A discrete-choice approach to modeling food shopping behavior

Amy Hillier, Tony Smith, Carolyn Cannuscio, Allison Karpyn and Karen Glanz

Abstract:

Keywords: food access, supermarkets, GIS, choice, Philadelphia, food deserts, NEMS

Introduction

The fields of public health and transportation planning have both made important contributions to the study of food access in recent years, but rarely have their distinct approaches to conceptualizing and modeling been combined (Forsythe et al 2010). Public health researchers have introduced conceptual models and measures of food access (Glanz et al, 2005; Glanz et al, 2007) and documented the economic, racial, and geographic disparities that characterize urban food environments (Hilmers et al, 2012; Bader et al, 2010; Treuhaft & Karpyn 2010). Geographic information systems (GIS) have been central to these efforts, allowing researchers to map where people live, the attributes of their neighborhoods, and the location and type of food sources (Charriere et al, 2010, Forsythe et al, 2010). These efforts have popularized the concept of “food deserts”—areas lacking sources of healthful foods (Bornstein, 2012; Kolata, 2012; Yaccino, 2011; Beaulac et al, 2009; Institute of Medicine, 2009; The New York Times, 2009)—and built support for geographically-targeted interventions. But in relying primarily on descriptive approaches to understanding food access, they stop short of modeling the complexities of food shopping behavior and are of limited utility in understanding how households choose food stores, how food store choice influences the foods they purchase, and how food item choice influences health (Kerr et al 2012; Hillier et al, 2011; Thompson et al, 2011; Forsythe et al, 2010; Laska et al, 2010; Ohls et al, 1999).

Transportation planning scholars, on the other hand, have led the way in conceptualizing and measuring accessibility, more generally, by focusing on the choices that individuals make regarding transportation mode choice (Dugundju & Gulyás, 2008; Ulfarsson & Shankar, 2008), travel path (Zhan & Ferreira, 2008) and destination (van den Berg et al, 2010; Tay & McCarthy 1996). Though much of this work has focused on non-work trips, little of this thinking has been applied to research on food shopping (Handy and Clifton, 2001; Pellegrini et al, 1997).

This article brings these distinct approaches together through a behavioral model of food shopping. First, we present a conceptual model of food store choice that is ecological in that it considers the interaction among individual, household, and neighborhood-level attributes with food store attributes and spatial in that it considers the proximity of food stores to where people spend time as well as where they live. We test our conceptual model with data from a door-to-door survey about food shopping using a discrete choice model in the form of a conditional logit model. In doing so, we provide empirical evidence for an
array of factors other than distance from home—namely price, selection, and quality of food and convenience to places other than home where people spend time—that need to be addressed in order to ensure access to healthful food. This article concludes with recommendations for the systematic collection of food shopping behavior data, refinement of the conceptual and statistical models presented here, and specific suggestions for further collaboration between planners and public health professionals to address barriers to accessing healthful foods.

**Distance-based approaches to measuring food access**

Analyses of food access have used a wide range of measures (Charriere et al 2010), but for the most part they use GIS to assess the geographic proximity of residents to food outlets (Gordon et al 2011; Satler et al 2011; Walker et al 2010; Giang et al 2008; Raja et al 2008). This has resulted in descriptive products like the interactive mapping systems developed by the US Department of Agriculture (USDA 2012) and Esri, Inc. (Richardson 2010) that assess food access for small geographic areas across the country.

Because they employ such a wide range of methods to measure access, it is difficult to compare results across studies (Forsyth et al, 2010; Neutens et al 2010; Apparicio et al 2008). More importantly, these models are limited conceptually because they do not account for the many factors other than distance that create barriers to access and, as a consequence, may ignore important sources of social inequity (Krukowski 2012, in press; Weiss et al 2011; Neutens et al 2010). Similarly, they do not account for how individual-level and neighborhood attributes might mediate the effect of distance to a food store, leaving no room for individuals to act as agents, even if those choices are constrained, and for those choices to differ from their neighbors’ choices. Finally, these distance-based descriptive models do not account for actual behavior (Crane & Daniere 1996) which several studies have found generally involves driving to a food store other than the one closest to home (Hillier et al, 2011; Thompson et al 2011; Ohls et al, 1999; Laska et al, 2010; Handy and Clifton 2001). Recent studies of food access have called for incorporating individual and neighborhood-level attributes that might mediate distance (Bader et al 2010) and for models based on individual-level (disaggregate) shopping behavior (Forsythe et al 2010; USDA 2009).

**Behavioral models of accessibility**

Transportation researchers have focused more broadly on accessibility, defined as the “spatial distribution of potential destinations, the ease of reaching each destination, and the magnitude, quality, and character of the activities found there” (Handy and Niemeier 1997: 1175). Travel cost, mode choice, and destination choice have been the three main variables in models of accessibility. The cost of travel, measured both in time and money, is considered the main resistance or impedance while the destination is considered the main attraction or motivation (Koenig 1980). Through the concept of spatial disaggregation, transportation planners have acknowledged that individuals and groups within a specified zone do not necessarily experience transportation options and potential activities the same way, particularly across socioeconomic groups (Handy and Niemeier 1997). For example,
research has shown that most households travel by car for their food shopping (Hillier et al, 2011; Thompson et al, 2011; Ohls et al, 1999). All households that can afford to own a car do not necessarily use them in the same way. If households that have more drivers than vehicles, household members may need to negotiate vehicle use (Clifton 2004). Lower-income car-owners may face problems relating to the age and unreliability of their cars (Blumenberg 2008). People who do not own cars will often borrow them, get a ride from a friend or relative, or take a taxi in order to do their food shopping (Clifton 2004; Clifton 2000).

