H., Goerg, T., Ziegler, A., Hinney, A., Barth, N., Mayer, H., & Remschmidt, H. (2000). Epidemic obesity: Are genetic factors involved via increased rates of assortative mating? *International Journal of Obesity*, *24*(3), 345–353.

Hedrick, P. W. (2016). Assortative Mating and Linkage Disequilibrium. *G3: Genes—Genomes—Genetics*, *7*(1), 55–62.

Herrera, B. M., Keildson, S., & Lindgren, C. M. (2011). Genetics and epigenetics of obesity. *Maturitas*, *69*(1), 41–49.

Huang, H., Yan, Z., Chen, Y., & Liu, F. (2016). A social contagious model of the obesity epidemic. *Scientific Reports*, *6*(1), 37961.

Huangfu, Y., Palloni, A., Beltrán-Sánchez, H., & McEniry, M. C. (2023). Gene–environment interactions and the case of body mass index and obesity: How much do they matter? (A. Bovell-Benjamin, Ed.). *PNAS Nexus*, *2*(7), pgad213.

Jacobson, P., Torgerson, J. S., Sjostrom, L., & Bouchard, C. (2006). Spouse Resemblance in Body Mass Index: Effects on Adult Obesity Prevalence in the Offspring Generation. *American Journal of Epidemiology*, *165*(1), 101–108.

Kempf, E., Landgraf, K., Stein, R., Hanschkow, M., & et al. (2022). Aberrant expression of agouti signaling protein (ASIP) as a cause of monogenic severe childhood obesity. *Nature Metabolism*, *4*(12), 1697–1712.

Keyfitz, N. (1971). On the momentum of population growth. *Demography*, *8*(1), 71–80.

Keyfitz,N. (1977). *Introduction to the Mathematics of Population (with revisions)*. Addison-Wesley Series in Behavioral Science.

Khera, A. V., Chaffin, M., Wade, K. H., Zahid, S., & et al. (2019). Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*, *177*(3), 587–596.e9.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., & et al. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424–428.

Levy, D. T., Mabry, P. L., Wang, Y. C., Gortmaker, S., & et al. (2011). Simulation models of obesity: A review of the literature and implications for research and policy: Simulation models of obesity. *Obesity Reviews*, *12*(5), 378–394.

Ling, C., & Rönn, T. (2019). Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metabolism*, *29*(5), 1028–1044.

Loos, R. J. F., & Yeo, G. S. H. (2022). The genetics of obesity: From discovery to biology. *Nature Reviews Genetics*, *23*(2), 120–133.

Macchi, E. (2022). *Worth your weight: Experimental evidence on the benefits of obesity in low-income countries*.

McAllister, E. J., Dhurandhar, N. V., Keith, S. W., Aronne, L. J., & et al. (2009). Ten Putative Contributors to the Obesity Epidemic. *Critical Reviews in Food Science and Nutrition*, *49*(10), 868–913.

Mills, M. C., Barban, N., & Tropf, F. C. (2020). *An Introduction to Statistical Genetic Data Analysis*. MIT Press.

Morshed, A. B., Kasman, M., Heuberger, B., Hammond, R. A., & Hovmand, P. S. (2019). A systematic review of system dynamics and agent-based obesity models: Evaluating obesity as part of the global syndemic. *Obesity Reviews*, *20*(S2), 161–178.

Neel, J. V. (1999). Diabetes mellitus: A "thrifty" genotype rendered detrimental by "progress"? 1962. *Bull World Health Organ*, *77*(8), 694–703, discussion 692–3.

Pampel, F. C., Denney, J. T., & Krueger, P. M. (2012). Obesity, ses, and economic development: A test of the reversal hypothesis. *Social science & medicine (1982)*, *74*(7), 1073–1081.

Popkin, B. M. (2011). Contemporary nutritional transition: Determinants of diet and its impact on body composition. *The Proceedings of the Nutrition Society*, *70*(1), 82–91.

Popkin, B. M., Corvalan, C., & Grummer-Strawn, L. M. (2020). Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet*, *395*(10217), 65–74.

Power, C., Pouliou, T., Li, L., Cooper, R., & Hyppönen, E. (2011). Parental and offspring adiposity associations: Insights from the 1958 British birth cohort. *Annals of Human Biology*, *38*(4), 390–399.

Preston, S. H., & Campbell, C. (1993). Differential Fertility and the Distribution of Traits: The Case of IQ. *American Journal of Sociology*, *98*(5), 997–1019.

