Determinants of Dietary Choice in the US: Evidence from Consumer Migration

Stefan Hut*
Brown University
March 2019

Abstract
I study the evolution of the quality of grocery purchases among migrants to learn how changes in the environment affect dietary choice. Using detailed household level panel data on food purchases I find that healthfulness of grocery purchases is very persistent in the short-run. Three to four decades after moving, however, households have closed about half of the gap in healthfulness between the origin and destination area. The results suggest that dietary habits are highly persistent, but may eventually shift in the face of different local environments.

Keywords: dietary choice, health, migration, applied microeconomics.

*Brown University Department of Economics. Brown University Population Studies and Training Center. I am very grateful to Emily Oster and Jesse Shapiro for their valuable guidance and feedback. I thank Anna Aizer, Daniel Bjorkegren, Andrew Foster, John Friedman, Ryan Kessler, Nicholas Reynolds, Bryce Steinberg, Matthew Turner and participants in the Brown Applied Microeconomics Seminar and the Health Breakfast for their comments and suggestions. I would like to thank to Bart Bronnenberg, Jesse Shapiro, and AC Nielsen for granting me access to the consumer life history survey, and to Jesse Shapiro for supervising my work with this survey. Researcher(s) own analyses calculated (or derived) based on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.
1 Introduction

One-in-three adult Americans are obese, and another one-third is considered overweight\(^1\). Obesity is among the top three preventable causes of death in the US (Danaei et al., 2009) and medical costs associated with adult obesity were estimated at $147bn in 2008 (Finkelstein et al., 2008). Poor diet is shown to be an important driver of differences in obesity rates (Cutler et al., 2003, Park et al., 2015) and significant funds have been allocated to promoting healthy diets.

Evidence to date suggests that dietary choices are generally difficult to change, however. Diets appear persistent both in the face of changes in the local environment or food options (Atkin, 2016; Alcott et al., 2017) and in response to changes in household-specific circumstances (Oster, 2017; Hut and Oster, 2018). Most of the evidence is based on specific stimuli or changes in food options, however, and has been limited to a relatively short time-horizon of a few years.

In this paper I study the broader impact of the local environment on dietary choice - foods purchased for at home consumption in particular - over both a short- and long time horizon. To assess how changes in the environment shape nutritional quality I document changes in grocery purchases among households that move within the US. The results confirm earlier findings that dietary choice is sticky, but also provides suggestive evidence that the quality of grocery purchases may be more malleable over the course of several decades.

In order to measure healthfulness of household grocery purchases over time I combine the Nielsen HomeScan panel, a dataset of household food purchases, with product level nutrition facts. The Nielsen data contain Universal Product Code (UPC) level grocery purchases for a large number of households across the US over time, making it uniquely suited to tracking household food purchases over time and across geography. I construct four measures of healthfulness of grocery purchases: (1) the nutrient ratio, which is a composite measure based on the nutrients in all purchases; (2) the expenditure score, a composite measure based on food group spending shares; (3) the share of purchases in soda and sugar-based categories and (4) the share of purchases in vegetables, fruits and whole grains.

To measure healthfulness of grocery spending at the area-level I construct the same measures by ZIP code and state based on grocery store sales in that area in the Nielsen

\(^1\)https://www.cdc.gov/obesity/data/adult.html
Scantrack data. These data contain weekly sales and prices by UPC for over 30,000 retailers in the US. The measures confirm that the nutrient content of foods purchased for at home consumption varies substantially across geography, in a way that correlates highly with reported obesity rates.

The movers analysis consists of two parts. In the primary analysis I examine relatively short-run changes in the healthfulness of grocery purchases for households that moved within the time frame of the HomeScan panel. The key advantage of this set of recent movers is that I observe their purchases before and after a move, which allows me to study within-household changes in the nutrient content of grocery purchases as a function of the change in area-level healthfulness. Second, I study the evolution of food purchases among households that have been in the destination area for a longer period of time. In this part of the analysis I rely on survey data on current and past states of residence for a subset of the Nielsen HomeScan panelists from Bronnenberg et al. (2012). Together with information on the years since the move, I use these movers to study how current food purchases relate to the average food purchases in the origin and destination over the course of several decades. This part of the analysis relies on stronger assumptions, which I explicitly discuss and examine in the paper. Given these limitations, however, the long-run analysis at best provides suggestive evidence of how the quality of grocery purchases might evolve over a longer period of time.

The short-run analysis of recent movers shows that healthfulness of household grocery purchases is virtually unchanged in the first two years following a move. I identify recent movers in the panel and infer the moving month from changes in household residential ZIP codes and their primary store’s ZIP code. I use these movers to study how grocery purchases change around a move using an event-study design with household fixed effects. This closely follows the approach taken in other papers that examine consumer migration, most notably Finkelstein et al. (2016). The results demonstrate that the quality of household grocery purchases is sticky in the short-run: at most 3 percent of the gap in healthfulness between the origin and destination is closed in the first 24 months following a move. To illustrate how movers help to deal with issues of endogenous sorting in this context, I show that a naïve cross-sectional analysis of a household’s own food purchases on their neighbors’ food purchases would gravely overestimate the effect of place in the short-run.

The key identifying assumption for the short-run analysis is that the timing of a household’s move does not coincide exactly with other shocks that are correlated with both the
healthfulness of a household’s own grocery purchases and with the difference in healthfulness between the origin and destination. Income changes or changes in household composition are the most obvious candidates for such a shock. I show that while a subset of moves coincide with household income shocks, income changes are not predictive of the change in area level healthfulness. In addition, move years are more likely than non-move years to coincide with household composition changes, but the results are robust to limiting to households who do not show such a change. Furthermore, I argue that any other unobserved changes in lifestyle or household characteristics would likely generate an upward bias in the estimated effects of moving, which seems less concerning given that I do not find a change in the quality of grocery purchases in the short run.

The long-run analysis of movers, who moved before entering the panel, shows that over time the quality of food purchases converges about halfway to the average in the destination area. For this part of the analysis I rely on within-household variation in the years since the move to study how healthfulness of food purchases evolves over several decades, relative to the origin and destination state. This is similar to the approach in Bronnenberg et al.’s (2012) study of brand preferences. The results show that migrants that have been in their destination state for three to four decades on average have closed about 50% of the gap in healthfulness between the origin and destination state.

Because I do not observe grocery purchases prior to the move for this set of movers, these results should be interpreted with more caution. A key assumption underlying the long-run analysis is that households do not selectively migrate and re-migrate. While the use of within-household (as opposed to cross-sectional) variation addresses some of these concerns, the results could be biased if the decision to relocate is correlated with a household’s changes in food purchase behavior over time. For example it is possible that households that decide to stay in an area are more open to changing their behaviors, including the quality of their food purchases. Such selection would imply that the estimates of food purchase changes in the data would overstate the change we would expect to see in the general population.

A second limitation of the long-run analysis is that I do not observe diet in the origin state at the time of the move. While not perfect, I use current healthfulness of grocery purchases in the origin state as a proxy for past healthfulness. I test this assumption by comparing area-healthfulness in 2006 and 2016 in my data. In expectation past area-level healthfulness equals today’s area-level healthfulness, suggesting geographic patterns in diet
quality have been relatively stable over time. In addition, I show that state-level obesity rates between 1990 and today have remained stable in relative terms.

In the last section I discuss the implications of these results for the potential drivers underlying the observed geographic variation in the quality of food purchases. The results rule out two potential drivers, including large and fixed differences in inherent food preferences across space and short-run constraints imposed by the retail environment. Since the store environment changes discretely around the move, the results appear inconsistent with factors such as local product availability being immediate constraints for the healthfulness of grocery purchases. This is further exemplified by the absence of a short-run change in the quality of purchases upon moving to or from areas defined by the USDA as food deserts. Instead, the patterns of change suggest that dietary habits shift slowly over time. This suggests that individuals have food habits and preferences that are persistent, but that over time such habits are susceptible to changes in the local environment. These changes may be driven by neighbors or peers, through social interactions, or by long-run effects of the retail environment or product marketing. Future empirical work could attempt to examine these channels more directly by relying on social network data.

