HEALTHINESS OF HOUSEHOLD FOOD EXPENDITURE IN URBAN AND PERI-URBAN KENYA: HOW MUCH IS EXPLAINED BY A SPATIAL MEASURE OF THE FOOD ENVIRONMENT

By

Ian Fisher

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food and Resource Economics- Master of Science

2023

ABSTRACT

Both the Kenyan government and international development organizations have labeled problems of food access and availability a pressing policy issue. Due to many rapidly changing forces such as the nutrition transition, urbanization, and food systems change, consumers in low-income settings face a plethora of food purchasing choices, in terms of what to buy and where to buy it. These choices reveal themselves in the food expenditure behavior of consumers in response to their rapidly changing food environment (FE). People rely more on markets now than on own production, and thus food is becoming more purchased and processed to varying degrees of healthiness (Kenya National Food and Nutrition Security Policy 2011). Based on these themes and using a cross-section dataset from household-level and food environment surveys, food expenditure data is regressed against a proximity-to-outlet measure for various outlet types. The main result of this paper is estimated distance elasticities that measure the responsiveness of household's food shopping expenditure to variations in distances to different types of food outlets. The results generally show that household's location characteristic matters more than household's average distance to outlets when it comes to predicting healthiness of food purchase. The paper also finds heterogeneity in these distance elasticities by characteristics such as main shopper age as well as household poverty probability and location. This research contributes a new application of the distance elasticity concept previously featured in international trade literature. It motivates future studies on food environment metrics within urban and peri-urban settings of low- and middle-income countries (LMICs) to better understand what drives food shopping behavior and how to increase food expenditures toward more healthy food options.

This thesis is dedicated to Cam, Mom, Dad, and Colin.
Thank you for all your love and support.

ACKNOWLEDGEMENTS

Research for this study is supported by the "Support for Applied Research and Analysis in Kenya and East Africa" (SARA-KEA) project with funding from USDA-FAS under grant No. FX21TA-10960R001 and FX22TA-10960R002. Data used for this research was collected with partial support from two grants- 1) the SARA-KEA project, and 2) the Feed the Future Innovation Lab for Food Security Policy, Research, Capacity, and Influence (PRCI) with funding from the United States Agency for International Development under grant No. 7200AA19LE00001.

The author would like to thank his committee of Dr. Mywish Maredia, Dr. David Ortega, Dr. Tom Reardon, and Dr. Dave Tschirley for their mentorship and support throughout the entire research process. Additionally, he would like to thank Dr. Ayala Wineman from MSU, Ph.D. students Nahian Bin Khaled, Jiawen Liu, and Myat Thida Win from MSU, Dr. Timothy Njagi from Tegemeo Institute (Kenya), and Aisha Sophia from IPA for their consultation and assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION 1: INTRODUCTION

1.1: The importance of food environments for better nutrition

Ensuring availability of and access to healthy, nutritious food is a pressing issue of modern economic and agricultural development. These are key agenda items that many developing countries face alongside challenges of malnutrition and food insecurity. The Kenyan government, for example, noted the importance of the availability and access to adequate, diverse, and healthy diets by households to obtain proper nutrition levels and made this a priority in their 2011 National Food and Nutrition Security Policy strategy document (Kenya Ministry of Agriculture 2011). At the global scale, this priority is also reflected in the United Nation's Sustainable Development Goals #2 of ensuring access and availability to safe, nutritious, and sufficient food for all people year round as well as in the goal of eradicating all forms of malnutrition (FAO 2020).

Concurrently, ongoing trends such as the nutrition transition, urbanization, and food systems transformation have significantly altered the landscape of food environments. The nutrition transition has been well-documented as rapid changes over time in dietary patterns towards more processed foods, a decline in activity levels due to economic changes, and a rise in non-communicable diseases (NCD's) such as cardiovascular disease (Popkin 2018 & Monteiro et al. 2013). Studies show that 61-83% of the food the global middle class consumes is purchased, and processed food occupies 70-80% of food expenditure for the middle class (Tschirley et al. 2015). What makes this problem particularly pressing, especially in urban populations, is the negative nutritional impacts on the poor (Perez-Escamilla et al. 2018 & Stevano, Johnston, & Codjoe 2019). Urbanization is linked with changes in diet and food composition in low- and middle-income countries (LMICs) (Popkin 1999 and Ruel et al. 2017). The United Nations reports that in Africa, the areas experiencing the fastest urban growth have less than 500,000 inhabitants, "which account for 62% of the continent's urban population" (United Nations 2014). With a shift to urban food environments, consumers have access to a plethora of food options that both promote healthy and unhealthy food for those who can afford them. But, for the urban poor, often the unhealthy food options are the most affordable and more easily available spatially (Hawkes, Harris, & Gillespie 2017 & Crush & Battersby 2017). All the above trends fall within a larger food system transformation that is characterized by, for example, expansion of the food retail sector through supermarkets (Reardon et al. 2003), changes in the broad food system through rising processing and wholesale segments (Popkin and Reardon 2018), and the rise of convenience food options such as fast food (Fraser et al. 2010). This sample of trends reflects how food quality, access, and availability are rapidly changing both at local and international levels.

The current context of health and nutrition in Kenya reflects the various complexities affecting urban and peri-urban households. Per capita food availability of cereals over the past three decades has declined by more than 10% (Kenya National Food and Nutrition Security Policy 2011). Most Kenyan diets are still based on staple crops, such as maize, that lack diversity of nutrients. Generally, studies have found that Kenyan households source their food from a variety of outlets, with 82% purchasing in grocers, 79% in supermarkets, and 69% in kiosks (Owuor 2018). Kiosks are defined in this context as small, free-standing structures selling food along highly trafficked areas. Small shops and Kiosks are used almost daily by many urban and peri-urban consumers. On the food vendor side, the most commonly sold food, according to research, is sweet and confectionary foods (29% of vendors), raw vegetables (28% of vendors), fried starches (23% of vendors), and fruits (21% of vendors) (Busse et al. 2022).

Additional studies find mixed results about the current state of healthiness of the Kenyan population and food situation especially during COVID-19. The national rate of stunting among children aged 0-59 months decreased by 1.6% per annum from 1993 to 2014 and was correlated with wealth quintiles and poverty levels (Republic of Kenya 2021). During the COVID-19 pandemic, it is reported that 58% of Kenyans “were consuming inadequate quantities and compromising on diet quality,” which reflects changes in both food environment and food security (Republic of Kenya 2021). Food availability was generally sufficient; 80% of households reported availability of widely purchased food such as rice, vegetables, and onions, but at higher prices (Republic of Kenya 2021).

Access involves geographic relationships between households and food sources such as the location of the food supply and ease of getting to that location while availability is the adequacy of the supply of healthy food, such as the presence of certain types of food outlets (Clapp et al. 2022 & Caspi et al. 2012). Food environments, which can be defined as “the collective physical, economic, policy, and sociocultural surroundings, opportunities, and conditions that influence people’s food and beverage choices and nutritional status” (Turner et al. 2018 and Swinburn et al. 2013), conceptualize the spatiality and interlinked nature of access and availability and have direct impacts on both the supply and demand dynamics of the food system (Global Panel 2017). In addition, enacting policies that support nutrition-sensitive food systems and food environments can help address malnutrition (FAO 2016). Other research suggests that characteristics of the food environment are strongly linked with degree of prevalence of obesity and diabetes (Babey et al. 2018) through the existence of food deserts (Wrigley 2016). Generally, there is opportunity for policy to help nudge consumers' dietary patterns in more healthy directions by influencing their food environments, such as prices, availability, and presentation of healthier options (Hawkes et al. 2015).

Research that has been conducted across the continent of Africa on the topic of food environments and nutrition covers a wide range of topics with diverse findings. In Cape Town, researchers found that those in lower socioeconomic neighborhood classes bought fruits and vegetables less frequently than those of higher classes. It is reported that lack of mobility and low food choice presented challenges for these neighborhoods (Odunitan-Wayas et al. 2018). Generally, the type of food available in the neighborhood and convenience are shown to be important physical-level factors with accessing healthy foods, as well as general economic access (Osei-Kwasi et al. 2021 & Pradeilles 2021). Over a 50-year period in Africa, food environments show an "increasing availability of energy, animal products, fruit and vegetables, vegetable oils, sugar and sweeteners but a decrease in animal fats" (Holdsworth et al. 2019).

Distance to food outlets, a key measurement of accessibility, is an important metric of determining healthiness of food environments. Research shows that proximity to outlets of various levels of healthiness is a strong predictor of consumption frequency from each respective outlet type (Athens, Duncan, Elbel 2016). Other results show that distance to food outlets predicts healthy food perceptions (Barnes et al. 2015). Access as measured by distance from household to food outlets is a core dimension of food security and ensures that populations have the appropriate geographical advantage to find healthy and nutritious food (Clapp et al. 2022).

When considering outlet distance, the healthiness of food is not homogenous across outlet types in the context specifically in Kenya. In the following sections, the data studied in this research confirms that outlets are heterogeneous in terms of healthiness of food offered. We also find in the literature that in Kenya, outlets offer varying amounts of healthy and unhealthy foods (Green et al., 2019), such as the food environment surrounding primary schools (Gewa et al., 2021). Outlet types differ in healthiness, and thus distance is an important determinant of access to healthy foods.

The main research question that will be answered is: how does proximity of the household to the nearest outlet of each type affect the healthiness of food expenditure decisions? This question will be investigated according to two main goals in the paper. The first objective is to investigate how the access dimension of the food environment, measured by proximity to different types of food outlets that sell different types of foods, might explain the healthiness of food expenditure for those who make the primary food shopping decisions for the household. The second goal is to investigate if there are heterogenous effects across socioeconomic and demographic groups in the data sample.

1.2: Review of the literature

While consumer choice theory emerged during the late 19th century and early 20th century (Auspitz and Lieben 1889, Cassel 1923 for example), literature on this topic began to take root in the 1930's until the 1980's. Key contributions during this time associated household income levels with the type of food eaten. For example, Engel's law (Zimmerman 1932), which states that percentage of income allocated for food purchases is inversely related to household income, and Bennett's law, which states that as incomes rise, people substitute calorie-dense starchy staple foods with nutrient dense foods (Bennett 1941), were both popular contributions. Then, moving into the 1960's, the literature focused on household choices as a function of time, meaning there is a cost of time in consumption decisions (Becker 1965). Research also focused on the goods demanded themselves, with emphasis on how the intrinsic properties and services of goods influenced consumers' choices that ultimately gave birth to hedonic price formation (Lancaster 1966, Rosen 1974, and Ladd and Zober 1977).

Moving into the 1980's, the literature surrounding consumer choice moved beyond classical economics to include demographic characteristics of consumers. Literature finds demographic variables to be significant determinants of household consumption patterns (Pollack and Wales 1981 and Heien et al. 1989). Other literature pointed to demographic variables concerning household role, finding that income earned by wives working full time did not alter their food consumption behavior outside the home (Kinsey 1983). More broadly, empirical evidence showed that the amount of effort put into shopping is related to stage of family life cycle, sex, and other demographic variables (Slama and Tashchian 1985). Generally, transformations in global food consumption patterns was attributed to socio-economic and demographic changes (Senauer et al. 1986).

Beginning in the 1990's and early 2000's, applied research in the areas of nutrition, economics, and health pointed towards a documented change in dietary behavior and food consumption leading to a rise in noncommunicable disease. This trend is especially documented in low-income countries (Popkin 1999 and Popkin 2004). The literature attributes this change to social, political, economic, and environmental forces, as well as a demographic as well as epidemiologic transition (St-Onge, Keller, & Heymsfield, S.B. 2003 & Popkin 1993). For example, many households have moved from one to two working parents, leading to greater time constraints to make food consumption decisions. This was empirically shown to be the case in Sri Lanka (Senauer et al. 1991). In another context in Bahrain, it was shown that consumption of traditional foods declined while processed food demand has risen (Musaiger 1990). Attributing factors include increasing income and literacy, changing household structure, and the influence of mass media.

Much of the literature during the early 2000's pointed towards a growing body of work on the role environmental/spatial factors have on the economics of diet and nutrition. Researchers described that this rise in obesity was poorly explained by psychological and social factors at the individual level but could be more holistically explained through the complex social and built environment influences access to affordable, healthy food (Glanz et al. 2005). Generally, the food environment has influence over what consumers purchase and encompasses "availability, affordability, convenience, and desirability of various foods" and its study and measurement can help shed light on effectiveness of agricultural interventions (Herforth and Ahmed 2015). Greater access and availability of less healthy foods relates to poorer dietary outcomes, and there is evidence for associations between price, availability, and dietary intake specifically in the U.S. (Black, Moon, and Baird 2004).

An important sub-wave and consideration with food environment research is how to quantify and show measurable effects of the spatial environment on the economics of nutrition and dietary behavior. One branch studies shelf space as an important metric. One study found that shelf-space availability of unhealthy snack foods was positively associated with BMI, or a 100-meter increase in shelf-space of these foods was associated with additional 0.1 increase in BMI (body mass index) points (Rose et al. 2009). Supermarkets are a specific food outlet type of interest within this body of work since they devote more shelf-space to unhealthy snacks than to fruits and vegetables (Farley et al. 2009). In other work, different shelf-space measures were compared, but it was determined that measures only explain a small fraction of variation in consumer purchase behavior (Tschirley et al. 2021).

A second important sub-wave of food environment metrics literature concerns density or investigating if the concentration of healthy or unhealthy food outlets impacts dietary choice and nutrition economics. Studies find that consumers living near fast-food restaurants had lower fruit and vegetable consumption as well as higher body mass indices (Kruger et al. 2013). Relatedly, "the lower the ratio of fast-food restaurants and convenience stores to grocery stores and produce vendors near people's homes, the lower the odds of being obese" (Spence et al. 2009).

One particular spatial measure of the food environment is proximity or distance to food outlets, and literature has been mixed on the validity of this metric as a determinant of shopping pattern and dietary behavior. On one hand, some find that proximity of supermarkets makes a more diverse diet more accessible, but also encourages consumption of equally accessible energy-dense, highly processed foods (Hawkes 2008 and Laraia et al. 2004). General supermarket usage is inversely related with distance (Koech, Chege, and Bett 2022). But similar studies on the topic of supermarkets found that easy access to supermarkets was associated with increased use of fruits per household (Rose and Richards 2004) or even that there was no association between distance to nearest supermarket and fruit or vegetable consumption (Pearson et al. 2005). Other research points to associations between household proximity to grocery stores, convenience stores, and other retail outlets and consumption of sugar-sweetened beverages (SSBs). But it is also acknowledged that, in addition to access, personal and social values are also affecting food choice, so it is challenging to isolate the effect of proximity exclusively (Laska et al. 2010). For instance, greater availability of fresh vegetable such as close proximity to healthy food outlets is correlated with higher intake of vegetables (Bodor et al. 2008). But the casual link between access and intake is complex. Other work within this sub-wave of the literature points to an inverse relationship between distance to small food stores (convenience stores) and fruit, juice, and low-fat vegetable consumption (Jago et al. 2007). Consumption of fast food also follows this inverse trend with distance (Sharkey et al. 2004).

It is also well-noted in the literature that proximity to food outlets is highly heterogeneous by demographics and socioeconomic status. In Detroit, Michigan, access to fresh produce was significantly lower in predominantly African American communities than in other communities and neighborhoods where African American reside were on average 1.1 miles further from nearest supermarket compared to White neighborhoods (Zenk et al. 2006 and Zenk et al. 2005). Other studies show that in neighborhoods of higher deprivation, there is greater access to unhealthy foods (Black, Moon, Baird, 2014 and Sharkey et al. 2011). Further results confirm that, in the context of higher income countries, those residing in low socioeconomic neighborhoods have higher BMI (body mass index) than high socioeconomic neighborhoods, higher density of small grocery stores is associated with higher BMI among women, and closer proximity to supermarkets correlates with higher BMI among women (Wang et al. 2007).

Distance elasticity, or the responsiveness of a measure to a change in distance, is a term appearing commonly in trade literature but less so in food environment studies. Distance elasticity can be used to study the effect of distance on migration (Schwartz 2023), the responsiveness of trade flows to changes in travel distances to ports (Berthelon and Freund 2008), and the declining effect of distance on international trade (Bergstrand, Larch, and Yotoy 2015). Distance elasticity also appears in agricultural trade specific research (Wang, 2000 and Ebeke and Etoundi 2016). Yet the topic of distance elasticity is hardly seen in the food environment or food choice literature applied to proximity. In one of the few recent papers on the topic, it was found that for the developing country of Vietnam, a 10% decline in supermarket distance to a household increased consumer demand for supermarkets by about 7% (Mergenthaler, Weinberger, and Qaim 2009). There is much still to be learned about how responsive consumers are to changes in outlet distances in terms of food expenditure.

1.3: Research contribution

Given the presented strands of the literature, this research fills several knowledge gaps. First, it focuses on a more descriptive measure of linear outlet distance, disaggregating distances by outlet type specifically within the context of urban/peri-urban Kenya. While some studies have looked at distance or proximity measures of the food environment (Laska et al. 2010; Rose and Richards 2004; Bodor et al., 2008), researchers have not studied how distances to specific outlet types, such as street vendors vs. supermarkets vs. vegetable sellers, etc., might affect the healthiness of food expenditures. This is an important topic to study because better understanding the linkages between food expenditure and specific outlets in the food environment allow for more informed policy response and insight into the role proximity to outlet plays in healthiness of food purchases. Distance to food outlets represents an opportunity cost of time needed to acquire food, and this research seeks to answer how this opportunity cost across different outlet types translates to healthiness of expenditure decisions.

In addition, this research applies distance elasticity, or the responsiveness of the household main shopper expenditure to changes in outlet distance, to food environment research. Much of the literature that uses distance elasticity concept is in trade (Berthelon & Freund 2008 & Bergstrand et al. 2015), with a few exceptions who apply the metric to substitution patterns between stores (Chenarides and Jaenicke 2016) and to analysis of demand systems from modern supply chains (Mergenthaler, Weinberger, and Qaim 2009). Few previous studies have looked at the margin of how additional distance to outlet affects healthiness of expenditure.