Qasim, A., Turcotte, M., de Souza, R. J., Samaan, M. C., & et al. (2018). On the origin of obesity: Identifying the biological, environmental and cultural drivers of genetic risk among human populations. *Obesity Reviews: An Official Journal of the International Association for the Study of Obesity*, *19*(2), 121–149.

Rohde, K., Keller, M., la Cour Poulsen, L., Blüher, M., & et al. (2019). Genetics and epigenetics in obesity. *Metabolism*, *92*, 37–50.

Sermondade, N. (2012). Obesity and Increased Risk for Oligozoospermia and Azoospermia. *Archives of Internal Medicine*, *172*(5), 440.

Sharp, G. C., Salas, L. A., Monnereau, C., Allard, C., & et al. (2017). Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: Findings from the pregnancy and childhood epigenetics (PACE) consortium. *Human Molecular Genetics*, *26*(20), 4067–4085.

Shoham, D. A., Hammond, R., Rahmandad, H., Wang, Y., & Hovmand, P. (2015). Modeling Social Norms and Social Influence in Obesity. *Current Epidemiology Reports*, *2*(1), 71–79.

Silvestris, E., de Pergola, G., Rosania, R., & Loverro, G. (2018). Obesity as disruptor of the female fertility. *Reproductive Biology and Endocrinology*, *16*(1), 22.

Speakman, J. R. (2013). Evolutionary Perspectives on the Obesity Epidemic: Adaptive, Maladaptive, and Neutral Viewpoints. *Annual Review of Nutrition*, *33*(1), 289–317.

Speakman, J. R. (2018). The evolution of body fatness: Trading off disease and predation risk (R. K. Suarez & H. H. Hoppeler, Eds.). *Journal of Experimental Biology*, *221*(Suppl_1), jeb167254.

Speakman, J. R., Djafarian, K., Stewart, J., & Jackson, D. M. (2007). Assortative mating for obesity. *The American Journal of Clinical Nutrition*, *86*(2), 316–323.

Swan, S. H., & Colino, S. (2021). *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development, and Imperiling the Future of the Human Race*.

Swinburn, B. A., Sacks, G., Hall, K. D., McPherson, K., Finegood, D. T., Moodie, M. L., & Gortmaker, S. L. (2011). The global obesity pandemic: Shaped by global drivers and local environments. *The Lancet*, *378*(9793), 804–814.

Uchiyama, R., Spicer, R., & Muthukrishna, M. (2021). Cultural Evolution of Genetic Heritability. *Behavioral and Brain Sciences*, 1–147.

Vazquez, C. E., & Cubbin, C. (2020). Socioeconomic status and childhood obesity: A review of literature from the past decade to inform intervention research. *Current Obesity Reports*, *9*(4), 562–570.

Vickers, M. (2014). Early Life Nutrition, Epigenetics and Programming of Later Life Disease. *Nutrients*, *6*(6), 2165–2178.

Walter, S., Mejía-Guevara, I., Estrada, K., Liu, S. Y., & Glymour, M. M. (2016). Association of a Genetic Risk Score With Body Mass Index Across Different Birth Cohorts. *JAMA*, *316*(1), 63.

Wang, K., Li, W.-D., Zhang, C. K., Wang, Z., & et al. (2011). A Genome-Wide Association Study on Obesity and Obesity-Related Traits (Z. Zhao, Ed.). *PLoS ONE*, *6*(4), e18939.

Waterland, R. A., & Jirtle, R. L. (2003). Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. *Molecular and Cellular Biology*, *23*(15), 5293–5300.

Wells, J. C. K. (2011). An evolutionary perspective on the trans-generational basis of obesity. *Annals of Human Biology*, *38*(4), 400–409.

Wells, J. C. K. (2012). The evolution of human adiposity and obesity: Where did it all go wrong? *Disease Models & Mechanisms*, *5*(5), 595–607.

Willyard, C. (2014). Heritability: The family roots of obesity. *Nature*, *508*(7496), S58–S60.

Yengo, L., Robinson, M. R., Keller, M. C., Kemper, K. E., & et al. (2018). Imprint of assortative mating on the human genome. *Nature Human Behaviour*, *2*(12), 948–954.

Zietsch, B. P., Verweij, K. J. H., Heath, A. C., & Martin, N. G. (2011). Variation in Human Mate Choice: Simultaneously Investigating Heritability, Parental Influence, Sexual Imprinting, and Assortative Mating. *The American Naturalist*, *177*(5), 605–616.

