Web Appendix: Constructing an agent-based model of obesity in England

Abdulrahman M. El-Sayed, DPhil,[1,2,3] Lars Seemann, MS,[4] Peter Scarborough, DPhil,[1] and Sandro Galea, MD, DrPH[2]

1. Department of Public Health, University of Oxford, Oxford, UK
2. Department of Epidemiology, Columbia University, New York, NY, USA
3. College of Physicians and Surgeons, Columbia University, New York, NY, USA
4. Department of Physics, University of Houston, Houston, USA

Section 1. Summary and Introduction

The agent-based model employed here was intended to simulate the social transmission of obesity among a representative cohort in England throughout the life course between ages 18 and 70 years. The aim was to create a model to study how social factors operating at the levels of the individual and her environment in a socially networked reality might interact to produce population obesity in England, rather than to forecast the absolute prevalence of obesity into the future, as others have done.[1] Therefore, the process of framing, parameterizing, and constructing this model, as well as the interpretation of its output were focused around the nature of *relationships* and *interactions* underlying the production of obesity.

The model (heretofore, "Obesagent") featured a cohort of 10,000 agents of similar age (similar to a birth cohort). Initial conditions for the baseline model were as follows: each agent was stochastically assigned gender, ethnicity, social class, and education-level consistent with correlated distributions of each characteristic in England, such that, for example, the proportion of South Asian, Black, and White agents mimicked that of the English population, overall. Each agent was nested within one of four spatial contexts, representing different ethnic density (i.e., proportion non-white) and deprivation (i.e., proportion manual social class) compositions, and was placed in these contexts preferentially by demographic characteristics (ethnicity and social class). Agents were also nested within a social network with preferential mixing between demographically similar agents. Moreover, a proportion of the agents were assigned obesity at outset based on demographic and neighborhood allocation, such that the distribution of obesity was similar to that among the 18 year-old population in England. At the outset, each agent represented an 18 year-old individual, and aged by one year each time step. At each time-step, each agent's risk for developing obesity in that time-step, as a function of its gender, ethnicity, social class, education, social contacts, and social context (residence in areas defined by ethnic density and social deprivation) was calculated and implemented.

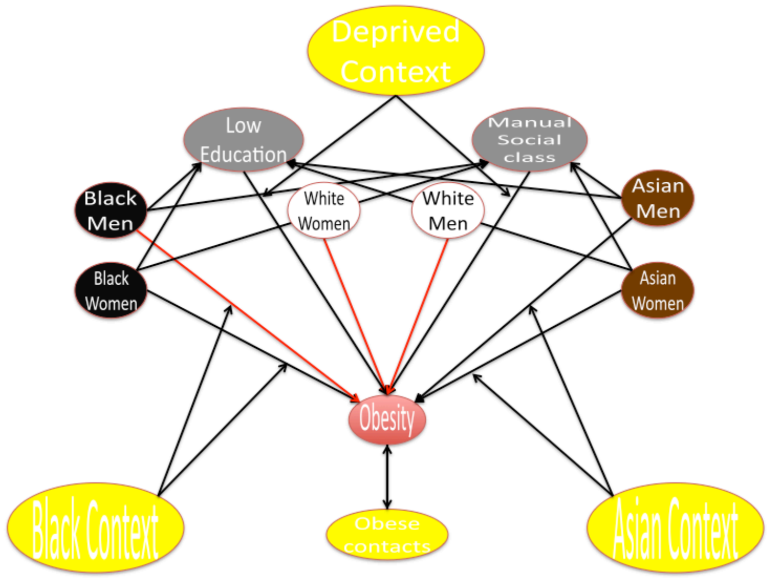
The following describes the process of constructing Obesagent. We describe first the process of designing the conceptual framework that underpinned the model. Second, we describe the Obesagent parameterization protocol. And third, we describe the process of implementing the model.

Section 2. Conceptual Framework

Obesagent was informed by a conceptual framework that is concerned with the social transmission of obesity in England. This framework attempted to account for the social transmission of obesity at several levels of influence, including individual, spatial, and network social determinants of obesity, while limiting its scope to those factors that were 1) measurable and operationalizable, and 2) demonstrated predictors of obesity in the epidemiologic literature. This framework was developed, for the most part, from the findings of two systematic reviews carried out by the authors previously.[2,3] However, so as not to exclude potential determinants of obesity that have not yet been studied in the UK, and therefore, that were excluded from the literature reviewed, we also drew upon available literature in other contexts. For example, social networks were demonstrated to play a role in the etiology of obesity using data from the Framingham Heart Study in the US.[4] Similarly, the conceptual link between ethnic density and differential obesity risk among ethnic minority and majority groups comes from findings that demonstrated this link among Hispanics in the US[5] (although Health Survey for England data were used to derive this parameter, as described below). It is important to note, however, that while we included ethnic and socioeconomic contexts as predictors of obesity in the Obesagent model, we did not intend to model the complex social production of these factors themselves, for which our parameterization was not equipped.