This interest in the influence of individual and household attributes, such as car ownership, on travel mode and destination choices led to the development of disaggregate choice, or behavioral, modeling (McFadden, 1974, Lerman, 1979). These models now commonly used by economists and transportation experts to model choices made by people and organizations among a finite, exhaustive, and mutually exclusive set of alternatives. The statistical models that fall within the "discrete choice" category are based on a common theoretical framework (McFadden, 1974) and estimate the probability of an individual making a particular choice based on attributes of that individual and the alternatives they are considering. The transportation literature also provides a critique of home-based measures of accessibility and points to multipurpose trips and trip chaining as important considerations (Kryzek, 2003; Ewing, et al 1994; Lerman, 1979).

A small number of marketing studies have used discrete choice models to understand food shopping behavior (Briesch et al, 2009; Bell et al, 1998; Fotheringham, 1988; Arnold et al, 1981). Discrete choice experiments, where study participants are exposed to a number of hypothetical scenarios, or choice sets, are common within the health care and health economics literature (Ryan et al, 2008), but discrete choice models have not been used in public health studies of food access.

**Toward a conceptual model of food store choice**

In Figure 1, we present a spatial-ecological model of food store choice that builds on the extensive public health research on food access and transportation research on accessibility and choice. Consistent with the dominant public health approach to understanding food access, the model is ecological in that it considers interactions among individual, household, and neighborhood factors that can facilitate or constrain access. Like GIS-based maps of food deserts, our model is spatial in that it considers the distance between where people live and where they shop, but rather than focusing exclusively on the environment immediately surrounding one's residence, it considers where daily activities take place, taking into account, for example, that trips may originate at work rather than at home. Finally, consistent with transportation models of origin and destination, it emphasizes choice by considering the food store alternatives individuals consider and the particular store attributes that attract or repel them. Each of these components of the conceptual model is discussed in turn.

[Figure 1 about here]
Individual and household factors

Food store choice reflects, in part, different individual and household attributes including car ownership, income, gender, age, race/ethnicity and nationality, and food preferences (Thomas 2011; Dunkley et al 2004). Car ownership has received the most attention from transportation planners. People who own vehicles usually use them to shop for food because it saves time, allows them to travel further, makes it possible to coordinate their shopping with other errands (trip chaining), and because it is easier to transport heavy or large amounts of groceries by car (Handy 1996). People who do not own cars will often borrow them, get a ride from a friend or relative, or take a taxi in order to do their food shopping (Clifton 2004; Clifton 2000).

Lack of car ownership is generally associated with socio-economic status. Low-income households also differ in that they are more likely to prioritize cost-savings and use comparison shopping (Webber et al 2010; Dunkley et al 2004). Lack of a vehicle, long work commutes, family responsibilities or other time-consuming commitments of low-income households may result in higher “opportunity cost” of food shopping travel (Dunkley et al 2004). Education level may also influence store choice. Yoo et al (2006) found that people with more education were less likely to choose convenience stores. Knowledge about nutrition may also influence store choice. Race/ethnicity, nationality, language, religion, gender and age have also been shown to be associated with store choice and shopping patterns (Krukowski et al, 2012; Bluebird et al 2011; Chaufan & Constantino 2011; Franzen et al 2010; Gittelsohn et al 2010; Rose et al 2010; Morland and Filomena 2008; Yoo et al 2006; Ayala et al, 2005; Handy and Clifton 2001).

Neighborhood environment

Attributes of neighborhoods relating to their urban form, including street connectivity, land use mix, and residential density, have been shown to influence food shopping behavior (Kerr et al 2012). There is some evidence that crime and traffic hazards might also influence store choice, particularly for people who walk or use public transportation (Bader et al, 2010; Odoms-Young et al, 2009; )

The availability of transit at both the origin and destination of a food shopping journey is another frequently acknowledged factor in food store access. Even though most households drive or are driven to do their food shopping, access to public transportation may be relevant to the remaining percentage as well as those exercising what Clifton (2004) calls a “secondary mobility strategy”—a back-up plan when driving is not viable. Walking to transit as well as to food stores introduces a number of new potential impediments including the safety and comfort of the built and social environment (Bader et al 2010).

Activity space

Our conceptual model considers the geographic area in which shoppers spend time rather than just the area around a shopper’s home as being relevant to convenience. Proximity of food stores to work, school, child care and other shopping has been shown to be important...
(Krukowski et al, in press; Kerr at al, 2012; Thomas 2011; Webber et al, 2010; Clifton 2004), relating directly to the transportation concept of “trip chaining.” Recent behavioral health research has given greater attention to the places where people spend time—their “activity space environment” (Zenk et al 2011a) and their “space-time behavior” (Sarloos et al 2009)—in order to understand the impact of the built environment since aggregate or population-level data makes it difficult to identify a causal process. This approach is more familiar within the fields of behavioral geography and transportation, which make use of disaggregate data such as activity diaries and activity-based models (Sarloos et al 2009).