Supplemental Material

The code to reproduce the analyses performed in the paper is available in the repository https://github.com/palloni/obesity-leslie-abm.

1 Description of Leslie Matrices

1.1 Matrix of assortative mating

The general form of the 16x4 mating matrix N is the following:

$$N = \begin{pmatrix} p_{11} & \cdots & p_{14} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & p_{21} & \cdots & p_{24} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & p_{31} & \cdots & p_{34} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & p_{41} & \cdots & p_{44} \end{pmatrix} \tag{S1}$$

where p_{ij} is the probability for an individual of obesity category i to mate with another from obesity category j, as defined in equations (5) and (6) of the main text.

1.2 Matrix of fertility

The fertility matrix R is 16x16 diagonal. Its entries are the fertility rates (children per woman) for the 16 classes of couples:

$$R = \begin{pmatrix} f_{11} & 0 & \cdots & 0 \\ 0 & f_{12} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & f_{44} \end{pmatrix} \tag{S2}$$

where f_{ij} is the mean number of children produced by a couple in which one individual is in obesity category i and the other in j. The values of f_{ij} are defined in equation (7) of the main text.

1.3 Heritability matrix

The heritability matrix, $H(t)$ is of size 16x4. Its entries are the obesity categories-specific probabilities of generating offspring in each obesity category:

$$H(t) = \begin{pmatrix} h_{111} & h_{121} & \cdots & h_{441} \\ h_{112} & h_{122} & \cdots & h_{442} \\ h_{113} & h_{123} & \cdots & h_{443} \\ h_{114} & h_{124} & \cdots & h_{444} \end{pmatrix} \tag{S3}$$

where h_{ijk} is the probability that a child of couple whose members are in the obesity categories i and j, belongs to the obesity category k. The entires h_{ijk} are computed either using Mendelian segregation rules or from PRS and the BMI-PRS relationship as defined in section 3.1.3 of the main text.

When we use Mendelian segregation rules we can only determine two different phenotypes, obese and non-obese. We then classify the non-obese into three non-obese BMI categories ($k \leq 3$) using information on parental obesity. If a non-obese offpsring has parents in obesity categories i and j we define the probabilities of belonging to obesity categories $k \leq 3$ as follows.

Let q_1 be the conditional probability of a non-obese offspring to a (i, j) couple be in category $k = 1$. We define this as

$$p(k = 1|non - obese_{i,j}) = q_1 = 0.1 - 0.015 \cdot (i + j - 2) \tag{S4}$$

Define q_2 as

$$q_2 = 0.125 \cdot (i + j - 1) \tag{S5}$$

and then

$$p(k = 2|non - obese_{i,j}) = (1 - q_1) \cdot (1 - q_2) \tag{S6}$$

with

$$p(k=3|non-obese_{i,j}) = (1-q_1)\cdot q_2 \tag{S7}$$

2 The ABM model

Let $t = 0, 1, ...$ be an index for generation number, $BMI_{i,t}$ the Body Mass Index of agent i in generation t and $PRS_{i,t}$ the BMI's Polygenic Risk Score of agent i in generation t. We start with a population of N_o agents at time t.

2.1 Generation-specific PRS

For agents in the initial generation, $(t = 0)$, the value of the Polygenic Risk Score, PRS_{it} is defined by a random normal variate with mean 1 and standard deviation 0, $\varepsilon(i; 0, 0.7)$. For an offspring i in subsequent generations, $t > 1$, the PRS is calculated as:

$$PRS_{off} = \frac{PRS_{par1} + PRS_{par2}}{2} + \eta(\mu = 0, \sigma = 0.7) \tag{S8}$$

where $PRS^{f}_{i,t-1}$ and $PRS^{m}_{i,t-1}$ are the PRS's of the generation (t-1) father's and mother's PRS, respectively.

2.2 Baseline BMI

The BMI for an i who is an offspring in generation t is given by

$$BMI = 27.796 + 1.071 \times PRS + \varepsilon(\mu = 0, \sigma = 4.55) \tag{S9}$$

2.3 Fertility Differentials

As in the case of the matrix approach, we assume that the number of a couple's offspring in generation t is a function of the combination of parental obesity categories in generation $t-1$:

$$f_{ij,t} = 2 + (i + j - 6) \cdot \frac{\varphi}{2} \tag{S10}$$

where $\varphi \in [0,1]$ is the parameter for the size of fertility differentials. The number of children for each class of couples when $\varphi = 1$ are displayed in Table S1[29].

