An agent-based simulation model to explore how population composition, residential segregation, social influence and environmental disparities produce racial/ethnic disparities in diet

Background
Racial/ethnic minorities suffer from high rates of obesity and related chronic disease.\textsuperscript{1-3} While physical (in)activity plays a part, these disparities are at least partially driven by well-documented disparities in diet.\textsuperscript{4} High rates of obesity among Mexican Americans (~78%) may, in part, be driven by the high level of consumption of sugar-sweetened beverages among that population.\textsuperscript{1} Between-group differences in fast food consumption may also play a part.\textsuperscript{5-7}

Improving diet quality is of critical public health importance, but doing so is difficult because diet is influenced by multilevel factors at the environmental (e.g., access to food stores), household (e.g., family composition, income), individual (e.g., preferences), and interpersonal (e.g., social influence) levels.\textsuperscript{8-15} Addressing disparities is even more difficult, because racial/ethnic groups vary markedly in many of the characteristics (e.g., age, income, educational attainment) that most strongly influence diet patterns. For example, the Latino population as a whole is younger, poorer, and less well-educated than the White population. Thus, dietary disparities could be driven by cultural differences (e.g., food preferences), environmental inequities in food access, differences in the group-level distribution of socio-demographic characteristics (i.e., population composition), or some combination thereof. Furthermore, due to the interrelated effects of residential, economic, and school-based segregation, individuals are likely to live, work, and play with others who share their racial/ethnic and economic characteristics. As a result, the influence exerted through homogeneous social networks may help consolidate health behaviors and exacerbate even small health disparities.

Over the last ten years, health researchers have increasingly used agent-based models to explore and disentangle the multilevel factors that influence health behaviors like smoking, alcohol consumption, and diet.\textsuperscript{16, 17} In the sociological field, recent work has used agent-based models to understand how population structure (i.e., the relative size of minority groups and within- and between-group disparities in income) can generate racial/ethnic residential segregation given housing price differences across neighborhoods. The handful of models that have explored diet have provided tremendous insight into the complex systems that produce racial/ethnic disparities. For example, Auchincloss et al (2011) used a stylized model with a simplistic population, environment, and behavior rules to explore how a limited number of model conditions (i.e., residential segregation and food environment disparities) can interact to generate diet disparities across neighborhoods. Other models have explored the plausible effects of interventions, including increased access to healthy foods in poor and minority neighborhoods, tax breaks (or incentives) for healthy (or unhealthy) food, and improving social norms.

In this study, we seek to add to the emerging public health and sociological literature in this field by presenting results of an empirically-driven agent-based model that explores racial/ethnic disparities in consumption of sugar-sweetened beverages, fast food, and total calories. The model is driven by population data from the U.S. Census Bureau (e.g., the American Community Survey), health data from the 2007-2010 National Health and Nutrition Examination Survey, and the published literature. The purpose of the model is to explore the following questions:

1) To what extent do initial conditions (i.e., population composition, segregation, and environmental disparities) contribute to racial/ethnic disparities in diet?

2) To what extent does residential segregation (and, consequently, segregation in schools and workplaces) create racial/ethnic homophily within social networks?

3) Do racially/ethnically homogeneous social networks consolidate health behaviors within racial/ethnic groups and exacerbate between-group disparities?
Methods

Model Overview

The purpose of the proposed ABM is to explore how multilevel influences shape food purchasing behaviors at supermarkets, corner stores, fast food restaurants, and sit-down restaurants. Specifically, we will use the ABM to examine how racial/ethnic disparities in diet are generated from residential segregation, disparities in the food environment, between-group differences in population composition (e.g., income, education) and social influence. The primary agents in the model are individuals. Individuals make daily decisions about whether or not to consume sugar-sweetened beverages and fast food, as well as about total calorie consumption. The model environment is stylized but realistic, and includes a continuous 500 x 500 space with 5,000 households and approximately 10,000 individuals. Each individual is assigned a race/ethnicity of either White, Black, or Latino; the relative proportion of each racial/ethnic group varies across model runs.18, 19 Household composition (e.g., one adult, couple, single parent, two parents) and socio-demographic characteristics (e.g., income, education) are generated by the model based on distributions representative of the U.S. population as determined from the American Community Survey. Supermarkets and fast food restaurants are distributed across the model space in proportions drawn from the published literature.20