**Store attributes**

Because shoppers in urban areas perceive options among food stores, the attributes of those stores influence their choice. Relevant attributes include the type and size store, price, selection, and quality of products, and store cleanliness, safety and customer service (Krukowski et al, 2012; Weber 2010; Wang et al, 2008; Yoo et al 2006; Handy and Clifton 2001). Several qualitative studies have offered insight into the social dynamics in food stores that may influence choice, including concerns about the sale of alcohol and drugs, fear of violence, and differences in ethnicity among shoppers and store owners (Zenk et al 2011b; Cannuscio et al 2010). The attributes of other food shoppers—including race/ethnicity and income—may also influence choice.

**Summary of conceptual model**

The basic model to be developed is a probabilistic model of store choice. Since different types of people have different shopping preferences, our model leaves room for interaction effects. For example, individuals who do not own a vehicle might shop differently based on the availability of public transit, and individuals who prioritize cost savings may be willing to travel further distances than those who prioritize convenience. Our data collection efforts thus focus on both attributes of individuals and stores.

**Methods**

We tested our conceptual model using data collected for a study about food access and physical activity conducted within six contiguous ZIP codes in West and Southwest Philadelphia. The study area is predominantly black (75.2%) with poverty rates slightly above the citywide average (27.6%, compared to 24.2%) and a homeownership rate somewhat below the city’s average (47.0%, compared to 54.1%). All procedures were approved the University of Pennsylvania Institutional Review Board.

**Individual attributes and store choice**

We conducted a door-to-door survey about food shopping and physical activity on 30 randomly selected blocks within the study area during the summer of 2010. Eligible participants were adults who were the primary food shoppers for their households, could speak and understand English, and did not have physical limitations that precluded being physically active (Hillier et al in review).
We collected demographic data, including sex (recorded by interviewer), race/ethnicity (reported by participant), and information about whether they were working full or part time, were a student full or part time, and received public assistance including SNAP or cash benefits, or owned one or more vehicles. The location of the store where participants reported doing most of their food shopping, the location of the place where they reported spending most of their time when they were not at home, and the centroid of the residents' block (used as a proxy for home address since home address was not recorded with survey results) were geocoded in ArcGIS 10.1.

**Neighborhood attributes**

Census tract-level data on racial composition was obtained from the 2010 US Census. Tract-level data on poverty was obtained from the pooled 2005-2009 American Community Survey. 2010 census tract-level data for aggravated assaults and narcotics-related arrests per 10,000 were obtained from the Philadelphia Police Department. The lowest tract-level crime rate within the city was used as a proxy for crime rate for suburban areas. The number of bus stops (including each bus line and each direction) within ¼ mile of block centroids and chosen stores and the shortest network distance from block centroids and chosen stores to the nearest trolley, subway, or high speed or regional rail station were calculated in ArcGIS 10.1 using shapefiles from the Southeast Public Transit Association (SEPTA). Stores were then categorized as being within ¼ mile of a trolley, subway or train stop or not.

**Food store attributes**

Food stores within the study area were identified using a list of stores that are authorized to accept Supplemental Nutrition Assistance Program (SNAP benefits and visited to confirm their location and status. All food stores, excluding bakeries, produce stores, and fruit and vegetable trucks, were surveyed using the Nutrition Environment Measure Survey for Stores (NEMS-S) developed by Glanz et al. (2007). The NEMS-S instrument is used to identify food items available in retail establishments, including both regular items (such as whole milk or ground beef) and more healthful alternatives (such as reduced fat milk and lean beef or turkey). NEMS-S also measures the quality of fresh fruits and vegetables (acceptable/ unacceptable) and price. It has been shown to be associated with diet (Franca et al., 2009) and to have high inter-rater and test-retest reliability (Glanz et al., 2007). The standard NEMS-S instrument was modified by including items typically found in small stores such as single-serving beverage options and canned and frozen vegetables.

Trained student and community research assistants completed NEMS-S surveys at 336 stores located within the study area. Survey participants identified an additional 31 stores outside the study area where they did most of their food shopping, so these were surveyed, as well. Scores for each store were calculated based on the availability, quality, and price of food items consistent with the original NEMS-S scoring system. A score for relative price of healthful foods was developed by comparing the price of conventional items, such as whole fat milk or white bread, with healthful alternatives, such as reduced-fat milk or whole-grain
bread. Data collected through NEMS-S was also used to construct an absolute price measure based on the price of a ½ gallon of milk (least expensive variety) and 24-oz loaf of whole grain bread and a selection measure based on the varieties of fresh fruits, fresh vegetables, and whole grain bread.

The selling square footage of food stores was obtained from TradeDimension's Retail Site Database.

