An Agent-Based Model of Income Inequalities in Diet in the Context of Residential Segregation

Amy H. Auchincloss, PhD, MPH, Rick L. Riolo, PhD, Daniel G. Brown, PhD,
Jeremy Cook, BA, Ana V. Diez Roux, MD, PhD

Appendix A

What Is an Agent-Based Model?

In brief, an agent-based model (ABM) is a computational model that represents a system of interacting discrete micro-entities (people, organizations) that create an artificial society.[1] Each agent has a set of characteristics (e.g., age, gender, income), which may be different from all other agents. Agents exist within an environment and follow programmed rules that use each agent's characteristics and information available to it to decide what each agent does at each time step. From these micro behaviors emerges the behavior of an aggregate system(s). ABMs can feature dynamics, feedbacks, and adaptation processes that are difficult to represent in traditional statistical models. ABMs can range from very simple, abstract models[2] sometimes referred to as "stylized" or "toy" models, to sophisticated models that have high degrees of realism and typically make use of high-resolution empirical data.[3] A few primers exist for agent-based modeling[1,4,5], including a discussion[6] of the utility of ABMs when examining place effects on health.

Objectives of the Current Study

In this study, a very simple, stylized ABM was constructed in order to explore synergies between where people live, healthy food resources in their community, income constraints, and healthy food preferences. At the start of this investigation, extreme scenarios of economic residential segregation and spatial clustering of healthy food stores were imposed. The computational model was used to identify which scenario showed income differentials in diet (as observed in previous empirical work) and could thus serve as a tool for examining the ways that segregation can contribute to income disparities in diet. Then, simple experiments were run to test whether pricing and preference factors were capable of reducing income differentials in diet generated by segregation.

Appendix B: Outline of Computer Program for Food Environment Model (High- Level Description)

INITIALIZATION

Set parameters

Create the grid for households and stores

Place households:

  Make the specified number of households

  Place each one on a random open square until grid is full

  Assign household income using input settings

Place stores:

  Make the specified number of stores

  Place each one on a random square with no store already placed there (OK if square has a household).

  If we're clustering all stores

    make sure the store is placed in the cluster region.

  Assign price and quality based on settings

For each store:

  Reset variables

For each household:

  Reset variables

  Randomly determine initial quality preference based on the specified parameter

EACH TIME STEP

For each household:

  Calculate utility for all stores (multiply scores by weights, sum, and add in random noise)

  Choose the store with highest utility

  Update statistics based on store chosen (total quality, distance traveled, money spent)

Every 15 time steps:

  Stores calculate and reset statistics (profit, # customers high and low income, # turns high and low quality)

  The model calculates all statistics (average health, distance, price, quality preference) by averaging all households' and stores' statistics

If stores are not static:

  Every 30 steps, choose a store to close:

Choose cheap vs expensive randomly based on the proportion of stores that are initially specified as cheap

If we are allowing a random store to close

Choose whether a random (3% chance) or the worst performing (97% chance) store will close

otherwise

choose the worst performing store to close

If random, select a random store of chosen price that has been open more than 180 time steps

If worst performing:

Find the store with least number of customers

If that store has been open more than 180 time steps, select it

Otherwise, no store closes

Tell the selected store to close: set store's closure variable to true and reset its counters

For each store:

Keep track of number of turns that a store location has not had an open store and has had an open store

If store location has not had an open store > 180 time steps:

Store opens

Store assumes a food store type opposite of the previous store

Store assumes a food store price that is the same as the previous store (or if systematic price differential is activated, then store assumes food store price that is linked to the food store type)

Households calculate and reset statistics (health, average money spent, average distance traveled)

Households update their quality preference based on everyone's new statistics (based on network)

Appendix C: Household Behaviors

Each time step, each household selected a store to shop at and shopped for food (a time step could be thought of as about every 2–3 days corresponding to food shopping frequency in empirical studies[7, 8]). Households chose which store to shop at by ranking the stores on four dimensions: price of food at the store, distance to the store, the household's habitual behavior, and the household's preference for healthy foods.