This paper relates to a literature on drivers of metabolic habits, and consumer dietary intake more specifically. In particular, my paper relates closely to Allcott et al. (2017), which shows that food availability is not a primary driver of the observed variation in diet quality. Allcott and et al. (2017) also use movers to examine the effect of place on household diet in the short-run. Similar to my results, they find that household diet is virtually unchanged in the short-run. I build on this by examining long-run changes in diet following a move, and I provide suggestive evidence that household diet appears more malleable over the course of several decades.

Second, this paper relates to a literature on the persistence of food preferences. My paper is most closely related to Bronnenberg et al. (2012), which documents strong evidence of persistent brand preferences among migrants. I use the same scanner data and measurement framework. While Bronnenberg et al. (2012) study migrants primarily to examine how past experiences affect current brand choices, I exploit migration to study how diet responds to a change in one’s environment. In addition, I explicitly examine the costs of persistent food preferences by studying nutrition and related health outcomes. Atkin (2016) studies nutrition among migrants in India, and also finds strong persistence of diet by showing that
people are willing to give up valuable nutrients to maintain their preferred diet. I add to this by studying dynamics through changes in diet as a function of the amount of time spent in the destination, and also by studying the US context where diet quality issues are different from those in India.

Section 2 introduces the scanner and nutrition data and describes the measures of nutritional quality of grocery purchases used in this paper. I also show descriptive facts of the geographic variation in nutritional quality across space. Section 3 outlines how I identify movers in the data and the empirical strategy for both short- and long-run movers. Section 4 presents the results. Section 5 discusses several potential underlying drivers of the observed changes in grocery spending and related policy implications. Section 6 concludes.

2 Data

2.1 Consumer Purchase Data

The main data set I use in this paper is the Nielsen HomeScan Panel. The data was made available through the Kilts Center at the University of Chicago Booth School of Business. It contains transaction-level purchase records for 158,830 households between 2004 and 2015. Panelist households have an optical scanner at home and are directed to scan all items with a bar code that they purchase, regardless of the store. The data therefore contains purchases from a wide range of stores, including grocery stores, super-centers, convenience stores, drug stores, and so on. Purchases are recorded at the Universal Product Code (UPC) level, and the data set includes quantities purchased as well as prices that are recorded by the panelists or drawn from Nielsen store-level data, where available.

In addition to purchase data, the HomeScan data set contains annual household demographic characteristics and a limited set of store characteristics. Demographic variables recorded include household income, household head age and employment, household composition, race, and the ZIP code of residence. For each trip, the data set contains an ID for the store and store chain at which the purchases were made, as well as the 3 digit ZIP code of the store. The residential and store ZIP codes are particularly important in the context of this paper. As I explain in Section 3.1 I use both of these variables to identify movers in my panel as well as the month in which households move.
The panel structure and the level of detail and broad coverage in terms of both food products and stores make the Nielsen data well-suited to studying the evolution of household grocery purchases over time. The Nielsen data is set up to be a representative sample of the US in terms of basic demographics. Appendix Table A.1 shows basic demographics for the sample used here relative to the US overall. The Nielsen sample is very similar on most demographics, although the racial composition is more heavily white. The sample is roughly representative of the US population in terms of movers as well. Approximately 32% of household heads in the Nielsen life history survey had moved out of their birth state, compared to about 36% in the US as a whole.

Data Limitations

There are a number of limitations to the data.

First, the data only contain at home food purchases. The results in this paper can therefore only shed light on the effect of the local environment on food purchases for at-home consumption. According to the National Health and Nutrition Examination Survey (NHANES), Americans on average consume approximately 34% of their calories away from home, with 25% being in restaurants\(^2\). The majority of calories consumed would therefore be captured in this paper. However, the results cannot be generalized to overall diet unless the share of calories consumed away from home remains constant around the move, and the nutrient content of calories consumed away from home changes in the same manner as that of calories consumed at home.

Second, I only observe purchases and not consumption, and it is likely that there is at least some wastage. One way that I address this limitation is to base the measures of healthfulness of grocery purchases (detailed in section 2.2) on nutrients per calorie, rather than total calories or nutrients. As long as wastage rates do not vary by the healthfulness of food items, healthfulness of purchased foods will equal healthfulness of consumed foods.

Third, coverage of fresh produce is limited to items with a UPC bar code. A subset of households also record items that do not have a standard UPC as “random weight” items. I do not use these items in the main analysis for two reasons. First, restricting only to the subset of households that record such items reduces the number of migrants identified in the panel substantially. Second, many of these categories are too broad to infer meaningful nutritional

information. Other papers looking at nutritional outcomes using this data, including Allcott et al. (2017), exclude random weight items from their analysis for similar reasons.

2.2 Measuring Healthfulness of Grocery Purchases

In order to measure healthfulness of food purchases I combine the Nielsen purchase data with UPC-level nutrition facts. The nutrition data includes the USDA Branded Food Products Database and the USDA National Nutrient Database for Standard Reference. I supplement these data with nutrition facts for food items on Shopwell.com, Walmart.com, and Labelinsight.com. Altogether, the data provides a direct UPC match for approximately 45% of UPCs and 75% of sales in the data.

For the remaining items that do not have a direct match with the nutrition data, I undertake an imputation procedure similar to that outlined in Dubois, et al. (2014). This procedure has two rounds. First, I impute the nutrition values using the average within product module, size type, brand, flavor, and formula (as defined by Nielsen). Many of the items that remain unmatched are store brand items. For these items I take the average within product module, size type, variety, type, formula, and style (i.e., drop the brand requirement). After imputation, approximately 7.6% of the purchases in the data remain unmatched. These items are not included in the calculation of the nutrition based outcome measure.

The primary measure of the healthfulness of grocery spending used in the paper is called the nutrient ratio. This is an index measure of healthfulness that captures the extent to which a household’s grocery purchases deviate from the nutrient composition recommended in the federal Dietary Guidelines for Americans (DGA). For each household \( h \) in month \( t \), the nutrient ratio is defined as:

\[
Nutrient Ratio_{ht} = \frac{\sum_{j \in J_{\text{Healthful}}} (pc_{jht} \cdot pc_{DGA}^j)}{\sum_{j \in J_{\text{Unhealthful}}} (pc_{jht} \cdot pc_{DGA}^j)}
\]  

(1)

where \( j \) indexes nutrients, \( pc_{jht} \) denotes the amount of nutrient \( j \) per calorie in household \( h \)’s grocery purchases in month \( t \), and \( pc_{DGA}^j \) is the amount of nutrient \( j \) in the DGA recom-
mended diet per calorie consumed\(^4\).

This index measure of nutritional quality has been commonly used in the public health literature (Drewnowski 2005, Drewnowski 2010). Each of the the ratios \( \frac{p_{j,ht}}{p_{j,DGA}} \) for nutrient \( j \) represents the Nutritional Quality Index (NQI) for that nutrient. To construct a comprehensive measure of nutritional quality I combine the NQIs into a single index by taking the ratio of the sum of healthful to unhealthful NQIs. Nutrients in the unhealthful category include those for which the DGA recommends an upper bound (total fat, saturated fat, sodium, and cholesterol) and nutrients in the healthful category include those for which the DGA recommends a lower bound (fiber, iron, calcium, vitamin A, and vitamin C). A higher ratio implies a higher quality diet.

A limitation of the nutrient ratio is that the nutrition data that it relies on has a lower coverage rate of UPCs in the initial years of the HomeScan panel. This is caused by the fact that there is substantial turnover in UPCs and products over time. Because the nutrition data was collected in 2017, coverage worsens the further I go back in time. Appendix Table A.2 shows coverage as a share of sales and UPCs over time.

In addition to the nutrient ratio I construct three alternative measures of healthfulness of grocery spending. These measures are all based on the share of total food spending on groups defined in the USDA “Thrifty Food Plan” (TFP). The TFP is one of four USDA-designed food plans specifying foods and amounts of foods that provide adequate nutrition. This approach to measuring nutritional quality has been used in related literature (Volpe et al., 2013; Oster, 2017; Hut and Oster, 2018).