[Table 1 about here]

**Conditional logit model**

Given a set of individuals (households) $i \in I$ and stores, $s \in S$, if the set of store alternatives relevant for individual, $i$, is denoted by $S_i \subseteq S$, then our *conditional logit model* takes the general form

$$P_i(s) = \frac{\exp(V_{is})}{\sum_{i \in S_i} \exp(V_{ir})}, \ s \in S_i, i \in I$$

where $P_i(s)$ denotes the probability that store $s$ is chosen by individual $i$ from set $S_i$. These choice probabilities are assumed to depend on the *value*, $V_{is}$, of each store $s$ to individual $i$. As in linear regression, these values are assumed to be representable as linear functions of a relevant set of store attributes, $(x_{sj} : j = 1, \ldots, J)$, such as size (retail floor space) and selection of produce at store $s$. These values may of course differ among individuals, depending on attributes, $(z_{ik} : k = 1, \ldots, K)$, such as the sex and race of the individual. Such value differences can be captured by interacting individual attributes with each store attribute. In addition, the spatial accessibility of stores to individuals is of central importance in the present study. Here the primary measure of accessibility is the travel distance from individual $i$’s residence to each store $s$, designated as *home distance*, $d_{i}(is)$. However, we are also interested in the distance to store $s$ from the place where $i$ spends the most time (such as job location), here designated as *place distance*, $d_{z}(is)$. As with store attributes, the value of these distance accessibilities may differ among individuals. For example, such distances may be less important for car owners. Such effects can again be captured by interacting these distances with individual attributes. Hence in the most general model considered here, values of stores for individuals are taken to be linear functions of the form:

$$V_{is} = \sum_{j=1}^{J} \beta_{j} x_{sj} + \sum_{k=1}^{K} \beta_{kj} z_{ik} x_{sj} + \sum_{h=1}^{2} \theta_{h} d_{h}(is) + \sum_{k=1}^{K} \theta_{hk} z_{ik} d_{h}(is)$$
where the first term on the right hand side involves store attributes together with individual interaction effects and the second term involves distances (residential and place) together with their individual interaction effects.

Following standard terminology, coefficients \( \beta_j \) and \( \theta_h \) are referred to the “main effects” for store attribute \( j \) and distance attribute \( h \), respectively. Similarly, for any given individual attribute, \( k \), coefficients \( \beta_k \) and \( \theta_{kh} \) are referred to as “interaction effects” between \( k \) and, respectively, store attribute, \( j \), and distance attribute, \( h \). To interpret these coefficients, note for example that the effects of store attribute \( j \) can be isolated by considering two hypothetical stores, \( s \) and \( s' \), that differ only with respect to attribute \( j \).

To capture the effects of a unit change in attribute, \( j \), suppose in addition that \( x_{sj} - x_{sj'} = 1 \). Then the relative likelihood of any individual \( i \) choosing store \( s \) versus \( s' \) is seen from (1) and (2) to be of the form:

\[
P_i(s)/P_i(s') = \exp\left[ \beta_j (x_{sj} - x_{sj'}) + \sum_{k=1}^{K} \beta_k z_{ik} (x_{sj} - x_{sj'}) \right] = \exp\left( \beta_j + \sum_{k=1}^{K} \beta_k z_{ik} \right)
\]

So in this context it is clear that “main effect”, \( \beta_j \), reflects that component of change in the relative likelihood of choosing \( s \) versus \( s' \) which is common to all individuals, \( i \).\(^1\)

Similarly, \( \beta_k \), reflects the additional component of change in this relative likelihood that is specific to individuals with \( k^{th} \) attribute level, \( z_{ik} \).\(^2\)

Parallel interpretations can be given to the distance parameters, \( \theta_h \) and \( \theta_{kh} \).

To fit this model, one of course requires data on store choices for individuals. Here we define the relevant store choice for each individual \( i \) to be that store where most food shopping is done. One additional assumption implicit in this model is that all stores, \( s = 1, \ldots, S \), are relevant options for each individual. But given the large number of stores in this study (370), individuals may not even be aware of all stores, let alone consider them as relevant options. Unfortunately, the perceived sets of relevant options for each individual are not available in the present data. So the approach used here assumes that the set of store options for each individual consists of the \( S \) store choices for all individuals. For further discussion of this assumption, see the section on “Goodness of Fit Problems” below.

**Store choices and choice sets**

---

\(^1\) Technically one should add “for all individuals for whom both \( s \) and \( s' \) are relevant options”. But since \( \beta_j \) is clearly independent of these particular option choices, we ignore this complication.

\(^2\) By taking logs in (3), these can also be interpreted as linear changes in “log odds”, similar to logistic regression. Alternatively, one can obtain interpretations in terms of “elasticities” and “cross-elasticities” of substitution, as for example in Section 3.6 of Train (2009).
To fit this model, one of course requires data on store choices for individuals. Here we define the relevant store choice for each individual $i$ to be that store where most food shopping is done. The specification of relevant choice set, $S_i$, for each individual $i$ is a much more difficult issue. Ideally, one would like to have data on the store-choice options actually perceived to be relevant by each individual. But since this data is typically not available (and indeed may not even be fully known to individuals themselves), it is necessary to define such sets exogenously. While one could simply include all 333 stores surveyed, it is clear individuals may not even be aware of all these stores, let alone consider them as relevant options. While there a numerous approaches to defining such relevant options, one appears to be particularly appropriate for this study. Given that the study area consists of 30 spatially distinct block units, one can reasonably to suppose that (based on word of mouth) store choices made by any members of a block unit should at least be considered as relevant options by all members. Based on this assumption, we take the relevant choice set, $S_i$, for each individual $i$ to be the set of all store choices made by individuals on $i$’s block.

