Detailed Description of HEBSIM

Appendix A: Agents and Attributes

HEBSIM includes two types of agents: households and food outlets. These agents operate in an environment defined by neighborhoods within a city. Table A1 gives an overview of all attributes in the model.

Environment

The city of Eindhoven, situated in the South-Eastern part of the Netherlands, was modeled. The city is approximately 88 km^2 and includes 116 neighborhoods of which 88 are residential neighborhoods. It is the fifth largest city of the Netherlands with 213,223 inhabitants in 2012 A grid was constructed using a GIS-file (raster-format) of the city of Eindhoven, which was obtained from Statistics Netherlands.[1] This file contains a rasterized representation of all neighborhoods of Eindhoven with actual sizes. Each grid cell is 10m x 10m in size and can be occupied by a household and a food outlet or a vacant (food and non-food) outlet. Unoccupied grid cells are considered ‘other areas’, such as parks and roads.

Distribution of Agents

Households, food outlets, and vacant outlets are distributed on the grid following empirical data. Locations of households and vacant outlets were determined by the number of households or vacant outlets per neighborhood obtained from Statistics Eindhoven.[2] The exact location within each neighborhood was randomly determined assuming that household locations and vacant outlet locations were clustered (default: 3 clusters per neighborhood). Locations of food outlets were actual locations based on a search in the Yellow Pages in July 2012. A location is defined by x and y coordinates. In 2012, Eindhoven had 97,523 households, 12 fruit vegetable stores, 33 supermarkets, 22 discount supermarkets, 160 fast foods, and 279 vacant outlets.

Household

Initial conditions of households' attributes are based on data on the population in the city of Eindhoven. Besides a location, each household has three other attributes: (1) household composition; (2) income level; and (3) preference for healthy food. Household attributes do not change during the simulation.

Household Composition

The model included four types of household compositions: single person, single parent, multi-person without children, and multi-person with children (Table A1). The household composition was assigned following an empirical distribution obtained from Statistics Eindhoven.[2] Household size was determined by the sum of the average number of adults and the average number of children per household. The average number of children in Eindhoven was on average 1.7 in 2012 (50,986 children in 29,384 households with children).[2] Household sizes were: 2.7 members in a single parent household, 2 members in a multi-person household without children, and 3.7 members in a multi-person household with children.

Income Level

Households were categorized into low or high income. Each household was assigned an income based on the distribution of income per neighborhood, obtained from Statistics Netherlands.[1] Low income was defined as households with a combined net household income below $31,777 per year.

Preference for Healthy Food

The preference for healthy food represents a household's attitude towards healthy food. The preference was expressed as a continuous number between 0 and 1, where 0 means prefers unhealthy food and 1 means prefers healthy food. We assumed that the average preference for healthy food differs between high- and low-income households. Based on results of Turrell et al.,[3] the mean preference for healthy food was set at 0.739 and 0.661 for high- and low-income households, respectively.

The preference for healthy food was randomly assigned to high- and low-income households following a Beta distribution with a mean that is equal to the mean preference for healthy food. The preference of high-income households was drawn from a Beta (4, 1.55) distribution (mean = 0.739), assuming that 5.5% have a score less than 0.4 and 16% a score higher than 0.9. The preference for low-income households was drawn from a Beta (4, 2.06) distribution (mean = 0.661), assuming that around 10% of the households have a score less than 0.4 and 7.5% a score higher than 0.9.

Food Outlet

The model distinguishes four types of food outlets: (1) fruit-vegetable stores; (2) supermarkets; (3) discount supermarkets; and (4) fast food outlets. Besides a location, each type of food outlet has 4 other attributes: (1) quality of food; (2) price level; (3) monthly costs; and (4) capital. The model allows the composition of food outlets in a city to change during the simulation. Dynamic processes cause food outlets to close and to start new food outlets (Appendix C). Each simulation run starts with a different number of food outlets allowing the system to arrive at equilibrium, which should be equal to the observed number of food outlets. The starting number of food outlets was randomly drawn from a Uniform distribution: fruit and vegetable stores U(7,17); supermarkets U(28,38); discount supermarkets U(17,27); fast food outlets U(130,190).