I generate the share of food expenditures on each of the 24 TFP food categories\(^5\) and use these spending shares to construct three measures. First, I compute the share of total spending going to obvious unhealthful food groups, including soft drinks, soda, fruit drinks and ades, and sugar, sweets, and candy. Higher values imply a less healthy diet. Second, I similarly calculate the total share of spending going to obvious healthful food groups, which include vegetables, fruits and whole grains. Higher values imply a healthier diet. Third, I use spending shares on all 24 categories to construct an index of nutritional quality of food spending called the expenditure score. The expenditure score measures the extent to

\(^4\)The recommendations can be found here: https://www.fda.gov/Food/IngredientsPackagingLabeling/LabelingNutrition/

\(^5\)The TFP categories are based on the Quarterly Food-at-Home Price Database (QFAHPD) classification. I thank Jessica Todd and Ilya Rahkovsky at the USDA for sharing their UPC to QFAHPD category concordance. See Todd et al. (2010) for more information on the methodology behind the QFAHPD.
which a household’s grocery purchases deviate from the expenditure shares recommended by the USDA Center for Nutrition Policy and Promotion’s Thrifty Food Plan. A higher expenditure score corresponds to a more healthful diet. For more details on the construction of this score and the division of TFP food groups into unhealthy and healthy categories, see Appendix B. Furthermore, correlations between all four measures of healthfulness can be found in Appendix B3.

2.3 Average Area-Level Diet Quality: Data and Stylized Facts

Grocery Store Sales Data

In order to examine how the quality of grocery spending changes upon moving to a more or less healthy place, I need to obtain a measure of the typical healthfulness in each area in the US. For this purpose I construct the same measures of the healthfulness of grocery spending described above by ZIP code and state using grocery store sales in the Nielsen Retail Measurement Services (RMS) data. These data were made available through the Kilts Center at the University of Chicago Booth School of Business. The data set contains weekly sales and quantities by UPC for approximately 38,000 participating retailers in over 100 retail chains across the US. Altogether, the data covers about 40% of all US grocery purchases. I use these data combined with the nutrition facts to construct the nutrient ratio, the expenditure score, the share of spending going to healthful foods, and the share of spending going to unhealthful foods at the area level.

Other Area-Level Healthfulness Data

I merge the Nielsen data with two other data sets. First, I use data on county-level obesity rates from the CDC Behavioral Risk Factor Surveillance System (BRFSS) as an alternative measure of area-level healthfulness of diet. BRFSS data are self-reported, however, which may lead to underreporting of weight and hence BMI. Studies comparing BRFSS to NHANES (which relies on physical examinations) have noted that prevalence of obesity and overweight is underreported by between 5.7 and 9.5 percentage points (Pierannunzi et al., 2013). Second, I merge the HomeScan with data from the USDA on “food deserts” by ZIP code, which are

---

6These data are made available as part of the Health Inequality project: https://healthinequality.org/data/.
defined as low income census tracts more than 1 (10) miles away from a supermarket in urban (rural) areas.

**Area-Level Healthfulness of Grocery Purchases: Stylized Facts**

Figure 1a shows the area-level nutrient ratio for each 3 digit ZIP-code in the US. There is substantial variation in nutritional quality of grocery purchases across the US. Areas with the highest nutritional quality of grocery sales include the the District of Columbia, Oregon, Colorado, Washington, and California. Conversely, the lowest nutritional quality can be found in Louisiana, Mississippi, Arkansas, Kentucky, and Alabama.

These patterns line up closely with observed obesity rates in the US. Figure 1b shows a scatter plot between the nutrient ratio and obesity rates at the state level. The correlation between these two measures is -0.74. At the 3 digit ZIP code level, this correlation is -0.39.\(^7\) While the area-level nutrient ratio is not a perfect measure, it is clear that it captures broad patterns of area-level nutritional quality and related obesity rates in the US. A literature in both economics and public health has similarly established a close relationship between changes in dietary intake and obesity rates over time. Cutler et al. (2003) find that during the rise of obesity rates since 1980, calories expended did not change significantly while calories consumed increased considerably. Bleich et al. (2008) and Swinburn et al. (2009) confirm the key importance of dietary intake in explaining the rise in obesity rates in the US. The findings shown here demonstrate that dietary intake and obesity rates are also strongly correlated across space.

Importantly, the geographic variation in healthfulness of grocery purchases explored here captures more than just spatial differences in observable individual and place characteristics, including income. The R-squared of a regression of the area-level nutrient ratio on a quadratic in median household income is 0.08, highlighting that there is substantial geographic variation in nutritional quality not explained by income (see Appendix table A.3). Furthermore, using the HomeScan panelists, Figure A.1 shows the coefficients of a regression of the household nutrient ratio on region dummies, both with and without controls for in-

---

\(^7\)This is likely lower in part because I observe obesity rates from the BRFSS at the county-level, and counties are not mapped 1:1 to 3 digit ZIP codes. I use a ZIP code to county crosswalk provided by the HUD, but this inevitably introduces measurement error. These correlations do not seem unreasonable given that obesity rates are also measured with error, and that this is correlating the mean nutritional quality of grocery purchases with the rate of adult obesity at an aggregate level.
come, education, household size, and race. There is substantial variation across space, even conditional on these controls.

In what follows, I examine how the nutritional quality of household grocery purchases evolves upon moving within the US, depending on the change in area-level healthfulness experienced when moving.

3 Empirical Strategy

Having established that there is substantial variation in the healthfulness of grocery purchases across space I examine how a household’s food purchases change upon moving to a more or less healthy place. I identify two sets of movers in the Nielsen data. The primary set of movers consists of households that moved while they were a part of the Nielsen panel. I use store and residential ZIP code changes to identify such households. The key advantage of this set of recent movers is that I observe their diet before and after a move. Second, I use a survey of consumer life histories to identify a set of movers that moved prior to entering the panel and have been in the destination area for a longer period of time. This part of the analysis relies on stronger assumptions, which I explicitly discuss and examine in this section.

3.1 Identifying movers

3.1.1 Identifying Recent Movers

I identify movers in the panel using household and store locations. Movers in the HomeScan panel are defined as households whose residential 3 digit ZIP code changed\(^8\) during the panel time frame\(^9\). There are a total of 6,271 households in the data that moved across 3 digit ZIP codes between 2004 and 2015. Table 1 contains summary statistics for non-movers (column 1) and the short-run movers (column 2). The groups are similar across most dimensions. The panelist movers are slightly more likely to belong to the top income group relative to non-movers, and have a lower mean household size.

---

\(^8\)Because I only observe store locations at the 3 digit ZIP code level in the Scanntrack data, and hence can only create area-level healthfulness measures at that level, I restrict to moves across 3 digit ZIP codes.

\(^9\)I restrict to households whose residential ZIP code changed at most once. The relatively short time frame of the HomeScan panel would otherwise make it hard to identify clear pre- and post-move periods.
Figure 2 shows the distribution of the change in area-level healthfulness, as measured by the nutrient ratio, experienced by the migrating households. The mean is close to zero and the distribution is approximately symmetric, meaning moves to healthier areas are about as common as moves to unhealthier areas. A subset of households experience substantial changes in area level healthfulness, keeping in mind the mean area-level nutrient ratio of 0.59 and standard deviation of 0.1.

Having identified who moves, I use changes in store ZIP codes to infer the month within which households move. I define the moving month to be the first month in which the household made a purchase at a store in the new residential ZIP code. Figure 3a plots the share of shopping trips in the origin and destination ZIP code in each month relative to the inferred move month. The figure shows that once people move they do the majority of their shopping in the new ZIP code. Prior to the move about 90% of recorded trips occurred in stores in the origin ZIP code. The month after the move households switch to purchasing in the destination ZIP code, which accounts for over 80% of trips\(^\text{10}\).

Figure 3b plots the number of shopping trips recorded in each month relative to the inferred move month. There is a clear dip in the number of shopping trips in the move month as well as a decline in the number of trips in the months leading up to the move. After the move, however, households appear to return to their previous scanning behavior.

The decline in shopping trips in Figure 3b likely has two reasons. First, households will start to prepare for the move. Households may have less time to cook, and the dip may reflect actual reductions in grocery store purchases. Alternatively people may scan their products less consistently. These are not problematic for my analysis as long as households do not change purchasing and scanning behavior differentially across food groups. This is a limitation I am unable to address. In addition, it should be less of a concern when comparing grocery purchases a year before and after the inferred move month. Second, some of the store ZIP codes are missing in the Nielsen data. This likely leads to cases where the actual move takes place slightly before the identified moving month. This too may explain part of the decline in the average number of trips in the months leading up to the identified move month. This is likely not a major issue, however. While the store ZIP code is missing for almost 50% of recorded trips in the Nielsen data, because households on average record almost 20

\(^{10}\)Note that an important share of the remaining 20% of trips are accounted for by moves across neighboring ZIP codes. The share of purchases in the destination ZIP code is over 90% when considering cross-state moves.
trips a month only 7% of household-months have no recorded store ZIP codes at all.