Web Figure 1 shows the Obesagent conceptual framework.

Web Figure 1. Conceptual framework employed in the construction of the Obesagent model



Black arrows represent a direct effect, while red arrows represent an inverse effect. All individual social factors (i.e., ethnicity, social class, and education) influenced obesity risk directly. Obesogenesis among network contacts also influenced obesity risk directly. Contextual factors (i.e., ethnic density and deprivation) interacted with individual ethnicity and socioeconomic metrics, respectively, to modulate obesity risk at the model outset.

Section 3. Parameterization

3.1 Available data for parameterization

A central feature of agent-based models (ABMs) is that they update in a time-dependant manner. Therefore, beyond isolating cross-sectional relations between factors in the model, a central requirement in parameterizing ABMs is isolating *time-dependent* relations between factors in the model.

This was particularly challenging for the purposes of this model, which sought to understand the social etiology of obesity, as there is a paucity of longitudinal datasets that include representative data about ethnic minority populations in England. Although data from the 1946 (National Survey of Health and Development[6]) or 1958 (National Child Development Study[7]) birth cohorts presented the ideal data structures for parameterization (i.e., longitudinal datasets from birth), ultimately, these data were not pursued because there have been substantial demographic changes in England since the mid 20th century, and data about ethnic minorities included in either survey do not accurately represent the ethnic minority population in England today.[8, 9]

Ultimately, data from a longitudinal cohort that included measurements of obesity in a representative ethnically diverse sample that reasonably represents the experience of ethnic minorities in England today were not available. Therefore, to parameterize this agent-based model, we used data from the 1999 and 2004 Health Surveys for England (HSEs),[10, 11] a series of repeat cross-sectional datasets, along with metrics from the literature. The 1999 and 2004 HSE featured modules that focused on the health of ethnic minorities in England, and thus were chosen for parameterization of our agent-based model. While longitudinal data includes information about the same individuals over time, thus best approximating the single cohort we sought to model, serial cross-sectional data, which includes data about the same population at several time points, was a satisfactory alternative in this regard. Parameterizing from serial cross-sectional data may, in fact, more accurately capture both intertemporal and population heterogeneity, which then may improve detection of potential emergent features arising in simulations run with the model.

3.2 Health Surveys for England

The HSEs are an annual series of population-based, cross-sectional datasets that have been commissioned every year since 1993. Each survey includes information about demographic indicators and measurements of height and weight and other biometrics. As each HSE features a module focused on a particular topic or population group, we chose to use data from the 1999 and 2004 HSEs because each featured a boost sample that collected data specifically about ethnic minorities in England. More information about the Health Surveys for England and their sampling and survey procedure is available elsewhere.[10-12]

3.3 Recoding and cleaning the data

First, we separately uploaded and cleaned the 1999 and 2004 datasets, which are freely available from the UK Data Archive (http://www.data-archive.ac.uk/). Area-level ethnic density and deprivation were important contributors to obesity in our conceptual framework. However, publicly available HSE data are anonymized to avoid identification of individual participants. This includes removing all area-level variables below identification of the Government Office Region (an administrative unit in England with a resident population between 2.5 and 8 million people) in which the participant resides. In order to parameterize the ethnic density and deprivation links from our conceptual model, we obtained data directly from the National Centre for Social Research, the institution that oversees the HSE data, about ward-level ethnic density (in 2001) in quintiles, as well as ward-level 2004 Index of Multiple Deprivation[13] in quintiles. Ideally, we would have used continuous data about both ethnic density and deprivation, however, there was a concern on the part of the National Centre for Social Research that further granularity in these data might compromise respondent anonymity.

We recoded exposure and outcome covariates of interest. Occupational social class was recoded into two categories: non-manual (professional, managerial, and skilled non-manual) vs. manual social class (skilled manual, semi-skilled manual, and unskilled manual). Educational attainment was recoded as three variables: Post-secondary diploma or greater; A-level to some post-secondary education; and O-level (or General Certificate of Secondary Education [GCSE]) or below. The latter two educational classifications correspond roughly to the US qualifications of Associate's degree, and high school diploma or General Educational Development (GED) qualification or lower, respectively. Ethnicity was categorized as “Black” (including Caribbean and African), “South Asian” (including Indian, Pakistani, Bangladeshi), or “White”. Any individuals not classified in one of these three ethnicity categories were removed (N=2697, 7.4% of observations). Both ward-level ethnic density and ward-level index of material deprivation were analyzed as binary variables: quintile four or greater vs. quintile three or lower. Only one metric of obesity was used: body mass index (BMI). BMI greater than or equal to 30 kg/m^2 was classified as obese.[16] All individuals without a BMI measure were removed from the dataset (N=5584, 21.6% of all observations). We considered using a waist/hip ratio (WHR) measure but there was not sufficient data to support parameterization of this measure (69.2% of WHR data were missing).