This research builds on microeconomic theory through a standard expenditure model, outlined in the following sections, that is analyzed empirically through multiple regression analysis. What this paper contributes to the discussion is adding outlet distance variation by specific outlet type to the healthy food expenditure function. The outcome of the model is also differentiated from previous literature to focus on healthy and unhealthy food expenditure, while most previous studies looked at general outlet proximity’s effect on nutrition/diet indicators. This paper looks separately at healthy and unhealthy expenditure while the majority of previous work looked at overall measures.

1.4: Research questions

The main research questions of this research are:

1. How does proximity of main shoppers’ household to the nearest outlet of each type affect food expenditure? What types of outlets in closest proximity are significant predictors of healthiness (or unhealthiness) of household food expenditure?
2. When looking at expenditures based on distance to outlet, what, if any, socioeconomic, geographic, demographic group heterogeneity exist within the sample?

1.5: Hypotheses

Based on the research questions and previous literature, this paper hypothesizes that healthiness of food offerings varies meaningfully across types of outlets, that healthiness of diets and food expenditure varies meaningfully across households, and that distance to various outlet types meaningfully influences the healthiness of household food expenditure.

A general hypothesis that will be tested is whether distance to outlets that sell healthy foods is negatively correlated with expenditures on healthy foods (Laraia et al. 2004). We categorize the following types of food outlets as selling more healthy foods- mama mbogas, small supermarkets, large supermarkets (for the case of Nairobi only in this paper), dukas, kiosks, cereal shops/poshomills, and depots/wholesalers- and hypothesize that *as distance increases from the main shopper's household location to each of these outlet types, healthy expenditure will decrease (or as % distance to these outlet types increases, % change in healthy food expenditure decreases)*. See Table 2 for a description of each outlet type. This Hypothesis follows similar results from both Rose and Richards (2004) as well as Koech, Chege, and Bett (2022). Evidence is mixed on whether proximity to supermarkets of various sizes leads to healthy or unhealthy outcomes, but little is known about the effect of supermarket distance change on expenditure. Similarly, we categorize the following types of food outlets as selling fewer healthy foods- street vendors, hawkers, bakeries, milk bars/milk atms, butcheries, informal prepared outlets, and hotel/restaurant outlets- and test if the proximity of the main shopper's location to these outlet types is associated with the healthiness of food expenditure.

As noted in the literature review section, proximity to food outlets of various degrees of healthiness is discovered to be heterogenous by demographics (Zenk et al. 2006 and Zenk et al. 2005) and socioeconomic status (Black, Moon, Baird 2014 and Sharkey et al. 2011). A secondary hypothesis of this paper that will be tested is that *different socioeconomic and demographic segments of the population experience healthy access to the home food environment differently, or that group heterogeneity exists with regards to distance elasticity to outlets*.

1.6: Importance of research questions

Generally, the research questions for this paper align with national government goals, as highlighted by the Kenya National Food and Nutrition Security Policy of 2011. When it comes to food availability and access, the objective is to "increase the quantity and quality of food available and accessible in order to ensure that all Kenyans have an adequate, diverse, and healthy diet" (11). This paper conducts research on how food availability as part of the larger food environment ties to distance to outlet types of varying healthiness.

In addition, the research questions focused on in this paper relate to both improving knowledge of measuring the food environment generally and specifically better understanding the food environment context of urban and peri-urban areas of Kenya. Food availability, which is focused on in this paper, is an important arm of ensuring food security and it is important to measure and study how responsive consumers are to changes in distance to nearest food outlets. If, for example, during the Covid-19 pandemic, there is a sudden closure of food outlets near to households, it might be important to see how food expenditure might change as a result as distances are disrupted in terms of healthiness. Or, in another example, how might a new healthy or unhealthy outlet close to a household alter their expenditure on healthy and/or unhealthy foods for their household.

Last, the paper focuses on the potential heterogeneous nature of the food environment through proximity to outlets and healthy food expenditure amount. Studies show that interactions in the food environment differ by socioeconomic and demographic group; what about the urban and peri-urban Kenya context? Do household main shoppers experience the food environment in terms of geographic range to food outlets differently as they purchase food for their households? This paper seeks to answer these questions and shed new light on what effect distance has on expenditure decisions.

1.7: Roadmap

The remainder of this article is organized as follows. Section 2 presents the theoretical framework. Section 3 outlines context of the study. Section 4 outlines the data types and sampling method. Section 5 presents the descriptive statistics of the data. Section 6 details the methods used in this study and the approach of regressions used. Section 7 reports the regression results. Section 8 provides a summary, some concluding remarks, and implications of this research.

SECTION 2: THEORETICAL FRAMEWORK

The theoretical foundation of this work rests on the prior works of microeconomic theory on consumer choice and preferences (Mas-Colell, Whinston, Green 1995 & Nicholson and Snyder 2008). Household main shoppers (those who make most of the food shopping decisions for the household) are assumed to follow the axioms of rational choice: completeness, transitivity, and continuity. The ceteris paribus assumption is also followed: main shoppers are only choosing among food categories at one point in time, with all other factors (such as psychological attitudes, peer group pressure, personal experiences, cultural environment, etc.) held constant.

The utility function of the household level aggregate (the main shopper of the household) is comprised of each food category type x_1, x_2, etc. and vector z, which is comprised of all other factors that influence utility (held constant in this instance). For simplicity, goods $x_1, x_2, \ldots x_n$. are a mix (bundle) of both healthy and unhealthy foods that each contribute towards raising utility levels through consumption. Each combination of these goods differs in terms of cost, nutrition, and utility (satisfaction). This utility function is represented as equation 1 below:

$$U_{HH} = f(x_1, x_2, x_3, \ldots, z) \tag{1}$$

Utility in this instance is the level of satisfaction gained from buying and using goods and services (Nicholson and Snyder, 2008). Essentially, these functions represent how households and individual main shoppers rank certain bundles of goods over others. We assume that more of any food category is preferred to less, but with some minimum threshold due to nutrition requirements and a budget constraint. Thus, prices of each food are a key input into this utility equation. Time is also a constraining resource that must be accounted for, since there is a cost of time in consumption decisions (Becker, 1965). The consumer choice (household) problem in this instance is that main shoppers are seeking to find the highest utility for themselves and their household, subject to a budget constraint, nutritional needs, and time constraints. Distance to a food outlet is directly related to cost of time: outlets further away from household locations cost more time to shop at and purchase food from.

The simultaneous dual problem alongside utility maximization is expenditure minimization. The goal is reframed to be allocating income (to satisfy cost of food requirements), nutritional requirements, and time to achieve a minimum utility level, or satisfaction level, while minimizing expenditure. For the household main shopper, their expenditure is a function of food category prices, utility (or satisfaction) of the main shopper and general household utility. To capture the opportunity cost of time, outlet distance is added to the expenditure function. This is based on the idea that how households choose to expend resources depends in part on the spatial distribution of food outlets of varying degrees of healthiness and the cost of time it takes to visit these destinations. Since households derive utility from both healthy and unhealthy foods, as defined by Fung et al. (2018) and Bromage et al. (2021), both healthy and unhealthy food prices are incorporated.

Equation 2 below represents this simplified dual expenditure minimization problem linked with utility maximization (Nicholson and Snyder 2008). Total expenditures per capita *Expendpercap* of i healthy and unhealthy foods are represented in equation 2 where p_i is the ith good price and x_i is the ith good. The goal is to minimize expenditure per capita subject to a minimum level of utility U_{HH} that is above some threshold y:

$$\text{minimize } Expendpercap = f(p_1 * x_1 + p_2 * x_2 + \cdots + p_i * x_i) \quad (2)$$
$$s.t. U_{HH} = f(x_1, x_2, x_3, \dots, z) > y$$

Starting here, we will use the above framework and add outlet distances as an additional input into the expenditure function. The function, where px_i represents the ith unhealthy or healthy food price, U_{HH} represents the desired utility level of the household is represented as equation 3. The left-hand-side variable is expenditure per capita. Vector v represents all other factors that have influence on expenditure (such as nutrition levels, etc.):

$$TotalFoodExpenditure = f(px_1, px_2, \dots, px_i, outletdistance_j, v) \quad (3)$$

For each of the above equations, the key variable of interest is, ceteris paribus, the effect of $outletdistance_j$, or the distance to the jth outlet type, on the dependent variable listed. In general, we assume that the above expenditure function, in accordance with microeconomic theory, follows the properties of homogeneity, non-decreasing in prices, and concavity in prices (Nicholson and Snyder, 2008).

Healthy food expenditure is defined as the summation of the value of all healthy food categories purchased over a month period. Unhealthy food expenditure is defined in similar manner. The ratio of total healthy to unhealthy food expenditure is simply a ratio of healthy over unhealthy food expenditure. The goal of this paper is to investigate the effect of outlet distances on the margin on healthy food expenditure. Further, this paper seeks to assess the effect of outlet distances on households' expenditure mix between healthy and unhealthy foods.

SECTION 3: SAMPLING METHOD AND DATA

3.1: Sampling method

As mentioned above, the study area geographically encompasses 2 cities in Kenya: urban/peri-urban Nairobi and urban/peri-urban Kisumu. Administratively, Kenya is divided into counties, sub-counties, and locations. For statistical sampling purpose, the Kenyan National Bureau of Statistics (KNBS) further divides locations into enumeration areas (EAs) which are designated as either urban or rural. As per this designation, urban and peri-urban Nairobi consists of EAs classified by KNBS as 100% urban. Kisumu consists of a mix of urban and rural enumeration areas, with 93% of EAs in urban Kisumu designated as urban and 17% of EAs in peri-urban Kisumu designated as urban. Figure 1 displays the region and specific cities of Kenya surveyed, while figures 2 and 3 display the location of EAs selected.

A multi-stage sampling design was followed to conduct the household and food environment surveys. For each of these four areas (i.e. urban and peri-urban Nairobi and Kisumu), 2019 census data was used to construct a super index to characterize each administrative location based on a household level asset index, dwelling score, communication index, education index, and employment index. Based on the results across the population, locations were grouped according to this super-index into quartiles. For urban Nairobi, the top quartile was discarded since it was not comparable to other areas and due to a low expected survey response rate. The remaining three quartiles in urban Nairobi were then divided into four new quartiles. These quartiles characterize locations in each study area into four groups: Quartile 1=low income, Quartile 2=low/mid income, Quartile 3=mid income, and Quartile 4=mid/high income.

A two-stage cluster random sampling method was used to identify enumeration areas for household and food environment surveys. In the first stage, the goal was to select 8 locations per study area, and then in the second stage select 2 EAs per location. In urban Kisumu there were a total of 8 locations and in peri-urban Kisumu there were 7 locations. All were therefore included in this study in the first stage. In urban and peri-urban Nairobi, there were many more locations. Thus, 2 locations per quartile were randomly selected, resulting in a total of 8 locations each in urban and peri-urban Nairobi.

From these 31 locations (i.e., 8 each in urban Nairobi, peri-urban Nairobi, and urban Kisumu, and 7 in peri-urban Kisumu), 2 enumeration areas (EAs) were randomly selected by the staff at KNBS. This process resulted in the selection of 16 EAs in urban Nairobi, 16 EAs in peri-urban Nairobi, 16 EAs in urban Kisumu, 14 EAs in peri-urban Kisumu. After the specific EAs were selected, a full listing of all the households was done. In the final stage, we randomly selected 23-28 households per EA to reach a target of about 375 households per study area. However, due to field implementation challenges and survey non-responses, surveys were successfully conducted with he18-24 households per EA. In each selected household, interviews were conducted with the household main shopper and few randomly selected adult household members.

Figure 1: Map of regions of Kenya surveyed

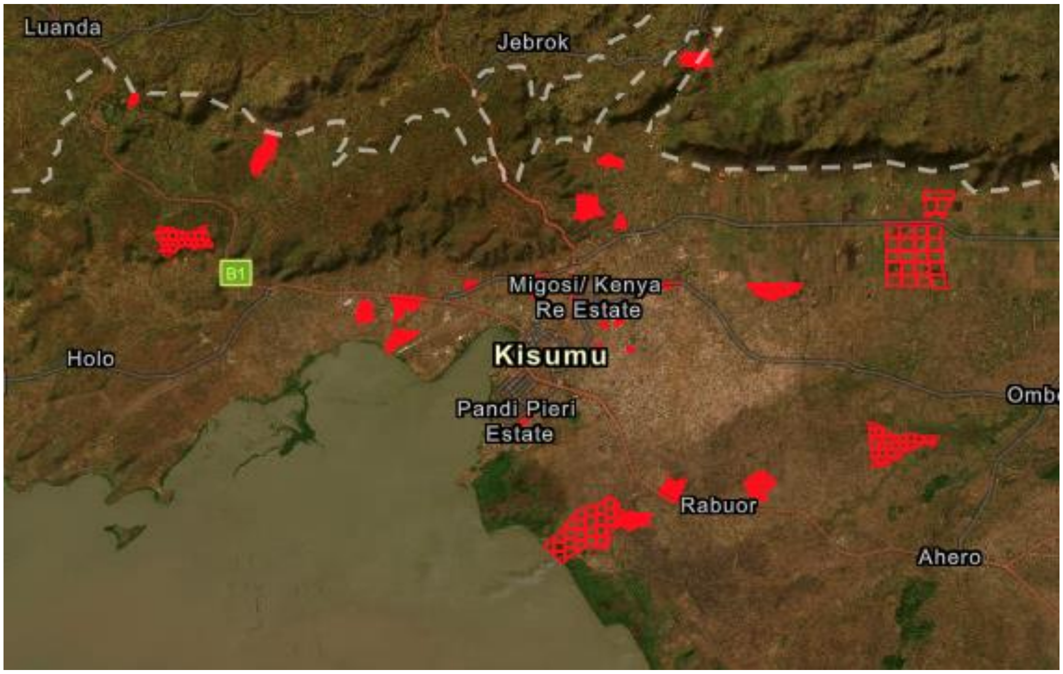

Figure 2: Map of enumeration areas in Kisumu

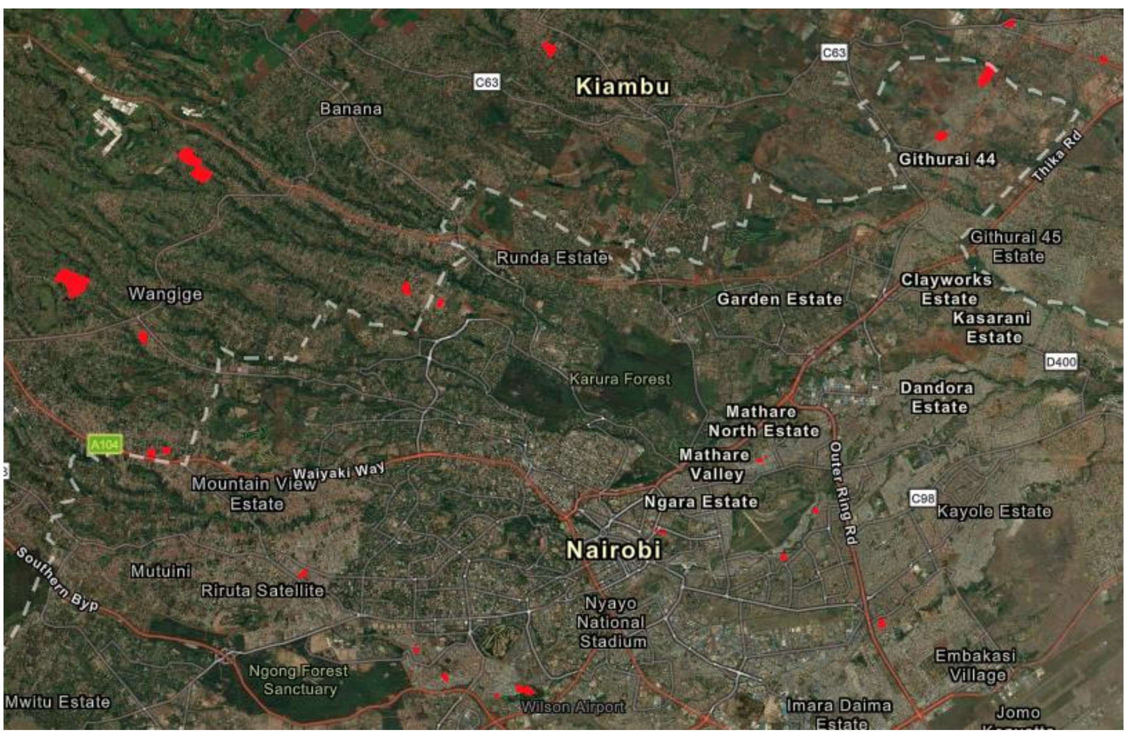

Figure 3: Map of selected enumeration areas in Nairobi (shown in red)

3.2: Data

Data for this paper come from two sources—the household survey and the food environment census survey. These surveys were conducted across 61 EAs in urban and peri-urban Nairobi and Kisumu in April-May 2022 with the purpose of studying the local food environments.

First, main shoppers, or those that conduct most of the shopping for the household, were interviewed using a questionnaire consisting of 2 modules that focus on demographic, economic and geographic characteristics of the main shopper and individual members of the household. Key elements of the household (main shopper) survey include a detailed household roster, a food shopping module, and food values. The food shopping module captured the food items purchased by the household yesterday, the previous 6 days, and over the last month. For each food item, data was collected on the outlet from where it was purchased and its GPS location. An additional feature of the survey is the inclusion of questions that help construct a poverty likelihood score (PLS) for each survey respondent (Schreiner, 2018). Results from this metric are used to stratify the sample into two groups – below and above the median poverty score.

Second, a food environment census of various food outlet types operating their business in the home food environment of households was conducted. The home food environment is defined as a radius around the mean center, or centroid, of each enumeration area (EA). For Nairobi urban and peri-urban, as well as Kisumu urban, the radius is defined to be 0.4 kilometers. But, for Kisumu peri-urban, the radius is defined to be 0.6 kilometers. Each outlet was asked about what food groups (as defined by Bromage, 2021) are sold. For a look at the different food groups, see Table 1 below. GPS location coordinates were also collected for each outlet, along with food prices. Both unprepared and prepared foods are captured in the survey along with sales information.