Note that because of these challenges with identifying the precise move month for each household I will conduct the analysis at the quarterly level. Monthly, bi-annual and annual figures are included in Appendix A.

3.1.2 Identifying Long-Run Movers

In order to examine the long-run evolution of diet I merge the HomeScan with a survey of panelist life histories. The survey was administered in 2008 as part of Nielsen’s monthly PanelViews survey. The questionnaire asked individuals in each household about their birth state and current state of residence. Respondents who no longer lived in their birth state were asked to report the age at which they moved out of the birth state and the number of years since the move. In addition, the questionnaire asked respondents to identify the head of the household and the household’s primary shopper. The survey was sent to 75,221 households. A total of 80,077 individuals in 48,951 households responded, implying a response rate of 65 percent. The surveys were completed between September 13, 2008, and October 1, 2008.

Given that the HomeScan data records purchases at the household level, in order to examine long-run effects I need to select one individual in each household whose life history I use in the analysis. In the rest of this paper, I call this individual the household’s primary shopper. I follow the selection procedure outlined in Bronnenberg et al. (2012). I first restrict to individuals born in the US. Then, I apply the following criteria until a single individual is left in a household: (1) keep only the primary shopper(s) if at least one exists, (2) keep only the household head(s) if at least one exists, (3) keep only the female household head if there exists both a male and female household head, (4) keep the oldest individual, (5) drop responses that appear to be duplicate responses by the same individual, (6) select one respondent randomly. Table A.8 shows the number of primary shoppers identified in each stage. Stage 1 identifies 92% of the primary shoppers in the household. The results are robust to restricting only to this subset of households.

I use the reported birth date to compute an individual’s age. Again following Bronnenberg et al. (2012), I restrict to individuals whose gap between their age and the sum of the number of years lived in the birth state and the current state does not exceed 5 years. In addition, for those cases where the gap is negative (i.e. the total years lived in the birth and current state exceed the respondent’s age), I either recode the number of years lived in the current
state to be the difference between their age and the years lived in the birth state if the gap does not exceed 2 years, or drop the household if the gap is more than 2 years. In addition, I exclude individuals whose age is below 18 or above 99.

My final sample consists of 38,059 households, of which 10,142 households are movers.

Table 1 contains summary statistics for non-movers (column 1) and life history survey movers (column 3). The groups are broadly similar, although the movers have slightly higher household income and total food spending. In addition, movers spend slightly less of their food budget on soda, and slightly more on fruits and vegetables.

Table 2 shows the migration patterns in this sample. About 40 percent of the moves are across states within the same census region. Across regions, we mainly observe moves from the Northeast and Midwest to the South and West of the US. Figure 5 shows the distribution of the respondents’ current age, age at move, and years since move. There is substantial variation across all three dimensions, making the sample well-suited to studying the long-run evolution of the nutritional quality of food purchases. Figure 6 shows the distribution of the change in area-level healthfulness, as measured by the nutrient ratio, experienced by the migrating households.

Limitations

A key limitation of the set of short-run movers is that these households are required to continue to use the Nielsen scanning device around the move. If certain households are more or less likely to continue to scan than others, this may affect the results. I use the set of long-run movers to assess whether particular types of moves, or moves by particular types of households, are especially likely to be lost in the Nielsen panel. The long-run mover sample does not suffer from this problem since these moves are captured retrospectively\textsuperscript{11}.

As shown in Table 1 the short-run and long-run mover samples are similar on most demographics. Long-run movers are slightly more likely to be high income households, and have somewhat higher total food spending. This suggests that attrition among short-run movers might be more concentrated among high-income households\textsuperscript{12}.

\textsuperscript{11}Note also that the rate of migration in the Nielsen life history survey is similar to that in the US as a whole. Approximately 32\% of the Nielsen households have a household head that moved out of their birth state, compared to 36\% of individuals in the US census.

\textsuperscript{12}Though this difference could also be generated by differences in the frequency of moving between low- and high-income households, since the short-run set of movers only captures households who have recently moved.
Table 2 compares the origin and destination region for the set of short-run (panel A) and long-run movers (panel B). Short-run moves are slightly more likely to occur within-region (45%) compared to long-run moves (40%). In addition coast-to-coast moves (North East to West or vice versa), while relatively uncommon in both sets of moves, are more common among long-run movers (4.8%) compared to short-run movers (2.7%). These results suggest that some longer distance moves may be particularly likely to be lost in the set of short-run movers based on the Nielsen panel.

I discuss further issues of selective migration for both sets of movers in the respective empirical strategy sections (3.3 for short-run and 3.4 for long-run moves).

3.2 Measurement Approach

To examine changes in nutritional quality of grocery purchases when households move, I follow a measurement approach outlined in Bronnenberg et al. (2012) and Finkelstein et al. (2016). Define for mover \( h \):

\[
y_{ht}^{scaled} = \frac{y_{ht} - \bar{y}_o(h)}{\bar{y}_d(h) - \bar{y}_o(h)}
\]

where \( y_{ht} \) is a household \( h \)'s nutritional outcome in month \( t \), and \( \bar{y}_o(h) \) and \( \bar{y}_d(h) \) are the mean nutritional outcomes in the origin and destination area, respectively.

This is a useful and intuitive measure because it serves as a summary of how migrants’ nutritional quality compares to the average in the origin and destination. If a migrant household’s nutritional quality is identical to the average in the origin, \( y_{ht}^{scaled} \) equals zero. Conversely, if a migrant household’s nutritional quality is identical to the average in the destination, \( y_{ht}^{scaled} \) equals one. As such, this measurement approach scales the outcome variable such that the direction and magnitude of the change in nutritional quality following a move are informative regardless of the origin and destination.

Following Bronnenberg et al. (2012), I model a household’s healthfulness of grocery purchases as a function of the origin and destination average, depending on the time (years or quarters) since the move, \( r_{ht} \):

\[
y_{ht} = f(r_{ht})\bar{y}_d(h) + (1 - f(r_{ht}))\bar{y}_o(h) + \eta_{ht}
\]

where \( \eta_{ht} \) is an error term that is i.i.d. mean zero conditional on \( r \).
Combining (2) and (3) and re-arranging yields:

\[ y_{ht}^{scaled} = f(r_{ht}) + \frac{\eta_{ht}}{\bar{y}_d(h) - \bar{y}_o(h)} \]  

(4)

This is the main estimating equation, where \( f() \) depends on the exact specification (see below). I estimate equation (4) using weighted least squares\(^{13}\).

### 3.3 Short-Run Event Studies

The main analysis uses an event study design to examine relatively short-run changes in food purchases for movers identified in the HomeScan panel. I estimate a version of equation (4), parametrizing \( f \) with dummies for \( r_{ht} \) pooled in one-quarter bins. I also include quarter-year fixed effects and include household fixed effects to estimate within household changes. The event study equation for the short-run analysis is:

\[ y_{ht}^{scaled} = \alpha_h + \tau_t + \theta_{r(h,t)} \cdot 1_{r(h,t)} + \epsilon_{ht} \]  

(5)

where \( \alpha_h \) are household fixed effects and \( \tau_t \) are quarter-year fixed effects. The coefficients of interest are given by the vector \( \theta_{r(h,t)} \), which represent the change in \( y_{ht}^{scaled} \) in the quarters relative to the move. These coefficients have a simple interpretation: they represent the share of the gap between the origin and destination that the household closes in the quarters around the move.

I use the nutrient ratio as the main measure of the quality of food purchases in this analysis. In addition, I run several alternative event study regressions. First, as alternative measures of nutritional quality, I examine changes in the expenditure score, and the share of quarterly food spending on healthy and unhealthy food categories upon moving. Second, as a robustness check, I use county-level obesity rates as an alternative measure of area-level healthfulness. In the latter case I follow Finkelstein et al. (2016) by estimating an event study equation that interacts quarters relative to move with the change in in area level healthfulness\(^{14}\), as follows:

\(^{13}\)The weights are \( (\bar{y}_d(h) - \bar{y}_o(h))^2 \), which is inversely proportional to the variance of the error term. Intuitively, this weighting approach assigns the highest weight to moves with more substantial changes in area-level healthfulness.