Web Table 1 shows descriptive statistics about all exposure and outcome variables considered among all adults surveyed in the 1999 and 2004 HSEs. This table demonstrates the prevalence of missing data among covariates of particular relevance to our central inferences. Among 25,837 observations from both 1999 and 2004 HSEs, 6.7% of ethnicity data were missing, along with 15.8% of BMI data and 69.2% of WHR data.

Web Table 2 shows descriptive statistics after deleting 5,584 observations with missing data about ethnicity or BMI. This table describes data included in the parameterization of Obesagent. Web Table 3 shows bivariate chi-square analyses between each exposure covariate and obesity (BMI >30 kg/m^2).

Web Table 1: Univariate statistics about all exposures and outcomes among all adults aged 18 years or older from the Health Surveys for England, 1999 and 2004

	N	%
Total	25,837	--
Ethnicity		
White	14,816	57.3
Black	3,015	11.7
South Asian	6,276	24.3
Other/Unknown	1,730	6.7
Body mass index obesity[1]		
Obese	4,473	17.3
Non-obese	17,279	66.9
Unknown	4,085	15.8
Sex		
Male	14,249	55.2
Female	11,588	44.9
Education		
University Degree	4,410	17.1
A-level +	5,654	21.9
O-level or less	6,183	23.9
Unknown	9,590	37.1
Social class		
Non-manual	12,840	49.7
Manual	10,903	42.2
Unknown	2,094	8.1
Waist-hip ratio obesity[2]		
Obese	2,592	10.0
Non-obese	5,369	20.8
Unknown	17,876	69.2
Ethnic density[3]		
>3rd quintile	6,911	26.8
<=3rd quintile	18,926	73.3
Index of Material Deprivation[4]		
>3rd quintile	12,103	46.8
<=3rd quintile	13,734	53.2

[1] Body mass index obesity was defined as body mass index > 30 kg/m^2
[2] Waist-hip ratio obesity was defined as waist-hip ratio >0.85 among women and >0.95 among men
[3] Ward-level ethnic density calculated in 2001 and categorized into quintiles by ward
[4] Ward-level Index of Material Deprivation calculated using the 2004 formulation[13] and categorized into quintiles by ward

Web Table 2: Univariate statistics about all exposure and outcomes covariates from 20,253 observations with available ethnicity and body mass index data from the Health Surveys for England, 1999 and 2004

	N	%
Total	20,253	--
Ethnicity		
White	12,767	63.0
Black	2,322	11.5
South Asian	5,164	25.5
Body mass index obesity[1]		
Obese	4,347	21.5
Non-obese	15,906	78.5
Sex		
Male	11,005	54.3
Female	9,248	45.7
Education		
University Degree	3,428	16.9
A-level +	4,520	22.3
O-level or less	5,090	25.1
Unknown	7,215	35.6
Social class		
Non-manual	10,096	49.9
Manual	8,685	42.9
Unknown	1,472	7.3
Waist-hip ratio obesity[2]		
Obese	2,115	10.4
Non-obese	4,261	21.0
Unknown	13,877	68.5
Ethnic density[3]		
>3rd quintile	5,904	29.2
<=3rd quintile	14,349	70.9
Index of Material Deprivation[4]		
>3rd quintile	9,758	48.2
<=3rd quintile	10,495	51.8

[1] Body mass index obesity was defined as body mass index > 30 kg/m^2
[2] Waist-hip ratio obesity was defined as waist-hip ratio >0.85 among women and >0.95 among men
[3] Ward-level ethnic density calculated in 2001 and categorized into quintiles by ward
[4] Ward-level Index of Material Deprivation calculated using the 2004 formulation[13] and categorized into quintiles by ward

Web Table 3: Bivariate chi-square tests between each exposure covariate and obesity (body mass index [BMI]>30 kg/m^2) among 20,253 observations with available ethnicity and BMI data from the Health Surveys for England, 1999 and 2004

	Obese[1] (%)	p
Total	21.5	--
Ethnicity		<0.001
White	22.0	
Black	29.0	
South Asian	16.9	
Sex		<0.001
Male	18.6	
Female	23.8	
Education		<0.001
University Degree	15.7	
A-level +	18.7	
O-level or less	21.6	
Unknown	25.9	
Social class		<0.001
Non-manual	19.5	
Manual	23.5	
Unknown	22.8	
Ethnic density[2]		0.043
>3rd quintile	21.1	
<=3rd quintile	22.4	
Index of Material Deprivation[3]		0.078
>3rd quintile	22.0	
<=3rd quintile	20.9	