Quality of Food

The quality of food refers to the level of healthy food sold in a food outlet, which is either mostly 'healthy' or mostly 'unhealthy'. Healthy food outlets mainly sell healthy, while unhealthy food outlets mainly sell unhealthy food. In the model, we chose to categorize fruit vegetable stores and supermarkets as 'healthy', and discount supermarkets and fast foods as 'unhealthy'. This (crude) assumption was based on findings in the literature[4,5] and a Dutch consumer report.[6] These have shown that people shopping at supermarkets eat and buy healthier food compared to those shopping at discount supermarkets.

Price Level

The price level is defined as the average price level of a food outlet. Food outlets were classified into 'cheap' or 'expensive'. Following a Dutch consumer report,[6] the price level of fruit vegetable stores and supermarkets was set to 'expensive' and the price level of discount supermarkets and fast food outlets was set to 'cheap'.[6]

Monthly Costs

Monthly costs cover all sorts of costs of a food outlet, such as rent. The monthly costs remained unchanged during the simulation and only differed per type of food outlet (i.e., fruit vegetable stores, supermarket, discount supermarket, and fast food). Since actual costs were unknown, this parameter was calibrated such that the number of food outlets in the steady state matches the observed number of food outlets in Eindhoven.

Capital

Each food outlet has capital that increases with revenue and decreases with monthly costs. At the start of the simulation or when a new food outlet is started, each food outlet is also assigned a starting capital to protect a food outlet from directly closing down during times when it does not have enough customers. We assumed that food outlets may survive up to 5 months without customers. To account for heterogeneity, starting capital was randomly determined by a random draw between 0 and 5 times the monthly costs (Uniform distribution).

Table A1. Overview of Agent's Attributes

Household				
Location	*Fixed location on grid (x and y coordinates)*			
Household composition	*'Single-person' or 'Single parent' or 'Multi-person without children' or 'Multi-person with children'*			
Income level	*'Low income' or 'High income'*			
Preference for healthy food	*Range: 0 to 1 (0 = prefer unhealthy food; 1= prefer healthy food); Random draw from Beta distribution*			
Food outlet				
	Fruit vegetable store	*Supermarket*	*Discount Supermarket*	*Fast food outlet*
Location	*Location may change during the simulation (x and y coordinates)*			
Quality of food	*'Healthy'*	*'Healthy'*	*'Unhealthy'*	*'Unhealthy'*
Price level	*'Expensive'*	*'Expensive'*	*'Cheap'*	*'Cheap'*
Monthly costs	*Monthly costs are calibrated*			
Capital	*Starting capital was drawn from a Uniform (0, maximum capital) distribution. Maximum starting capital is equal to 5 times the monthly costs;*			

Appendix B: Behavioral Rules of Households

In HEBSIM, two behaviors of households were modeled: food shopping and fast food visits. Behavioral rules of households were based on literature and quantified using empirical data. The multi-attribute utility theory was used to guide interactions between households and food outlets and to determine behaviors.[7,8]

Food Shopping

Based on a Dutch consumer report[6],we assumed that households shop for food 3 times a week. At each food shopping event, each household selects a food outlet. Households can only shop for food at fruit vegetable stores, supermarkets, or discount supermarkets. In the Netherlands, fruit and vegetable stores are considered complementary to supermarkets or discount supermarkets.[9] We therefore assumed that each household visits a supermarket or discount supermarket at least once a week. Decisions on where to shop for food are based on three determinants: distance to a food outlet, price of a food outlet, and a household's preference for healthy food. These determinants have been pointed out to be important factors for food behaviors or the decision what to eat.[10,11]

In the model, selecting a food outlet for food shopping is determined by assigning a utility to all eligible food outlets using a utility function (Equation 1). The food outlet with the highest utility at that time is selected.