\(^{14}\)Finkelstein et al. (2016) study health care utilization and interact year relative to move dummies with the destination - origin change in average utilization rates.
\[ y_{ht} = \alpha_h + \tau_t + \theta_{r(h,t)} \cdot \delta_h + \epsilon_{ht} \tag{6} \]

Depending on the specification, \( y_{ht} \) is a household \( h \)'s nutrient ratio, expenditure score, or share of food spending on a given food category. \( \delta_h \) represents the destination-origin area-level healthfulness change, which is either the change in ZIP code healthfulness, or the change in county-level obesity rates. The regression also includes household fixed effects, \( \alpha_h \), and quarter-year fixed effects, \( \tau_t \). The coefficients of interest are \( \theta_{r(h,t)} \) which measure the change in \( y_{ht} \) in quarters around the move scaled relative to \( \delta_h \).

Key assumptions and limitations

Two key assumptions underlie the analysis.

The first assumption is that there are no shocks to the nutritional quality of grocery purchases that coincide exactly with the timing of the move and that are correlated with the change in area-level healthfulness between the origin and destination. This would be violated for example if households move to a healthier area for the purpose of adopting a healthier lifestyle. Unless such a lifestyle change occurs very suddenly, one would expect this to show up as pre-trends in the event study analysis. As shown in section 4.1, I find no evidence of such pre-trends.

The second assumption is that the timing of the move does not coincide with other changes that are correlated both with a household’s own nutritional quality as well as origin and destination area-level healthfulness. The most obvious examples of violations of this assumption are income shocks and changes in household composition. I examine each in turn.

Income shocks would be problematic if a positive income shock coincides with moves to healthier areas, and income itself positively affects nutritional quality. While a subset of moves coincide with income changes, table A.6 shows that there is no correlation between income changes and the change in area-level healthfulness. In addition, I do not find sig-

\(^{15}\)Note that while not exactly equivalent, empirically a specification of equation (6) in which \( y_{ht} \) is the household’s nutrient ratio and \( \delta_h \) is the destination-origin ZIP code nutrient ratio yields coefficients that are very similar to the event study results produced using equation (5) when \( y_{ht}^{scaled} \) is the scaled nutrient ratio.

\(^{16}\)In addition, I also find that household size changes surrounding the move are not correlated with the change in area-level healthfulness.
significant changes in total expenditure around the move. If moves to healthier areas coincided with a positive income shock then we might expect total food spending to increase as well. Figure A.2 shows that a one-unit increase in the area-level nutrient ratio is associated with at most a $3 increase in monthly household spending. For the median mover that moves to a healthy area this would imply an increase in spending of about $0.30 a month, which is negligible given that average monthly food spending among movers is about $170 per month (see Table 1).

Moves may also coincide with other life events, such as marriage, divorce, a child leaving the household, or changes employment including job loss, retirement, etcetera. Nielsen panelists report their household size and composition, as well as marital status, on an annual basis. Table A.7 examines whether the identified move year coincides with such life events, and compares this to a randomly picked year for non-movers. The table shows that move years are more than twice as likely to coincide with life events. This is not unexpected: people move for a reason. However, I show in Figure A.3 that changes in the quality of grocery purchases are robust to restricting to moves that do not coincide with such events. To the extent that I observe all relevant life events this will address the concern.

To the extent that there is any other endogeneity of moving that is unobserved, I argue that this would likely generate an upward bias in the estimated place effects. Unobserved changes in lifestyle or household characteristics that would cause people to move to healthier places would likely also lead them to purchase healthier foods. The estimated effect would therefore be an upper bound on the true effect of place on diet quality. The fact that I do not find statistically or economically significant changes in household diet the first two years after the move (see section 4.1) makes this less of a concern.

### 3.4 Long-Run Analysis

The second part of the analysis considers the long-run evolution of grocery purchases relative to the origin and destination area. I estimate equation (4) from section 3.1 parametrizing $f$ with dummies for $r_{ht}$ pooled in annual bins. I follow the short-run analysis and use the panel structure of the data to examine within-household changes in diet. Unlike the short-run analysis I do not observe all years relative to the move for a given household, however.

---

17 This is not unexpected given the result in Hut and Oster (2019) that household diet does not change significantly around these changes in household composition.
Instead, I estimate year-since-move effects for each household for the years that I do observe them in the panel, and combine these estimates to construct an estimate of the overall within-household evolution of diet over time. To see the intuition behind this approach, consider that the life history survey contains the years since the move for each migrant $h$ in 2008, denote this $r(h, 2008)$. I observe most migrant households for multiple years in the panel, in which case I also observe their nutritional quality in other relative year(s) $r(h, t), t \neq 2008$. I use this to estimate the within household change in nutritional quality over time. The estimating equation is therefore:

$$y_{ht}^{scaled} = \alpha_h + \tau_t + \theta_{r(h,t)} \cdot 1_{r(h,t)} + \epsilon_{ht} \quad (7)$$

where $\alpha_h$ are household fixed effects, $\tau_t$ are year fixed effects, and $\theta_{r(h,t)}$ are the coefficients on indicator variables for the years since the move.

**Key assumptions and limitations**

In order for the changes in $y_{ht}^{scaled}$ across the years since the move – captured by $\theta_{r(h,t)}$ – to be informative about a given migrant’s evolution of diet quality over time, two assumptions have to be met.

The first is that any selection into migration and re-migration is constant over time. In other words, grocery purchases of households who migrated $T$ years ago in $X$ years should look like grocery purchases of households who migrated $T + X$ years ago. The results could be biased if the decision to relocate is correlated with a household’s changes in food purchase behavior over time. For example it is possible that households that decide to stay in an area for a longer period of time are more open to changing their behaviors, including the quality of their food purchases. Such selection would imply that the estimates of food purchase changes in the data would overstate the change we would expect to see in the general population.

The second assumption is that in expectation, past area-level healthfulness equals today’s area-level healthfulness. I test this assumption by comparing the area-level nutrient ratio in 2006 to 2016\(^{18}\) in the Nielsen scantrack (RMS) data. Figure 7a shows a scatter plot of the nutrient ratio by state in these two years. The nutrient ratio in 2016 and 2006 are not exactly equal. For the assumption to hold, however, they should be equal in expectation.

\(^{18}\)These are the first and last years of Nielsen Scantrack data available at this point in time.
In other words, the slope of the line of fit should be equal to 1 and the constant equal to 0. As shown, the slope coefficient for the linear regression is 0.83 (95% CI: 0.64 - 1.02) and the constant is 0.06 (95% CI: -0.06 - 0.18). I cannot reject the hypothesis that the slope equals 1 and the constant equals 0.

To examine the stability of area healthfulness for a longer time horizon I look at changes in state-level obesity rates using data from the CDC Behavioral Risk Factor Surveillance System (BRFSS). I use obesity rates as a proxy for the quality of grocery purchases - as discussed in Section 2.3, state level nutrient ratios are highly correlated with obesity rates (correlation = -0.74). The first year in which a large number of states were surveyed in BRFSS is 1990, so I compare state-level obesity rates in 1990 and 2017. Figure 7b shows that obesity rates in 2017 and 1990 are highly correlated (correlation = 0.75). The slope coefficient in a regression of 2017 on 1990 obesity rates is 1.07 (95% CI: 0.75 - 1.43), and I cannot reject that this equals 1. The constant is 0.07 (-0.02 - 0.16), capturing the overall increase in obesity rates over time across the US.

Altogether, while not perfect, these results indicate that current area-level healthfulness appears to be a good predictor of past area-level healthfulness.

4 Results

4.1 Changes Among Recent Movers

Figure 4a shows the primary event study plot for changes in the nutrient ratio in a period of 4 years around a move. The coefficients plotted are $\theta_{r(h,t)}$ from equation (5). I normalize the coefficient for the quarter prior to the move to 0, i.e. $\theta_{r(h,t)=-1} = 0$. The coefficients can be interpreted as the share of the gap between the origin and destination that is closed by the household upon moving. The results show that the nutrient ratio does not change substantially in the short-to-medium run. Pooling the coefficients allows me to rule out changes in nutritional quality of more than 3 percent of the destination - origin gap with 95 percent confidence. The results are further summarized in Table 4, column 1.