[1] Obesity defined as body mass index > 30 kg/m^2
[2] Ward-level ethnic density calculated in 2001 and categorized into quintiles by ward
[3] Ward-level Index of Material Deprivation calculated using the 2004 formulation[13] and categorized into quintiles by ward

3.4 Parameterizing the initial population of Obesagent

To parameterize the initial population, a series of cross-tabulations using data about the population aged 18-25 in 1999 was carried out. Both socioeconomic indicators (educational attainment and occupational social class) were cross-tabulated with one another stratified by ethnicity. Using the results of these cross-tabulations, agents were stochastically assigned to ethnicity, occupational social class, and educational attainment. Fifty percent of the population was assigned to female gender, and the remaining 50% were assigned to male gender.

Each agent was allocated to a space representing a “neighborhood”. To model segregation, neighborhoods were distinguished by ethnic density and deprivation. There were four neighborhoods simulated in the model: one representing each of the resulting squares from the cross-tabulation between both social class categories and minority vs. majority ethnicity. Agents were allocated to a neighborhood based on ethnicity and social class. To model a segregated population akin to that in England, each agent was 75% likely to be placed into a neighborhood corresponding to its social class with 25% of agents being placed without consideration of social characteristics. Similarly, each agent was 75% likely to be placed into a neighborhood corresponding to its ethnicity (minority vs. majority) with 25% of agents being placed at random.

We operationalized contextual effects on obesity risk as inputs to initial risk for obesity at model outset, rather than as exposures that contributed to obesity risk through time. Beta coefficients from a multilevel model of the influence of the interaction between ethnic density and ethnic minority status, and adjusted for area-level deprivation were used to calculate predicted probabilities of obesity by ethnic minority status in an ethnically dense context relative to a less dense context. Similarly, beta coefficients from a multilevel model of the influence of the interaction between manual social class and ward-level deprivation, and adjusted for ethnic density were used to calculate the predicted probability of obesity in a deprived context by social class (see Web Table 4, below).

Ethnicity was then cross-tabulated with obesity risk. Using data from this cross-tabulation, and then the predicted probabilities of obesity by neighborhood context and ethnic and social class identity in the model, initial obesity risk was stochastically assigned among the population.

3.5 Building the Social Network

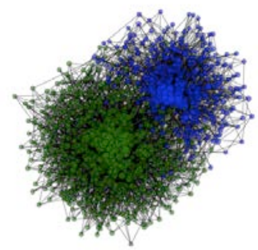
Each agent was situated in a social network. So as not to bias the outcomes of the model by network topology, we implemented and applied three separate social networks: a segregated scale free network, a clustered network, and a segregated Erdős–Rényi network. Each of these three network topologies incorporates different empirical aspects of human social networks.

The segregated scale-free network was generated using a biased preferential attachment growth model[17] to create a scale free (Barabási–Albert) social network (a network characterized by an uneven degree of connectedness among agents) with assortative, or non-random, mixing. This type of network is thought to most accurately approximate the behavior of real-world social networks, and is commonly used in stochastic social network simulations.[16, 17]

The network was initialized from a seed network comprised of a small number of agents. Agents were connected to the network sequentially. Each new agent added to the network was connected to four existing agents with a probability of connecting to an existing agent that was proportional to the number of connections that existing agent already possessed. The mean number of connections among all agents in the network was eight.

Moreover, an additional bias was included to preferentially connect agents with like characteristics. While 25% of new agents to the network were connected without regard to their characteristics (i.e., ethnicity, social class, education), 50% of new agents to the network were restricted to connecting with existing agents of similar ethnicity, again with a probability of connecting to existing agents with the same ethnicity that was proportional to the number of connections that that agent already possessed. And 25% of new agents added to the network were restricted to connecting with existing agents of similar ethnicity and social class, again with a probability of connecting to an existing agent with the same ethnicity and social class that was proportional to the number of connections that that agent already possessed. Web Figure 2 shows a visualization of the social network constructed.

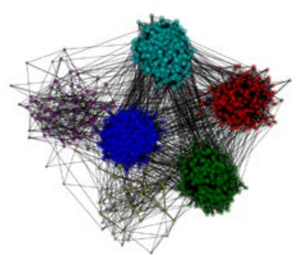
Web Figure 2: A visualization of the scale free (Barabasi-Albert) social network with assortative mixing created via a biased preferential attachment growth model among 10,000 agents



Green nodes represent ethnically “White” agents, and blue nodes represent ethnic minority agents

The clustered network was generated by establishing a link between two agents randomly with a probability *p* that depends on the neighborhood assigned to each agent. If the agents do not share the same neighborhood, the probability of a connection is $p=1/N$, where N is the population size. The probability of establishing a connection is higher if two agents share the same neighborhood ($p=50/N$), thereby introducing a bias that reflects the neighborhood structure, as illustrated in Web Figure 3.