Justification for the Dimensions Used by Households to Rank Stores

Observational studies and survey data from government and industry sources were used to guide agent decision-making rules for generating plausible behaviors. Agent behaviors were tested against available data to make sure that intuitive and known behaviors were consistently represented, such as high-income households spend more on food[9, 10] and travel at least as far or farther than low-income households.[11, 12]

Price of Food at the Store:

Price is a well-known strong determining factor in food shopping behavior—two thirds of U.S. consumers rank "good value for money" as the most important consideration when choosing a grocery store.[13] This is particularly true for low-income shoppers who spend double the share of their household income on food as high-income shoppers,[14] even though quantity of food purchased is similar.[9,10]

Distance to the Store

Distance (or proximity) to stores is a well-known consideration for shopping. This metric was simplified by considering only the distance from the household and so ignored the fact that distance may be less burdensome if travel is combined with other required trips (e.g., to the workplace).

Household's Habitual Behavior

Habitual behaviors are frequently included in agent modeling since our histories (or experiences) can influence current decisions, and habitual reactions themselves tend to be important in decision making.[5, 15] There is limited information on food store loyalty across income groups and contexts, but some work found that shoppers return to a single supermarket for more than 60% of all their food purchases.[16, 17] Thus, preference for where the household previously shopped was included under the assumption that people tend to repeat previous behaviors and shop for food at a familiar store.

Household's Preference for (Un)Healthy Foods

Preference for (un)healthy foods was included because it is often hypothesized to be a predictor of dietary behaviors and a contributor to social inequalities in diet; and experiments were run to test the capacity for household food preferences to alter the income differential in diet that segregation generated.

Ranking Stores

Utility Function

Each household ranked all stores and selected the store with the highest value of utility. The utility function was an additive form of the Cobb-Douglas function[18] utilizing both scores and weights. The additive form permitted a low score in one dimension to not affect scores in other dimensions (i.e., no dimension dwarfed the others). If two scores were multiplied or a score in one dimension was otherwise made to affect the other dimensions, a low score in one dimension would make it difficult for a household to choose that store, even if the other scores were very high.

In Equation 1, *i* is the household, *k* is the dimension, and ε is random noise (random variable, $\mu = 0$, $\sigma = 0.05$).

$$Utility_i = \left(\sum_{k=1}^{4} score_{ki} * weight_k \right) + \varepsilon_i \qquad [1]$$

Random noise was added to represent bounded rationality (households do not always act with perfect rationality[19]).

The weights and scores were set through experimentation by iteratively testing and changing model rules to adhere to *a priori* criteria: high-income households should spend more on food[9,10] and travel at least as far as low-income households.[11, 12] High-income households tend to travel farther to shop for food than do poorer households,[11, 12, 20, 21] likely due to a combination of car ownership and their ability to pay for household-related labor, which may afford them more time for shopping. In turn, increased travel becomes linked to price since highly mobile shoppers can more easily comparison shop over a wide area to obtain the best price and match their preference for particular foods.[11, 12, 21]

Weights

For convenience, the weights were normalized so they add to 1.0; thus, they have meaning only relative to each other (e.g., making one weight larger has the same effect as reducing the others). The weights were constant parameters for all households (did not vary by household income). Weights for the analysis reported in the text were set to: distance 0.5, habit 0.1, price 0.2, preference for (un)healthy foods 0.2. Sensitivities to alternate weighting and scoring for the utility function and size and household/store density of the grid were examined (see Sensitivity to Weights, below).

Scores

Table C1 shows the functions for weighting and scoring the inputs for utility. Scores for price and distance were allowed to vary by household income because the intent was to match existing evidence that high-income households pay more for food and travel farther than low-income households. Each score was on a scale from 0 to 1, where 1 was the most preferred score.

Price score. In the current model, both low- and high-income households preferred a cheap store; thus, all households scored a cheap store as 1. To account for the fact that rich households were comparatively insensitive to price, rich households in the model scored an expensive store as 0.8 (i.e., close to 1) whereas poor households scored it 0.1 (i.e., close to 0).