$$Utility_j = w_{distance} \times distance\ score_j + w_{price} \times price\ score_j + w_{preference} \times preference\ score_j + \varepsilon_j \qquad \text{Equation 1}$$

The utility function includes determinants of food behavior and a random noise (ε_j). Each determinant was assigned a score and a weight. Here, the utility of food outlet *j* is determined by the distance to a food outlet, the price level of a food outlet, and the preference. All scores were assigned based on available data. The weights (i.e. $w_{distance}$, w_{price} and $w_{preference}$) indicate the relative importance of the determinants in the decision-making. Random noise (ε_j) represents bounded rationality, which causes people to deviate from the model-derived optimal choice.[12] Random noise was drawn from a Normal distribution with a mean equal to zero and a user-defined SE. Weights and user-defined SE were calibrated such that the proportion of households that visited fruit vegetable stores, supermarkets, and discount supermarkets matched the data obtained from the GLOBE study (Appendix D).

Scores

Scores for food shopping were based on the attributes of the household itself and the attributes of the food outlet. All scores were scaled such that the average score is 1.

Distance Score

Each household calculates the Euclidian distance to a food outlet. The distance score of food outlet *j* is defined as the inverse of the distance to food outlet *j*:

$$Distance\ score_j = \frac{1}{(distance_j)} \qquad \text{Equation 2}$$

Price Score

The price score depends on a food outlet's price level and household's income level. We assumed that high and low income groups value cheap and expensive food outlets differently. Table B1 shows the price scores as used in the model. These scores were obtained from the 2011 wave of the Dutch GLOBE study.[13] Respondents could indicate the importance of different aspects of food on a scale of 1-4, ranging from 'not important at all' to 'very important'. Price scores were based on two of those aspects: importance of cheap food and importance of non-expensive food. A high score on the importance of cheap food indicates that cheap food is important. This score was used as a proxy to value cheap food outlets. A high score on the importance of non-expensive food indicates that non-expensive food is not important. This score was used as a proxy to value expensive food outlets.

Table B1. Price Score by Income

		Income	
		High	Low
Price	Cheap	2.21	2.84
	Expensive	2.43	1.84

*Price scores are calculated from Dutch GLOBE study.[13]

Preference Score

The preference score depends on the quality of a food outlet and the household's preference for healthy food. We assumed that households with a high preference for healthy food value 'healthy' food outlets more than 'unhealthy' food outlets, while households with a low preference for healthy food value ‘unhealthy’ food outlets more than ‘healthy’ food outlets. The preference score of food outlet j was determined as follows:

$$Preference\ score_j \begin{cases} preference\ for\ healthy\ food, & if\ j = healthy \\ 1 - preference\ for\ healthy\ food, & if\ j = unhealthy \end{cases} \qquad \text{Equation 3}$$

Fast Food Visits

Fast food visits were considered daily. The same approach was used to determine fast food visits. Here, each household assigns a utility to visiting fast food and to a utility to eating or cooking at home. The same utility function was used, but it only includes two determinants: (1) price level of fast food and eating at home, and (2) preference for fast food and eating at home (Equation 4). These have been pointed out to be important determinants of fast food.[10] We further assumed that the preference for fast food was also influenced by the availability of fast food in the neighborhood.[14-17]

$$Utility_j = w_{preference} \times preference\ score_j + w_{price} \times price\ score_j + \varepsilon_j \qquad \text{Equation 4}$$

The utility of fast food is determined by the preference for fast food and the price score of fast food. Similarly, the utility of eating at home is determined by the preference score of eating at home and the price score of eating at home. Weights and the random noise were calibrated such that the proportion of households that visits fast food matched empirical data (Appendix D).