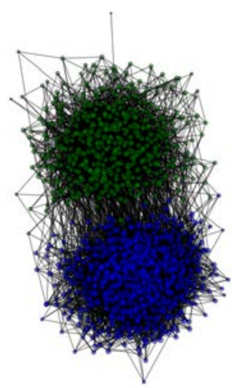
Web Figure 3: Visualization of the clustered social network a biased algorithm, which assigns a higher probability to connections by neighborhood



Colors represent different neighborhoods

The segregated Erdős–Rényi network was generated using a modified Erdős–Rényi model where the probability of a link between two agents is conditioned on the agents' ethnicity. If either two agents belong to a minority or both agents are ethnically "white", they have a higher probability of connecting than in the case when one agent belongs to a minority and the other agent is ethnically "white". The resulting social network is segregated based on ethnicity, which is illustrated in Web Figure 4.

Web Figure 4: Visualization of the segregated Erdős–Rényi network



Green nodes represent ethnically "white" agents, and blue nodes represent ethnic minority agents. Although the figure looks very similar to the segregated scale-free network, the underlying random graph generator model is different and results in a network with different characteristics. For examples, whereas the node degree distribution of the scale-free network follows a power-law, the node degree distribution of the segregated Erdős–Rényi is binomial.

Using serial cross-sectional datasets to assess time-dependent relations

The HSEs are non-contiguous, and do not collect data about the same individuals at different time points. Therefore, they pose a challenge to parameterization in a model that is intended to yield inference about time-dependent relations between social factors and obesity.

Such time-dependent parameters relating factors of interest and obesity in the Obesagent model that accounted for age effects in the trajectory of obesity with time were needed. To produce these parameters, we first created several iterations of the dataset to capture age effects in changes in obesity risk by year by social factors of interest. These iterations divided the full dataset into several age delimited datasets, such that the first dataset included those aged 18-30; the second included those aged 31-40; the third included those aged 41-50; the fourth included those aged 51-60, and the fifth included those aged 61-70.

Second, we created a variable that differentiated each resulting dataset by age and cohort (i.e., HSE 1999 vs HSE 2004). This binary variable differentiated between those who were in the younger half of each age-delimited dataset from the 1999 survey and those who were in the older half of each age-delimited dataset in the 2004 survey (for example, differentiating between the population aged 31-35 in the HSE 1999 and those aged 36-40 in the HSE 2004). Individuals who were not in either age group were then removed from the dataset. This protocol, therefore, allowed us to account for a cohort effect within each age-delimited dataset.

3.6 Parameterizing the updating protocol

In parameterizing the updating protocols, or the process of calculating obesity risk among each agent at each time-step (representing one year in time), we attempted to capture age effects with time. We did so by including five age-delimited datasets, described above, and isolating age-dependent ethnic and socioeconomic parameters reflecting the differences in obesity risk along each stratum between 1999 and 2004. However, because, as described above, most parameters were derived from data about one population in only two time points, we were unable to capture cohort effects beyond the first twelve time-steps of the model (agent ages 18-30 years), which would theoretically correspond to the years for which data was available (1999-2004).

Baseline agents for each update were the ethnic majority (White) and socioeconomically most advantaged (non-manual and university degree) groups;

changes in all other ethnic and socioeconomic groups were proportional to changes in the baseline group. Therefore, we first calculated changes in obesity risk among these groups over time in all age strata. To consider changes in obesity risk among these groups, we used logistic regression (logit) models of the standard form below (Equation 1) restricted to each group individually.

$$\text{Equation 1. Logit(p)} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \ldots + b_kX_k$$

For example, to assess changes in obesity risk between those in non-manual social classes aged 18-25 in 1999 and then aged 25-30 in 2004, we restricted the analysis to non-manual social class individuals in the corresponding age-delimited dataset (construction of which was described above), and the following logit model was fit (Equation 2):

$$\text{Equation 2. Logit(bmiobese)} = b_0 + b_1\text{(young99)} + b_2\text{(A-level+)} + b_3\text{(GCSEorlower)} + b_4\text{(SouthAsian)} + b_5\text{(Black)}$$

By similarly carrying out analyses within each age-delimited dataset (18-30; 31-40; 41-50; 51-60; or 61-70), and restricting ethnicity, educational attainment, or occupational social class based on the analysis of interest, we were able to isolate changes in obesity risk by year per age group for the ethnic majority and socioeconomically advantaged groups by using the beta coefficients from these logit models (such as Equation 2) to calculate predicted probabilities of obesity among each group in each year (1999 and 2004) (for a demonstration of predicted probability calculations, see Web Table 4, below). These predicted probabilities were then used to calculate the change in probability of obesity in each group between 1999 and 2004. These five-year changes were used to calculate corresponding one-year changes in obesity risk in the ethnic majority and each socioeconomically advantaged group between 1999 and 2004.