Distance score. Both high- and low-income households equally preferred a closer store, but because travel was more of an obstacle for low-income households, poor households gave a distant store a lower score than a high-income household. For example, if there was no distance between a household and store (they share the same grid cell), households gave the store a score of 1 (maximum score) and with increases in distance they scored stores <1. But for a far away store, the score varied for low- and high-income households: the distance score was 1 – (number of grid cells between residence and store)/30 for low-income households and 1 – (number of grid cells between residence and store)/130 for rich households.

Habit score. Households remembered the store(s) they visited during the past 5 time steps. They gave a score of 1 (highest) to the store they last shopped at and decreased the score by 1/5 for each model step since they last visited that store.

Preference score. Preference scores were just the household's preference: a healthy store got the household preference; an unhealthy store got a score of 1 – household preference. Household food preference was a continuous attribute 0 to (preference for unhealthy food=0, preference for healthy food=1). Preference was either randomly assigned or assigned by household income in our "preference experiments" (unhealthy preference was assigned the lower end of the preference range [0.0–0.6] and healthy preference the upper end of the range [0.4–1.0] with some overlap between possible preferences).

Table C1. Functions for weighting and scoring the inputs for utility

	Weight[a]	Score[b]	
		High-income	Low-income
Distance	0.5		
Distance between household and store is 0		1	1
Distance between household and store >0		$1 - \frac{d}{30}$	$1 - \frac{d}{30}$
Habit	0.1		
Same store that household shopped at on previous time step		1	1
Different store (*N steps* = number of time steps since the household last shopped at the store)		$1 - \frac{N\ steps}{5}$	$1 - \frac{N\ steps}{5}$
Price	0.2		
Prices at the food store are cheap		1	1
Prices at the food store are expensive		0.8	0.1
Food preference	0.2		
Store sells healthy food		Preference	Preference
Store sells unhealthy food		1 – Preference	1 – Preference

[a]These weighting parameters were used in the main results reported in the manuscript. For alternate weights, see Appendix Table E1 and Appendix Figures E1–E3.
[b]Score for distance and price were iteratively selected by primarily relying on the desegregated scenario (random scenario, Figure 1 S1) under the criteria that high-income households should spend more on food and travel at least as far or farther than low-income households.

Appendix D: Store Behaviors

Highly stylized store dynamic behaviors were incorporated so that stores could respond to customer demand and households could make new choices for where they shop. Changes in store availability were programmed (store closures and openings) as were changes in type of food sold at the store.

Store Closures

Schedule

Every 30 steps, a store closes. Cheap or expensive stores will close in proportion to their relative frequency: if one quarter of the stores were cheap, one quarter of the time a cheap store closed. Since all households preferred a cheap store, stratifying in this way allowed cheap stores to close. In order to represent stores going out of business due to a variety of reasons, 97% of the time the worst-performing store, as measured by number of customers, of a particular price level closes; and a slight chance (3% random chance) that a random store of that price level would be chosen was programmed. There is not much information in the literature on the length of time that food stores stay in business, in part due to regional and temporal heterogeneity in the suitability of retail food environments. Simple rules were programmed such that about 10% of stores closed every 180 steps (this can be roughly thought of as about 10% per year which was likely a conservative approximation of food store turnover[22]). In order to give each store an opportunity to attract customers and make a profit, the rules were programmed so that no store open for less than 180 steps may close (such stores were not considered when making random choices or searching for the worst-performing store of that price level).

New Stores

Location

The primary experiments reported in the text used a "move-out/move-in" scenario: when low-performing stores closed, the location remained vacant for 180 time steps, and then a new store moved into the old store's location. This scenario forced households to make new decisions about where to shop.

Food Type

Previous work found high correlation over time between the type of foods being sold within particular neighborhoods.[23] Thus, in the base experiments, most new stores sold the same type of food as the old store but allowance was made for some changes: there was a 10% chance that the new store would change the type of food it sold.

Price

When healthy food type was not linked to food price, new stores sold food at the same price as the previous store, which took into account the assumption that price of food was at least partly linked to location.[24] The price level stayed the same unless experiments for "healthy food relative pricing" were enabled (experiments 2-i, 2-ii, 3-i, 3-ii, 3-iii, 3-iv), in which case store prices followed store type. For instance, if healthy food stores were set as expensive, and a cheap, unhealthy food store closed; then the new store could be a healthy food store which would mean that it would automatically be made expensive.