The option with the highest utility was chosen. If visiting a fast food outlet was chosen, a household will also choose which fast food outlet to visit. Since all fast food outlets only differ from each other in location, this decision was only based on distance. Using the same utility function, a utility is assigned to each fast food outlet only based on distance:

$$Utility_j = distance\ score_j + \varepsilon_j \qquad \text{Equation 5}$$

The fast food outlet with the highest utility was chosen. The score for distance is defined as the inverse Euclidian distance (See also Equation 2).

Scores

Scores for fast food visits were based on the attributes of the household itself and the attributes of the fast food visits and eating at home. All scores were scaled such that the average score was 1. These scores were assigned as follows:

Preference Score

We assumed that eating or cooking at home is a healthier than visiting a fast food outlet. So the preference score for eating at home is equal to the households' preference for healthy food:

$$Preference\ score\ for\ eating\ at\ home = preference\ for\ healthy\ food \qquad \text{Equation 6}$$

We assumed that visiting fast food is unhealthy and is influenced by the availability of fast foods in the neighborhood. Households living in a neighborhood with a lot of fast food outlets have a higher preference for fast food compared to households living in neighborhoods with no fast food outlet. Several studies have shown that the availability of fast food outlets is associated with food purchasing behavior and BMI.[14-17] The preference for fast food is therefore calculated as:

$$Preference\ score\ for\ fast\ food = (1 - preference\ for\ healthy\ food) + availability\ score \quad \text{Equation 7}$$

$$Availability\ score_i = \frac{number\ of\ fast\ food\ outlet\ in\ neighborhood_i}{total\ number\ of\ fast\ foods} \quad \text{Equation 8}$$

Price Score

The price score depends on the price level of cooking or eating at home and fast food visit. A fast food visit was considered ‘cheap’. Eating or cooking at home could be either ‘cheap’ or ‘expensive’ depending where the household did their latest food shopping. For example, if a household shopped at a discount supermarket the price level of cooking at home was also considered ‘cheap’. Price score were assigned following Table B1.

Appendix C: Behavioral Rules of Food Outlets

Food outlets can respond to decisions of households in two ways: closing a food outlet or starting a new food outlet.

Closure

Every 30 days, a food outlet can close. The decision to close is determined by a food outlet's capital, which increases with every customer (revenue) and decreases with monthly costs. A food outlet's capital accumulates if revenue exceeds monthly costs and depletes if monthly costs exceed revenue. If capital falls below zero, the food outlet closes. The location of the closed food outlet becomes vacant.

Revenue is determined by the number of customers taking into account household sizes. We assumed that each household member has a weekly food budget of 1. Thus, the total food budget of a household is equal to the household size. The food budget is spent on food shopping and fast food visits. In the model, households shopped for food three times a week and could visit a fast food outlet up to six times a week. The minimal amount that a household could spend at a food outlet was therefore defined as the food budget divided by nine:

$$Min. amount\ spent\ at\ food\ outlet_i = \frac{1}{9} \times household's\ food\ budget \qquad \text{Equation 9}$$

We assumed that households spend less at fast food outlets and fruit vegetable stores compared to supermarkets and discount supermarkets.[6] Also, households that visit fast food outlets frequently may spend less on food shopping. The revenue (R) obtained by a visit of a household (i) is therefore calculated as:

$$R_i = Min. amount\ spent\ at\ food\ outlet \quad \text{, if fast food or fruit vegetable store}$$

$$R_i = \frac{Food\ budget - \left(\begin{matrix} min.amount\ spent\ at\ food\ outlet \times \\ [fast\ food + fv\ store\ frequency]_i \end{matrix} \right)}{[supermarket + discount\ supermarket\ frequency]_i} \quad \text{, if supermarket or discount supermarket}$$

Equation 10

New Food Outlet

Every 30 days, new food outlets can be started at a vacant location. The decision to start a food outlet is determined by an empirical probability distribution per type of food outlet. The monthly probability to start a new food outlet per type of food outlet was calculated using data of the Chamber of Commerce.[18] The data include the fraction of total food outlets that started a new food outlet and the number of new food outlets per month per type of food outlet from April 2011 to May 2013 in the province of Noord-Brabant, in which Eindhoven is situated. The number of food outlet (per type) in Eindhoven was multiplied with the fraction of total food outlets that started a new food outlet to obtain the number of new food outlets in the city of Eindhoven. Using this number, we calculated the monthly probability to open 0 new stores, 1 new store or 2 new stores per type of food outlet (Table C1).