Notably, there is no immediate discontinuous jump at the time of the move. Given that the supply-side environment changes immediately upon moving, if such factors were important drivers of nutritional quality we would expect to see an immediate change following
the move. As I further elaborate on in section 5, the absence of such a jump can be interpreted as prima facie evidence that the supply-environment is not a primary driver of nutritional quality in the short-run.

Appendix Figure A.5 demonstrates that the results are similar for the three alternative grocery purchase quality measures. The change in expenditure score around the move is shown in Figure A.5a. By the end of the second year after the move, the change in the household expenditure score is equal to approximately 3 percent of the gap between the origin and destination, though it is not significant. In addition, figures A.5b and A.5c examine changes in food spending shares on unhealthful (soda and candy) and healthful (whole grain products and fruits and vegetables) product categories. For the unhealthy share, I find small but significant changes in the direction of the destination (when pooling the coefficients). However, I can reject changes in spending shares exceeding 0.6 percentage points with 95 percent confidence, which is small in magnitude given that the average household allocates approximately 23 percent of their food spending to this category. Spending on healthy food groups is virtually unchanged.

I present several alternative specifications as robustness checks in Appendix Section A.3. First, I examine changes in grocery purchases surrounding a move using county-level obesity rates as an alternative measure of area-level healthfulness. These results are similar to the main results using the area-level nutritional quality measures. There is no immediate change in the quality of grocery purchases in the quarter after the move, and the results similarly imply that nutritional quality is virtually unchanged in the short-run. Second, I explore a less parametric specification in which I explore changes in the quality of grocery purchases when moving to a healthier area, defined as any move in which the destination - origin ZIP code diet quality change is positive. The results are qualitatively similar, in the sense that nutritional quality does not change substantially in the first 2 years after the move. Lastly, I show that restricting to movers that move across state lines does not change in the results (Figure A.9).

**Comparison to naive cross-sectional approach**

To illustrate the importance of using movers in this analysis, I compare the main findings to a naive cross-sectional regression of a household's own healthfulness of grocery purchases on average purchases in their area. To ease comparison, I first re-run the above regressions but
pooling all months before the move and all months after the move to obtain a differences-in-differences estimate. This estimate represents the change in nutritional quality after relative to before the move, scaled by $\delta_h$. The coefficients can be found in Panel A of Table 3.

A cross-sectional approach to estimating the effect of neighbors has been commonly used in related literature. Christakis and Fowler (2007), for example, regress a person’s obesity status on the obesity status of their friends, spouses, and neighbors, controlling for age, gender, and education level. In addition the authors put in lags of a person’s own obesity status and the other’s obesity status.

I run the following cross-sectional regression of a household’s own nutritional quality on the quality of grocery sales in their ZIP code in the Nielsen Scantrack data:

$$y_{hjt} = \tau_t + \beta \cdot \bar{y}_j + X_{ht}\gamma + \epsilon_{hjt} \tag{8}$$

where $y_{hjt}$ is a household $h$ living in area $j$’s diet quality in month $t$, and $\bar{y}_j$ is the quality of grocery purchases in stores in area $j$ at time $t$. I follow Christakis and Fowler and include controls $X_{ht}$ which consist of age and education of the household head(s)$^{19}$.

First, I estimate a basic version of this regression which includes only a contemporaneous measure of the quality of grocery purchases in area $j$, $\bar{y}_j$. Second, I follow Christakis and Fowler and include lags of a household’s own nutritional quality and area-level quality of grocery purchases, estimating an autocorrelation regression. Third, I estimate the basic contemporaneous regression but including household fixed effects. In the latter specification the effects are identified only off of those households who experience a change in the area-level quality of grocery purchases, i.e. movers. In each case I report $\beta$, which tells us how average quality of grocery purchases in area $j$ affects the quality of a household’s own grocery purchases.

Panel B of Table 3 shows the estimated $\beta$ coefficients from equation (8) for the four measures of diet quality. In the most basic specification the estimated coefficient ranges from 0.37 for the nutrient ratio to 0.44 for the healthy food share. Each unit increase in average nutritional quality in the area is therefore associated with approximately a 0.4 unit increase in a household’s own nutritional quality. The Christakis and Fowler autocorrelation specification reduces this coefficient by about half, yielding an increase of approximately 0.16

---

$^{19}$I take the mean age and maximum education level if there are two household heads. I do not control for gender because this analysis is at the household level.
in a household’s own nutritional quality for every one unit increase at the ZIP code level. Lastly, the household fixed effects specification generates a coefficient of about 0.05.

These results demonstrate that a basic cross-sectional approach gravely overestimates the effect of place, at least in the short-run. Including lags of own and neighbors’ purchases addresses some of these concerns, but for most outcomes the movers estimates are considerably smaller. Note that the naive cross-sectional results are more consistent with the long-run findings (see next section).

4.2 Long-Run Evolution of Grocery Purchases

Having shown that the quality of grocery purchases does not change meaningfully in the first few years after a move, I consider whether nutritional quality might change more over the course of several decades.

Figure 8a shows the within-household change in the nutrient ratio with respect to years since move, by plotting the $\theta_{r(h,t)}$ coefficients from equation (9). The results suggest that migrants’ nutritional quality slowly converges toward the average in the destination. After approximately 40 years, a migrant that has lived in the destination area for approximately 40 years has closed approximately 50% of the gap in diet quality between the origin and destination. The pattern of change is roughly linear, suggesting a gradual convergence of diet quality toward the destination area over time. Note that the results are consistent with the earlier finding in section 4.1 that nutritional quality does not change substantially in the early years after the move. Figure 8b pools the coefficients into years-since-move bins. The coefficients on year-since-move-bins can also be found in column (2) of Table 4.

Appendix section A5 shows similar plots for the expenditure score (figure A.10a), the healthy share (figure A.10b) and the unhealthy share (figure A.10c). The expenditure score and healthy share plots show even stronger evidence of within-household convergence of diet quality over time. Selection seems more problematic when examining the unhealthy share in food purchases. The figure does not show evidence of within-household changes in the direction of the destination for this outcome.

As discussed in section 3.4, this analysis relies on stronger assumptions both related to selective migration and how healthfulness of diet changed across areas in the US over the last 50 years. For this reason, the results at best provide suggestive evidence of how household diet might evolve over a longer time horizon.
5 Discussion

The results in this paper suggest that healthfulness of grocery purchases is virtually unchanged in the short-run, but that over time purchases may converge to the average in the destination area. In this section I discuss potential explanations for these results, including the roles of access to foods, marketing, and social interactions.

There is a large literature primarily in public health that has established a cross-sectional relationship between the retail environment and food purchases (see Larson et al. (2009)). Accordingly, recent policies including the U.S. Healthy Food Financing Initiative introduced in 2010 have focused on bringing grocery stores and other healthy food retailers to areas denoted as “food deserts” - defined as areas with low availability or high prices of healthy foods. More recent findings in economics suggest that the retail environment does not seem to be a primary driver of differences in dietary choice, however, both across SES neighborhoods (Allcott et al., 2017) and countries (Dubois et al., 2014). I present two pieces of evidence that similarly show that differences in such supply-side factors are unlikely to explain the variation in nutritional outcomes across geographic areas in the US, at least in the short-run.

The primary piece of evidence in support of this interpretation is the absence of an immediate change in diet in the months following the move. Since the retail environment changes immediately when a household moves, if the local availability and prices of healthy foods were an important driver of dietary choice we should expect to see an immediate effect after the move. The fact that household diet is virtually unchanged in the short run, and more importantly does not show a discontinuous jump in the period after the move, is evidence that short-run constraints imposed by supply side factors such as availability and prices are unlikely to explain the observed variation in diet.