Appendix E: Sensitivity

Sensitivity to Weights

Table E1 shows the weighting scheme used for the results reported in the manuscript (primary weighting parameters #1) and alternate weighting schemes in order to evaluate sensitivity to weights used in the utility function. For each dimension, alternate weighting schemes were tested within a reasonable bracketed range that still showed that higher-income households pay more for food and travel at least as far or farther than low-income households in the random scenario (S1) and index scenario (S6). Bracketed ranges were: distance 0.2–0.5, habit 0.0–0.5, price 0.1–0.4, and preference 0.1–0.3. Figures E1–E3 show income differentials in diet and absolute diet levels for 11 weighting schemes applied to the index scenario and experiments 1-i, 1-ii, 3-i, 3-ii, 3-iii, and 3-iv. Price experiments (2-i, 2-ii, absent preferences) are not shown because they are very similar to the results already reported in the manuscript.

Different weighting schemes did not change interpretation of the results except for experiment 3-ii (Figure E3): those with high incomes prefer healthy food, which is cheap, and those with low incomes prefer unhealthy food, which is expensive. High-income diet stayed relatively stable but low-income diet was very dependent on the utility weights. The sensitivity to weighting was due to low-income households being surrounded by the food that they preferred (unhealthy food) but not having sufficient income to buy it (since it was expensive). Thus, outcomes were very different, depending on small modifications to the weights for distance, price, and preference.

Sensitivity to Food Store Changes (Change in Store Location and Food Type)

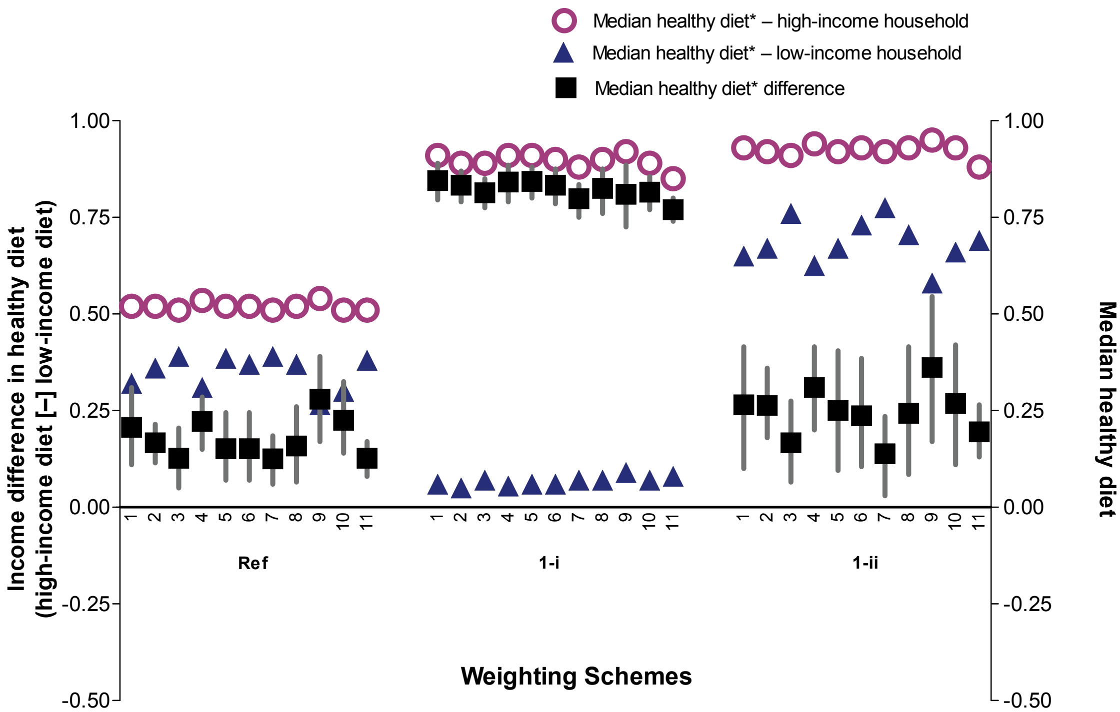
Tests were conducted of sensitivity to the store-change behavior rules that were used in the base scenario (as described above in *Store Closures* and *New Stores*). The scenarios are shown in Figure E4. The base scenario is change-rate #4. Lower rates of store changes than in the base scenario were tested: a static scenario where stores never closed or changed (labeled #1); an urban decay or planned shrinkage scenario[25] where stores closed and were not replaced so there were fewer and fewer stores (labeled #3); and move-out/move-in scenarios where low-performing stores closed, the location remained vacant for 180 time steps, and then a new store moved into the old store's location (labeled #2, 4, 5). Higher rates of store change than in the base scenario #4 were tested by assigning a 30% chance (labeled #5) and >40% chance that the new store would sell a different type of food (not shown).

