Methods S1: Statistical methods

An individual's body mass index (BMI) is defined as:

,

where and correspond to individual's weight and height, respectively. provides a simple measure of a person's “fatness” or “thinness”. Although is measured on a continuous scale, it is grouped in the following five categories:

1. : <18.5 (underweight)
2. from 18.5 to 24.99: (healthy weight)
3. from 25 to 19.99: (overweight)
4. from 30 to 39.99: (obese)
5. : (morbidly obese)

Let =1,...,5, denote group (lower and larger values of correspond to groups with lower and larger values of on the continuous scale, respectively). Let be the prevalence of individuals with values that correspond to group at time . To ensure that takes values within [0,1], we model by

 (1)

An alternative way to estimate is through a logistic regression model with percentage for group as the outcome, and time, , as the single explanatory variable:

 (2)

By solving equation (2) for we obtain

Equations (1) and (2) are mathematically identical, but (1) provides greater numerical stability. Fitting separate models for the estimation of for each group g, does not guarantee that

.

Thus, to ensure that the estimates of prevalence of individuals in all groups sum up to 1 for each time , we estimate the prevalence of individuals with values that correspond to group at time by

 (3)

**Micro simulation – BMI growth model**

The distribution of BMI in the population is estimated using regression analysis stratified by both sex (S) and age group (A=0-9, 10-19,20-29,…,70-79, 80+). The fitted models are extrapolated to forecast the distribution of BMI groups in the future. For each sex-and-age-group stratum, the set of cross-sectional, time-dependent, discrete distributions of BMI groups, , is used to manufacture BMI growth models for individual members of the population. This is done in a way that guarantees that the cross-sectional BMI group distributions obtained by simulation under the growth models match the BMI group distributions of the observed data. The details are as follows;

For each such discrete distribution in , there is a continuous counterpart. Let denote BMI in the continuous scale and let be the probability density function of for age group and sex at time . Then

 (4)

Equations (3) and (4) both refer to the same quantity. However, equation (4) uses the definition of a probability density function to express the age-and-sex-specific percentage of individuals in BMI group g at time t. Equation (3) gives an estimate of the this quantity using model (1) for all g=0,..,5. The cumulative distribution function of is

 (5)

At a time t, a person with sex belonging to the age group is said to be on the –th percentile of this distribution if Given the cross-sectional information of , it is possible to simulate longitudinal trajectories by forming pseudo cohorts within the population. A key requirement for these sets of longitudinal trajectories is that they reproduce the cross-sectional distribution of BMI groups for any year with available data. The method adopted here and in the earlier Foresight report[[1]](#footnote-1) is based on the assumption that people’s BMI changes throughout their lives in such a way that they always stay on the same BMI percentile. As they age, individuals move from one age group to another and their BMI changes so that they remain on the same percentile but of a different distribution. This rule is not too far from the truth; and has as a result that relatively fat people stay relatively fat and relatively thin people stay relatively thin. Crucially it meets the important condition that the cross-sectional BMI group distributions obtained by simulation match the BMI group distributions of the observed data. This can be seen as follows:.

When the population’s BMI distributions by sex and age are known for all years (by extrapolation of fitted model (3)), a person who is in age group and who grows ten year older will at some time move into the next age group and will have a BMI that was described first by the distribution and then at the later time by the distribution . If the BMI of that individual is on the pth percentile of the BMI distribution, his BMI will change from to so that

 (6)

 (7)

Where is the inverse function of the cumulative distribution function of . Equation (7) guarantees that the transformation taking the random variable to ensures the correct cross-sectional distribution at .

The micro simulation both generates individuals from the BMI distributions of the set and, once generated, grows the individal’s BMI in a way that is also determined by the set . It is possible to implement equation (7) as a suitably fast algorithm.

**Micro simulation: Birth, disease and death models**.

Simulated people are generated with the correct demographic statistics in the simulation’s start-year. In this year women are stochastically allocated the number and years of birth of their children – these are generated from known fertility and mother’s age at birth statistics (valid in the start-year). If a woman has children then those children are generated as members of the simulation in the appropriate birth year.

In the course of their lives, simulated people can die from one of the BMI related diseases that they might have acquired or from some other cause. The probabilities that a person of a given age and gender dies from a cause other than a BMI related disease are calculated in terms of known death and disease statistics valid in the start-year and are held constant over the course of the simulation. The death rates from BMI related diseases will change as a consequence of the population’s changing BMI distribution.

1. Foresight report [↑](#footnote-ref-1)