Table S1: Mean number of children by couple as a function of parents' BMI category

BMI group (i) *(parent 1)*	BMI group (j) *(parent 2)*	Number of children
1	1	0
1	2	0.5
1	3	1
1	4	1.5
2	1	0.5
2	2	1
2	3	1.5
2	4	2
3	1	1
3	2	1.5
3	3	2
3	4	2.5
4	1	1.5
4	2	2
4	3	2.5
4	4	3

2.4 Assortative Mating

The way to determine the number of random pairs and, consequently, the number of endogamous pairs is calculated as follows:

[29] This table is useful to compute relation between the values of the input parameter φ and empirically observed quantities. See Section 3.2 of this Supplement

$$\text{Coup}_{rand} = (1 - \omega) \times \frac{N_t}{2} \tag{S11}$$

where Coup_{rand} are the number of random couples created, N_t is the total sample and ω measures the intensity of assortative mating. For instance, when $\omega = 0$, the number of random couples created will be half the sample, so the entire sample will be randomly mated. On the other hand, when $\omega = 1$ the value of Coup_{rand} will equal to 0 and every couple will be formed endogamously, by individuals from the same obesity category.

3 Empirical justification of parameters

The value of the input parameters we used throughout are not educated guesses but correspond closely with empirical observation.

3.1 Assortative mating

The values of the input parameter ω partially reflects empirical values based on information from two sources. The first was the creation of artificial data set containing the distribution of BMI for 10,000 females. Their BMI was a normal variate with mean 27 and a standard deviation of 3. We then computed the BMI distribution of their 'partners' using predicted values from a regression with slopes, β, that varied between .10 to .70, normally distributed errors, ε with 0 and standard deviations ranging from 1 to 5. In all, we computed 35 different partners' BMI distributions. The R^2 values of the predictive regressions ranged from .004 to .81. In each case, we computed two additional quantities: (i) the χ^2 value of a cross-tabulation of females and their partners using the four obesity categories we have employed throughout and (b) the odds ratios, OR_{ij} of females in the i^{th} category of having partners in the j category.

With the artificial values of p_{ii} and π_i available from each of the 35 pairs of BMI distributions, we computed the value of ω using the expression

$$\omega = (p_{ii} - \pi_i)/(1 - \pi_i) \tag{S12}$$

We then estimate three regression models to sequentially express ω as a function of β, R^2 and χ^2.

The final step consists of searching the recent literature and identifying estimates of either *beta*, R^2, χ^2, or combinations thereof. Estimates of ω's consistent with a study's parameter(s) are then calculated using one of the three regression models defined before.

Column 5 of S2 displays estimates of ω obtained from each of the studies identified in the first column using the parameter estimate in column 4. The range spanned by omega estimates is [.06-0.25] or the first 25 percent of the range of omega used in simulation. Thus, our inference regarding the *modest* direct role of AM shaping additive vertical genetic heredity (via increased LD or homozygosity) and, more generally, the global obesity epidemic, is probably an overstatement.

3.2 Differential fertility

To establish an empirically plausible range for the differential fertility parameter ω we relied on the observed relation between maternal obesity and number of children ever born (controlling for maternal age). We used information for multiple sources: 33 African Demographic Health Surveys, the US National Longitudinal Survey of Youth (NLSY) as well as the National Health and Nutrition Examination Survey (NHANES). The populations included in our surveys represent a very broad range of countries at different stages of the obesity epidemic, from those in which it is not yet discernible to those that have attained relatively high values (though not the highest). Admittedly, this is not an ideal data set because in all cases it lacks information on the father or partner. But, alas, there are no national data that include anthropometry of both members of a couple couple. To approximate the values of the parameter φ we first estimated regressions of the number of children ever born using a dummy variable for obesity and controls (age and education levels).

Table S2: Observed within-couple correlations of BMI/weight and omega

Study	**Population**	**Type of measure**	**Parameter estimate**	**Omega**
Ajslev et al. (2012)	37,792 pairs (Copenhangen)	BMI distributions	-	0 -.012
Allison et al. (1996)	Multiple studies	Couples' correlation weight	0.10-0.33	0.062-0.206
Hebebrand et al. (2000)	128-150 couples German National Nutritional Survey	BMI distributions	-	0.16-0.20
Katzmarzyck et al. (2002)	1341 parents Canadian Fitness Survey (1981)	BMI rank correlation	0.14	0.087
Sjaarda & Kutalik (2022)	51664 couples UK Biobank	Weight correlation	0.25	0.155
Speakman et al. (2007)	42 couples Northeast Scotland	BMI	0.33	0.206
Authors' estimate from HRS	1,568 couples in 2006 wave	BMI distributions	-	0.06
Authors' estimate from HRS	All couples - all waves (up to 2020)	BMI distributions	-	0.11
Authors' estimate from DHS India 2019-20	38,857 couples	BMI distributions	-	0.12

Estimates from Allison et al (1996) correspond to correlation of partners' weight (not BMI). The range of values (0.10-0.33) includes Allison's et al. own and those and those from 29 different studies in sub-populations from USA, UK, Italy, Sweden, Norway, Denmark, Brazil, Peru, Israel (see Table 1 in Allison et al. (1996)).