When considering the literature on healthiness of diets and food outlets, it is essential to define 'healthy' food. This definition is based on two previous works by Fung et al., 2018, and then Bromage et al., 2021. Foods are classified as healthy according to historical literature measuring associations between consumption of foods and risk of non-communicable diseases and nutrient contribution of each food group (Fung et al., 2018). Food groups are classified as healthy and unhealthy as shown below in Table 1. According to Bromage et al. (2021), there are two types of unhealthy food categories: unhealthy at any consumption level and unhealthy at an excessive consumption level. For the sake of simplicity, this analysis combines 'unhealthy in excess'-labeled food groups with 'healthy'-labeled food groups. The reason for this being that in a developing country like Kenya, both high fat dairy and red meat consumption will be limited due to budget constraint and mostly in the healthy quantity range. Table 1 describes each food type category that are used in the analysis, along with their healthy/unhealthy classification assigned by Bromage et al. (2021). All in all, there are 16 'healthy' food groups, 2 'unhealthy in excessive amounts' food groups, and 7 'unhealthy' food groups. This makes a total of 25 food groups in all. Also noted in the table are descriptions for each food group.

Table 1: Food group information

Food group type	Healthiness classification	Description
Citrus fruits	Healthy	Whole fruits in the genus *Citrus*
Deep orange fruits	Healthy	Whole fruits (not including juice or spreads) containing 20 or less retinol equivalents/100 g
Other fruits	Healthy	Whole fruits not belonging in other fruit categories (not including coconuts)
Dark green leafy vegetables	Healthy	Leafy vegetables containing 120 retinol equivalents/100 g.
Cruciferous vegetables	Healthy	Vegetables in the family *Brassicaceae*
Deep orange vegetables	Healthy	Nontuberous vegetables containing less than 120 retinol equivalents/100 g.
Other vegetables	Healthy	Vegetables not belonging in the other vegetable categories
Legumes	Healthy	Legumes and foods derived from legumes, such as tofu and soymilk. Does not include bean sprouts (classified as "other vegetables") or groundnuts (classified as "nuts and seeds")
Deep orange tubers	Healthy	Tuberous vegetables containing more than 120 retinol equivalents/100 g. (includes variants biofortified with vitamin A)
Nuts and seeds	Healthy	Nuts, seeds, and products derived from nuts and seeds, such as nut-based butters (but not oils). Also includes groundnuts. Seeds that are used as spices are included when used in their whole (not powdered) form
Whole grains	Healthy	Whole grains and whole-grain products. Does not include products with significant amounts of added sugar (classified as "sweets and ice cream")
Liquid oils	Healthy	All types of oils that are liquid at room temperature, regardless of fatty acid profile (this includes palm olein, liquid palm kernel oil, and liquid coconut oil). Does not include oil used to deep fry foods that are purchased, but does include oil used to deep-fry foods prepared at home

Table 1 (cont'd)

Food group type	Healthiness classification	Description
Fish and shellfish	Healthy	Fish (whether processed or unprocessed) based on phylogenetic classifications (including sharks, eels, and rays), and other seafood high in n3 fatty acids (including shellfish, jellyfish, cetaceans, and pinnipeds, but not echinoderms). Includes organs
Poultry and game meat	Healthy	Unprocessed poultry and game, including a range of undomesticated animals and bush meat, e.g., primates, rodents, canines, felines, marsupials, leporids (rabbits and hares), wild boar, bats, bears, semiaquatic mammals (including otters and beavers), undomesticated ungulates, reptiles (aquatic and terrestrial), and amphibians. Includes organs
Low fat dairy	Healthy	Reduced or naturally low-fat dairy products (2 or less percent milk fat). Includes flavored milk, and milk or cream added to coffee or tea
Eggs	Healthy	All types of eggs. Does not include mayonnaise
High fat dairy (in milk equivalents)	Unhealthy in excessive amounts	High fat milk and dairy products (more than 2% milk fat). Includes flavored milk, and milk or cream added to coffee or tea. Does not include butter or clarified butter. This category also does not include ice cream and whipped cream
Red meat	Unhealthy in excessive amounts	Unprocessed red meat belonging to domesticated animals (i.e., not game), including organs. "Red" classification is not based on color but on nutritional characteristics, and thus includes pork and lamb
Processed meat	Unhealthy	Processed red meat, poultry, or game, including organs, and excluding fish and seafood. Processing is defined per International Agency for Research on Cancer: "salting, curing, fermentation, smoking, or other processes to enhance flavor or improve preservation."
Refined grains and baked goods	Unhealthy	Refined grains and refined grain products. Does not include products with significant amounts of added sugar, which should instead be classified as "sweets and ice cream"
Sweets and ice cream	Unhealthy	Sugar-sweetened foods that are not beverages; includes sugar and other caloric sweeteners added to other foods and drinks. Whipped cream also classified in this category

Table 1 (cont'd)

Food group type	Healthiness classification	Description
SSBs (sugar sweetened beverages)	Unhealthy	Sweetened drinks that do not contain any fruit juice at all. Includes, e.g., sodas, energy drinks, sports drinks, and beverages made using low-calorie sweeteners, such as diet sodas. Sweetened tea and coffee, and dairy or cereal-based drinks are not included
Juice	Unhealthy	Unsweetened or sweetened drinks that are at least partly composed of fruit juice. This category also includes fruit smoothies made from whole fruit
White roots and tubers	Unhealthy	Tuberous vegetables with less than 120 retinol equivalents/100 g. Includes flours such as potato or cassava flour
Purchased deep fried foods	Unhealthy	Deep fried foods fried in an amount of fat or oil sufficient to cover the food completely. Only deep-fried foods that are purchased (i.e., not prepared at home) are classified in this group. Foods in this category are "double classified" and should be classified as belonging to the purchased deep fried group as well as the food group to which the food normally belongs if not purchased and deep fried

Note: adapted from Bromage et al. (2021)

In Table 2 below, the different types of outlets specified in this study are presented, 14 in all, along with descriptions for each. There are several dimensions to heterogeneity in outlet types in terms of healthiness of food offering, geographical location, as well as modernity. The number of each type of outlet in the food environment census is also presented. Overall, there are 6,533 outlets surveyed.

Table 2: Food outlet type information and number of outlets enumerated across the four study areas

Outlet Type	Description	# of outlets in survey
Small supermarket	Any self-service food outlet with 1-4 cash registers	30
Large supermarket	Any self-service food outlet with more than 4 cash registers	146
Duka (e.g. small traditional shop)	Traditional (not self-service) food outlets with permanent, constructed quarters	1712
Kiosk	Typically small, free-standing, "semi-movable" with rudimentary or transient structure such as shipping containers located along thoroughfares	463
Mama mboga	Vegetable seller/vendor (similar to street vendor, but specialized in fruits and veg)	1288

Table 2 (cont'd)

Outlet Type	Description		# of outlets in survey
Street vendor	A seller located outside a market on the streets and selling from a mobile structure or from ground		926
Hawker	Seller that sells items on foot (walking from place to place) may sometimes have a cart		89
Depot/wholesale	An outlet that primarily sells goods in bulk to either retailers or consumers directly		59
Milk bar/milk atm	Outlets that primarily sell dairy products either in a restaurant bar or simple mechanized nozzle		72
Hotel/restaurant	An outlet selling prepared food for consumption on the premises, featuring permanent construction		469
Informal prepared food	Same as street vendor or kiosk—but specialized in selling prepared foods ready to be eaten, rudimentary infrastructure		709
Cereal shops and posho mills	Specialize in selling dry grains, mainly cereals and pulses. some of them also value add and sell flours from these grains.		249
Bakery	An outlet that primarily sells baked goods		6
Butchery	An outlet that primarily sells red or white meat		315
		Total:	6533

Note: from Food environment census survey

3.3: Outlet healthiness characteristics

Using the food environment census data, outlet healthiness characteristics are calculated in terms of the average number of healthy and unhealthy food groups on offer per outlet type. Definitions of food group healthiness are applied from Bromage et al. (2021). We see that on average, small supermarkets, dukas, and mama mboga outlets have the greatest number of total food groups on offer. When it comes to healthy food groups sold, small supermarkets, mama mbogas, and dukas on average have the highest number of healthy food groups on offer. On the other hand, small supermarkets, dukas, and kiosks have the unhealthiest food groups on average on offer. In the final far-right column, we see that mama mbogas have the highest average of the ratio of total number of healthy food groups to total number of unhealthy food groups.

The main hypotheses of this paper are constructed using the healthy and unhealthy food group characterization highlighted in Table 3. We find that the food outlets that have the highest average total absolute number of healthy food groups, and therefore selling more healthy foods are mama mbogas, small supermarkets, dukas, kiosks, cereal shops/poshomills, and depots/wholesalers. A similar process is used with the average total absolute number of unhealthy food groups to determine outlets that tend to sell more unhealthy food items. Admittingly, we do find some 'unhealthy'-labeled outlets, such as street vendors and hawkers, to have a higher average healthy-to-unhealthy food group ratio than some 'healthy'-labeled outlets, such as supermarkets and dukas. While these relative statistics do provide a different perspective, the absolute counts of food groups per outlet type better describe the actual degree of healthiness of food types on offer at various outlets. This absolute counts of of food group will be used in this paper to characterize outlets as relatively healthy and unhealhty.

Overall, we find that there is heterogeneity in food offerings across outlet types, with healthy food group averages ranging from 0.32-5.80, while unhealthy food group averages range from 0.10-4.10. Small supermarkets, as expected, have both the healthiest and unhealthy food groups on offer on average across all outlet types.

Table 3: Outlet healthiness characteristics

Outlet type	average number of all total food groups (25 total)	Average total number of healthy food groups (16 total)	Average total number of unhealthy in excess food groups (2 total):	Average total number of unhealthy food groups (7 total):	Average of ratio of total number of healthy food groups to total number of unhealthy food groups
Small supermarket	8.9	4.47	0.33	4.1	1.09
Duka (e.g. small traditional shop)	7.33	3.29	0.41	3.63	0.91
Kiosk	5.41	3.17	0.16	2.11	1.50
Mama mboga	6.3	5.8	0.009	0.49	11.84
Street vendor	2.1	1.56	0.02	0.53	2.94
Hawker	1.34	0.82	0	0.34	2.41
Depot/wholesale	4.36	2.03	0.25	2.07	0.98
Milk bar/milk atm	1.92	0.51	0.97	0.43	1.19
Hotel/restaurant	1.12	0.32	0.06	0.75	0.43

Table 3 (cont'd)

Outlet type	average number of all total food groups (25 total)	Average total number of healthy food groups (16 total)	Average total number of unhealthy in excess food groups (2 total):	Average total number of unhealthy food groups (7 total):	Average of ratio of total number of healthy food groups to total number of unhealthy food groups
Informal prepared food vendor	0.84	0.33	0.02	0.49	0.67
Cereal shop/poshomill	2.99	2.32	0.02	0.65	3.57
Butchery	1.32	0.39	0.84	0.1	3.90
Bakery	2.67	0.67	0.17	1.83	0.37

Scale: green (most) to red (least)
Source: food environment survey (2022)
Note: data unavailable for large supermarkets

3.4: GIS data

In our household (HH) survey, we collected GPS coordinates of household location and food outlets where households shopped. The household location data was converted from .csv files to a point vector layer in ArcGIS Online. In addition to household location, we also collected GPS coordinates of 3 different types of food outlets: outlets where food was purchased within 24 hours, 6 days before that, and bulk purchases in the last one month. Figures 4-5 below are maps of plots of households for both Kisumu, Kenya and Nairobi, Kenya, as well as maps of households and outlets. Data is from the food environment census:

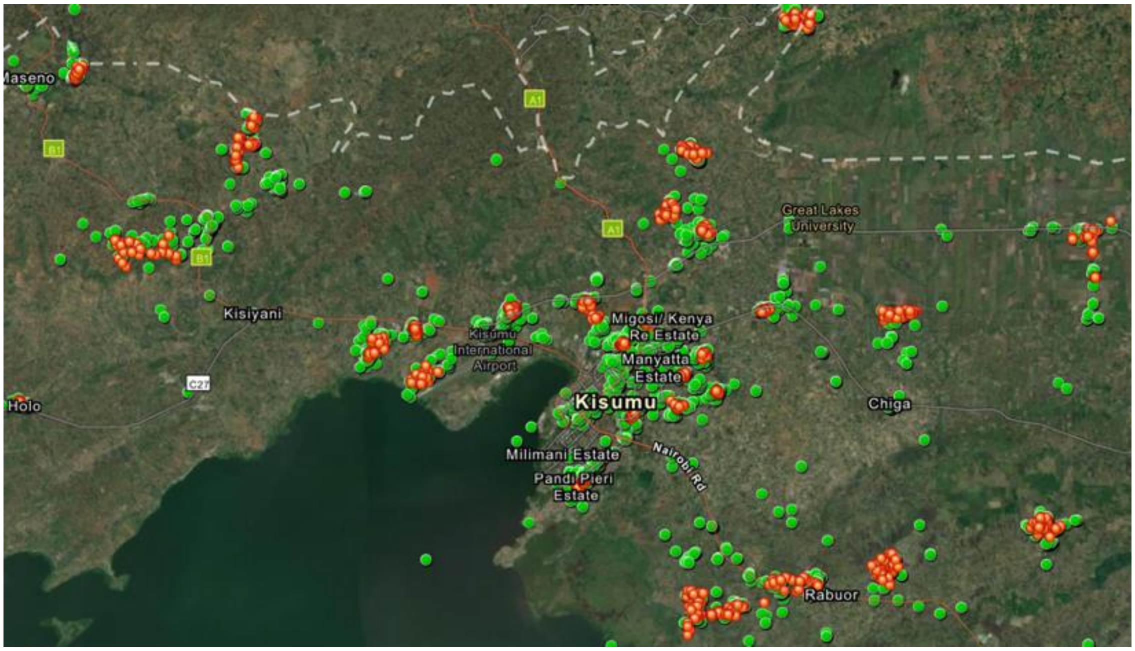

Figure 4: Map of Kisumu, Kenya, households and all outlets in their 'home' food environment

Note: using home food environment census survey

<u>Figure 4 legend:</u>
Orange dots: households
Green dots: outlets

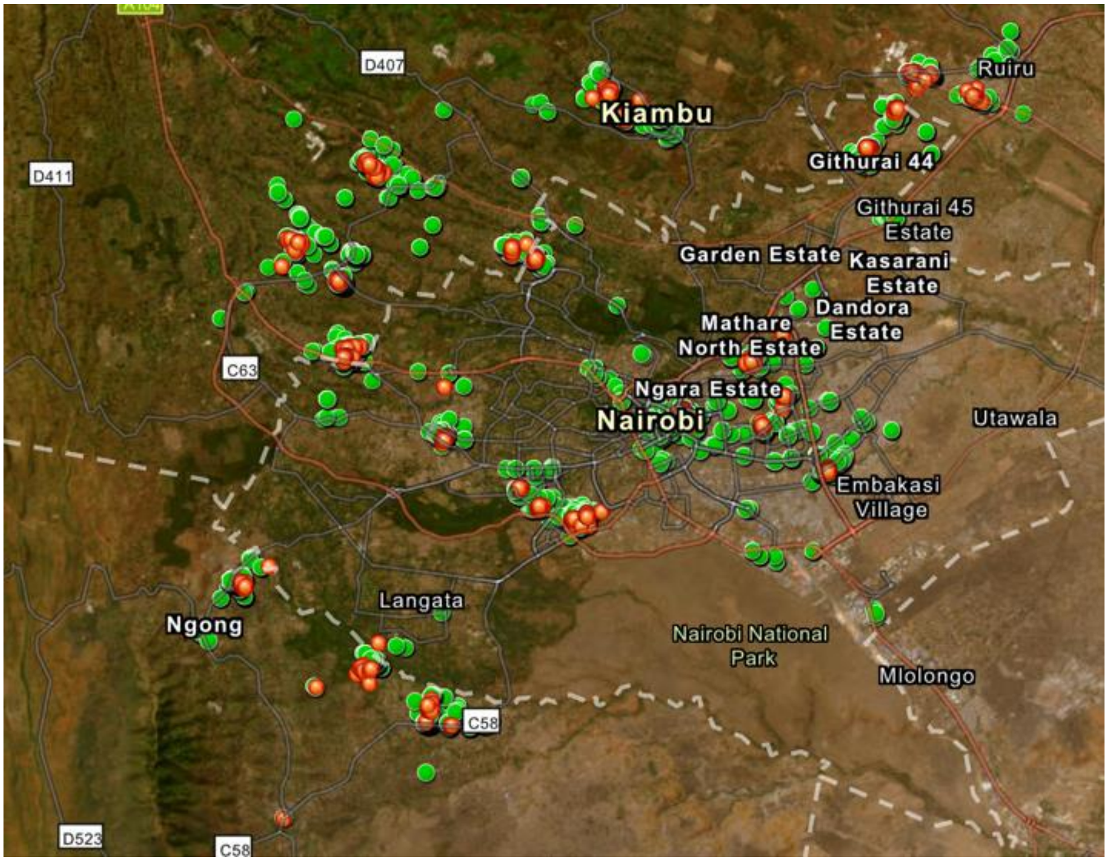

Figure 5: Map of Nairobi, Kenya, households and all outlets in their 'home' food environment

Note: from home food environment census survey

Figure 5 legend:
Orange dots: households
Green dots: outlets

The above GIS data in figures 4-5 comes from the food environment census survey, an exogenous measure of actual locations of outlets. But main shoppers also had an opportunity to report outlet locations where they shop (endogenous). In figures 6-7 below are maps of these outlet locations. We find that both reported and actual outlet locations are very similar- we expect there to be some overlap but not entire overlap due to the surveying:

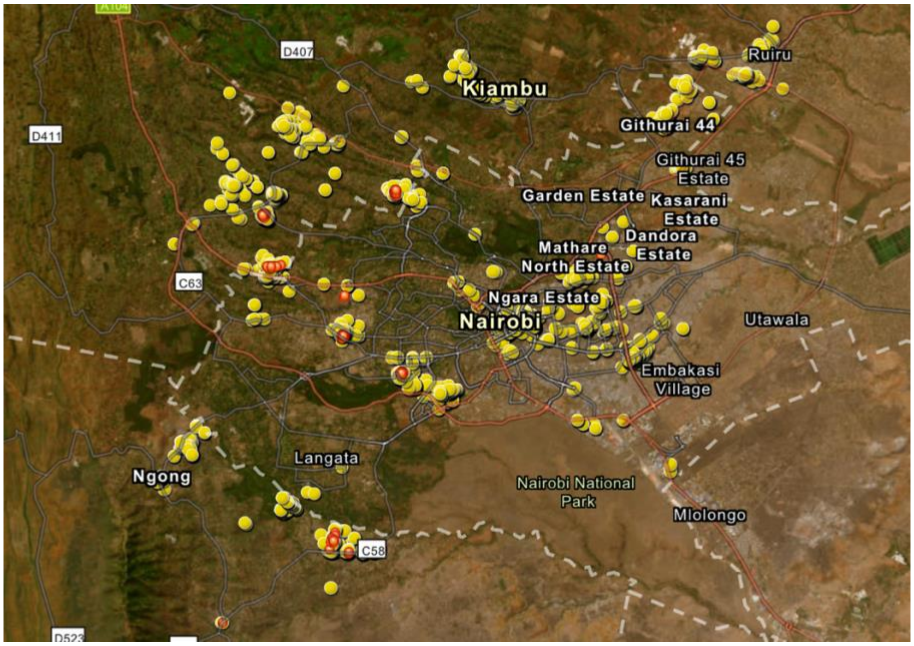

Figure 6: Map of Nairobi, Kenya, outlet and household locations

Note: from main shopper survey

<u>Figure 6 legend:</u>
Orange dots: households
Yellow dots: outlets

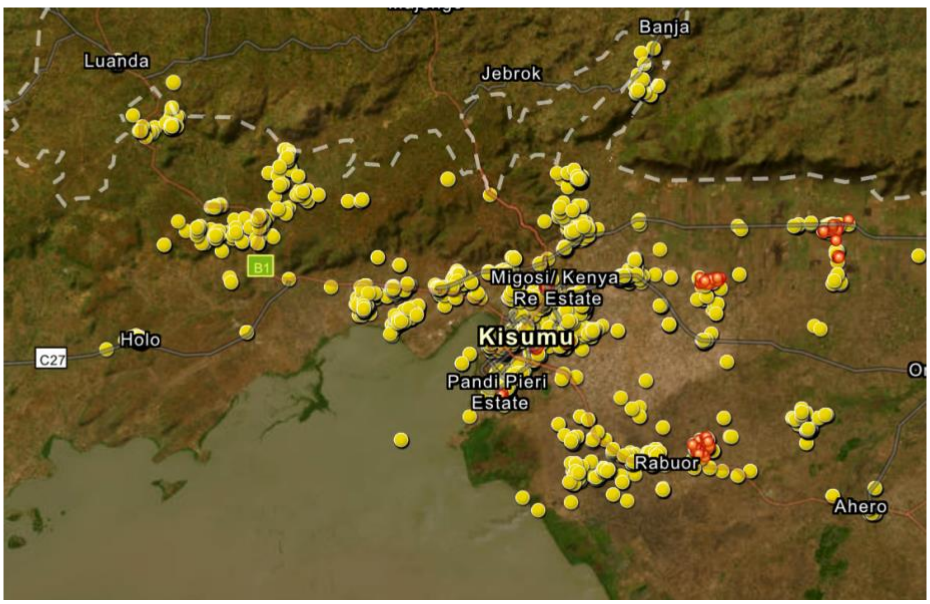

Figure 7: Map of Kisumu, Kenya, outlet and household locations

Note: using main shopper survey

<u>Figure 7 legend:</u>
Orange dots: households
Yellow dots: outlets

After the data was initially plotted, GIS tools were used to analyze distances between household points and outlet locations of the home food environment data. Starting with household main shopper survey data on household location, as well as home food environment census data on outlet location, maps were created plotting HHs and each type of outlet as separate layers. Then, using ArcGISOnline, the "Proximity-Find Nearest" tool was used to find the straight-line distance in kilometers from each household to the nearest of each type of outlet. These distances, at the household level, are then merged with the main shopper survey data at the household level. Any missing distance values per observation are replaced with the sample mean of outlet distance (11 occurrences). Below in figure 8 is a sample map of the "Proximity-Find Nearest" tool for a small group of households and duka outlets in Nairobi:

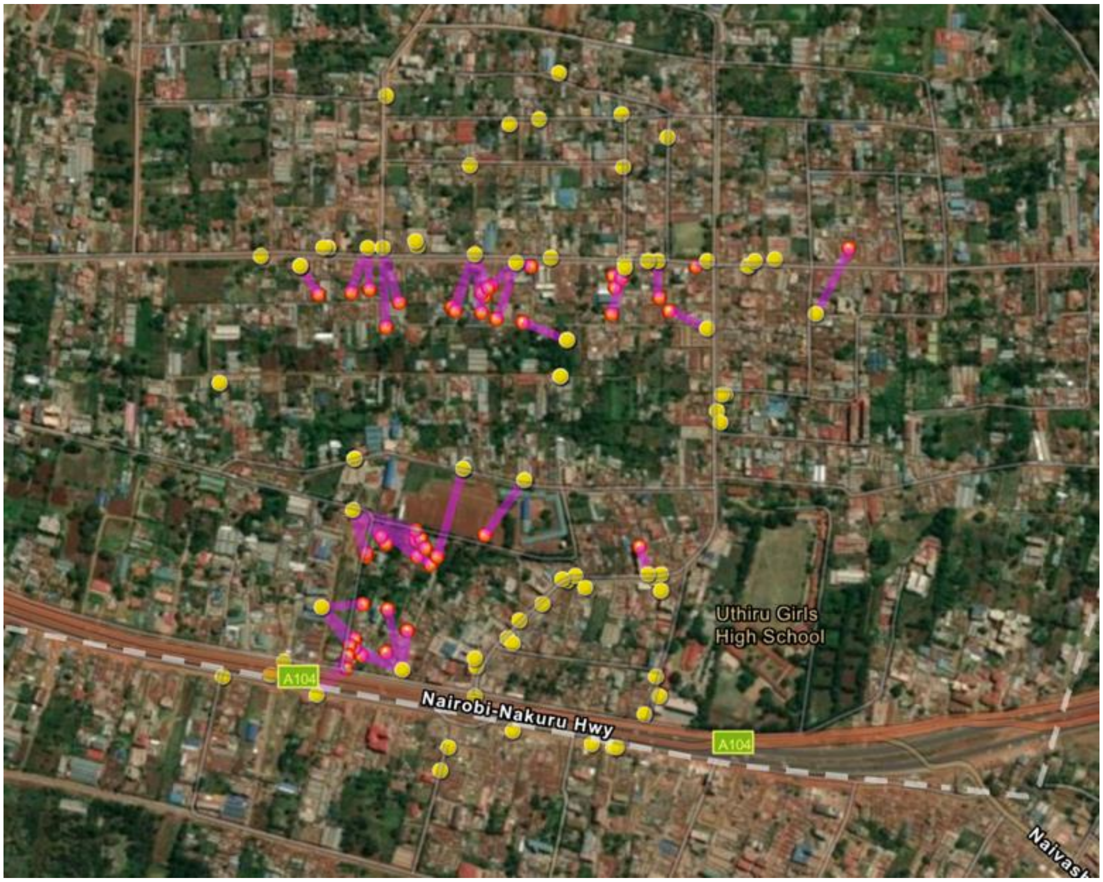

Figure 8: Map of household distance to nearest Duka near Kinoo area, Nairobi

<u>Figure 8 legend</u>
Orange dots: households
Yellow dots: outlets (dukas in this case)
Pink lines: line tracing the nearest distance in km for each household to the nearest outlet (duka)

3.5: Survey weights

Population weights are used in all analyses, such that the sample is representative of the population of urban and peri-urban Nairobi and Kisumu. These weights account for the likelihood of selection of locations in the first stage of sampling; the likelihood of selection of enumeration areas in the second stage of sampling; and the likelihood of selection of households in the third stage of sampling.

The population of Nairobi is about 10 times larger than the population of Kisumu. For this reason, descriptive statistics of the pooled sample are likely to resemble those of Nairobi, the larger city. However, we report statistics disaggregated by city and urban status to capture the differences in these subpopulations.

We also ensure to cluster standard errors at the EA level once we arrive at the regression step, since multistage sampling techniques were used.

3.6: *Extreme values of expenditure data*

It was detected that extreme values exist for the expenditure main shopper data. Any expenditure values lying beyond the 99^{th} percentile of the distribution are dropped from the analysis in an effort to clean extreme values.

SECTION 4: DESCRIPTIVE STATISTICS OF THE DATA

Tables 4 and 5 present household and main shopper characteristics from the survey. The average household size across the sample is 3, with variation from 3-4 on average across regions. The region with the highest average income is Nairobi urban, with 14,131 KES (Kenyan Shillings) while Kisumu peri-urban has the lowest reported average income 8,434 KES. The exchange rate between US Dollars and Kenyan Shillings, as of June 1, 2023, is 1 KES to 0.0073 US Dollars. Across all regions, owing a bicycling is much more common for households than owning a car or truck. Across the entire sample, about 32% of main shoppers are male while 68% are female, while the average age is 36 years. Main shoppers have on average 13 years of education, with the majority holding self-employed or wage labor occupations on average. As seen in Table 4, price is by far the most valued food value across the sample, with almost 80% of main shoppers reporting. Nutrition is also valued highly by 46% of households.

Table 4: Household and main shopper characteristics

Variable	Overall (mean and SD)	Nairobi u. (mean and SD)	Nairobi p.u. (mean and SD)	Kisumu u. (mean and SD)	Kisumu p.u. (mean and SD)
Household size (mean)	3 (1.85)	3 (1.87)	3 (1.66)	4 (3.31)	4 (2.42)
Household income per month (KES)	12,524 (14332.17)	14,131 (15773.39)	10,806 (11975.41)	10,021 (12374.1)	8,434 (12182.05)
Poverty score	61 (12.41)	61 (12.74)	61 (11.64)	58 (13.21)	57 (12.76)
Number of children (age < 18) in household	1 (1.43)	1 (1.47)	1 (1.27)	2 (1.68)	2 (1.81)
Household owns a bicycle	15%	17%	12%	13%	17%
Household owns a car	7%	9%	6%	3%	0.5%
Household owns a truck	2%	3%	1%	0.2%	0.2%
Household has 2 main shoppers	38%	35%	43%	38%	48%
Main shopper gender female	68%	66%	71%	75%	70%
Main shopper age (years)	36 (12.03)	35 (10.56)	37 (13.60)	34 (12.58)	36 (15.23)
Main shopper education (years)	13 (6.10)	14 (6.44)	13 (5.44)	12 (5.51)	14 (7.58)

Table 4 (cont'd)

Variable	Overall (mean and SD)	Nairobi u. (mean and SD)	Nairobi p.u. (mean and SD)	Kisumu u. (mean and SD)	Kisumu p.u. (mean and SD)
% of main shoppers who value food nutrition as most important	46%	41%	54%	43%	43%
% of main shoppers who value food taste as most important	33%	32%	35%	30%	33%
% of main shoppers who value food convenience as most important	26%	28%	22%	26%	29%
% of main shoppers who value food price as most important	80%	80%	78%	82%	89%
% of main shoppers who value food nutrition as most important	34%	34%	31%	39%	46%
% of main shoppers who value food perishability as most important	33%	32%	35%	31%	33%
Sample size n	1496	368	354	382	392

Note: data weighted with survey population weights
Note: Poverty score is from Schreiner (2018). Responses are collected on 10 indicators and then used to estimate consumption-based poverty rates. Values signify the likelihood that the individual experiences poverty. Scores range from 0-100, with lower scores indicating higher poverty likelihood.
Note: under occupation, 'other' includes too young, retired, unpaid worker/volunteer

Table 5: Main shopper occupation percentages

Occupation type	Overall (%)	Nairobi u. (%)	Nairobi p.u. (%)	Kisumu u. (%)	Kisumu p.u. (%)
Farming	4	1	7	2	17
Self-employed	29	31	25	35	20
Wage labor	27	26	29	22	19
Salaried work	19	20	19	12	6
Student	4	3	3	9	25
Unemployed	15	16	14	18	11
Other	3	3	3	2	1

Using the definitions of 'healthy', 'unhealthy in excessive amounts', and 'unhealthy' from Bromage et. al. (2021), aggregate average food expenditure amounts are calculated for healthy and unhealthy groups. The overall mean monthly food expenditure per capita is 2,867 KES while average healthy food expenditure per capita is 1225 KES and average unhealthy food expenditure per capita is 1215 KES. The ratio of healthy to unhealthy expenditure is also calculated and will serve as a key dependent variable in the model. See Appendix D for, per food group, how main shoppers chose to allocate their monthly food expenditure in KES among categories of various healthiness.

Table 6: Food expenditure descriptive statistics monthly per capita (KSH)

Variable	Overall (mean and SD)	Nairobi u. (mean and SD)	Nairobi p.u. (mean and SD)	Kisumu u. (mean and SD)	Kisumu p.u. (mean and SD)
Total	2,379 (1,624.76)	2,350 (1,541.16)	2,485 (1,758.97)	2,101 (1,571.57)	2,099 (1,417.62)
Healthy	1,322 (1,002.04)	1,329 (980.13)	1,333 (1,044.14)	1,167 (945.62)	1,307 (978.58)
Unhealthy	973 (697.23)	983 (710.46)	997 (679.53)	830 (700.85)	781 (626.52)
Sample size n	1425	352	335	362	376

Note: population weights applied
Note: for all variables, it is assumed 'unhealthy in excess' is grouped with the 'healthy' category
Note: Observations with extreme values greater than 99th percentile are dropped

In Table 7, main shopper outlet expenditure statistics are presented. The self-reported average distance (in kilometers) to each outlet varies from 0.24 km (milk bar/milk atm) to 9.55 km (depot). The top three closest outlets are: milk bar/milk atm (0.24 km), Kiosk (0.30 km), and mama mboga (0.30 km). The top three farthest outlets are: depot (9.55 km), wholesale (4.32 km), and hotel/restaurant (3.71 km). The average amount expended at each outlet per capita per month is also displayed. By far, duka outlets have the highest amount expended (2931.71 KES/mo.). In the final right-hand-side column, the top 5 items purchased at each outlet, in terms of % of households, is presented.

Table 7: Main shopper's (MS) self-reported outlet distances and expenditure statistics

Outlet type	MS self-reported distance (km) on average (mean and SD)	Amount spent at outlet per capita per month in KES (mean and SD)	Top 5 items purchased at outlet (% HHs) on average
Small supermarket	2.02 (5.82)	292 (1,538.48)	1. Sugar (white, granulated or lump) 2. Vegetable oil 3. Maize meal/flour, sifted 4. Rice (white, milled, polished grain) 5. Wheat flour (refined, fortified, sifted)
Large supermarket	5.2 (15.33)	941 (2,739.26)	1. Sugar (white, granulated or lump) 2. Maize meal/flour, sifted 3. Vegetable oil 4. Rice (white, milled, polished grain) 5. Wheat flour (refined, fortified, sifted)
Duka (e.g. small traditional shop)	0.50 (3.06)	2,932 (4,281.77)	1. Sugar (white, granulated, or lump) 2. Vegetable oil 3. Maize meal/flour (sifted) 4. Milk (cow, whole, fresh) 5. Bread (white)
Kiosk	0.30 (2.26)	259 (946.05)	1. Tomato (red, ripe) 2. Kale (Sukuma wiki) 3. Vegetable oil 4. Onion (mature, red skinned, peeled) 5. Sugar (white, granulated, or lump)
Mama mboga	0.30 (0.90)	1004 (1,862.63)	1. Tomato (red, ripe) 2. Kale (sukuma wiki) 3. Onion (mature, red skinned, peeled) 4. Cabbage (leaf head, white) 5. Onion (spring, raw)

Table 7 (cont'd)

Outlet type	MS self-reported distance (km) on average (mean and SD)	Amount spent at outlet per capita per month in KES (mean and SD)	Top 5 items purchased at outlet (% HHs) on average
Street vendor	1.49 (3.08)	98 (660.25)	1. Sardines 2. Basic Mandazi 3. Milk (cow, whole, fresh) 4. Tomato (red, ripe) 5. White Chapati
Hawker	0.58 (4.69)	15 (185.51)	1. Kunde (cowpeas, leaves, picked) 2. Mrenda (jute mallow, picked leaves) 3. Sardines 4. Watermelon 5. Managu (leafy green vegetable)
Depot	9.55 (28.76)	31 (309.32)	1. Rice (white, milled, polished grain) 2. Yellow beans dry raw 3. Vegetable oil 4. Gram (green, dry) 5. Milk (cow, whole, fresh)
Wholesale	4.32 (10.62)	199 (1,083.887)	1. Vegetable oil 2. Maize meal/flour (sifted) 3. Rice (white, milled, polished grain) 4. Sugar (white, granulated, or lump) 5. Wheat flour (refined/fortified/sifted)
Milk bar/milk atm	0.24 (0.42)	62 (285.30)	1. Milk (cow, whole, fresh) 2. Milk (cow, whole, fermented) 3. Milk (cow, condensed, skimmed, sweetened) 4. Milk (cow, powder, skimmed) 5. Milk (cow, powder, whole)
Hotel/restaurant	3.71 (11.88)	46 (545.66)	1. White Chapti 2. Ugali from refined maize flour 3. Red beans stew 4. Bean stew 5. Chai ya Maziwa (mixed tea), with sugar

Table 7 (cont'd)

Outlet type	MS self-reported distance (km) on average (mean and SD)	Amount spent at outlet per capita per month in KES (mean and SD)	Top 5 items purchased at outlet (% HHs) on average
Informal prepared food	2.32 (23.04)	187 (557.75)	1. Basic Mandazi 2. White Chapati 3. Boiled beans 4. Githeri (fresh maize and dry beans) 5. Githeri (dry maize and dry beans)
Posho mill or cereal shop	0.78 (2.12)	108 (477.0)	1. Whole maize meal flour 2. Maize grain (white variety, whole, dry) 3. Millet (finder, flour) 4. Cornflour (from maize starch) 5. Rice flour
Butchery	2.45 (27.91)	376 (1,072.18)	1. Beef (with bones) 2. Beef (medium fat, without bones) 3. Matumbo (tripes) 4. Beef (high fat, without bones) 5. Beef (lean)

Note: population weights applied
Note: data source is main shopper census survey

In the final descriptive table below, average distances to each outlet type are displayed as collected in the food environment census. We see among all households and all outlets, the average distance to any nearest food outlet is 0.08 km. Some variation in distance is noted between urban and peri-urban households.