After calculating yearly changes in obesity risk among these groups over time at all age strata, we calculated changes in obesity risk among ethnic minority and socioeconomically-deprived groups relative to them. To do this, we used logit models of the standard form (Equation 1) to isolate the regression coefficients from models of obesity by the interaction between an input variable and the variable that differentiated between those who were in the younger half of the 1999 survey and those who were in the older half of the 2004 survey. Parameter estimates yielded from these Logit models were then used to calculate predicted probabilities of obesity by input variables for each group in each year.

For example, to estimate the effects of manual social class relative to non-manual social class on obesity risk between 1999 and 2004 among those aged 31-40, we fitted the following Logit model (Equation 3) using the corresponding age-delimited dataset (restricted to those aged 31-40):

Equation 3. Logit(Obese)= b_0 + b_1(Manual Social Class) + b_2(Differentiation Variable) + b_3(Manual*Differentiation) + b_6(South Asian) + b_7(Black) + b_8(A-level+) + b_9(GCSE or Lower)

The parameter estimates yielded from this model were then used (Web Table 4) to calculate predicted probabilities of obesity among each ethnic group in each year.

Web Table 4: Predicted probabilities of obesity by social class among those 31-35 years old in 1999 and 36-40 years old in 2004 from the Health Surveys for England, 1999 and 2004

	Beta	Manual 1999	Manual 2004	Non-manual 1999	Non-manual 2004
Intercept	-1.711	1	1	1	1
Manual Social Class	0.208	1	1	0	0
Survey Differentiation Variable	-0.313	1	0	1	0
Manual*Differentiation	-0.007	1	0	0	0
South Asian	-0.226	0	0	0	0
Black	0.374	0	0	0	0
A-level+	0.333	1	1	1	1
GCSE or lower	0.431	0	0	0	0
logit(p)		-1.06028	-0.7394	-1.2604	-0.947
p*100		25.7	32.3	22.1	27.9

Predicted probabilities (in %) were calculated using values in the “Beta” column and number values in each class/year column (e.g., Manual/1999, Manual/2004, Non-manual/1999, and Non-manual/2004) according to Equation 3 and then exponentiating the resulting number and multiplying by 100.

Using differences by group relative to the ethnic majority group and the socioeconomically least-deprived groups in predicted probabilities by year, we created time-dependent parameters of association between each exposure (i.e., Black or South Asian ethnicity; manual social class; and/or A-level+ or GCSE or lower educational attainment) and obesity representing the relative risk of developing obesity in the next five years compared to the corresponding ethnic or socioeconomic group (i.e., White ethnicity; non-manual social class; or university degree educational attainment). Because each time step in the final ABM represents one year, and not five, as was represented in the above example, it was necessary to scale appropriately. To do so, we calculated the fifth-root of the five-year relative risk parameter the calculation of which we described above.

In line with broader systems theory, which implies that even weak relationships may have important implications in outcomes, it is important to note that at no point did we consider measures of significance (i.e., confidence intervals or p-values). Applying an arbitrary cutoff measure of significance on associations, as a threshold for their inclusion in the model would have, by definition, excluded weaker markers of association, and therefore weaker relationships, which could plausibly influence model outcomes.

3.7 Parameterization from the literature

The Obesagent model featured only one parameter that was drawn from the literature. The parameter for the relation between obese contacts and subsequent risk for developing obesity was derived from a 2007 study by Christakis and Fowler,[4] which assessed the networked spread of obesity within a densely interconnected social network. This study used data from the Framingham Heart Study to assess the increase in risk for obesity in a given time period if a social contact was to become obese in the previous time period. The authors demonstrated that those whose network contacts developed obesity in the last time-period had higher risk for developing obesity in the subsequent time-period as compared to individuals without contacts who had themselves recently developed obesity.[4] Using the study findings, we derived the parameter for the relative risk of developing obesity if a network contact had developed obesity in the last time-step of 1.16.[4] Incorporation of this parameter into the Obesagent model is described below.

Section 4. Model implementation and output

4.1 Update rules and Monte Carlo Simulations

At each time step, the probability of developing obesity during that time step was determined independently for each agent in the population as a multiplicative function of that agent's ethnicity, social class, education social context, and whether or not a network contact had developed obesity in the previous time-step. After calculating the probability of developing obesity in that time-step among all agents, a random number between zero and one was generated for each agent simultaneously. If that number was lower than that agent's calculated probability of developing obesity in that time-step, the agent progressed to obesity in that time-step; if it was higher, the agent did not progress to obesity during that time-step.

Model output was recorded in obesity prevalence by year, both overall as well as stratified by ethnicity, social class, and education. Each time-step update was implemented 100 times, and the outputs were averaged to produce mean outcome measures as well as 95% confidence intervals.