**Model Specifications and Results**

*Individual and household attributes*

Of the 622 residents who were contacted in person, 82.6% completed a survey. Of the 514 surveys completed, 63 did not include information about where they did their food shopping or data about the store they chose was not available so those participants were eliminated from these analyses. Among the remaining 451 participants used in the analyses, 73.4% were non-Hispanic Black, 66.1% were women, 15.7% were students, 61.9% were employed at least part time, 30.8% were receiving some type of public assistance including SNAP/food stamps or cash assistance, and 67.0% owned a vehicle.

*Attributes of chosen stores*

Nearly 3 out of 5 participants (59.4%) chose one of six stores from three different national chain supermarkets. Participants overwhelmingly chose national chain supermarkets (86.2%), with much smaller proportions choosing national discount supermarkets like Save-A-Lot and Aldi (11.5%), small locally-owned grocery stores (1.3%) or discount department stores like Target or Wal-mart (0.9%). In total, 70 stores were identified by at least one participant as the store where they do most of their food shopping. However, NEMS-S and TradeDimension data were not available for 14 of these, either because they

---

3 For additional discussion of such choice-set identification issues, see for example Fotheringham (1988) and Pelligrini (1997).

4 As one additional restriction, we require that each choice set contain at least three store alternatives. This necessitated combining four small blocks with their nearest-neighbor blocks to meet this restriction, while at the same time preserving spatial locality as much as possible. The resulting choice sets ranged in size from 4 to 11 stores.
were not primarily food stores and NEMS-S surveys were not complete or because they could not be identified based on the name and address provided by the survey participant. The remaining 56 store were used as food store alternatives, and only the 451 individuals who chose one of these 56 stores was included in these analyses.

The NEMS availability scores for these 56 stores ranged from 4 to 36, with an average of 27.3 while the NEMS overall scores ranged from 4 to 51, with an average of 35.3. Scores for relative price ranged from -4 to 11, with an average of 2.4 while absolute price ranged from $3.74 to $6.98, with an average of $5.52. Selection scores ranged from 0 to 28, average, with an average of 21.3. With respect to size, these stores ranged between 1,000 to 104,000 square feet of floor space, with an average of 30,370 square feet.

With respect to public transportation access, the number of bus stops within ¼ mile of each chosen store ranged from 0 to 138 stops, with an average of 30.0. This includes multiple stops in both directions for the same bus line. In addition, 23 of these 56 stores were within ¼ mile of a trolley, subway or train stop. The distance to one of these transit stops ranged from 0 and 4 miles, with an average of 0.5 miles. With respect to neighborhood attributes around the chosen stores, the demographic and crime attributes of their census tracts varied widely, with an average of 45.6% Black. In addition, average aggravated assault rates and narcotics-related arrest rates were, respectively, 102.4 per 1,000 residents and 139.3 per 1,000 residents.

**Distance to chosen stores**

The distance between study participants’ home and store choice (Home distance) ranged from 0.05 and 28.9 miles, with an average of 1.9 miles (median of 1.2 mile). Participants frequently chose to shop at a store other than the closest store. This Extra distance ranged from 0 to 28.5 miles, or an average of 1.5 (median of 1.0). For 294 of the 451 study participants, data were also available on the place where they spent most time. For these participants, the distances from place to store choice (Place distance) ranged from 0.01 to 20.92 miles, with an average of 2.95 miles (median of 2.07 miles).

**Conditional logit models**

A number of alternative specifications of expression (2) above were explored. Here we report results for only two of these models, and simply sketch our other findings. Model 1 might be designated as a “standard” model of spatial choice in terms of accessibility. In particular, home distance is used as the basic measure of accessibility to shopping. In contrast, Model 2 attempts to compare the relative effects of place distance versus home distance, and hence uses both these measures of accessibility. Since all other variables are

---

5 Here it should also be mentioned that all conditional logit models in this study were estimated using the Matlab program, conditlogit.m, written by James LeSage, which is available on line as part of his suite of Matlab programs at http://www.spatial-econometrics.com.
the same in both models, we shall develop Model 1 in full detail, and then discuss Model 2 mainly in terms of its differences from Model 1.

**Variables used in Model 1**

The initial set of store-attribute variables used for analysis included a number of variables that exhibited inflated standard errors due to collinearity effects, namely store size, NEMS availability scores, full service supermarket status, and product selection. Among these, NEMS availability was considered to be of most importance for our present purposes. Further experimentation showed that store size could also be included without seriously affecting standard errors. So the final set of store attributes used in the initial model consisted of: store size (in square feet), NEMS availability, relative price, and absolute price.

The individual-attribute variables used were mostly indicator variables, consisting of: Black resident, Female, Student, Employed, Public assistance, and Auto. The single continuous variable used was Extra distance. None of these variables exhibited serious collinearity problems.

**Results of Model 1**

The results of this model are shown in Table 1 below, where for example Store size-Black resident denotes the interaction effect between store attribute, store size, and individual attribute, Black resident.