Table 8: Linear distance (km) and counts within the home food environment

Outlet type	Count in home FE	Entire sample (km)	Nairobi u. (km)	Nairobi p.u. (km)	Kisumu u. (km)	Kisumu p.u. (km)
All outlets	6533	0.08 (0.25)	0.08 (0.19)	0.08 (0.33)	0.07 (0.09)	0.15 (0.15)
Informal prepared food outlet	709	0.24 (0.57)	0.13 (0.21)	0.36 (0.70	0.15 (0.12)	0.84 (1.68)

Table 8 (cont'd)

Outlet type	Count in home FE	Entire sample (km)	Nairobi u. (km)	Nairobi p.u. (km)	Kisumu u. (km)	Kisumu p.u. (km)
Milk bar/milk atm	72	1.47	0.43	2.42	1.01	8.84
Depot/wholesale	59	1.87 (2.76)	0.43 (0.58)	3.54 (3.19)	2.33 (2.01)	6.63 (4.05)
Hawker	89	1.10 (1.64)	0.91 (1.19)	1.10 (1.93)	2.25 (1.49)	2.54 (2.90)
Street vendor	926	0.40 (1.08)	0.24 (0.42)	0.62 (1.59)	0.25 (0.35)	0.81 (1.82)
Mama mboga	1288	0.18 (0.51)	0.11 (0.21)	0.23 (0.56)	0.12 (0.12)	0.93 (1.68)
Kiosk	463	0.26 (0.67)	0.13 (0.21)	0.34 (0.70)	0.35 (0.44)	1.41 (2.28)
Duka	1712	0.12 (0.34)	0.09 (0.20)	0.12 (0.33)	0.12 (0.11)	0.50 (1.16)
Small supermarket	30	1.90 (2.75)	0.86 (0.81)	2.71 (3.38)	2.61 (1.58)	9.00 (2.94)
Large supermarket[a]	146	1.51 (1.27)	1.16 (0.68)	2.18 (1.75)	--	--
Hotel/restaurant	469	0.62 (1.57)	0.24 (0.40)	1.17 (2.33)	0.42 (0.46)	1.33 (2.33)
Poshomill or cereal shop	249	1.87 (2.61)	0.50 (0.66)	2.99 (3.60)	0.84 (0.82)	3.14 (2.78)
Butchery	315	0.43 (1.03)	0.26 (0.41)	0.37 (0.70)	0.58 (0.50)	3.69 (3.34)

a: large supermarket statistics for Kisumu urban and peri-urban are currently under development.
Note: values in parentheses are standard deviation

SECTION 5: METHOD AND APPROACH

5.1: Regression analysis

To test hypotheses with regards to how well outlet distances explain the variation in healthy/unhealthy food expenditure, multivariate regression approach is used. Accordingly, ordinary least squared (OLS) estimates are calculated. Since we are interested in responsiveness of main shopper healthy expenditure to changes in outlet distance, we use a log-log functional form. After accounting for missing data, our analytical sample is 1495 households, although some variation is noted across regressions.

Our key dependent variable of interest is the log of household main shopper's healthy monthly expenditure, while the key independent variables of interest are the log of each type of outlet distance as calculated using the food environment survey (not self-reported distance). In order to avoid undefined values with taking the log of zero expenditure, 1 is added to each observation to boost the sample size after taking the log. We also include controls for individual main shopper characteristics, household-level characteristics, and location dummies as well. In our model, these are represented as vectors of specific controls:

Table 9: Details of vectors for model

Vector Name	Controls included
$HH\_characteristics_i$	HHsize, poverty score, # of children, if own bicycle, if own car, if own truck
$MS\_characteristics_i$	gender, age, education, occupation, food values (nutrition, taste, convenience, price, availability, perishability)
$Location\_controls_i$	Region, EA (enumeration area), census wealth index location strata
$Log\_outlet\_distance_{ij}$	Log of linear distance from main shopper household I to nearest shopped outlet of type j: informal prepared outlet, milk bar/milk atm, depot/wholesaler, hawker, street vendor, mama mboga, kiosk, duka, small supermarket, hotel/restaurant, poshomill/cereal shop, butchery

This research assumes that prices for a food product will be the same for all households within an enumeration area. In this case, we used a 0.4 km radius circle around the household weighted center of an enumeration area to be considered the home food environment for all main shoppers living in that EA. Thus, by controlling for EA, we are also controlling for all factors that are constant within the EA, such as weather infrastructure, wages, policies, and prices. In addition, prices are self-reported by the main shoppers and thus are endogenous to the model and not suitable for use as a control.

Due to the presence of heteroskedasticity, standard errors are clustered at the EA (enumeration area) level. As indicated earlier, survey weights are applied to the data. These weights account for the likelihood of selection of location in the first stage of sampling; the likelihood of selection of enumeration areas in the second stage of sampling; and the likelihood of selection of households in the third stage of sampling. An OLS estimator will be used to estimate the coefficients of the model. Our general empirical model that will be estimated for the ith household is as follows:

$$\log\_health\_expend\_percap_i = \alpha + \beta_1 \log\_outlet\_distance_i + \delta_1 HH\_characteristics_i + \delta_2 MS\_characteristics_i + \delta_3 Location\_controls_i + \varepsilon_i \quad (4)$$

For this regression, we will first utilize a parsimonious model only incorporating the distance variable on the right-hand side, and then add household controls, followed by main shopper controls, and finally location controls. Our main dependent variables that will be utilized is the log of healthy food expenditure per capita. Again, we are utilizing a log-log functional form to test for significance of distance elasticity measures.

5.2: Group heterogeneity tests

To determine if there are heterogeneous effects among different groups in the sample, we perform a test for heterogeneity in terms of regression coefficients. For each set of groups, a fully saturated model is estimated interacting all right-hand side variables with the group dummy. Then, a Wald test is run, testing if the coefficients of the two groups are equal. Any significance of the interacted outlet distance variables is identified. The null hypothesis in this case is that the groups are the same, with the alternative being that groups are different:

H_0: groups' coefficient is the same:

$$\beta_i = \beta_j \text{ or } \beta_i - \beta_j = 0$$

H_0: groups' coefficient is different:

$$\beta_i \neq \beta_j \text{ or } \beta_i - \beta_j \neq 0$$

We are interested in testing for the group differences represented in the table below:

Table 10: Identification of dummy variables for heterogeneity test

Characteristic	Variable	Description
Age	$young_i$	• 1 if main shopper's age is less than 34 • 0 if main shopper's age is greater than or equal to 34
Gender	$male_i$	• 1 if main shopper's gender is male • 0 if main shopper's gender is female
Location (urban vs. peri-urban)	$urban_i$	• 1 if household location is urban • 0 if household location is peri-urban
Poverty probability	$lowpov_i$	• 1 if household's chance of poverty is greater than 59 • 0 if household's chance of poverty is less than or equal to 59
Strata location (based on census wealth index)	$poor_i$	• 1 if household belongs to a location in quartiles 1 or 2 (more poor) • 0 if household belongs to a location in quartiles 3 or 4 (more rich)
Transportation (car)	car_i	• 1 if main shopper's household owns a car • 0 if main shopper's household does not own a car

These variables are presented as a binary variable created using the median of the dataset. For age, 34 years is the median, thus any main shopper with an age less than 34 is considered 'young' in this context. With poverty status, the median probability is 59%, thus any main shopper with less than 59% is considered high poverty probability.

The regression model that is run is a fully saturated model with each dummy variable multiplied by each right-hand side term in the model. Each model 4-7 are specified below as multi-variable linear regressions with the interaction term. In this case, any of our three dependent variables may be used as our LHS variable. Since the log of monthly healthy expenditure per capita provides an ideal summary measure of expenditure of all range of healthiness, this variable is used as our LHS variable.

5.3: Robustness check

Early in the analysis, an assumption was made about the classification of foods into healthy and unhealthy expenditure categories. For the main analysis, it was decided that 'unhealthy in excess' food items should belong to 'healthy,' because households of low income might not choose to spend their resources on these luxury items and instead allocate their income to staple foods. While this assumption might prove to be valid, it is important to consider other courses of action, such as including 'unhealthy in excess' items with the 'healthy' category.

Thus, the dependent variables for the model change, but the independent variables do not. In a separate check, 'unhealthy in excess' is excluded entirely from the analysis.

Due to lack of data on the listing of large supermarkets in Kisumu urban and peri-urban, they are excluded from the main regression. However, as a robustness check, we explore how results vary if large supermarkets are included in the regression for the subset of the sample in Nairobi where this data are available.

SECTION 6: RESULTS

6.1: Regression results and interpretation

Table 11 shows the regression results of the OLS estimation for the dependent variables of interest: log of healthy food expenditure per capita. Columns A-D represents models with different combination of controls added to the model. Column A represents the parsimonious model, Column B includes household controls, C adds main shopper controls in addition, and D is the full model. The key variables of interest are the logged outlet distances, with estimated coefficients representing outlet distance elasticities with respect to healthy/unhealthy food expenditure per capita. Distance elasticity of food expenditure measures how 'distance sensitive' main shoppers are in terms of their food purchases.

From Table 3 in section 3.3, it was noted that the top 5 outlets with the highest average number of healthy items are mama mboga (5.8), small supermarket (4.47), duka (3.29), kiosk (3.17), and cereal shop/poshomill (2.32). On the other hand, the top 5 outlets with the highest average number of unhealthy items are small supermarket (4.1), duka (3.63), kiosk (2.11), depot/wholesale (2.07), and bakery (1.83). As we unpack the results, this classification will be referenced.

For Table 11, which uses the log of healthy food expenditure per capita as the dependent variable, we find as expected the same coefficient but of greater magnitude: for a 1% increase in distance to nearest cereal shop/poshomill, healthy expenditure increases by 0.34% on average for the fully saturated model. As the results generally move from only the parsimonious model in column A to the fully saturated model in column D, we find several significant variables. But, once location controls are added in column D, much of this significance disappears. One possible explanation for this interesting result is that once location is controlled for, much of the variation in healthiness of expenditure is attributed to where people live and their characteristics, not distance to specific outlets.

Regression results are robust for select outlet types in columns A-C of Table 11, before adding location controls to the model. We find that a 1% increase in the distance to the nearest milk bar/milk atm is correlated with a 0.10% increase in monthly healthy food expenditure per capita on average. Thus, as households move closer to milk bars/milk atms, they are incentivized to decrease their monthly healthy food expenditure. We also find a consist negative and significant coefficient on hawkers: a 1% increase in the distance from households to the nearest hawker results in a 0.06% decrease in monthly healthy food expenditure per capita on average. Mama mboga outlet distance and hotels and restaurants distance also have a similar inverse relationship with respect to healthy expenditure.

Our main hypothesis, that as distance increases from the main shopper's household location to each of the healthy outlet types (mama mbogas, small supermarkets, large supermarkets, dukas, kiosks, cereal shops/poshomills, and depots/wholesalers), healthy expenditure share will decrease, is partially confirmed. If we focus on column C in Table 11, which controls for household and main shopper characteristics, we find negative, significant coefficients for mama mboga outlets only. Variables that do have a negative coefficient but are not significant are dukas and depots/wholesalers. Surprisingly, small and large supermarkets, kiosks, and cereal shops/poshomills have positive signs, which indicates that as distance increases from households to the nearest of these outlet types, healthy expenditure increases.

Table 11: Regression results with the log of healthy monthly expenditure per capita as dependent variable

	A	B	C	D
Log of informal prepared distance	0.10	0.04	0.05	-0.02
	(0.06)	(0.06)	(0.06)	(0.10)
log of milk bar/milk atm distance	0.11**	0.082**	0.10***	-0.17
	(0.04)	(0.04)	(0.03)	(0.11)
log of depot/wholesaler distance	-0.07	-0.02	-0.02	-0.14
	(0.05)	(0.04)	(0.03)	(0.18)
log of hawker distance	-0.08**	-0.05*	-0.07***	0.03
	(0.03)	(0.03)	(0.02)	(0.06)
log of street vendor distance	-0.05	-0.05	-0.05	-0.06
	(0.05)	(0.05)	(0.05)	(0.10)
log of mama mboga distance	-0.07*	-0.11***	-0.09**	0.01
	(0.04)	(0.04)	(0.04)	(0.07)
log of kiosk distance	-0.02	0.07	0.0922**	0.09
	(0.05)	(0.05)	(0.04)	(0.06)
log of duka distance	0.02	0.01	-0.02	-0.04
	(0.05)	(0.05)	(0.05)	(0.07)
log of small supermarket distance	0.00	-0.01	0.03	-0.07
	(0.05)	(0.05)	(0.04)	(0.14)
log of hotel/restaurant distance	-0.10*	-0.12**	-0.13***	-0.09
	(0.05)	(0.05)	(0.04)	(0.09)
log of butchery distance	-0.07	-0.02	-0.03	0.00

Table 11 (cont'd)

	A	B	C	D
	(0.06)	(0.05)	(0.05)	(0.09)
log of cereal/poshomill distance	0.12***	0.11**	0.09**	0.34*
	(0.05)	(0.04)	(0.04)	(0.18)
Household controls	No	Yes	Yes	Yes
Main shopper controls	No	No	Yes	Yes
Location controls	No	No	No	Yes
Observations	1,429	1,429	1,429	1,429
R-squared	0.05	0.13	0.17	0.22

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

6.2: Heterogeneity test results and interpretation

Regression results for the heterogeneity test are presented below for each of the 6 group characteristics: age, gender, location, poverty probability, wealth class, and transportation via car. For each regression, note that the dependent variable was selected to be the log of healthy food expenditure per capita. For the intermediate stage regressions results, please see Appendix B.

We see that across all results, there is select detected heterogeneity among groups. When comparing young and old main shoppers, distance elasticities differ only for mama mboga outlets. Therefore, we see that the responsiveness of expenditure to a change in distance to mama mbogas is different when specifically looking at the age of main shoppers. Next, when comparing distance elasticities among male and female groups of main shoppers, we find no difference between groups. Yet, the location in which main shoppers live matters; urban and peri-urban group differences are found to be significant for informal prepared outlets and small supermarkets. Significant differences in distance elasticity with respect to food expenditure are also present between main shopper who belong to low and high chance of poverty households; street vendor outlets and small supermarkets are significant. Testing for differences in wealth, we find dukas to be significant. Finally, if we test for differences in households with and without a car, we find large supermarkets to be significant.

The results presented in this section demonstrate the nuances of food access across different groups. Distance elasticities with respect to food expenditure are far from homogenous as proven in the results.

Table 12: Wald test results for difference in coefficient value

Variable	Age less than or greater than/equal to median	household poverty probability less than vs. greater than/equal to median	urban vs. peri-urban location	male vs. female shopper	location strata (quartiles 1-2 vs. 3-4)	HH owns a car (Yes vs. No)
			P-value			
Log of informal prepared distance	0.68	0.26	0.08***	0.97	0.30	0.65
Log of milk bar/milk atm distance	0.38	0.67	0.20	0.58	0.16	0.002***
Log of depot/wholesaler distance	0.87	0.23	0.91	0.20	0.84	0.74
Log of hawker distance	0.94	0.77	0.76	0.10	0.63	0.42
Log of street vendor distance	0.94	0.003***	0.12	0.24	0.16	0.17
Log of mama mboga distance	0.06***	0.35	0.82	0.54	0.36	0.97
Log of kiosk distance	0.20	0.18	0.33	0.67	0.89	0.07***
Log of duka distance	0.35	0.93	0.75	0.71	0.005***	0.64
Log of small supermarket distance	0.08***	0.12	0.008***	0.14	0.15	0.35
Log of hotel restaurant distance	0.16	0.88	0.26	0.52	0.15	0.20
Log of butchery distance	0.59	0.71	0.85	0.33	0.72	0.59
Log of cereal shop/poshomill distance	0.65	0.90	0.13	0.40	0.88	0.56

*** if p-value is < 0.1

6.3: Robustness check results

Appendix D highlights regression results for the previously outlined robustness checks. Again, the purpose of these checks was to test the robustness of the results under different classifications of the dependent variable. Rather than adding 'unhealthy in excess' to the 'healthy' category, the two checks performed first assign 'unhealthy in excess' to 'unhealthy,' then excluded 'unhealthy in excess' categories entirely from the analysis.

The regressions in Appendix D show that results are generally robust and show similar patterns. Signs and magnitudes of coefficients are comparable; significance also follows a similar pattern. We even continue to see the same phenomenon with location controls; much of the coefficient significance is dropped once accounting for household location characteristics.

The second part of our robustness checks shows how the results change if large supermarkets are included. This regression analysis is based on the sub-sample of main shoppers who reside in Nairobi. Table 13 outlines regression results and it shows that results are robust across all 4 specifications A-D. We find similar significance, magnitude, and sign of coefficients as well as ascending R-squared values. The log of large supermarket distance is not found to be significant for any of the specified models A-D.