The estimated effect of the dummy variable, β and φ are related by the following expressions:

$$\beta_{max} = \varphi \times 3 \tag{S13}$$

$$\beta_{min} = \varphi \times .5 \tag{S14}$$

where β_{min} and β_{max} are the minimum and maximum values consistent with a value of φ.

Table S3: Estimates of Relations between Children Ever Born and Maternal Obesity

Country	Source	Population	β estimate	Adult Female Prevalence
India	DHS	All	-0.381	0.07
		Low Ed	-0.181	
		High Ed	0.05	
Turkey	DHS	All	0.604	0.41
		Low Ed	0.054	
		High Ed	0.496	
Africa[1]	DHS	All	-0.792	0.017-0.150
		Low Ed	-0.371	
		High Ed	0.206	
Asia[1]	DHS	All	-0.157	0.010-0.044
		Low Ed	-0.237	
		High Ed	0.168	
USA	NLSY 1997	All	0.246	0.37
	NHANES	All	0.323	0.37

1. Africa includes 33 DHS samples and Asia 8 DHS samples.
2. All regression coefficients are significant at $p < 0.001$.
3. Source of obesity prevalence estimates: https://ncdrisc.org/obesity-prevalence-ranking.html

Table S3 displays values of β from sample surveys of some populations and subpopulations. The figures in this table confirm that the value of φ in the middle of the range we are using is consistent with minimum and maximum β values of .25 and 1.5 respectively. Because the range of β values is approximately [.20,.50], they are consistent with φ in the range [.067, 1].[30].

An important feature of the table is the association between the magnitude of β's and the

[30]The maximum and minimum values were computed using Table S1

population prevalence of obesity (last column). In particular, there is a strong positive relation (close to that observed in the US) among females with highest education in countries with the lowest prevalence of obesity (Africa, India). This is consistent with the idea that as the obesity epidemic advances, there is a transition from a negative relation between obesity and fertility to a positive one.

3.3 Penetrance and GxE effects

The parameter p is defined as the excess penetrance (relative to a baseline) and we have employed it to reflect possible GxE interactions. The values we employ, [0-.20], are lower than those that have been observed in several high income countries. In the US for example, empirical estimates of p are in the range of .06 to 1.10 (Huangfu, Palloni, Beltrán-Sánchez, & McEniry, 2023). In particular, data from HRS suggests a p value of the order of .40, twice as large as the maximum we use as input. From this we may conclude that our simulation probably underplay the impact that GxE interactions may have on the population distribution by obesity categories.

4 Additional results

Two additional results of the simulations are described below

1. **Heritability** To what an extent does the genetic heritability of the phenotype (BMI) changes across generations and how are those changes related to alternative regimes of AM, FD and PE? To investigate this issue we compute standard heritability ratios as

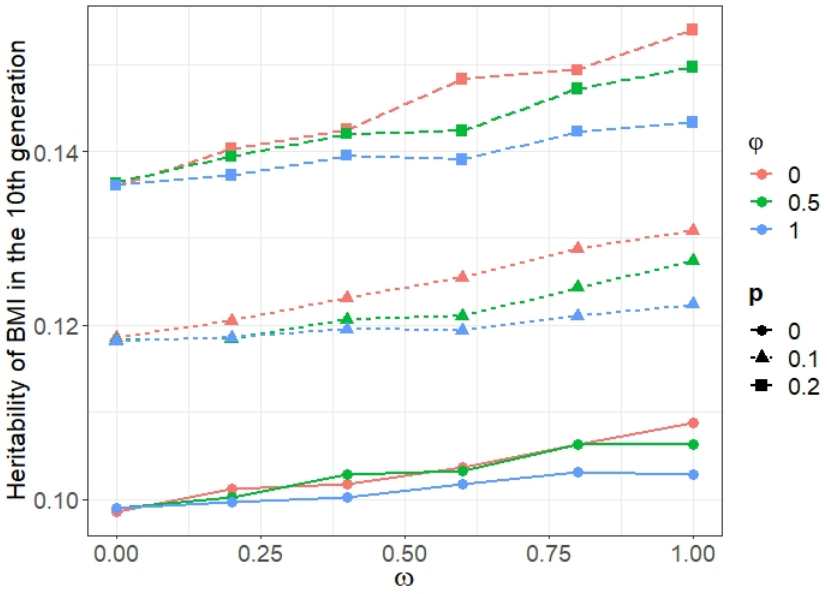
$$H^2 = [\alpha \times (1+p)]^2 \times (V(PRS)/V(BMI)) \tag{S15}$$