The single most significant main effect is the negative effect of home distance, suggesting that physical proximity is indeed a major factor for participants. In addition, there were a number of interaction effects that suggest differing degrees of importance to different types of individuals. Here the strong positive significance of home distance-auto clearly suggests that closeness is less important for car owners. In addition, the positive significance of Home distance-Female suggests that closeness is also less important to females. On the other hand, the negative significance of Home distance-Student suggests that closeness is more important for students. The other significant main effect is Absolute price, indicating that overall, participants chose stores with higher prices. This is consistent with the choice among Black participants for large local supermarkets where prices tend to be slightly higher. This is also consistent with the strong negative interaction between Absolute price and Extra Distance, indicating that those seeking lower prices tend to travel well beyond the closest supermarket. In addition, this pattern is consistent with the lack of general significance of NEMS availability. Here the strong positive interaction between NEMS availability and Extra Distance again suggests that those seeking such availability tend to travel well beyond the closest store.

[Table 2 about here]

---

6 Experimentation with simulated data shows that conditional logit models tend to be even more sensitive to collinearity effects than linear regression.
Comparative Results of Model 2

The results of Model 2 are shown in Table 1 alongside results from Model 1. All variables are identical except for the inclusion of Place distance along with interaction effects paralleling those for Home distance. While Home distance is somewhat less significant, this may be part be due to the smaller sample size (as mentioned in footnote 7 above). But the key effect here is the much stronger significance of Place distance. Notice also that the significance of all interactions with Home distance have disappeared, while the interaction of Auto with Place distance is now significant. At first glance, this might appear to suggest that Home distance and Place distance are strongly correlated, and that significance is simply shifting from one correlate to another. But in fact, the correlation between Home distance and Place distance in Model 2 is virtually zero (-0.007).

Turning to other variables, we see the same patterns of significant variables with the exception of price variables. Here Absolute price is no longer significant. This may again be partly attributable to the smaller sample size in Model 2. There is also a strong positive significance of the interaction between NEMS relative prices and Extra Distance that is explained primarily by collinearity effects. The correlation between this interaction variable and that between Absolute prices and Extra distance is quite high (0.75). Moreover, when this second interaction variable is removed, the interaction between NEMS relative prices and Extra Distance is no longer even weakly significant.

Public Transport Considerations

Given the focus on car ownership and distance in these models, it would appear that there is a conspicuous absence of variables related to public transportation in our final models. After considerable experimentation, we concluded that these variables were providing little additional explanatory power. With respect to bus stops, all stores in Philadelphia have at least seven bus stops within ¼ mile, so that there is little ability to discriminate between stores on this basis. Only about half the stores were within ¼ mile of a train, subway or trolley, but these stores were for the most part those closest to residences of the 451 participants. So again this variable provided no additional predictive value beyond Home distance, and only served to confound the other main effects.

Store Neighborhood Considerations

Aside from public transportation, the store-neighborhood variables describing local aggravated assault rates and narcotics-related arrest rates are also absent from our final

---

7 For purposes of comparison, Model 1 was rerun using only the 294 participants included in Model 2. Essentially all results were the same, with slightly lower levels of significance reflecting the smaller sample size. So only the full-sample model is reported here.

8 In the version of Model 1 run with the smaller sample (mentioned in footnote 7), Absolute prices was also insignificant. So the insignificance of this variable in Model 2 is also due in part to the particular composition of this smaller sample.

9 The only two stores with fewer bus stops were in New Jersey.
model. First it was found that these two variables were too highly correlated to provide separate information. so only narcotics-related arrest rates were studied further. But as with public transportation, this variable provided almost no additional information. The stores with highest local crime rates tended to be located in predominately Black areas close to many of the participant’s residences. Since this variable only served to confound the other key effects, it was decided to drop it from the analysis.

*Consideration of Goodness of Fit*

While there is no universally accepted measure of “fit” for conditional logit models, the measure reported in Table 2 is the simplest to interpret.\(^\text{10}\) This rate is defined to be the percent of choice situations in which the highest estimated choice probably corresponds to the alternative actually chosen. This is particularly appropriate in view of our definition of “store choice” as the most frequently chosen alternative. In conditional logit models there are many more choice alternatives (here ranging from 4 to 11), so the significance of these success rates should be compared with random predictions.\(^\text{11}\) The present success rates are seen to be more than two and a half times larger than would be expected by chance alone. While there is still much room for improvement, identification of significant explanatory variables is meaningful even when goodness of fit (in terms of R-square) is far from perfect.

*Discussion and Conclusion*

The empirical results provided support for many elements of the conceptual model, as shown in Figure 2 below. They show that distance from home is an important consideration, but other factors—including distance from places where people spend time—is also central to understanding food store choice. People who live in the same area often choose different food stores because their attributes and preferences vary. Our results indicate that sex, race, and car ownership are all factors that can mediate the effect of distance as different types of individuals prioritize different factors in store choice. Store attributes—store size, food selection, and food prices—were all statistically significant factors.

[Figure 2 about here]

Race had more of an effect on choice in this study than we expected. Black participants were more likely to choose larger stores while non-black participants were more likely to choose stores with more healthful foods available. It is unclear whether this reflects preferences or the types of stores located in areas where black versus white residents

\(^{10}\) Other standard measures are discussed in section 3.8.1 of Train (2009).