Table 13: Robustness check regression results including large supermarkets for Nairobi urban and peri-urban sub-sample

	A	B	C	D
Log of informal prepared distance	0.12*	0.05	0.05	-0.01
	(0.07)	(0.06)	(0.07)	(0.12)
log of millk bar/milk atm distance	0.14**	0.09**	0.10**	-0.15
	(0.05)	(0.04)	(0.04)	(0.11)
log of depot/wholesaler distance	-0.05	-0.02	-0.05	-0.28
	(0.06)	(0.05)	(0.04)	(0.22)
log of hawker distance	-0.08**	-0.05*	-0.07***	0.03
	(0.03)	(0.03)	(0.02)	(0.08)
log of street vendor distance	-0.05	-0.06	-0.05	-0.09
	(0.05)	(0.05)	(0.05)	(0.12)
log of mama mboga distance	-0.08*	-0.11**	-0.10*	0.03
	(0.04)	(0.05)	(0.05)	(0.08)
log of kiosk distance	-0.01	0.11**	0.14**	0.09
	(0.05)	(0.05)	(0.05)	(0.06)
log of duka distance	-0.02	-0.01	-0.04	-0.10

Table 13 (cont'd)

	A	B	C	D
	(0.06)	(0.06)	(0.06)	(0.08)
log of small supermarket distance	-0.01	-0.03	-0.01	-0.04
	(0.05)	(0.04)	(0.04)	(0.15)
Log of large supermarket distance	-0.07	0.03	0.03	-0.40
	(0.09)	(0.07)	(0.07)	(0.26)
log of hotel/restaurant distance	-0.16***	-0.16***	-0.16***	-0.16
	(0.05)	(0.04)	(0.04)	(0.10)
log of butchery distance	-0.04	-0.01	-0.01	0.05
	(0.08)	(0.07)	(0.06)	(0.11)
log of cereal/poshomill distance	0.15***	0.14***	0.14***	0.41*
	(0.05)	(0.04)	(0.04)	(0.20)
Household controls	No	Yes	Yes	Yes
Main shopper controls	No	No	Yes	Yes
Location controls	No	No	No	Yes
Observations	695	695	695	695
R-squared	0.06	0.15	0.19	0.22

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

SECTION 7: SUMMARY, CONCLUDING REMARKS, AND IMPLICATIONS

7.1: Summary

This paper analyzed the strength of the relationship between a simple food environment measure, linear distance to 14 different food outlets, and household main shoppers' healthy food expenditure in urban and peri-urban Kenya. The key takeaways, among all results, from this research are:

1. Once location is controlled for, much of the variation in healthiness of expenditure is attributed to where people live and their characteristics, not distance to specific outlets
2. The main hypothesis, that as distance increases from the main shopper's household location to each of the healthy outlet types, healthy expenditure will decrease, is partially confirmed. It was found that only for mama mboga outlets, who are street vegetable sellers, does increasing distance result in a decrease in healthy expenditure. Surprisingly, increasing distance to large and small supermarkets results in increases in healthy food expenditure, ceteris paribus.
3. Heterogeneity by group exists in the results: various model coefficients are statistically significant in terms of age, gender, location, poverty probability, and wealth location socio-economic strata. More research is needed to determine the nuances of this striking difference, especially since mama mboga outlets supply a plethora of healthy food groups.
4. From the descriptive statistics, we find that there is considerable heterogeneity in terms of what different types of foods outlets have on offer in terms of healthiness as well as what sampled main shoppers expend their resources on.
5. Robustness checks indicate that results are robust across variation definitions of the 'healthy' food category.

The main contribution of the paper is to first explore descriptive statistics on the healthiness of both food outlet offerings and main shopper food expenditure. Next, this research disaggregates distances to specific food outlets to create a more nuanced measure of the food environment in urban and peri-urban Kenya and apply the concept of distance elasticity. The results show that distance to some outlets relative to others can have an impact on how the main shopper responds in terms of food expenditure. It is also emphasized in the findings that how food groups are classified into 'healthy' and 'unhealthy' also can dictate results.

7.2: Study Limitations

This paper presented a simple measure of proximity for the food environment, one that only captures linear distance between two points. This simplified measure adequately measures proximity but may not reflect the reality of navigating city streets, such as a Manhattan distance would, or time of travel. The distance measure also does not factor in the mode of transportation, whether it be walking, driving own car, or taking public transportation. An additional limitation of this study is the assumption made about homogeneous prices; allowing for micro-variation in prices would help build more accurate model.

An additional limitation worth noting deals with the types of outlets that are included in this analysis. Due to data limitations, only smaller supermarkets, as opposed to larger operations, were able to be included in the analysis. Future work should investigate how distance elasticities of food expenditure might differ between various sizes of supermarkets, given the rise of supermarkets in urban and peri-urban areas of Kenya.

7.3: Concluding remarks and implications

The Kenyan government has made it clear that, among many policy-related issues, an important one is to ensure that households are able to purchase foods that contribute towards an adequate, diverse, and healthy diet (Kenya Ministry of Agriculture, 2011). Similar priorities have also been established by the United Nations and other global leadership (FAO et al. viii). This ability is partially driven by the nature of the food environment to provide access and availability of these foods. The research presented in this paper provides a critical perspective on the outlet-level differences in terms of consumer response to changes in proximity.

Generally, the food environment literature movement, marrying economics, nutrition, geography, among other fields, has shed new light on better understanding the spatial implications of consumers' surroundings on their well-being. It is even more important as a healthy food environment also lends itself to improved food security conditions. According to recent research, improving access to markets can influence consumption expenditure, household dietary diversity, and food security (Usman et al. 2022). Improved market access also expands the variety of foods available.

This research helps fill a knowledge gap in better understanding how those responsible for household food decisions might respond in terms of their food purchases to changes in outlet distances. This research also investigated how this responsiveness might differ based on economic, geographic, or gender differences. It is found that the degree of access measured through a sensitivity to changes in outlet distance has varying impacts on the amount of expenditure per capita dedicated to healthy foods. Most notably, results show that as the distance marginally decreases from main shoppers' households and small supermarkets, unhealthy food expenditure rises. Cereal mills and poshomills also have a similar effect: if distance to this type of outlet marginally decreases, healthy food expenditure decreases. As the food environment changes in terms of access measured through linear distance, so does the share of expenditure allocated to healthy and unhealthy food.

In response to these findings, greater care in the policy-making sphere of local and regional government must be taken to ensure food environments surrounding consumers can deliver food security and nutritional adequacy to people of all socioeconomic backgrounds. More research is needed on how the rise of supermarkets influences the food environment landscape, as well as the influence of other types of outlets such as cereal shops and poshomills. The results of this research show that unhealthy expenditure is sensitive and inversely related to the relative linear distance to small supermarkets; policymakers need to pay close attention to the influence of supermarkets, especially as this type of outlet supplies all ranges of both healthy and unhealthy food. Urban planning must consider the nature of different outlet types and assess the frequency with which consumers expend their resources. As an example, COVID-19 presented an interesting natural experiment to measure the impact of food outlet closures or travel restrictions on healthiness of household expenditure as well as diet itself.

Future research on this topic can address questions such as how people substitute the healthiness of their purchases if outlet distances vary, and what additional factors of outlets beyond distance might impact food expenditure. This research used a cross-sectional dataset, but future longitudinal studies can analyze trends over time in outlet placement and location relative to household food decisions. More advanced spatial econometric tools can be used as well to provide more refined estimates.

Additional questions can be asked about how price changes signal consumers to switch between outlets, and how food expenditure decisions shifts between healthy and unhealthy choices with price changes. Especially with recent current events of global conflict, climate change, and a pandemic, providing evidence to policy makers of the linkages between prices, the food environment, and nutrition can help inform key policy responses to these ongoing global challenges, in Kenya and beyond.

Lastly, future research can investigate how the entrance of a new healthy or unhealthy outlet in a food environment changes the welfare of both existing outlets and consumers. Ongoing urbanization and food system transformation in Kenya has led to rapid changes in the socioeconomic distribution of people across urban areas and how and where food is acquired. A better understanding of the nature of these dynamics will lend itself to more responsive policy, improved economic welfare, and most importantly, a more food secure population.

BIBLIOGRAPHY

Auspitz, R. & Lieben, R. (1889). Untersuchungen iiber die Theorie des Preises, Verlag von Duncker und Kumblot, Leipzig.

Athens, J.K., Duncan, D.T., & Elbel, B. (2016). Proximity to fast food outlets and supermarkets as predictors of fast food dining frequency. *Journal of the Academy of Nutrition and Dietetics*, 116(8): 1266-1275. DOI: 10.1016/j.jand.2015.12.022

Babey, S.H. et al. (2008). Designed for disease: the link between local food environments and obesity and diabetes. *UCLA Center for Health Policy Research*. https://escholarship.org/uc/item/7sf9t5wx

Barnes, T.L. et al. (2015). Geographic measures of retail food outlets and perceived availability of healthy foods in neighborhoods. *Journal of Public Health Nutrition*, 19(8): 1368-1374. DOI: 10.1017/S1368980015002864

Becker, G.S. (1965). A theory of the allocation of time. *The Economic Journal*, 75: 493-517. https://doi.org/10.2307/2228949

Bennett, M.K. (1941). International contrasts in food consumption. *Geographical Review*, 31: 365-376. https://doi.org/10.2307/210172

Bergstrand, J.H., Larch, M., & Yotov, Y.V. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78: 307-327. https://doi.org/10.1016/j.euroecorev.2015.06.003

Berthelon, M. & Freund, C. (2008). On the conversation of distance in international trade. *Journal of International Economics*, 75: 310-320. https://EconPapers.repec.org/RePEc:eee:inecon:v:75:y:2008:i:2:p:310-320

Black, C., Moon, G., & Baird, J. (2014). Dietary inequalities: What is the evidence for the effect of the neighborhood food environment? *Health & Place*, 27: 229-242. https://doi.org/10.1016/j.healthplace.2013.09.015

Bodor, J.N. et al. (2008). Neighbourhood fruit and vegetable availability and consumption: the role of small food stores in an urban environment. *Public Health Nutrition*, 11:413-420. DOI: 10.1017/S1368980007000493

Bromage S. et al. (2021). Development and validation of a novel food-based global diet quality score (GDQS). *Journal of Nutrition*, 151: 75-92. DOI: 10.1093/jn/nxab244

Busse, K.R. et al. (2022). Food vendors and the obesogenic food environment of an informal settlement in Nairobi, Kenya: a descriptive and spatial analysis. *Journal of Urban Health*, 100: 76-87. DOI: 10.1007/s11524-022-00687-7

Cassel, G. (1923). The Theory of social economy. Longdon: T. Fisher Unwin.

Caspi, C.E. et al. (2012). The local food environment and diet: A systematic review. *Health & Place*, 18: 1172-1187. https://doi.org/10.1016/j.healthplace.2012.05.006

Chenarides, L. & Jaenicke, E.C. (2016). Store choice and consumer behavior in food deserts: an empirical application of the distance metric method. Working Paper: 1-25. https://www.tsefr.eu/sites/default/files/TSE/documents/sem2016/food_eco/jaenicke_2.pdf

Clapp, J. et al. (2022). Viewpoint: The case for a six-dimensional food security framework. *Food Policy*, 106: 102164. https://doi.org/10.1016/j.foodpol.2021.102164

Crush, J. & Battersby, J. (2017). Rapid urbanization, urban food deserts, and food security in Africa. London.

Ebeke, C. & Etoundi, M.N. (2016). Does distance still matter for agricultural trade? Working paper. https://www.ifpri.org/publication/does-distance-still-matter-agricultural-trade

FAO et al. (2020). The state of food security and nutrition in the world 2020: transforming food systems for affordable healthy diets. Rome.. https://doi.org/10.4060/ca9692en

FAO (2016). Influencing food environments for healthy diets. Rome. https://www.fao.org/3/i6484e/i6484e.pdf

FAO (2006). Policy brief: food security- issue 2. Reliefweb. https://reliefweb.int/report/world/policy-brief-food-security-issue-2-june-2006

Farley, T.A. et al. (2009). Measuring the Food Environment: Shelf Space of Fruits, Vegetables, and Snack Foods in Stores. *Journal of Urban Health*, 86: 672-682. DOI: 10.1007/s11524-009-9390-3

Fraser, L.K. et al. (2010). The geography of fast food outlets: a review. *International Journal of Environmental Research and Public Health*, 7: 2290-2308. doi:10.3390/ijerph7052290

Fung, T.T., et al. (2018) International food group–based diet quality and risk of coronary heart disease in men and women. *Clinical Nutrition*, 107: 120-129. https://doi.org/10.1093/ajcn/nqx015

Gewa, C.A. et al. (2021). Food environments in and around primary school children's schools and neighborhoods in two urban settings in Kenya. International Journal of Environmental Research and Public Health, 18(10): 5136. DOI: https://doi.org/10.3390/ijerph18105136

Glanz, K. et al. (2005). Healthy nutrition environments: Concepts and measures. *American Journal of Health Promotion*, 19: 330-343. 10.4278/0890-1171-19.5.330

Global Panel (2017). Improving nutrition through enhanced food environments. Policy Brief No. 7. London, UK: Global Panel on Agriculture and Food Systems for Nutrition. https://glopan.org/sites/default/files/FoodEnvironmentsBrief.pdf

Green, M.A. et al. (2020). Investigating foods and beverages sold and advertised in deprived urban neighborhoods in Ghana and Kenya. *BMJ Open*, 10(6): n.p. DOI: 10.1136/bmjopen-2019-035680

Hawkes, C., Harris, J., & Gillespie, S. (2017). Changing diets: Urbanization and the nutrition transition. In 2017 Global Food Policy Report. Chapter 4. Pp 34-41. Washington, DC: International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/9780896292529_04

Hawkes, C. et al. (2015). Smart food policies for obesity prevention. *The Lancet*, 385: 2410-2421. doi: 10.1016/S0140-6736(14)61745-1

Hawkes, C. (2008). Dietary implications of supermarket development: a global perspective. *Development Policy Review*, 26: 657-692. https://doi.org/10.1111/j.1467-7679.2008.00428.x

Heien, D., Jarvis, L.S., & Perali, F. (1989). Food consumption in Mexico: Demographic and economic effects. *Food Policy*, 14: 167-179. https://doi.org/10.1016/0306-9192(89)90009-2

Herforth, A. & Ahmed, S. (2015). The food environment, its effects on dietary consumption, and potential for measurement within agriculture-nutrition interventions. *Food Security*, 7: 505-520. 10.1007/s12571-015-0455-8

Holdsworth, M. & Landais, E. (2019). Urban food environments in Africa: implications for policy and research. Proceedings of the Nutrition Society, 78: 513-525. DOI: https://doi.org/10.1017/S0029665118002938

Jago, R. et al. (2007). Distance to food stores and adolescent male fruit and vegetable consumption: mediation effects. *International Journal of Behavioral Nutrition and Physical Activity*, 4:35. doi:10.1186/1479- 5868-4-35

Kenya Ministry of Agriculture (2011). National food and nutrition security policy: 2011. *The Kenya Institute for Public Policy Research and Analysis*. http://repository.kippra.or.ke/handle/123456789/1620

Kinsey, J. (1983). Working wives and the marginal propensity to consume food away from home. *American Journal of Agricultural Economics*, 65: 10-19. https://doi.org/10.2307/1240332

Koech, K.K., Chege, C.G., & Bett, H. (2022). Which food outlets are important for nutrient dense porridge flour access by the base of the pyramid consumers? Evidence from the informal Kenyan settlements. *Sustainability*, 14: 12264. **https://doi.org/10.3390/su141912264**

Kruger, D.J. et al. (2013). Local Concentration of fast-food outlets is associated with poor nutrition and obesity. *American Journal of Health Promotion*, 28:340-343. DOI: 10.4278/ajhp.111201-QUAN-437

Ladd, G.W. & Zober, M. (1977). Model of consumer reaction to product characteristics. *Journal of Consumer Research*, 4:89-101. https://www.jstor.org/stable/2488715

Lancaster, K.J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74: 132-157. https://www.jstor.org/stable/1828835

Laraia, B.A. et al. (2004). Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Preventative Medicine*, 39: 869-875. DOI: 10.1016/j.ypmed.2004.03.018

Laska, M.N. et al. (2010). Neighbourhood food environments: are they associated with adolescent dietary intake, food purchases and weight status? *Public Health Nutrition*, 13: 1757-1763. DOI: 10.1017/S1368980010001564

Mas-Colell, A., Whinston, M.D. & Green, J.R. (1995). Microeconomic Theory. Oxford University Press.

Mergenthaler, M., Weinberger, K., & Qaim, M. (2009). The food system transformation in developing countries: a disaggregate demand analysis for fruits and vegetables in Vietnam. *Food Policy*, 34: 426-436. https://doi.org/10.1016/j.foodpol.2009.03.009

Monteiro, C.A. et al. (2013). Ultra-processed products are becoming dominant in the global food system. *Obesity Review*, 14: 21-28. DOI: 10.1111/obr.12107

Musaiger, A.O. (1990). Changes in food consumption patterns in Bahrain. *Nutrition and Health*, 6: 23-42. https://doi.org/10.1177/026010609000600403

Nicholson, W. & Snyder, C.M. (2008). Microeconomic theory: basic principles and extensions. Cengage Learning Press.

Odunitan-Wayas, F. et al. (2018). Food Purchasing Characteristics and Perceptions of Neighborhood Food Environment of South Africans Living in Low-, Middle- and High-Socioeconomic Neighborhoods. *Sustainability*, 10: 4801. https://doi.org/10.3390/su10124801

Osei-Kwasi, H.A. et al. (2021). The African urban food environment framework for creating healthy nutrition policy and interventions in urban Afric. *Plus One*, 16: e0249621. https://doi.org/10.1371/journal.pone.0249621

Owuor, S. (2018). Hungry cities report no. 11: The state of household food security in Nairobi, Kenya. Hungry Cities Partnership. https://scholars.wlu.ca/cgi/viewcontent.cgi?article=1034&context=hcp

Pearson, T. et al. (2005). Do 'food deserts' influence fruit and vegetable consumption? A cross-sectional study. *Appetite*, 45: 195-197. DOI: 10.1016/j.appet.2005.04.003

Perez-Escamilla, R. et al. (2018). Nutrition disparities and the global burden of malnutrition. *BMJ: Science and Politics of Nutrition*, 361: k2252. doi: https://doi.org/10.1136/bmj.k2252

Pollack, R.A. & Wales, T.J. (1981). Demographic variables in demand analysis. *Econometrica*, 49: 1533-1551. https://doi.org/10.2307/1911416

Popkin, B.M. & Reardon, T. (2018). Obesity and the food system transformation in Latin America. *Obesity Reviews*, 19: 1028-1064. https://doi.org/10.1111/obr.12694

Popkin, B.M. (1993). Nutritional patterns and transitions. *Population and Development Review*, 19: 138-157. https://doi.org/10.2307/2938388

Popkin, B.M. (1999). Urbanization, lifestyle changes and the nutrition transition. *World Development*, 27: 1905-1916. https://doi.org/10.1016/S0305-750X(99)00094-7

Popkin, B.M (2004). The nutrition transition: an overview of world patterns of change. *Nutrition Review*, 62: 140-143. DOI: 10.1111/j.1753-4887.2004.tb00084.x

Pradeilles, R. et al. (2021). Urban physical food environments drive dietary behaviours in Ghana and Kenya: A photovoice study. *Health and Place*, 71: 102647. https://doi.org/10.1016/j.healthplace.2021.102647

Reardon, T. et al. (2017). The rise of supermarkets in Africa, Asia, and Latin America. American *Journal of Agricultural Economics*, 85: 1140-1146. https://www.jstor.org/stable/1244885

Republic of Kenya (2021). An analysis of the national progress and household characteristics associated with stunting. *National Information Platform for Food Security and Nutrition*. https://nipfn.kippra.or.ke/download/an-analysis-of-the-national-progress-and-household-characteristics-associated-with-stunting/

Republic of Kenya (2021). Food situation during covid-19 pandemic. *Kenya National Bureau of Statistics*. https://www.knbs.or.ke/?wpdmpro=food-situation-during-covid-19-pandemic

Rose, D. & Richards, R. (2004). Food store access and household fruit and vegetable use among participants in the US Food Stamp Program. *Public Healthy Nutrition,* 7: 1081-1088. DOI: 10.1079/PHN2004648

Rose, D. et al. (2009). Neighborhood Food Environments and Body Mass Index The Importance of In-Store Contents. *American Journal of Preventative Medicine*, 37: 214-9. DOI: 10.1016/j.amepre.2009.04.024

Rosen, S. (1974). Hedonic price and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82: 34-55. https://www.jstor.org/stable/1830899

Ruel, M.T. et al. (2017). Urbanization, Food Security and Nutrition. In: de Pee, S., Taren, D., Bloem, M. (eds) Nutrition and Health in a Developing World . Nutrition and Health. Humana Press, Cham. https://doi.org/10.1007/978-3-319-43739-2_32

Schreiner, M. (2018). Simple poverty scorecard tool: Kenya. Scorocs, 1-69. https://www.simplepovertyscorecard.com/KEN_2015_ENG.pdf

Schwartz, A. (1973). Interpreting the effect of distance on migration. *Journal of Political Economy*, 81: 1153-1169. https://www.jstor.org/stable/1830643

Senauer, B. (1991). *Food trends and the changing consumer*. Eagan Press.