4.2 Software

All data cleaning, recoding, and analysis was done using SAS 9.2. The Obesagent model was constructed using the Python programming language (releases 2.4 to 2.7) with the NumPy, Matplotlib, and NetworkX packages. Python is an open-source programming language that is commonly used in agent-based modeling applications.[19]

4.3 Baseline output

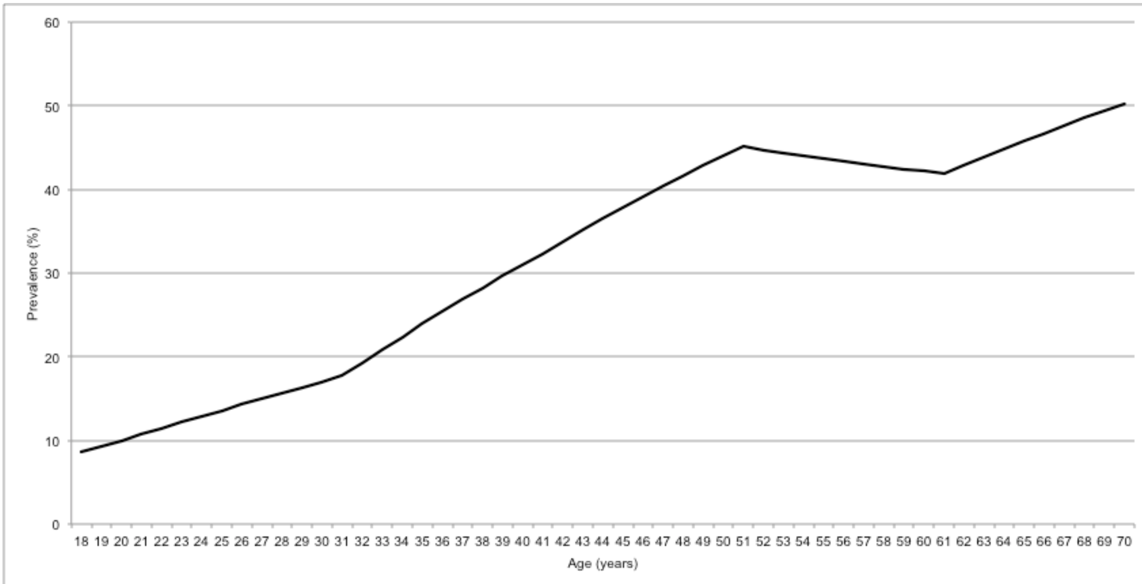
Web Figure 5 shows Obesagent estimates of obesity prevalence by age among 10,000 agents representing a cohort born in 1981 in England at baseline. By comparison, Web Figure 6 shows obesity prevalence among 20,253 adults aged 18-70 with available ethnicity and BMI data by age from both general and ethnic boost samples of the 1999 and 2004 Health Surveys for England, this represents all data used to parameterize the Obesagent model. In general, obesity prevalence in both figures increased steadily until the age of 51, decreased slightly between 51-60, and increased again between 61-70.

While we were originally concerned with the idiosyncratic shift in obesity prevalence at age 50 in our baseline output, concerned that it was an artifact of the modeling procedure, finding the same shift in obesity prevalence when stratified by age in our raw data from the Health Surveys for England suggests that our modeling procedure was, in fact, faithful to the data used to parameterize it. Web Figures 5 and 6 demonstrate the mirroring of this shift in our model relative to the data from which it was built.

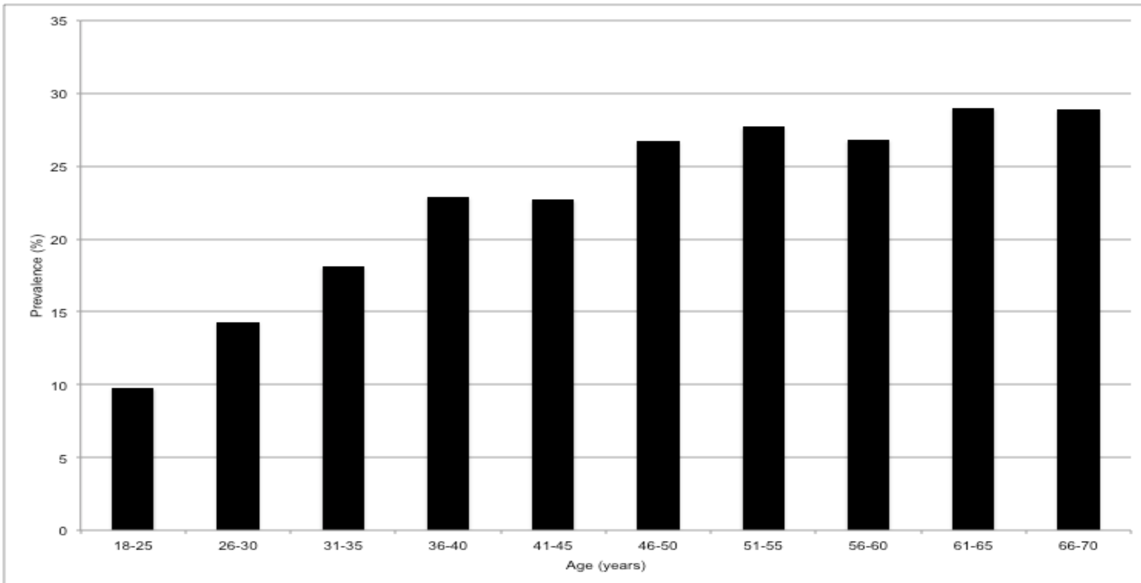
Moreover, while the character of the trajectory in obesity prevalence through the life course predicted by Obesagent was similar to that from human data from 1999 and 2004, Obesagent predicted substantially higher prevalence of obesity by age (average obesity prevalence between ages 66 and 70 predicted by Obesagent was 48.3%, compared to 28.9% in human data from 1999 and 2004).