\(^{11}\) To compute this rate for a series of choice situations, \(i = 1, \ldots, N\), where the number of alternatives in each situation, \(i\), is given by \(n_i = \#S_i\), the expect success fraction in each situation \(i\) under random predictions is simply \(1/n_i\). So the overall expected random success rate, SR, is given (in percentage terms) by \(SR = \frac{100}{N} \sum_{i=1}^{N} 1/n_i\). In the present case, with \(N = 451\) choice situations, this yields the value \(SR = 14.197\).
lived. Sex was a significant factor, as well. Consistent with the findings of Handy and Clifton (2001), men were more likely to choose stores close to home and women were more likely to travel beyond the closest store. In our survey, men were almost twice as likely to say they chose their food store based on proximity and convenience, and women were more likely to cite other factors such as price, quality, and selection.

Socioeconomic status was not significant in the conditional logit models, either when measured as households receiving any form of public assistance or not or when combining household income and education level to create a composite SES variable. The variable for Black resident was collinear with Public assistance, but even when race was removed from the model, SES was still not significant. Thomas (2011) found some difference in store choice behavior among food secure and food insecure households, but the difference was less than he expected and food secure and food insecure households held similar perceptions of food stores in their neighborhood.

As expected, car owners tended to travel further than non-car owners for their food shopping. Overall, distance from home was a significant factor, but it was less important for car owners. Study participants would travel further to stores with lower prices and greater availability of healthful foods, but all things being equal, stores closer to home were preferred. The results from the final model provide support for the relevance of activity space since distance from the place where study participants spend time was significant and, when added to the model, took away some of the explanatory power of distance from home. When explaining the reason they chose a particular store, some participants cited proximity to work or family members; others simply said “convenience” which could have included either convenience based on places where they spend time as convenience from home.

**Strengths and Limitations**

Overall, these analyses demonstrate that it is possible to model store choice behavior in order to identify the factors such as prices and availability of healthful foods that, together with distance and convenience, may serve as barriers or facilitators to food access. They also demonstrate that how people access food varies based on their individual and household attributes, suggesting that based on sex, race, and car ownership, individuals perceive the food landscape differently and that a single map cannot assess food access for all residents.

An important strength of this study was the combination of food shopping behavior data, collected through a resident survey, with detailed data about food stores, collected through field visits. While primary data collection is more time-consuming and costly than reliance on administrative data, we demonstrate that it is feasible to collect such data for a relatively large area and that such data contributes new understanding to how and why people choose to shop at certain food stores.
The predictive value of the conditional logit models might be improved by incorporating additional variables, particularly relating to store attributes. These might include the design of food stores, the availability, cost, and type of parking, and pedestrian infrastructure (Handy and Clifton 2001). A number of audit instruments exist for measuring selection and price at food stores, but no such validated measures of customer service exist. The percent of sales at a store paid for with SNAP benefits might provide some insight into class-based preferences. The marketing literature identifies store advertising practices, including weekly circulars (Bodapati & Srinivasan, 2006), and the number and types of brands (Briesch et al, 2009) as additional factors in store choice. Check-out options (express lines and self-checkout), the availability and type of grocery bags (membership fees such as those required for wholesale stores), in-store amenities such as a bank or ATM, coupon policies, special sales and promotional policies, and authorization to accept federal WIC and SNAP/food stamp benefits might also be incorporated into future models.

Refining other variables could also help improve the model. The measures of food item price were based on the data collected through the NEMS-S instrument and only included milk and whole grain bread since these were two items available for all stores. A more robust measure of price would include more items, particularly produce and meats which may vary more significantly. Alternative measures of activity space should also be explored. The location where people spend time when they are not at home is a relatively limited proxy for all of the places where people may spend time, and it does not take into consideration the routes by which individuals travel. A food store located along the way home from work might be just as convenient as a store located near work. Activity space can be measured using travel diaries and GPS (Kerr et al, 2012; Zenk et al, 2012), but these are expensive and time-consuming approaches. Simpler approaches might include a series of survey questions that can approximate where people spend time and the main routes by which they travel.

The predictive value might also be improved by refining the set of stores identified as alternatives. Most shoppers perceive fewer choices than are actually available (Robinson and Vickerman, 1976; Fotheringham, 1988) and “choice set specification” may have significant impact on parameter estimates as well as goodness of fit measures (Pelligrini et al, 1997). Future studies should try to identify these perceived options, either by asking residents the stores they ever shop at or the stores from which they choose. By modeling only the store where residents reported doing most of their food shopping, this study ignores that fact that most people shop at multiple stores.

Ultimately, we only modeled the primary choice of food store. The marketing literature supports the concept of habitual behavior and shopper loyalty to a primary store (US Grocery Shopper Trends, 2009; Bell et al, 1998; Rhee and Bell; 2002), but individuals and households shop at different outlets for different types of shopping (Hirsch & Hillier, in review; other references?). Future models might model shopping behavior for different types of shopping. Finally, this study takes the first step of modeling food store choice, but it does not connect food store choice to the foods residents purchase and eat or their health outcomes. Future research must take the next steps in identifying causal pathways.
between food shopping, food purchases, consumption, weight status, and health outcomes (Rose and Bodor; 2010; USDA, 2009). A limited number of studies have analyzed data about food shopping behavior collected automatically through store loyalty cards (REFERENCES) and SNAP and WIC program participation (REFERENCES). These types of redemption data could prove critical to making these links between store choice, food choice, and health outcomes.