Table C1. Monthly Probability of Starting a New Food Outlet for Fruit Vegetables Stores, Supermarkets, Discount Supermarkets, and Fast Food Outlets

	Monthly probability			
	Fruit vegetable store	Supermarket	Discount supermarket	Fast food outlet
Number of new food outlets				
0	0.96	0.86	0.91	0.00
1	0.04	0.14	0.09	0.91
2	-	-	-	0.09

The location of a new food outlet will be selected in the neighborhood with the highest expected revenue. The expected revenue in a neighborhood is based on the total revenue of a type of food outlet in the past month:

$$Expected\ revenue_{j,n} = \frac{Total\ revenue_{j,n}}{number\ of\ existing\ food\ outlet_{j,n} + 1} \qquad \text{Equation 11}$$

The exact location of the new food outlet is randomly determined at one of the vacant locations in the selected neighborhood.

Appendix D: Calibration

Calibration was performed through an iterative process until all model outcomes matched the data obtained from the Dutch GLOBE study and literature. Model outcomes include:

1. The proportion of households visited
    a. A fruit-vegetable store
    b. A supermarket
    c. A discount supermarket at least once in the past week
2. The proportion of households visited a fast food outlet
    a. 0 times
    b. 1-2 times
    c. >2 times in the past week
3. The proportion of households that shopped at one type of food outlet in the past week
4. The number of
    a. Fruit-vegetable stores
    b. Supermarkets
    c. Discount supermarkets
    d. Fast foods in the city of Eindhoven (See also Table 1 in article).

Parameters for which no empirical data were available were calibrated such that the model outcomes were consistent with the empirical data. Table D1 presents all calibrated parameters of the model and their optimal values.

The weights of food shopping and fast food behaviors and random noise were calibrated such that the proportion of households that visited food outlets matched the data from the GLOBE study (Table 1 in the article). From the calibration process it turned out that preference was the most important determinant followed by price and distance. This finding is in line with our knowledge from data and literature.[11,19] A recent discrete choice experiment among participant of the GLOBE study also showed that preference or taste is the most important determinant followed by price, and finally distance (results not shown). Also literature suggests that preference or taste is the most important determinant, followed by price or cost.[11,19]

The optimal values of the random noise after calibration were 0.8 and 0.6 for food shopping and fast food visiting, respectively. Since the scores of the determinants were on average 1, this indicates that a substantial part is explained by bounded rationality (approximately 40%). However, other determinants, such as lack of time, mood, social support, or social influences, may also play a role.[20,21]

Monthly costs per type of food outlet were calibrated to arrive at the observed number of food outlets accounting for the rate of new start-ups. Table D1 shows that supermarkets have the highest monthly costs, followed by discount supermarket and fruit vegetable stores. Fast food outlets have the lowest monthly costs. It is very reasonable to accept that this is also the pattern that is present in the real world. Supermarkets and discount supermarkets are in general larger and have more personnel and thus have higher monthly costs compared to fruit vegetable stores and fast food outlets.

Table D1. Values of Calibrated Parameters

Calibrated parameters		Values
Food shopping	Weight of distance	0.2
	Weight of price	0.35
	Weight of preference for healthy	0.45
Fast food visits	Weight of preference for fast food	0.6
	Weight of price	0.4
Random noise	Food shopping	0.8
	Fast food visits	0.6
Monthly costs	Fruit-vegetable store	4800
	Supermarket	15000
	Discount supermarket	8300
	Fast food outlet	1025

References

1. Statistics Netherlands (CBS). http://statline.cbs.nl/statweb/?LA=nl.

2. Statistics Eindhoven. http://eindhoven.buurtmonitor.nl/.

3. Turrell G. Socioeconomic differences in food preference and their influence on healthy food purchasing choices. *J Hum Nutr Diet.* 1998;11(2):135-149. http://dx.doi.org/10.1046/j.1365-277X.1998.00084.x.