Diet Behaviors

The ABM simulates the following three daily diet behaviors: 1) fast food consumption, 2) sugar-sweetened beverage consumption, and 3) total caloric intake. These behaviors are modeled within the ABM based on a series of equations that assign weights to individual and household characteristics, as well as the food environments:

\[
\hat{Y}_{ij} = I_i b_I + H_i b_H + E_i b_E + A_{ij} b_A
\]  

where \( I_i \) is a vector of individual characteristics (e.g., age, gender, educational attainment), \( H_i \) is a vector of household characteristics (e.g., income), \( E_i \) is a vector of environmental factors related to access food resources, \( A_{ij} \) is a random component with uniform distribution that reflects the preferences and attitudes of individual \( i \) towards outcome \( j \), and the \( b \)'s are vectors of weights of the importance of each factor for outcome \( j \). Outcomes in Equation 1 are predicted values of continuous outcomes for caloric intake and daily probabilities for binary outcomes.

To increase the realism and validity of the ABM, the parameter values (i.e., the \( b \)'s) in Equation 1 are ‘anchored’ to those derived from real-world data. For most parameters, values are derived using linear and logistic regression models using data from the 2007 to 2010 National Health and Nutrition Examination Survey. These models will take the following form:

\[
\hat{Y}_{ij} = I_i b_I + H_i b_H + E_i b_E + A_{ij} b_A + \epsilon
\]  

where, similar to Equation 1, \( I_i \) represents individual characteristics, \( H_i \) is household characteristics, \( E_i \) is the food and physical activity environment, \( A_{ij} \) is attitudes and \( \epsilon \) is the residual. For parameters that cannot be estimated directly using NHANES (e.g., food environment parameters), values are derived from the published literature and a calibration process is used to align model outcomes with those observed in the NHANES data.

Feedback

Social influence of family members and peers impacts diet and other health behaviors.21-24 To evaluate the effects of social influence, we will include feedback loops in the ABM. These
feedback loops, adapted from Yang (2011), update each individual’s attitudes ($A_{ij}$) based on the preferences of friends and neighbors. Each individual in the model will be assigned a group of three to five ‘friends’ with similar characteristics (e.g., age, neighborhood), as well as family members that share the same household. Each time step, a small proportion ($a = 0.001$) of an individual’s attitudes will be updated based on the preferences of their friends:

$$A_{ij}'(H) = (1 - a) \times A_{ij} + aA_{j}(F)$$

where $A_{ij}$ is an individual’s preference for $j$ behavior, $A_{ij}'$ is the updated or ‘new’ preference, and $F$ is the set of ‘friends’ or ‘family members’. This process will result in some, but not total, convergence of attitudes within social networks, thereby representing social influence.

**Simulation scenarios**

The calibrated ABM will be used to assess outcomes associated with various ‘what if’ scenarios related to segregation (i.e., residential, school, and workplace), food environment disparities, and population composition. The author will use the model to explore at least eight different scenarios characterized by high vs. low segregation, high vs. low food environment disparities, and high vs. low between-group differences in the age and income distributions (i.e., a 2 x 2 x 2 factorial design). These scenarios are highly relevant because they will help elucidate how segregation, environmental disparities, and population composition interact in the presence of social networks to produce racial/ethnic disparities in diet. The primary outcomes in each scenario will be mean daily caloric consumption and mean frequency of fast food and sugar-sweetened beverage consumption, stratified across racial/ethnic groups. Outcomes will be assessed across an average of 10 simulation runs.

**Progress to date**

Currently, the initial model has been developed in AnyLogic 7. This includes programming of two segregation mechanisms (a Schelling-type scenario that is static and a mechanism based on Bruch [2014] that includes residential mobility), population initialization based on distributions derived from the ACS, and behavioral rules governed by parameters derived from NHANES (as described above). In short, the model runs and I am ready to either start the calibration process and/or begin exploring the different simulation scenarios. I am including this information to convey my belief that, given progress to date, performing the proposed research before PAA is very feasible.

**Model snapshots**

On the following pages are two snapshots of the models, the first showing the population distributed across space and the second showing characteristics and state charts at the individual level.

---

1 There are two segregation scenarios because I am a little wishy-washy
Bibliography