where $V(PRS)$ and V(BMI) are the variance of the PRS and BMI.

Figure S1 displays the values of h^2 evaluated at the 10th generation for combinations of AM, FD and PE regimes. Its value ranges between .10 and .15. As expected, h^2 increases with the

strength of assortative mating. These increases appear to be similar for all combinations of PE and FD. Also as expected, the largest values are found for the highest levels of PE. The impact of FD is smaller but also in the expected direction, namely, as fertility differentials contract, h^2, increases slightly. Lastly, just as in the case of obesity prevalence and mean BMI, there is an interaction effect between fertility differentials and assortative mating.

Figure S1: Heritability of BMI in the 10th generation as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.



2. **Effects of interventions**: Suppose we are interested in evaluating the cost-efficiency of an intervention to reduce the prevalence of obesity. For example, we know that obesogenic environments may spike the penetrance of genetic variants associated with higher BMI or increased risks of obesity. One could design interventions that reduce or eliminate exposures to those environments, e.g. the equivalent of lowering the magnitude of the parameter α, that is, reducing p to bring α closer to the baseline value of genetic penetrance. Table S4 shows the effect of an intervention on prevalence of obesity on the 10th generation,. The intervention consists of reducing reducing exposure to obesogenic environments that translates into a reduction of p from 0.2 to 0 in the sixth and following generations. The table displays relative reductions under different combination of initial AM and FD regimes.

Obviously, in all combinations of regimes there are reductions as they range between 1 percent to 8 percent. The most salient ones take place under pre-intervention regimes of high assortative mating and fertility differentials.

Table S4: Prevalence of obesity from Leslie matrices with and without an intervention, as a function of ω & φ

		$\omega = 0$	$\omega = 0.4$	$\omega = 1$
$\varphi = 0$	With intervention	.323	.322	.323
	without intervention	.326	.326	.328
	Change	-1%	-1%	-2%
$\varphi = 0.5$	With intervention	.353	.366	.385
	without intervention	.365	.382	.411
	Change	-3%	-4%	-6%
$\varphi = 1$	With intervention	.382	.405	.438
	without intervention	.400	.429	.476
	Change	-4%	-6%	-8%

5 Tables and Figures

Figure S2: Mean BMI after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.