The income differential in diet was similar across no or few changes in store location and food type but became weaker under high rates of change when stores were allowed to change the type of food they sold (Figure E4, change-rates #4 and #5). Sensitivity analyses using food type changes >40% eventually led to large improvements in low-income diets such that the income differential in diet fell to 0 (results not shown). The disappearance of the differential was due to store segregation completely breaking down over time as more and more new stores opened and changed the type of food they sold.

Table E1. The weighting scheme used for the results reported in the manuscript (Weighting Scheme #1) and alternate weighting schemes

Weighting scheme #	Weights			
	Distance	Habit	Price	Preference
Primary weighting parameters[a]				
1	0.5	0.1	0.2	0.2
Sensitivity parameters[b]				
2	0.3	0.3	0.2	0.2
3	0.3	0.1	0.4	0.2
4	0.4	0.3	0.2	0.2
5	0.4	0.2	0.2	0.2
6	0.4	0.1	0.3	0.2
7	0.4	0.0	0.3	0.3
8	0.4	0.1	0.3	0.2
9	0.4	0.1	0.4	0.1
10	0.5	0.0	0.3	0.2
11	0.2	0.5	0.1	0.2

[a]Primary parameters (Weighting Scheme #1) is the weighting scheme used for the results reported in the manuscript.
[b]Alternate weighting schemes were within a reasonable bracketed range for each dimension that still were able to show that higher-income households pay more for food and travel at least as far or farther than low-income households.



1-i High income prefers healthy food, low income prefers unhealthy food
1-ii High income prefers healthy food, low income prefers healthy food

* Diet was derived from the average proportion of times the household shopped at a healthy food store

Figure E1. Income differentials in diet and absolute diet levels for 11 weighting schemes for the index scenario and experiments 1-i and 1-ii. Weighting Scheme #1 is the weighting scheme used for the results reported in the manuscript. See Table E1 for the 11 weighting schemes. No income differences in preference or prices; healthy food preference and price of food are randomly assigned.

Ref, referent experiment (Index Scenario 6, [S6])