An alternative test of this hypothesis is to take a direct measure of food availability by area. For this purpose I examine moves to and from ZIP codes identified by the USDA as “food deserts”. The USDA defines these areas as “neighborhoods that lack healthy food sources”, and more precisely as census tracts more than 1 (10) miles away from a supermarket in urban (rural) areas. I estimate the event study regression in equation (6), using an indicator for moving to a food desert or moving from a food desert as the interaction variable $\delta_h$. The coefficients $\theta_{r(h,t)}$ tell us how a household’s nutrient ratio changes in the month surrounding the move, upon moving to or from a food desert. Figure A.8 shows that there is
no immediate change in healthfulness of grocery purchases following a move to a food desert.

A second explanation for the variation in diet quality across space is that there are large differences in inherent food preferences that are fixed. In examining differences in diet quality by SES, Allcott et al (2017) conclude that higher income households have stronger preferences for healthy and weaker preferences for unhealthy nutrients. The empirical results in this paper suggest that such preferences may be malleable, at least over a long enough time period. If preferences were instead fixed, diet quality would be unchanged after a move even in the long-run.

A third explanation then, which appears more consistent with the empirical results, is that individuals have food habits and preferences that are persistent, but that over time such habits are susceptible to changes in the local environment. These changes could be driven by neighbors or peers, which may affect diet through social interactions, or by long-run effects of the retail environment or product marketing. The slow convergence of dietary choice to the destination area may reflect the fact that it takes time for individuals to establish a new social network after moving, and for the behavior of others to affect individual dietary choice. At the same time it may be that long-run exposure to products or product marketing affects dietary choice.

In theory one could attempt to distinguish between these channels by looking at differences in food access and marketing across areas.

First, I examine whether households display the same kind of convergence when examining healthfulness of local grocery access instead of local grocery sales. I construct an alternative measure of state-level healthfulness using national sales weights as opposed to local sales weights. Differences in nutrient ratios across areas therefore reflect what is available in a store locally (access) rather than how much of it is sold locally (tastes). The correlation between this measure and the area-level nutrient ratio is 0.82. As expected, convergence using this access-based measure of healthfulness (see Figure A.11) is similar to that observed in Figure 8. It is unclear what to take away from this result, however. The high correlation between these measures suggests that access is likely endogenous to local tastes: supermarkets will only include foods in their assortment if there is local demand for it. Absent an exogenous shock to access it therefore seems hard to use this evidence in support of either channel.

Second, I use the fact that the Nielsen RMS data contains information on the types of
products featured or on display for a subset of stores to examine this channel\textsuperscript{20}. Differential product marketing across space could be an alternative driver, if the healthfulness of marketed products is correlated with healthfulness of local diet. I use these data to construct the share of items that are featured or on display in each 3 digit ZIP code that are considered healthy according to the Thrifty Food Plan groups\textsuperscript{21}. Table A.4 shows the correlation between healthfulness of grocery purchases, as measured by the nutrient ratio, and the share of displayed or featured items that are considered healthful. A one standard deviation higher nutrient ratio is associated with a 0.2 percentage point higher share of healthy items among displayed items. The coefficient for featured items is insignificant. While this suggests that there is at best a weak correlation between healthfulness of purchased and marketed items, I cannot rule out that differential marketing could drive at least some of the the observed effects.

While difficult to distinguish empirically, these channels have different implications for policy. If social interactions were the main driver of dietary choice then interventions targeting the retail environment may not be effective at changing people's dietary choice. At present most of the policy interventions, including those targeting food deserts, are aimed at improving access. In this case it may be more helpful to promote neighborhood diversity including initiatives requiring low-income housing in more affluent neighborhoods. On the other hand, if long-run exposure to foods in stores eventually shifts people's habits then over time such interventions may help to improve the healthfulness of grocery purchases in the US. Such long-run effects are harder to capture empirically, however.

Future work could attempt to distinguish between these channels more directly. As more and more data on food consumption becomes available it might be possible to track households over a longer period of time. Further, social network data might help to distinguish peer effects from other effects related to the retail environment.

6 Conclusion

In this paper I document changes in the healthfulness of grocery purchases among migrants to study how the local environment affects dietary choice. I find that a migrant household’s

\textsuperscript{20}Note that coverage is incomplete, and these findings may be incorrect if there is non-random selection into which stores report this information.

\textsuperscript{21}See section 2.2 and Appendix B.
nutritional quality is virtually unchanged in the first two years following a move. Healthfulness of grocery purchases converges about halfway to the destination-area average after three to four decades, however. The results suggest that dietary habits are highly persistent, but eventually shift in the face of different local environments. I use these results to assess several potential drivers of the observed geographic variation in diet quality. Differences in diet across geography are unlikely to be explained by short-run constraints imposed by the retail environment or differences in inherent food preferences across space that are fixed. Instead, the results suggest that dietary choice may be shaped by neighbors or peers, through social interactions, or by long-run effects of the retail environment or product marketing.
References


Ventura, Alison K and John Worobey, “Early influences on the development of food preferences,” *Current Biology*, 2013, 23 (9), R401–R408.


Table 1: Household descriptives: non-movers and movers

<table>
<thead>
<tr>
<th></th>
<th>Non-Mover</th>
<th>Short-run Panel Mover</th>
<th>Long-run mover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Size</td>
<td>2.383</td>
<td>2.171</td>
<td>2.269</td>
</tr>
</tbody>
</table>

*Household Income Groups:*

<table>
<thead>
<tr>
<th>Income Category</th>
<th>Non-Mover</th>
<th>Short-run Panel Mover</th>
<th>Long-run mover</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25,000 USD</td>
<td>.1739</td>
<td>.167</td>
<td>.143</td>
</tr>
<tr>
<td>25,000-49,999 USD</td>
<td>.316</td>
<td>.305</td>
<td>.309</td>
</tr>
<tr>
<td>50,000-99,999 USD</td>
<td>.374</td>
<td>.385</td>
<td>.387</td>
</tr>
<tr>
<td>&gt;100,000 USD</td>
<td>.1351</td>
<td>.142</td>
<td>.159</td>
</tr>
</tbody>
</table>

*Food Spending:*

<table>
<thead>
<tr>
<th>Spending Category</th>
<th>Non-Mover</th>
<th>Short-run Panel Mover</th>
<th>Long-run mover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Food Spending</td>
<td>184.4</td>
<td>178.8</td>
<td>190.5</td>
</tr>
<tr>
<td>Share on Soft Drinks</td>
<td>.0525</td>
<td>.0501</td>
<td>.0474</td>
</tr>
<tr>
<td>and Soda</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share on Sugar, Sweets</td>
<td>.1848</td>
<td>.1851</td>
<td>.1824</td>
</tr>
<tr>
<td>and Candy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share on Whole Grains</td>
<td>.0212</td>
<td>.0226</td>
<td>.0224</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share on Fruits,</td>
<td>.0896</td>
<td>.0930</td>
<td>.0951</td>
</tr>
<tr>
<td>Vegetables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household-month obs</td>
<td>6801559</td>
<td>265444</td>
<td>829821</td>
</tr>
</tbody>
</table>

*Notes:* The table shows descriptive statistics for non-movers, short-run panelist movers, and long-run life history movers. The rows for income categories report the share of households within the given income category among non-movers and the two mover groups.
Table 2: Movers: origin and destination region (shares)

(a) Short-Run Movers

<table>
<thead>
<tr>
<th></th>
<th>North East</th>
<th>Midwest</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>0.043</td>
<td>0.015</td>
<td>0.098</td>
<td>0.016</td>
</tr>
<tr>
<td>Midwest</td>
<td>0.010</td>
<td>0.074</td>
<td>0.103</td>
<td>0.043</td>
</tr>
<tr>
<td>South</td>
<td>0.039</td>
<td>0.069</td>
<td>0.221</td>
<td>0.052</td>
</tr>
<tr>
<td>West</td>
<td>0.011</td>
<td>0.029</td>
<td>0.067</td>
<td>0.110</td>
</tr>
</tbody>
</table>

(b) Long-Run Movers

<table>
<thead>
<tr>
<th></th>
<th>North East</th>
<th>Midwest</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>0.078</td>
<td>0.025</td>
<td>0.148</td>
<td>0.042</td>
</tr>
<tr>
<td>Midwest</td>
<td>0.015</td>
<td>0.115</td>
<td>0.132</td>
<td>0.084</td>
</tr>
<tr>
<td>South</td>
<td>0.019</td>
<td>0.042</td>
<td>0.138</td>
<td>0.027</td>
</tr>
<tr>
<td>West</td>
<td>0.005</td>
<td>0.020</td>
<td>0.033</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Notes: The table shows the region of origin (rows) and the current region (columns) among movers. Panel A shows the share of migrants moving between each region for the short-run movers identified in the data (N=6,271). Panel B shows the share of migrants moving between each region for the households identified as long-run migrants in the consumer life history survey (N=10,142).
Table 3: Differences-in-Differences results and naive estimates