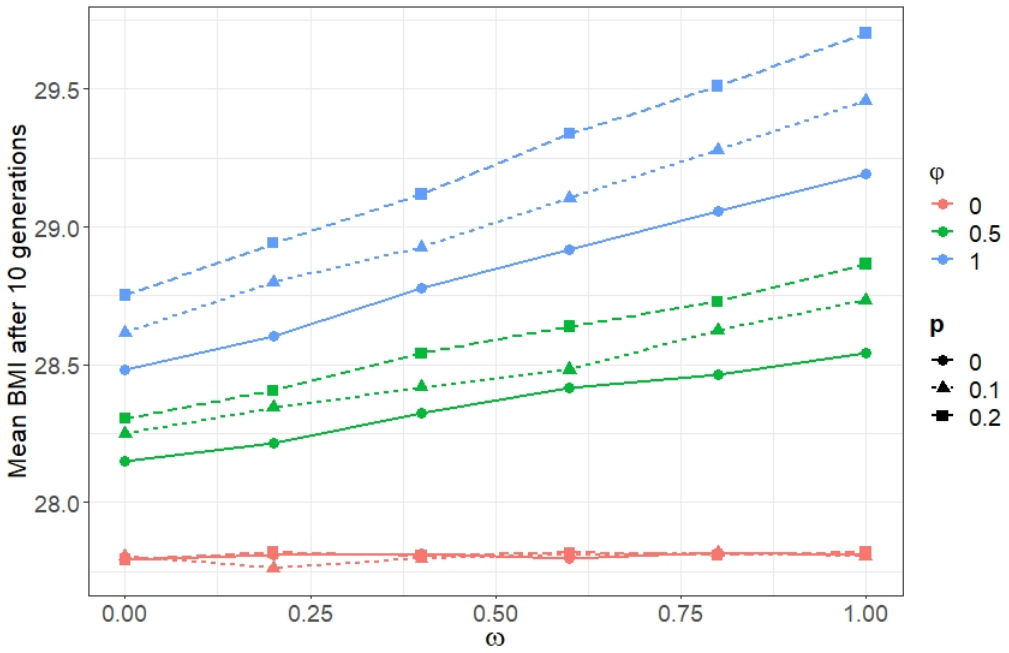


Figure S3: Standard deviation of BMI after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.

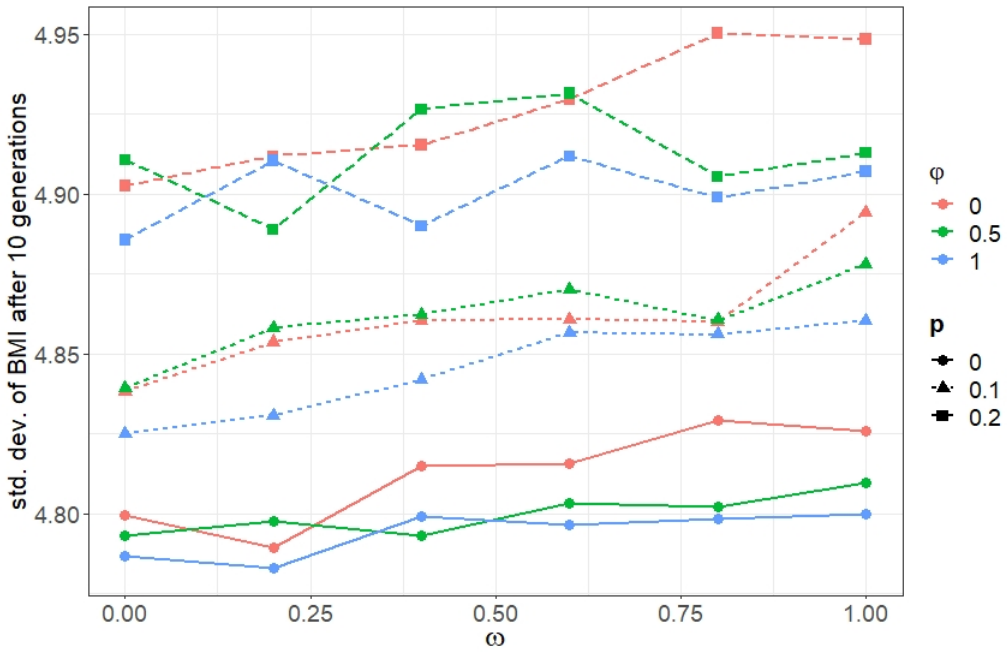


Figure S4: Mean PRS after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.

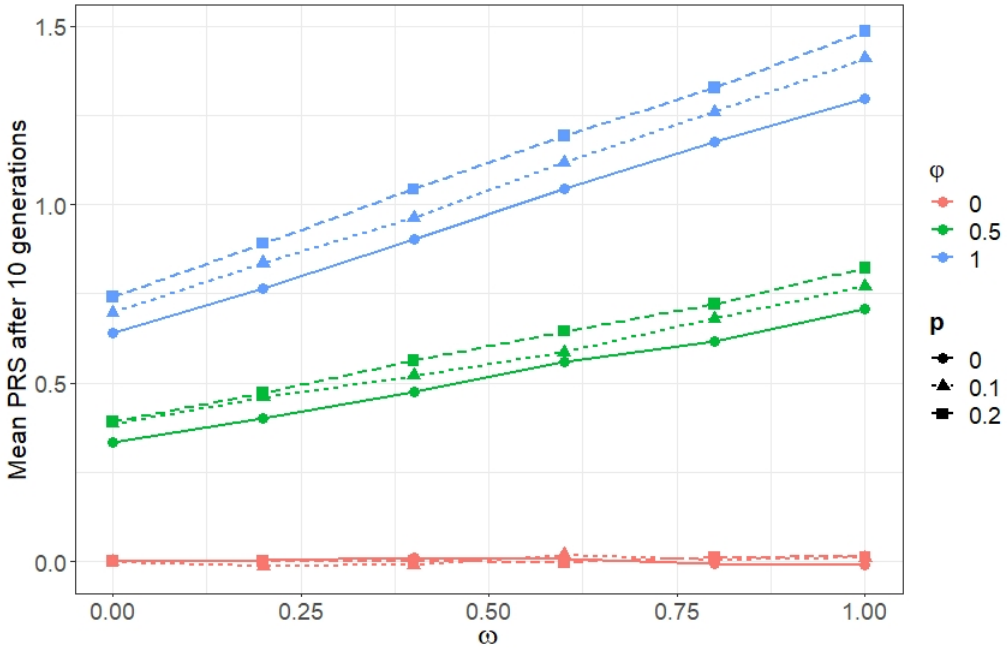


Figure S5: Standard deviation of PRS after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.