4. Chaix B, Bean K, Daniel M, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). *PLoS One*. 2012;7(4):e32908. http://dx.doi.org/10.1371/journal.pone.0032908.

5. Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health*. 2002;92(11):1761-1767. http://dx.doi.org/10.2105/AJPH.92.11.1761.

6. EFMI. Consumententrends 2011. Leusden: Centraal Bureau Levensmiddelenhandel; 2011.

7. Keeney RL, Raiffa H. *Decisions with multiple objectives: preferences and value tradeoffs*. New York: Wiley; 1976.

8. Dyer J. *Maut — Multiattribute Utility Theory*. In: *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer New York; 2005. p. 265-292.

9. Retail DB. Consumentenonderzoek 2013: Deloitte; 2013.

10. Story M, Neumark-Sztainer D, French S. Individual and environmental influences on adolescent eating behaviors. *J Am Diet Assoc*. 2002;102(3 Suppl):S40-51. http://dx.doi.org/10.1016/S0002-8223(02)90421-9.

11. Glanz K, Basil M, Maibach E, Goldberg J, Snyder D. Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *J Am Diet Assoc*. 1998;98(10):1118-1126. http://dx.doi.org/10.1016/S0002-8223(98)00260-0.

12. W.G. K. *Modelling human behaviour in agent-based models*. In: Heppenstall AJ, Crooks AT, See LM, Batty M, editors. *Agent-based models of geographical systems*. Springer; 2012. p. 167-179. http://dx.doi.org/10.1007/978-90-481-8927-4.

13. van Lenthe FJ, Kamphuis CB, Beenackers MA, et al. Cohort Profile: Understanding socioeconomic inequalities in health and health behaviours: the GLOBE study. *Int J Epidemiol*. 2014;43(3):721-730. http://dx.doi.org/10.1093/ije/dyt040.

14. Reitzel LR, Regan SD, Nguyen N, et al. Density and proximity of fast food restaurants and body mass index among african americans. *Am J Public Health*. 2014;104(1):110-116. http://dx.doi.org/10.2105/AJPH.2012.301140.

15. Chou SY, Grossman M, Saffer H. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *J Health Econ*. 2004;23(3):565-587. http://dx.doi.org/10.1016/j.jhealeco.2003.10.003.

16. He M, Tucker P, Gilliland J, Irwin JD, Larsen K, Hess P. The influence of local food environments on adolescents' food purchasing behaviors. *Int J Environ Res Public Health*. 2012;9(4):1458-1471. http://dx.doi.org/10.3390/ijerph9041458.

17. Dunn RA, Sharkey JR, Horel S. The effect of fast-food availability on fast-food consumption and obesity among rural residents: an analysis by race/ethnicity. *Econ Hum Biol*. 2012;10(1):1-13.

18. commerce Co. Brancheinformatie. www.ondernemersplein.nl/brancheinformatie/.

19. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. *Annu Rev Public Health*. 2008;29:253-272. http://dx.doi.org/10.1146/annurev.publhealth.29.020907.090926.

20. Steenhuis IH, Waterlander WE, de Mul A. Consumer food choices: the role of price and pricing strategies. *Public Health Nutr*. 2011;14(12):2220-2226. http://dx.doi.org/10.1017/S1368980011001637.

21. Kamphuis CB, van Lenthe FJ, Giskes K, Brug J, Mackenbach JP. Perceived environmental determinants of physical activity and fruit and vegetable consumption among high and low socioeconomic groups in the Netherlands. *Health Place*. 2007;13(2):493-503. http://dx.doi.org/10.1016/j.healthplace.2006.05.008.