**Implications for planning and public health**

This study provides support for the increasing interest and examples of collaboration between planning and public health. From a research perspective, the two fields bring different strengths that can contribute to better understanding of behavior in the context of food shopping as well as other areas such as medical care, housing, and social services. Planning scholars are poised to make important contributions to these efforts because of their expertise in collecting and analyzing disaggregate choice data. In addition to familiarity with discrete choice models, facility with GIS-based measures that incorporate distance and travel routes and activity data is essential. Public health researchers, on the other hand, have expertise in conceptualizing and measuring the food environment and linking food access to health outcomes. Together, planning and public health scholars can refine our understanding of the barriers and facilitators to accessing healthful foods and provide evidence to refine and potentially support a variety of food subsidies (REFERENCES) and place-based interventions such as corner store campaigns (REFERENCES), farmers’ markets (REFERENCES), urban farms and community gardens (REFERENCES), community-supported agriculture (REFERENCES), transportation improvements (REFERENCES) and tax credits (REFERENCES).

Supermarket tax credit programs have emerged as a primary state and federal strategy for improving food access and provide the strongest example of how planning tools and institutions are critical to meeting public health goals. President Obama and the US Congress established the National Fresh Food Financing Initiative (NFFFI) in 2010, allocating $45 million for the Community Development Financial Institutions Fund for the New Market Tax Credits fund, Department of Health and Human Services, and USDA’s Farmers’ Market Promotion Programs. The fiscal year 2012 budget allocated $33 million and President Obama’s fiscal year 2013 budget requests $285 million for NFFFI. Since Pennsylvania established its model Fresh Food Financing Initiative, sixteen states and three cities have passed legislation for their own FFFIs (Policy Link 2012; US Department of Health and Human Services 2010). Choice modeling like that presented in this article can help inform tax credit policies, including determining the best locations and additional criteria such as the availability of healthful foods, SNAP and WIC authorization, and competitive prices.

Guided by concerns for equity within urban areas and equipped with tools for modeling access and reshaping cities, planners are needed to work with public health researchers and practitioners to improve access to healthful foods and other essential resources and help realize the promise of cities.
ACKNOWLEDGEMENTS

This project was supported by Agriculture and Food Research Initiative Grant no. 2010-85215-20659 from the USDA National Institute of Food and Agriculture Human Nutrition and Obesity Program. We are grateful to Jana Hirsch, Auguste Dutcher, Alexandra Kaplan, Keith Gantt, Paul Bailey, Akim Cooper, Sharon McCollum, Katie Oberwager, Laura Riley, Michelle Fang, and Aliou Diallo for their assistant with field data collection, Latifah Griffin and Nicole Thomas for supervising the field staff, and the Philadelphia Area Research Community Coalition for their advice about conducting field work in West and Southwest Philadelphia.

REFERENCES

Aday LA, Andersen RM., 1974, “A framework for the study of access to medical care” Health Services Research Fall 208-220


Ewing R, Haliyr P, Page GW, 1994, “Getting around a traditional city, a suburban planned unit development, and everything in between” Transportation Research Record 1466 53-62

Fotheringham AS, 1988, “Consumer store choice and choice set definition,” Marketing Science 7(3) 299-310


Franzen L, Smith C, 2010, “Food system access, shopping behavior, and influences on purchasing groceries in adult Hmong living in Minnesota” American Journal of Health Promotion 24(6) 396-409


Hansen WG, 1959, ”How accessibility shapes land use,” Journal of the American Institute of Planners 25 73-76

Hensher DA, Stopher PR, eds, 1978 Behavioural Travel Modelling (Croom Helm, London)


Krukowski RA, McSweeney J, Sparks C, West DS, 2012 "Qualitative study of influences on food store choice" *Appetite*


Laraia BA, Siega-Riz AM, Kaufman JS, Jones SJ, 2004, Proximity distance to food stores and adolescent male fruit and vegetable consumption: Mediation effects of supermarkets is positively associated with diet quality index for pregnancy” *Preventive Medicine* **39** 869–75


Mehta NK, Chang VW, 2008, “Weight status and restaurant availability a multilevel analysis” American Journal of Preventive Medicine 34(2) 127-33


Penchansky R, Thomas JW, 1981, “The concept of access—definition and relationship to consumer satisfaction” Medical Care 19(2) 127–140

PolicyLink, 2012, Improving access to healthy food


Rose D, Richards R, 2004, “Food store access and household fruit and vegetable use among participants in the US Food Stamp Program” Public Health Nutrition 7(8) 1081–1088

Rose D, Bodor JN, Hutchinson PL, Swalm CM, 2010, “The Importance of a multi-dimensional approach for studying the links between food access and consumption” The Journal of Nutrition 140 1170–1174, 2010


van den Berg P, Arentze T, Timmermans H, 2010, Location-type choice for face-to-face social activities and its effect on travel behavior *Environment & Planning B: Planning & Design* 37(6) 1057-1075


Webber CB, Sobal J, Dollahite JS, 2010, “Shopping for fruits and vegetables: Food and retail qualities of importance to low-income households at the grocery store” *Appetite* 54(2) 297-303


Zenk SN, Odoms-Young AM, Dallas C, Hardy E, Watkins A, Hoskins-Wroten J, Holland L, 2011b, “You have to hunt for the fruits, the vegetables”: environmental barriers and adaptive strategies to acquire food in a low-income African American neighborhood” *Health Education Behavior* **38**(3) 282-92