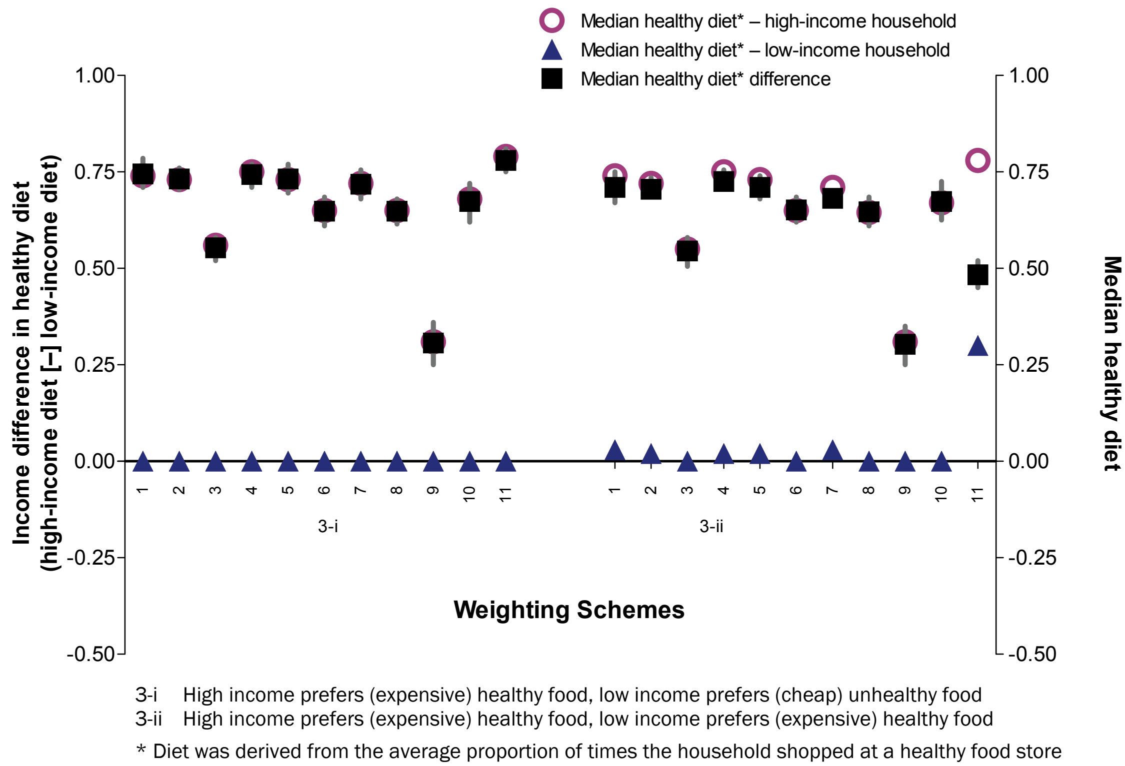


Figure E2. Income differentials in diet and absolute diet levels for 11 weighting schemes for experiments 3-i and 3-ii. Weighting Scheme #1 is the weighting scheme used for the results reported in the manuscript. See Table E1 for the 11 weighting schemes.