Sharkey, J.R. et al. (2011). Association between proximity to and coverage of traditional fast-food restaurants and non-traditional fast-food outlets and fast-food consumption among rural adults. *International Journal of Health Geographics*, 10:37. http://www.ij-healthgeographics.com/content/10/1/37

Slama, M.E. & Tashchian, A. (1985). Selected socioeconomic and demographic characteristics associated with purchasing involvement. *Journal of Marketing*, 49: 72-82. https://doi.org/10.2307/1251177

Spence, J.C. et al. (2009). Relation between local food environments and obesity among adults. *BMC Public Health*, 9:192. doi:10.1186/1471-2458-9-192

Stevano, S., Johnston, D., & Codjoe, E. (2019). The urban food question in the context of inequality and dietary changes: a study of schoolchildren in Accra. *The Journal of Development Studies*. 56. https://doi.org/10.1080/00220388.2019.1632434

St-Onge, M.P., Keller, K.L., & Heymsfield, S.B. (2003). *The American Journal of Clinical Nutrition*. 78: 1068-1073. https://doi.org/10.1093/ajcn/78.6.1068

Swinburn, B. et al. (2013). INFORMAS (International Network for Food and Obesity/non-communicable diseases Research, Monitoring and Action Support): overview and key principles. *Obesity Review*, 14: 1-12. doi: 10.1111/obr.12087

Tschirley, D. et al. (2015). The rise of a middle class in East and Southern Africa: implications for food system transformation. *Journal of International Development*, 27: 628-646. https://doi.org/10.1002/jid.3107

Tschirley et al. (2021). Measuring the Food Environment and its Impact on Diets: Alternative Metrics Deliver Very Different Results in Urban Nairobi. Working Paper.

Turner, C. et al. (2018). Concepts and critical perspective for food environment research: a global framework with implications for action in low- and middle-income countries. *Global Food Security*, 18: 93-101. https://doi.org/10.1016/j.gfs.2018.08.003

United Nations (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). *New York: UN Department of Economic and Social Affairs*.

https://www.un.org/en/development/desa/publications/2014-revision-world-urbanization-prospects.html

Usman et al. (2022). Market access, household dietary diversity and food security: Evidence from Eastern Africa. Food Policy, 113: 102374. https://doi.org/10.1016/j.foodpol.2022.102374

Wang, Z. et al. (2000). The impact of distance on us ag exports: an econometric analysis. Working paper. file:///Users/ianfisher/Downloads/dot_38943_DS1.pdf

Wang, M.C. et al. (2007). Socioeconomic and food-related physical characteristics of the neighborhood environment are associated with body mass index. *Journal of Epidemiology and Community Health*, 61: 491-498. DOI: 10.1136/jech.2006.051680

Wrigley, N. (2002). "Food Deserts" in British Cities: Policy Context and Research Priorities. *Urban Studies, 39: 2029*-2040. https://doi.org/10.1080/0042098022000011344

Zimmerman, C.C. (1932). Ernst Engel's Law of Expenditures for Food. *The Quarterly Journal of Economics*, 47: 78-101. https://doi.org/10.2307/1885186

Zenk, S.N. & Powell, L.M. (2008). US secondary schools and food outlets. *Health Place*, 14: 336-346. DOI: 10.1016/j.healthplace.2007.08.003

Zenk, S.N. et al. (2005). Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *American Journal of Public Health*, 95: 660-667. DOI: 10.2105/AJPH.2004.042150

APPENDIX A: FULL REGRESSION RESULTS

Table 14: Full regression results with the log of healthy expenditure per capita as dependent variable

	A	B	C	D
Log of informal prepared outlet distance	0.10	0.04	0.05	-0.01
	(0.06)	(0.06)	(0.06)	(0.10)
log of millk bar/milk atm distance	0.11**	0.08**	0.10***	-0.16
	(0.04)	(0.03)	(0.03)	(0.11)
log of depot/wholesaler distance	-0.06	-0.03	-0.03	-0.16
	(0.06)	(0.04)	(0.03)	(0.18)
log of hawker distance	-0.08**	-0.05*	-0.06***	0.01
	(0.03)	(0.03)	(0.02)	(0.07)
log of street vendor distance	-0.05	-0.05	-0.05	-0.08
	(0.05)	(0.05)	(0.05)	(0.10)
log of mama mboga distance	-0.07**	-0.10**	-0.09*	0.01
	(0.03)	(0.04)	(0.04)	(0.07)
log of kiosk distance	-0.02	0.07	0.09**	0.08
	(0.05)	(0.05)	(0.04)	(0.06)
log of duka distance	0.01	0.02	-0.01	-0.05
	(0.05)	(0.06)	(0.05)	(0.07)
log of small supermarket distance	0.00	-0.01	0.03	-0.06
	(0.05)	(0.05)	(0.04)	(0.14)
log of large supermarket distance	-0.03	0.06	0.04	-0.36
	(0.09)	(0.06)	(0.06)	(0.27)
log of hotel/restaurant distance	-0.11**	-0.11**	-0.12***	-0.11
	(0.05)	(0.05)	(0.04)	(0.08)
log of butchery distance	-0.06	-0.02	-0.03	0.003
	(0.06)	(0.05)	(0.05)	(0.09)
log of cereal/poshomill distance	0.13**	0.10**	0.09*	0.34*
	(0.05)	(0.05)	(0.04)	(0.18)
hhsize		-0.03	-0.03	-0.05
		(0.03)	(0.04)	(0.04)
pov_score		0.021***	0.02***	0.02***
		(0.00)	(0.00)	(0.01)
children		-0.002	-0.05	-0.04
		(0.04)	(0.04)	(0.04)
hh_bicycle		0.38***	0.36***	0.38***
		(0.05)	(0.06)	(0.06)

Table 14 (cont'd)

A	B	C	D
hh_car	0.18	0.14	0.14
	(0.13)	(0.13)	(0.14)
hh_truck	0.38*	0.325	0.46**
	(0.21)	(0.22)	(0.22)
2.shopper_gender		0.10	0.08
		(0.11)	(0.10)
3.shopper_gender		-0.03	-0.01
		(0.14)	(0.14)
age_shopper		0.00	0.00
		(0.00)	(0.00)
education_shopper		0.01	0.00
		(0.01)	(0.02)
2.occupation_MS_1		-0.17	-0.19
		(0.20)	(0.23)
3.occupation_MS_1		-1.00**	-0.85*
		(0.45)	(0.49)
4.occupation_MS_1		-0.10	-0.10
		(0.23)	(0.27)
5.occupation_MS_1		0.11	0.06
		(0.13)	(0.16)
6.occupation_MS_1		-0.08	-0.12
		(0.11)	(0.12)
7.occupation_MS_1		0.06	0.05
		(0.14)	(0.15)
nutrition1		0.04	0.03
		(0.13)	(0.14)
taste1		-0.05	-0.05
		(0.08)	(0.08)
convenience1		0.06	0.04
		(0.11)	(0.11)
price1		-0.11	-0.07
		(0.12)	(0.13)
availability1		-0.12	-0.13
		(0.09)	(0.10)
perishability1		0.18**	0.18**
		(0.08)	(0.08)
2.region			0.44

Table 14 (cont’d)

	A	B	C	D
				(0.42)
3.region				1.52
				(0.92)
4.region				0.98
				(0.60)
222205046.ea				-0.49
				(0.94)
222205058.ea				-1.52
				(1.04)
222205068.ea				-0.28
				(0.37)
222206019.ea				1.64***
				(0.56)
222206024.ea				1.39***
				(0.40)
222206029.ea				-0.06
				(0.69)
222206061.ea				-0.04
				(0.51)
222207021.ea				0.58
				(0.63)
222207038.ea				1.13*
				(0.66)
222211043.ea				0.17
				(0.40)
222211141.ea				1.02
				(1.06)
343403012.ea				0.49
				(0.51)
343403022.ea				0.89**
				(0.36)
343403023.ea				-0.52
				(0.68)
424201003.ea				-0.58
				(0.57)
424201006.ea				-0.85
				(0.61)

Table 14 (cont'd)

	A	B	C	D
424201008.ea				-1.09**
				(0.42)
424201014.ea				-1.31*
				(0.75)
424201018.ea				-1.21***
				(0.41)
424201021.ea				-0.57*
				(0.31)
424201024.ea				-1.57**
				(0.73)
424201026.ea				-1.30***
				(0.44)
424201056.ea				-1.04
				(0.68)
424201324.ea				-0.41
				(0.59)
424202001.ea				-1.03
				(0.78)
424202011.ea				-0.96
				(0.67)
424202020.ea				-1.14*
				(0.66)
424202036.ea				-1.05
				(0.65)
424202037.ea				-3.64***
				(0.69)
424202043.ea				-0.80
				(0.61)
424203003.ea				0.22
				(0.88)
424203005.ea				1.30
				(1.11)
424203006.ea				-0.45
				(0.56)
424203007.ea				-0.68**
				(0.31)
424203015.ea				-0.18

Table 14 (cont'd)

	A	B	C	D
				(0.55)
424203016.ea				-0.55
				(0.62)
424203025.ea				-1.25***
				(0.41)
424203033o.ea				-
424205003.ea				-0.63
				(0.65)
424205004.ea				-0.17
				(0.48)
424206003.ea				-0.37
				(0.43)
424206005.ea				0.77
				(0.51)
424206006.ea				0.65
				(0.44)
424206007o.ea				-
474701048.ea				0.26
				(0.43)
474701058.ea				0.74
				(0.53)
474702087.ea				-0.46
				(0.79)
474702089.ea				-0.46
				(0.77)
474703003.ea				0.05
				(0.49)
474703008.ea				0.35
				(0.52)
474704022.ea				0.35
				(0.47)
474704049.ea				0.08
				(0.40)
474704064.ea				0.71**
				(0.32)

Table 14 (cont'd)

	A	B	C	D
474704068.ea				-0.17
				(0.53)
474705003.ea				0.60
				(0.47)
474705062.ea				0.28
				(0.76)
474706029.ea				-0.05
				(0.48)
474706072.ea				-0.21
				(0.57)
474710012.ea				0.02
				(0.31)
Household controls	No	Yes	Yes	Yes
Main shopper controls	No	No	Yes	Yes
Location controls	No	No	No	Yes
Observations	1,429	1,429	1,429	1,429
R-squared	0.046	0.126	0.168	0.219

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX B: HETEROGENEITY TEST RESULTS

Table 15: Regression results for Wald test: age less than or greater than/equal to median

	A
0b.young#c.log_infprep_straightlinedistkm	0.04
	(0.05)
1.young#c.log_infprep_straightlinedistkm	0.10
	(0.13)
0b.young#c.log_milkba_straightlinedistkm	0.12***
	(0.04)
1.young#c.log_milkba_straightlinedistkm	0.06
	(0.06)
0b.young#c.log_depwhole_straightlinedistkm	-0.01
	(0.06)
1.young#c.log_depwhole_straightlinedistkm	0.01
	(0.07)
0b.young#c.log_hawker_straightlinedistkm	-0.05*
	(0.03)
1.young#c.log_hawker_straightlinedistkm	-0.05
	(0.05)
0b.young#c.log_streetv_straightlinedistkm	-0.07
	(0.04)
1.young#c.log_streetv_straightlinedistkm	-0.07
	(0.10)
0b.young#c.log_mamamboga_straightlinedistkm	-0.17***
	(0.04)
1.young#c.log_mamamboga_straightlinedistkm	-0.04
	(0.05)
0b.young#c.log_kiosk_straightlinedistkm	0.12**
	(0.06)
1.young#c.log_kiosk_straightlinedistkm	0.03
	(0.05)
0b.young#c.log_duka_straightlinedistkm	0.03
	(0.05)
1.young#c.log_duka_straightlinedistkm	-0.04
	(0.08)
0b.young#c.log_smalls_straightlinedistkm	-0.07
	(0.06)
1.young#c.log_smalls_straightlinedistkm	0.06

Table 15 (cont'd)

	A
	(0.07)
0b.young#c.log_hotelrest_straightlinedistkm	-0.07
	(0.06)
1.young#c.log_hotelrest_straightlinedistkm	-0.18***
	(0.06)
0b.young#c.log_butch_straightlinedistkm	-0.04
	(0.07)
1.young#c.log_butch_straightlinedistkm	0.01
	(0.08)
0b.young#c.log_cer_posho_straightlinedistkm	0.13**
	(0.06)
1.young#c.log_cer_posho_straightlinedistkm	0.10
	(0.07)
Household controls	Yes
Main shopper controls	Yes
Location controls	Yes
Observations	1,429
R-squared	0.17

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

Table 16: Regression results for Wald test: household poverty probability less than vs. greater than/equal to median

	A
0b.lowpov#c.log_infprep_straightlinedistkm	0.16
	(0.10)
1.lowpov#c.log_infprep_straightlinedistkm	0.01
	(0.10)
0b.lowpov#c.log_milkba_straightlinedistkm	0.09*
	(0.05)
1.lowpov#c.log_milkba_straightlinedistkm	0.06
	(0.05)
0b.lowpov#c.log_depwhole_straightlinedistkm	0.05
	(0.06)
1.lowpov#c.log_depwhole_straightlinedistkm	-0.05

Table 16 (cont'd)

	A
	(0.06)
0b.lowpov#c.log_hawker_straightlinedistkm	-0.08**
	(0.03)
1.lowpov#c.log_hawker_straightlinedistkm	-0.09**
	(0.04)
0b.lowpov#c.log_streetv_straightlinedistkm	-0.21***
	(0.06)
1.lowpov#c.log_streetv_straightlinedistkm	0.02
	(0.06)
0b.lowpov#c.log_mamamboga_straightlinedistkm	-0.03
	(0.07)
1.lowpov#c.log_mamamboga_straightlinedistkm	-0.11**
	(0.04)
0b.lowpov#c.log_kiosk_straightlinedistkm	0.11**
	(0.05)
1.lowpov#c.log_kiosk_straightlinedistkm	0.04
	(0.05)
0b.lowpov#c.log_duka_straightlinedistkm	-0.02
	(0.07)
1.lowpov#c.log_duka_straightlinedistkm	-0.01
	(0.08)
0b.lowpov#c.log_smalls_straightlinedistkm	-0.04
	(0.05)
1.lowpov#c.log_smalls_straightlinedistkm	0.06
	(0.07)
0b.lowpov#c.log_hotelrest_straightlinedistkm	-0.15**
	(0.06)
1.lowpov#c.log_hotelrest_straightlinedistkm	-0.14**
	(0.05)
0b.lowpov#c.log_butch_straightlinedistkm	0.01
	(0.07)
1.lowpov#c.log_butch_straightlinedistkm	-0.03
	(0.08)
0b.lowpov#c.log_cer_posho_straightlinedistkm	0.14*
	(0.08)
1.lowpov#c.log_cer_posho_straightlinedistkm	0.13*
	(0.07)
Household controls	Yes

Table 16 (cont'd)

	A
Main shopper controls	Yes
Location controls	Yes
Observations	1,429
R-squared	0.17

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

Table 17: Regression results for Wald test: urban vs. peri-urban location

	A
0b.urban#c.log_infprep_straightlinedistkm	0.18*
	-0.09
1.urban#c.log_infprep_straightlinedistkm	-0.05
	-0.09
0b.urban#c.log_milkba_straightlinedistkm	0.00
	-0.05
1.urban#c.log_milkba_straightlinedistkm	0.10
	-0.06
0b.urban#c.log_depwhole_straightlinedistkm	-0.01
	-0.08
1.urban#c.log_depwhole_straightlinedistkm	-0.02
	-0.07
0b.urban#c.log_hawker_straightlinedistkm	-0.08***
	-0.03
1.urban#c.log_hawker_straightlinedistkm	-0.09**
	-0.04
0b.urban#c.log_streetv_straightlinedistkm	-0.13**
	-0.06
1.urban#c.log_streetv_straightlinedistkm	0.03
	-0.07
0b.urban#c.log_mamamboga_straightlinedistkm	-0.06
	-0.05
1.urban#c.log_mamamboga_straightlinedistkm	-0.04
	-0.07
0b.urban#c.log_kiosk_straightlinedistkm	0.11**
	-0.05

Table 17 (cont'd)

	A
1.urban#c.log_kiosk_straightlinedistkm	0.02
	-0.07
0b.urban#c.log_duka_straightlinedistkm	-0.04
	-0.06
1.urban#c.log_duka_straightlinedistkm	0.00
	-0.08
0b.urban#c.log_smalls_straightlinedistkm	0.10*
	-0.06
1.urban#c.log_smalls_straightlinedistkm	-0.09**
	-0.04
0b.urban#c.log_hotelrest_straightlinedistkm	-0.17***
	-0.04
1.urban#c.log_hotelrest_straightlinedistkm	-0.04
	-0.10
0b.urban#c.log_butch_straightlinedistkm	0.02
	-0.05
1.urban#c.log_butch_straightlinedistkm	0.00
	-0.09
0b.urban#c.log_cer_posho_straightlinedistkm	0.10**
	-0.05
1.urban#c.log_cer_posho_straightlinedistkm	-0.04
	-0.07
Household controls	Yes
Main shopper controls	Yes
Location controls	Yes
Observations	1,429
R-squared	0.19