This was expected. Agents in our model represented a cohort born in 1981 that aged through time. Therefore, agents aged between 66 and 70 years old in the Obesagent model, for example, were not intended to represent humans in the same age group between 1999 and 2004, but rather humans in this age group between 2065 and 2069. As obesity prevalence has been predicted to rise substantially in the UK between these time periods,[1] the difference in obesity prevalence estimated by Obesagent and human data from 1999 and 2004 is warranted.

Web Figure 5. Obesity* prevalence by age among 10,000 agents representing a cohort born in 1981 in England



Web Figure 6. Obesity prevalence by age among 20,253 adults aged 18-70 with available ethnicity and body mass index date by age from both general sample and ethnic boost samples of the 1999 and 2004 Health Surveys for England



References

1. McPherson K, Marsh T, Brown M. *Tackling obesities: Future Choices–Modelling future trends in obesity and the impact on health*. London: UK Government Office for Science, 2007.both adult reviews

2. El-Sayed AM, Scarborough P, Galea S. Ethnic inequalities in obesity among children and adults in the UK: a systematic review of the literature. *Obes Rev*. 2011.12(5): e516-34.

3. El-Sayed AM, Scarborough P, Galea S. Unevenly distributed: a systematic review about socioeconomic inequalities in adult obesity in the United Kingdom. *BMC Public Health.* 2012. 12:18.

4. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357(4):370-79.

5. Wen M, Maloney TN. Latino residential isolation and the risk of obesity in Utah: The role of neighborhood socioeconomic, built-environmental, and subcultural context. *J Immigr Minor Health*. 2011. [Epub ahead of print]

6. Medical Research Council. National survey of health and development. Swindon: Medical Research Council. Available at: http://www.nshd.mrc.ac.uk/

7. Centre for Longitudinal Studies. National Child Development Study. Available at: http://www.cls.ioe.ac.uk/studies.asp?section=000100020003

8. Owen D. Profile of black and minority ethnic groups in the UK. Warwick: University of Warwick; 2006. Available at: http://www2.warwick.ac.uk/fac/cross_fac/healthatwarwick/publications/occasional/ethnicprofile.pdf.

9. Office for National Statistics. Population size: 7.9% from a minority ethnic group. Available at: http://www.statistics.gov.uk/CCI/nugget.asp?ID=273.

10. Department of Health. Health survey for England 1999: The health of minority ethnic groups. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4009393

11. Sproston K, Mindell J. Health Survey for England 2004. Volume 2: Methodology and documentation. Leeds: The Information Centre; 2006. Available at: http://www.ic.nhs.uk/webfiles/publications/healthsurvey2004ethnicfull/HealthSurveyforEngland210406_PDF.pdf

12. The Health and Social Care Information Centre. Health Survey for England. London: Department of Health. Available at: http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england

13. Noble M, Wright G, Smith G, Dibben C. Measuring multiple deprivation at the small-area level. *Environ Plann A*. 2006;38(1):169-85.

14. Keyes A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. *J Chronic Dis.* 1972;25(6):329-43.

15. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509-12.

16. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Åberg Y. The web of human sexual contacts. *Nature*. 2001;411(6840):907-8.

17. Ebel H, Mielsch L, Bornholdt S. Scale-free topology of e-mail networks. *Phys Rev E*. 2002;66(3):35103.

18. Erdős P, Rényi A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences.* 1960:5:17–61.

19. Railsback SF, Lytinen SL, Jackson SK. Agent-based simulation platforms: Review and development recommendations. *Simulation*. 2006;82(9):609-23.