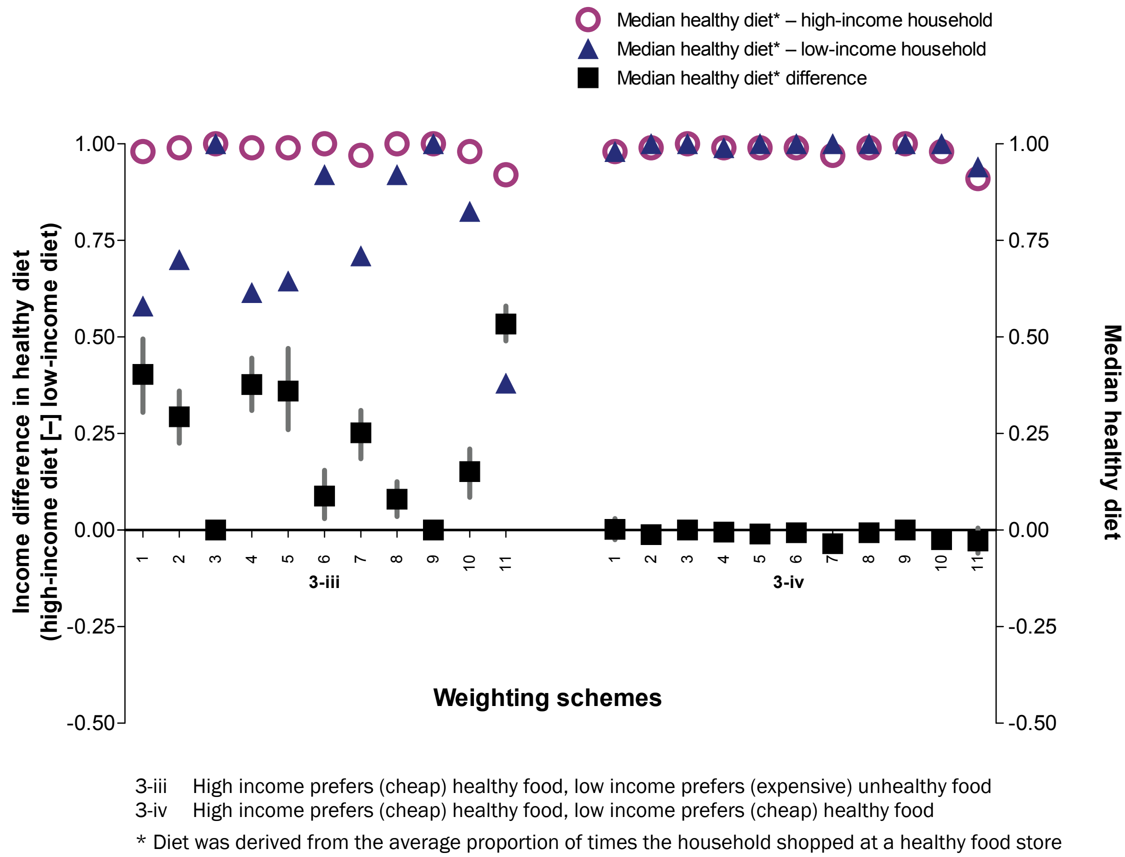


Figure E3. Income differentials in diet and absolute diet levels for 11 weighting schemes for Experiments 3-iii and 3-iv. See Table E1 for the 11 weighting schemes. Weighting Scheme #1 is the weighting scheme used for results reported in the manuscript. Preference for healthy food and food store prices are randomly assigned (they do not vary by income or healthy food store).

No change	#1	Stores do not change (stores never close and never change type of food)
	#2	Stores close and new stores open but always sell the same type of food
	#3	Urban decay; stores close and do not re-open
	#4	Stores close and 10% of the time change the type of food they sell.
High rate of change	#5	Stores close and 30% of the time change the type of food they sell

* Diet was derived from the average proportion of times the household shopped at a healthy food store

Figure E4. Income differentials in diet and absolute diet levels for five behavior algorithms using various intensities for change-rates of stores: 'no change' to 'high rate of change'. Change-rate #4 was used for the base experiments reported in the manuscript. Preference for healthy food and food store prices were randomly assigned (they did not vary by income or healthy food store).

Appendix F: Validity

The model presented was a very stylized model that was not intended to be realistic in the sense of being quantitatively calibrated to data. As a tool for enhancing our intuition and stimulating questions, the model had reasonable face validity, and qualitative patterns of the income differential in diet were largely insensitive to alternate parameterizations within a reasonable bracketed range. Simple robustness tests were conducted by contrasting the model's emergent endpoints with systems observed in the real world and by running the model multiple times while systematically varying initial conditions and parameters in order to assess the stability of results.[26–28] The abstract model had relatively few moving parts and dynamics; thus, when unexpected results were found, it was possible to trace the process backward to the algorithms and emerging behaviors to determine whether surprises were due to computer programming errors, path-like dependencies,[29] or truthful and informative properties.

Acknowledgements

The Food Environment Model was developed using the Recursive Porous Agent Simulation Toolkit (Repast) version 3,[30] an open-source software framework for creating agent-based simulations initially developed by Social Science Research Computing at the University of Chicago and further developed by the RePast Organization for Architecture and Development and Argonne National Laboratory. Additional libraries and code were from the Center for the Study of Complex Systems (CSCS) at the University of Michigan, and were written in Java using a Windows/Eclipse environment. Further information on agent-based modeling with Repast is available at the CSCS website at www.cscs.umich.edu and at the Repast home page at repast.sourceforge.net.

Author contributions

AHA and ADR conceived of and designed the study; JC and AHA developed model algorithms and analyzed simulation results; JC and RLR wrote the code; RLR and DGB provided modeling expertise; AHA drafted the initial manuscript; ADR and AHA obtained funding; all critically revised/edited the manuscript and approved the final manuscript.