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

Table 18: Regression results for Wald test: male vs. female shopper

	A
0b.MS_male#c.log_infprep_straightlinedistkm	0.02
	(0.07)
1.MS_male#c.log_infprep_straightlinedistkm	0.02

Table 18 (cont'd)

	A
	(0.13)
0b.MS_male#c.log_milkba_straightlinedistkm	0.07**
	(0.03)
1.MS_male#c.log_milkba_straightlinedistkm	0.13
	(0.10)
0b.MS_male#c.log_depwhole_straightlinedistkm	0.02
	(0.05)
1.MS_male#c.log_depwhole_straightlinedistkm	-0.13
	(0.12)
0b.MS_male#c.log_hawker_straightlinedistkm	-0.03
	(0.03)
1.MS_male#c.log_hawker_straightlinedistkm	-0.17**
	(0.08)
0b.MS_male#c.log_streetv_straightlinedistkm	-0.05
	(0.04)
1.MS_male#c.log_streetv_straightlinedistkm	0.18
	(0.19)
0b.MS_male#c.log_mamamboga_straightlinedistkm	-0.08*
	(0.05)
1.MS_male#c.log_mamamboga_straightlinedistkm	-0.15
	(0.09)
0b.MS_male#c.log_kiosk_straightlinedistkm	0.10**
	(0.04)
1.MS_male#c.log_kiosk_straightlinedistkm	0.05
	(0.10)
0b.MS_male#c.log_duka_straightlinedistkm	-0.03
	(0.06)
1.MS_male#c.log_duka_straightlinedistkm	-0.08
	(0.13)
0b.MS_male#c.log_smalls_straightlinedistkm	0.01
	(0.05)
1.MS_male#c.log_smalls_straightlinedistkm	-0.11*
	(0.06)
0b.MS_male#c.log_hotelrest_straightlinedistkm	-0.10**
	(0.04)
1.MS_male#c.log_hotelrest_straightlinedistkm	-0.17*
	(0.09)
0b.MS_male#c.log_butch_straightlinedistkm	-0.01

Table 18 (cont'd)

	A
	(0.04)
1.MS_male#c.log_butch_straightlinedistkm	-0.12
	(0.13)
0b.MS_male#c.log_cer_posho_straightlinedistkm	0.08
	(0.05)
1.MS_male#c.log_cer_posho_straightlinedistkm	0.17*
	(0.10)
Household controls	Yes
Main shopper controls	Yes
Location controls	Yes
Observations	1,416
R-squared	0.20

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

Table 19:Regression results for Wald test: location strata (quartiles 1-2 vs. 3-4)

	A
0b.poor#c.log_infprep_straightlinedistkm	-0.03
	(0.08)
1.poor#c.log_infprep_straightlinedistkm	0.11
	(0.10)
0b.poor#c.log_milkba_straightlinedistkm	-0.01
	(0.06)
1.poor#c.log_milkba_straightlinedistkm	0.09*
	(0.05)
0b.poor#c.log_depwhole_straightlinedistkm	0.02
	(0.08)
1.poor#c.log_depwhole_straightlinedistkm	0.00
	(0.06)
0b.poor#c.log_hawker_straightlinedistkm	-0.08
	(0.07)
1.poor#c.log_hawker_straightlinedistkm	-0.04
	(0.04)
0b.poor#c.log_streetv_straightlinedistkm	-0.21***
	(0.08)

Table 19 (cont'd)

	A
1.poor#c.log_streetv_straightlinedistkm	-0.05
	(0.08)
0b.poor#c.log_mamamboga_straightlinedistkm	-0.10
	(0.06)
1.poor#c.log_mamamboga_straightlinedistkm	-0.01
	(0.08)
0b.poor#c.log_kiosk_straightlinedistkm	0.02
	(0.07)
1.poor#c.log_kiosk_straightlinedistkm	0.03
	(0.04)
0b.poor#c.log_duka_straightlinedistkm	0.19**
	(0.08)
1.poor#c.log_duka_straightlinedistkm	-0.12*
	(0.07)
0b.poor#c.log_smalls_straightlinedistkm	0.09
	(0.12)
1.poor#c.log_smalls_straightlinedistkm	-0.10***
	(0.04)
0b.poor#c.log_hotelrest_straightlinedistkm	0.05
	(0.10)
1.poor#c.log_hotelrest_straightlinedistkm	-0.10***
	(0.03)
0b.poor#c.log_butch_straightlinedistkm	-0.04
	(0.09)
1.poor#c.log_butch_straightlinedistkm	-0.08
	(0.06)
0b.poor#c.log_cer_posho_straightlinedistkm	0.15*
	(0.08)
1.poor#c.log_cer_posho_straightlinedistkm	0.16**
	(0.07)
Household controls	Yes
Main shopper controls	Yes
Location controls	Yes
Observations	1,429
R-squared	0.19

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

Table 20: Regression results for Wald test: transportation via car

	A
0b.car#c.log_infprep_straightlinedistkm	0.03
	(0.06)
1.car#c.log_infprep_straightlinedistkm	0.16
	(0.26)
0b.car#c.log_milkba_straightlinedistkm	0.08**
	(0.03)
1.car#c.log_milkba_straightlinedistkm	-0.21**
	(0.09)
0b.car#c.log_depwhole_straightlinedistkm	0.01
	(0.06)
1.car#c.log_depwhole_straightlinedistkm	0.06
	(0.15)
0b.car#c.log_hawker_straightlinedistkm	-0.07**
	(0.03)
1.car#c.log_hawker_straightlinedistkm	-0.17
	(0.12)
0b.car#c.log_streetv_straightlinedistkm	-0.06
	(0.05)
1.car#c.log_streetv_straightlinedistkm	0.09
	(0.09)
0b.car#c.log_mamamboga_straightlinedistkm	-0.08**
	(0.04)
1.car#c.log_mamamboga_straightlinedistkm	-0.07
	(0.11)
0b.car#c.log_kiosk_straightlinedistkm	0.10**
	(0.05)
1.car#c.log_kiosk_straightlinedistkm	-0.16
	(0.12)
0b.car#c.log_duka_straightlinedistkm	-0.02
	(0.05)
1.car#c.log_duka_straightlinedistkm	-0.08
	(0.10)
0b.car#c.log_smalls_straightlinedistkm	-0.02
	(0.05)
1.car#c.log_smalls_straightlinedistkm	0.22
	(0.25)
0b.car#c.log_hotelrest_straightlinedistkm	-0.13***

Table 20 (cont'd)

	A
	(0.04)
1.car#c.log_hotelrest_straightlinedistkm	-0.42*
	(0.23)
0b.car#c.log_butch_straightlinedistkm	-0.02
	(0.05)
1.car#c.log_butch_straightlinedistkm	0.08
	(0.18)
0b.car#c.log_cer_posho_straightlinedistkm	0.14***
	(0.05)
1.car#c.log_cer_posho_straightlinedistkm	0.04
	(0.16)
Household controls	Yes
Main shopper controls	Yes
Location controls	Yes
Observations	1,429
R-squared	0.16

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX C: ROBUSTNESS CHECK REGRESSIONS

Table 21: Regression results with the log of healthy monthly expenditure per capita as dependent variable with 'unhealthy in excess' added to 'unhealthy'

	A	B	C	D
Log of informal prepared distance	0.10	0.04	0.04	-0.01
	(0.07)	(0.06)	(0.06)	(0.11)
log of millk bar/milk atm distance	0.12***	0.10**	0.13***	-0.13
	(0.04)	(0.04)	(0.04)	(0.08)
log of depot/wholesaler distance	-0.09	-0.05	-0.05	-0.16
	(0.06)	(0.05)	(0.04)	(0.20)
log of hawker distance	-0.07**	-0.05	-0.06**	0.00
	(0.03)	(0.03)	(0.03)	(0.09)
log of street vendor distance	-0.06	-0.07	-0.06	-0.09
	(0.06)	(0.06)	(0.06)	(0.15)
log of mama mboga distance	-0.05	-0.08*	-0.07	0.06
	(0.04)	(0.04)	(0.05)	(0.08)
log of kiosk distance	0.05	0.12**	0.13**	0.16*
	(0.05)	(0.05)	(0.05)	(0.08)
log of duka distance	-0.05	-0.05	-0.07	-0.14*
	(0.06)	(0.06)	(0.06)	(0.08)
log of small supermarket distance	0.04	0.03	0.07	0.10
	(0.05)	(0.06)	(0.05)	(0.17)
log of hotel/restaurant distance	-0.10**	-0.12**	-0.12***	0.01
	(0.05)	(0.05)	(0.04)	(0.09)
log of butchery distance	-0.02	0.02	0.01	-0.05
	(0.06)	(0.05)	(0.05)	(0.09)
log of cereal/poshomill distance	0.07	0.06	0.04	0.20
	(0.05)	(0.05)	(0.04)	(0.16)
Household controls	No	Yes	Yes	Yes
Main shopper controls	No	No	Yes	Yes
Location controls	No	No	No	Yes
Observations	1,434	1,434	1,434	1,434
R-squared	0.03	0.09	0.13	0.19

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

Table 22: Regression results with the log of healthy monthly expenditure per capita as dependent variable with 'unhealthy in excess' excluded

	A	B	C	D
Log of informal prepared distance	0.10	0.04	0.04	-0.01
	(0.07)	(0.06)	(0.06)	(0.11)
log of millk bar/milk atm distance	0.12***	0.10**	0.13***	-0.13
	(0.04)	(0.04)	(0.04)	(0.08)
log of depot/wholesaler distance	-0.09	-0.05	-0.05	-0.16
	(0.06)	(0.05)	(0.04)	(0.20)
log of hawker distance	-0.07**	-0.05	-0.06**	0.00
	(0.03)	(0.03)	(0.03)	(0.09)
log of street vendor distance	-0.06	-0.07	-0.06	-0.09
	(0.06)	(0.06)	(0.06)	(0.15)
log of mama mboga distance	-0.05	-0.08*	-0.07	0.06
	(0.04)	(0.04)	(0.05)	(0.08)
log of kiosk distance	0.05	0.12**	0.13**	0.16*
	(0.05)	(0.05)	(0.05)	(0.08)
log of duka distance	-0.05	-0.05	-0.07	-0.14*
	(0.06)	(0.06)	(0.06)	(0.08)
log of small supermarket distance	0.04	0.03	0.07	0.10
	(0.05)	(0.06)	(0.05)	(0.17)
log of hotel/restaurant distance	-0.10**	-0.12**	-0.12***	0.01
	(0.05)	(0.05)	(0.04)	(0.09)
log of butchery distance	-0.02	0.02	0.01	-0.05
	(0.06)	(0.05)	(0.05)	(0.09)
log of cereal/poshomill distance	0.07	0.06	0.04	0.20
	(0.05)	(0.05)	(0.04)	(0.16)
Household controls	No	Yes	Yes	Yes
Main shopper controls	No	No	Yes	Yes
Location controls	No	No	No	Yes
Observations	1,434	1,434	1,434	1,434
R-squared	0.03	0.09	0.13	0.19

Household sample weights applied

Robust standard errors in parentheses clustered at EA level

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX D: SUPPLEMENTAL DESCRIPTIVE STATISTICS

Table 23: Food expenditure descriptive statistics by food group (KES/month per household)

Food group	Overall (mean, SD, and % of total)	Nairobi u. (mean, SD, and % of total)	Nairobi p.u. (mean, SD, and % of total)	Kisumu u. (mean, SD, and % of total)	Kisumu p.u. (mean, SD, and % of total)
Citrus fruits (healthy)	23.84 (105.41) 0.36%	25.46 (101.56) 0.36%	23.70 (116.84) 0.26%	16.05 (77.69) 0.25%	10.75 (67.90) 0.14%
Deep orange fruits (healthy)	15.22 (103.65) 0.23%	15.31 (111.43) 0.22%	15.09 (91.84) 0.26%	18.81 (122.54) 0.29%	10.05 (54.88) 0.13%
Other fruits (healthy)	88.97 (403.20) 1.36%	104.67 (464.74) 1.49%	71.08 (301.93) 1.23%	73.52 (428.27) 1.13%	73.52 (428.27) 0.96%
Dark green leafy vegetables (healthy)	479.56 (1432.20) 7.31%	555.31 (1778.03) 7.92%	362.62 (459.50) 6.27%	616.91 (2119.22) 9.51%	311.85 (383.72) 4.06%
Cruciferous vegetables (healthy)	106.88 (512.11) 1.63%	89.57 (166.326) 1.28%	144.77 (819.75) 2.50%	49.48 (115.54) 0.76%	64.70 (151.62) 0.84%
Deep orange vegetables (healthy)	10.44 (57.36) 0.16%	9.93 (51.05) 0.14%	13.37 (70.81) 0.23%	1.81 (20.42) 0.028%	0.06 (1.51) 0.00078%
Other vegetables (healthy)	564.86 (1890.68) 8.61%	660.29 (2508.88) 9.42%	442.65 (470.01) 7.66%	458.71 (505.66) 7.07%	484.20 (468.44) 6.31%
Legumes (healthy)	180.04 (466.06) 2.74%	161.56 (500.97) 2.30%	211.82 (384.58) 3.66%	142.51 (503.08) 2.20%	192.51 (594.32) 2.51%
Deep orange tubers (healthy)	0.34 (11.03) 0.0052%	0.59 (14.69) 0.0084%	0 (0.00) 0%	0 (0.00) 0%	0.40 (9.00) 0.0052%
Nuts and seeds (healthy)	8.55 (88.27) 0.13%	13.31 (110.40) 0.19%	1.70 (32.75) 0.029%	4.29 (37.50) 0.066%	11.02 (131.02) 0.14%
Whole grains (healthy)	280.30 (760.44) 4.27%	280.58 (783.40) 4.00%	228.72 (639.39) 3.96%	473.03 (1041.77) 7.29%	545.00 (980.75) 7.10%

Table 23 (cont'd)

Food group	Overall (mean, SD, and % of total)	Nairobi u. (mean, SD, and % of total)	Nairobi p.u. (mean, SD, and % of total)	Kisumu u. (mean, SD, and % of total)	Kisumu p.u. (mean, SD, and % of total)
Liquid oils (healthy)	797.98 (1244.04) 12.16%	853.07 (1133.13) 12.16%	692.95 (1095.99) 11.98%	877.82 (2150.32) 13.53%	914.02 (2219.21) 11.91%
Fish and shellfish (healthy)	194.11 (719.00) 2.96%	212.37 (644.33) 3.03%	77.32 (416.76) 1.34%	518.52 (895.75) 7.99%	672.04 (2198.01) 8.75%
Poultry and game meat (healthy)	15.73 (140.77) 0.24%	19.03 (154.59) 0.27%	12.06 (116.95) 0.21%	16.82 (184.07) 0.26%	0.44 (16.35) 0.0057%
Low fat dairy (healthy)	33.20 (223.41) 0.51%	36.14 (236.42) 0.52%	31.17 (201.37) 0.54%	33.97 (292.78) 0.52%	6.83 (70.39) 0.089%
Eggs (healthy)	117.50 (631.30) 1.79%	165.67 (832.75) 2.36%	46.39 (143.36) 0.80%	103.26 (289.73) 1.59%	120.19 (257.42) 1.57%
High fat dairy (in milk equivalents) (unhealthy in excessive amounts)	604.41 (891.79) 9.21%	640.72 (944.81) 9.14%	602.53 (816.58) 10.42%	400.74 (846.79) 6.18%	335.77 (778.24) 4.37%
Red meat (unhealthy in excessive amounts)	299.45 (821.51) 4.56%	317.36 (763.24) 4.53%	305.49 (936.84) 5.28%	169.28 (725.05) 2.61%	135.87 (453.43) 1.77%
Processed meat (unhealthy)	60.77 (309.66) 0.93%	59.61 (373.23) 0.85%	65.94 (201.62) 1.14%	46.55 (233.47) 0.72%	44.83 (224.61) 0.58%
Refined grains and baked goods (unhealthy)	1897.98 (2867.99) 28.93%	1977.53 (2423.10) 28.20%	1758.32 (1330.45) 30.41%	1514.51 (1748.22) 23.35%	2643.19 (1951.11) 34.43%
Sweets and ice cream (unhealthy)	464.21 (761.54) 7.08%	511.36 (900.69) 7.29%	368.88 (486.50) 6.38%	564.33 (623.46) 8.70%	575.83 (813.15) 7.50%

Table 23 (cont'd)

Food group	Overall (mean, SD, and % of total)	Nairobi u. (mean, SD, and % of total)	Nairobi p.u. (mean, SD, and % of total)	Kisumu u. (mean, SD, and % of total)	Kisumu p.u. (mean, SD, and % of total)
SSBs (sugar sweetened beverages) (unhealthy)	74.09 (387.20) 1.13%	64.08 (202.56) 0.91%	66.11 (213.02) 1.14%	110.17 (267.30) 1.70%	265.59 (1746.55) 3.46%
Juice (unhealthy)	4.27 (52.77) 0.065%	4.01 (47.64) 0.057%	5.05 (61.52) 0.087%	4.44 (56.66) 0.068%	0 (0.00) 0%
White roots and tubers (unhealthy)	126.14 (339.09) 1.92	110.40 (331.64) 1.57%	162.93 (370.01) 2.82%	59.68 (190.85) 0.92%	83.58 (235.95) 1.09%
Purchased deep fried foods (unhealthy)	112.14 (313.34) 1.71%	124.84 (347.02) 1.78%	71.52 (208.89) 1.24%	210.69 (433.73) 3.25%	197.77 (394.59) 2.58%
Total expenditure across all categories per household (KES/mo.)	6560.98 (6471.68)	7012.78 (7072.24)	5782.20 (4416.13)	6485.90 (5950.93)	7676.19 (12183.61)
Sample size n	1496	368	354	382	392

Note: population weights applied

Note: Observations with extreme values greater than 99th percentile are dropped