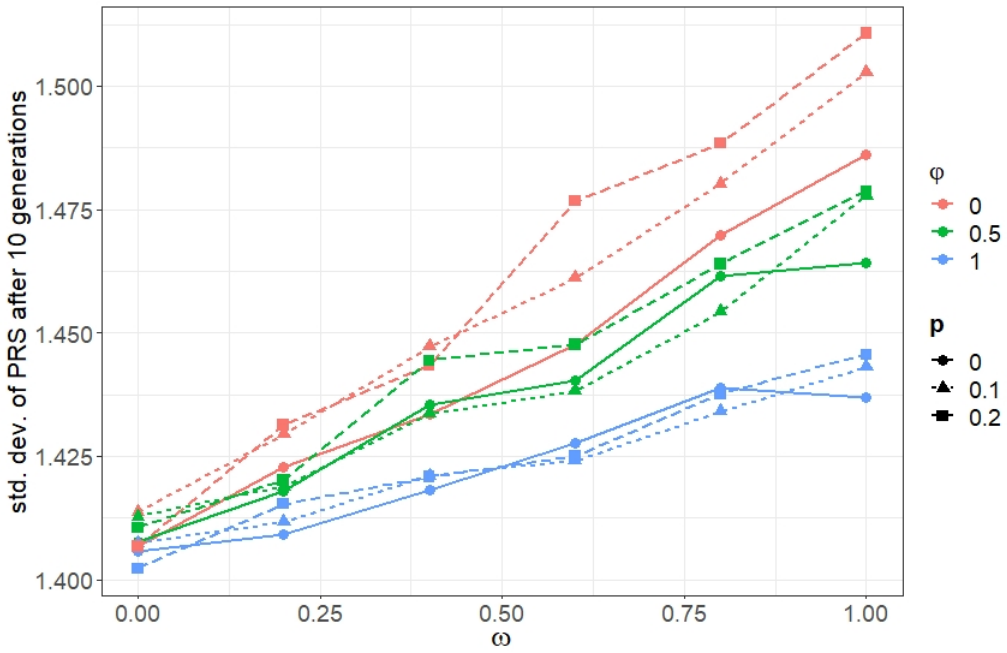


Table S5: Coefficients of regression of $\log(\sigma(\text{BMI}))$ and $\log(\sigma(\text{PRS}))$ for ABM and Leslie matrices

	$\log(\sigma(\text{BMI}))$		$\log(\sigma(\text{PRS}))$	
	ABM	**LM**	**ABM**	**LM**
Const.	1.570***	1.567***	0.361***	0.342***
$\beta(\omega)$	**0.004*** **	**0.007** **	**0.031*** **	**0.054*** **
$\beta(\varphi)$	0.001	-0.001	**-0.009*** **	-0.001
$\beta(p)$	**0.114*** **	**0.113*** **	0.034	0.012
$\beta(\omega \times \varphi)$	**-0.004*** **	-0.003	**-0.018*** **	**-0.029*** **
$\beta(\omega \times p)$	**0.014*** **	0.010	**0.065**	**0.075*** **
$\beta(\varphi \times p)$	-0.006	0.002	**-0.032*** **	-0.013
$\beta(\omega \times \varphi \times p)$	-0.005	-0.017	**-0.039*** **	**-0.065***
N	7,944	1,080	7,944	1,080
R^2	0.628	0.504	0.793	0.705
Adj. R^2	0.628	0.500	0.793	0.703
RSE	0.007	0.010	0.008	0.012
df	7,936	1,072	7,936	1,072

The N for LM models corresponds to 54 combinations of parameter values and 20 replicas for each of them. The N for ABM models corresponds to 27 combinations of parameter values and 300 replicas for each.

Figure S6: Mean PRS by BMI category after 10 generations as a function of AM (ω), differential fertility (φ), and genetic penetrance (p), Leslie matrices.