References for Appendixes A–F

1. Epstein JM, Axtell R. Growing Artificial Societies: social science from the bottom up. Washington DC: Brookings Institution and Cambridge MA: MIT Press, 1996.
2. Schelling T. Simple models of segregation. In: Micromotives and macrobehavior. New York: Norton, 1978.
3. Barrett CL, Eubank SG, Smith JP. If smallpox strikes Portland. Sci Am 2005;292(3):54–62.
4. Grimm V, Railsback SF. Individual–based modeling and ecology. Princeton NJ: Princeton University Press, 2005.
5. Miller JH, Page SE. Complex adaptive systems. Princeton NJ: Princeton University Press, 2007.
6. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol 2008;168(1):1–8.
7. Bawa K, Ghosh A. A model of household grocery shopping behavior. Marketing Lett 1999;10(2):149–60.
8. Yoo S, Baranowski T, Missaghian M, et al. Food-purchasing patterns for home: a grocery store-intercept survey. Public Health Nutr 2006;9(3):384–93.
9. Jekanowski MD, Binkley JK. Food spending varies across the U.S. Food Rev (Spotlight on the U.S. Food System) www.ers.usda.gov/publications/foodreview/jan2000/,2000;23(1).
10. Smallwood DM, Blisard N, Blaylock J, Lutz S. Food spending in American households, 1980–92. USDA, Economic Research Service, Statistical Bulletin No. 888 and ERS data product, October; 1994. Report No.: SB-888.
11. Dunkley B, Helling A, Sawicki DS. Accessibility versus scale: examining the tradeoffs in grocery stores. J Plan Educ Res 2004;23(4):387–401.
12. Clifton KJ. Mobility strategies and food shopping for low-income families: a case study. J Plan Educ Res 2004;23(4):402–13.

13. Nielsen Reports. "Good Value" is the top influencer of U.S. grocery store choice (a survey of internet users in 47 countries). 2007. us.nielsen.com/news/20071217.shtml
14. Kaufman PR, MacDonald JM, Lutz SM, Smallwood DM. Do the poor pay more for food? Item selection and price differences affect low-income household food costs. USDA, Food and Rural Economics Division, Economic Research Service, 1997. Report No.: AER-759.
15. Lindbladh E, Lyttkens CH. Habit versus choice: the process of decision-making in health-related behaviour. Soc Sci Med 2002;55(3):451–65.
16. East R, Harris P, Lomax W, Willson G, Perkins D. First-store loyalty to U.S. and British supermarkets: Occasional Paper Series No. 27. Kingston Business School, London, 1997.
17. NPD group. "Fast Checks" Study: consumers speak out on the U.S. economy, Port Washington NY, 2008, www.npd.com/press/releases/press_080228.html.
18. Cobb CW, Douglas PH. A theory of production. Am Econ Rev 1928;18(1):139–65.
19. Simon HA. Rationality in psychology and economics. J Business 1986;59(4 pt 2):S209–S224.
20. Auchincloss AH, Diez Roux AV, Brown DG, Erdmann CA, Bertoni AG. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. Epidemiology 2008;19(1):146–57.
21. Chung C, Myers SL. Do the poor pay more for food? An analysis of grocery store availability and food price disparities. J Consumer Affairs 1999;33(2):276–96.
22. Davis EE, Freedman M, Lane J, McCall BP, Nestoriak N, Park TA. Product market competition and human resource practices: an analysis of the retail food sector. Working Paper Series, Social Science Research Network, ssrn.com/abstract=717403.
23. Moore LV. Measuring the local food environment and its associations with diet quality. 2007. gradworks.umi.com/32/53/3253361.html.
24. Goldstein I, Loethen L, Kako E, Califano C. CDFI financing of supermarkets in underserved communities: A case study. The Reinvestment Fund, 2008. www.trfund.com/resource/downloads/policypubs/Supermarkets_Full_Study.pdf]:
25. Wallace R. Urban desertification, public health and public order: "planned shrinkage," violent death, substance abuse and AIDS in the Bronx. Soc Sci Med 1990;31(7):801–13.
26. Axtell R. Why agents? On the varied motivations for agent computing in the social sciences. Center on Social and Economic Dynamics, 2000.
27. Rykiel EJ. Testing ecological models: the meaning of validation. Ecological Modeling 1996;90(3):229–44.
28. Brown DG. Agent-based models. In: Geist H, editor. The Earth's changing land: An encyclopedia of land-use and land-cover change. Westport CT: Greenwood, 2006.
29. Page S. Path Dependence. Q J Polit Sci 2006;1:87–115.
30. North MJ, Collier NT, Vos JR. Experiences creating three implementations of the Repast agent modeling toolkit. ACM Transactions on Modeling and Computer Simulation 2006;16(1):1–25.