<table>
<thead>
<tr>
<th></th>
<th>Nutrient Ratio</th>
<th>Expenditure Score</th>
<th>Good Share</th>
<th>Bad Share</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b/se</td>
<td>b/se</td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>Panel A: DD results</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>after move * area level change</td>
<td>-0.0151</td>
<td>0.0196**</td>
<td>0.01266</td>
<td>0.0286***</td>
</tr>
<tr>
<td></td>
<td>(0.02465)</td>
<td>(0.00861)</td>
<td>(0.02128)</td>
<td>(0.00984)</td>
</tr>
<tr>
<td>after move * 1(to healthier area)</td>
<td>0.004</td>
<td>0.0197</td>
<td>-0.0005</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.00452)</td>
<td>(0.05372)</td>
<td>(0.00137)</td>
<td>(0.00193)</td>
</tr>
<tr>
<td>after move * obesity change</td>
<td>0.0253</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td>(0.02415)</td>
<td>(0.00366)</td>
<td>(0.00834)</td>
<td>(0.01415)</td>
</tr>
<tr>
<td>after move * 1(to lower obesity area)</td>
<td>-0.0039</td>
<td>-0.0085</td>
<td>-0.0006</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>(0.00453)</td>
<td>(0.03308)</td>
<td>(0.00139)</td>
<td>(0.00195)</td>
</tr>
<tr>
<td>Panel B: Naive results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive regression</td>
<td>0.3742***</td>
<td>0.369***</td>
<td>0.4391***</td>
<td>0.3912***</td>
</tr>
<tr>
<td></td>
<td>(0.0.0043)</td>
<td>(0.00308)</td>
<td>(0.00273)</td>
<td>(0.01390)</td>
</tr>
<tr>
<td>Naive with autocorrelation</td>
<td>0.163***</td>
<td>0.045***</td>
<td>0.074***</td>
<td>0.107***</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.0.0460)</td>
<td>(0.01625)</td>
<td>(0.00502)</td>
</tr>
<tr>
<td>Household Fixed Effects (Movers)</td>
<td>0.0511***</td>
<td>0.0609***</td>
<td>0.0906***</td>
<td>0.0348***</td>
</tr>
<tr>
<td></td>
<td>(0.01481)</td>
<td>(0.00460)</td>
<td>(0.00459)</td>
<td>(0.00502)</td>
</tr>
</tbody>
</table>

Notes: The unit is a household month. Panel A shows the differences-in-differences results using the movers strategy. The table reports the coefficients on the interaction term between an indicator for months after the move and the change in area-level healthfulness on move. Area-level healthfulness is measured in four ways: [1] a continuous variable that is the destination - origin area-level average of the diet measure, [2] an indicator for moving to a healthier area based on the diet measure, [3] a continuous variable that is the destination - origin area-level obesity rate, [4] an indicator for moving to a higher obesity rate area. Each regression includes year x month fixed effects and household fixed effects. Standard errors are clustered at the household level. The sample is restricted to all movers (N=6,271 movers).

Panel B shows the $\beta$ coefficient from equation (8) of a cross-sectional regression of a household’s healthfulness of food purchases on average healthfulness in the household’s 3 digit ZIP code. The regressions also control for household head education and age, and includes month and year fixed effects. Standard errors are clustered at the household level. The sample is all panelist households in the HomeScan (N=158,830 households).
### Table 4: Overview of Short-Run and Long-Run results

<table>
<thead>
<tr>
<th>Years Since Move</th>
<th>(1) Short-Run Analysis</th>
<th>(2) Long-Run Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>0-2 years</td>
<td>-0.007</td>
<td>0.0435</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.0792)</td>
</tr>
<tr>
<td>2-5 years</td>
<td>0.023**</td>
<td>0.0217</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>5-15 years</td>
<td>0.0141</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td></td>
</tr>
<tr>
<td>15-25 years</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td></td>
</tr>
<tr>
<td>25-35 years</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td></td>
</tr>
<tr>
<td>35-45 years</td>
<td>0.335**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.177)</td>
<td></td>
</tr>
<tr>
<td>45+ years</td>
<td>0.561***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The unit is a household year. The table summarizes the regression coefficients \( \theta_{r(h,t)} \) on years since move for the short-run (equation (5)) and long-run (equation (7)) analyses. The dependent variable is nutrient ratio \( \text{scaled}_{ht} \), the adjusted nutrient ratio in relative quarter \( t \). This measure is 0 when the migrant household’s nutrient ratio is identical to the average in the origin, and 1 when the migrant household’s nutrient ratio is identical to the average in the destination. The regressions contain year dummies and household fixed effects. Standard errors are clustered at the household level.
Figures

Figure 1: Area level nutritional quality of grocery food purchases in the US

(a) Nutrient Ratio by 3 Digit ZIP Code

(b) State-Level Nutrient Ratio and Obesity Rates

Notes: Panel A shows a map of the nutrient ratio computed in each 3 digit ZIP code using the Nielsen Scantrack data. A higher nutrient score (a darker shade) corresponds to more healthful grocery purchases. Panel B shows a scatter plot of the relationship between the state level nutrient ratio and obesity rates. The correlation between the nutrient ratio and obesity rates is -0.74.
Figure 2: Distribution of destination - origin difference in area-level nutrient ratio

Notes: The figure shows the distribution across movers of the difference in area-level nutrient ratio between the origin and destination. The sample is all short-run movers identified in the HomeScan panel (N=6,271 movers).

Figure 3: Shopping trips surrounding the move

(a) Share in origin and destination ZIP code  
(b) Number of recorded shopping trips

Notes: The unit of observation is a household month. The figure plots coefficients on months relative to move dummies. Panel (a) shows the share of recorded shopping trips in the origin and destination ZIP code. Panel (b) shows the number of monthly shopping trips recorded in the HomeScan panel. The sample is all short-run movers identified in the HomeScan panel (N=6,271 movers).
Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{c(h,t)}$ on indicators for quarters relative to move (equation 5). The coefficient for relative quarter -1 is normalized to zero. The dependent variable is $nutrient\ ratio_{ht}^{scaled}$, the adjusted nutrient ratio in relative quarter $t$. This measure is 0 when the migrant household’s nutrient ratio is identical to the average in the origin, and 1 when the migrant household’s nutrient ratio is identical to the average in the destination. The regressions contain quarter x year dummies and household fixed effects. Standard errors are clustered at the household level.
Figure 5: Life history survey - movers descriptives

(a) Age in 2008

(b) Age leaving birth state

(c) Years living in current state

Notes: The figure shows the distribution of age in 2008, age at move, and years since the move for all migrants in the life history survey (N=10,412 movers).

Figure 6: Distribution of destination - origin state difference in nutrient ratio

Notes: The figure shows the distribution across movers of the difference in area-level nutrient ratio between the origin and destination. The sample is all migrants in the life history survey (N=10,412 movers).
Notes: Subfigure (a) depicts a scatter plot of the nutrient ratio in the RMS data by state in 2016 and 2006. The y-axis is the state nutrient ratio in 2016, the x-axis is the state nutrient ratio in 2006. The line of fit slope coefficient is 0.83 (95% CI: 0.64 - 1.02) and the constant term is 0.06 (95% CI: -0.06 - 0.18). Subfigure (b) shows a scatter plot of BRFSS obesity rates in 2017 and 1990. The slope coefficient in a regression of 2017 on 1990 obesity rates is 1.07 (95% CI: 0.75 - 1.43), and the constant is 0.07 (95% CI: -0.02 - 0.16).
Figure 8: Long-run evolution of diet

(a) Annual changes

(b) Binned changes

Notes: The unit of observation is a household month. Subfigure (a) plots the coefficients $\theta_{t(h,t)}$ on indicators for years since move in equation (9) including household fixed effects. Subfigure (b) plots the same coefficients but binned into years-since-move groups. In both cases the dependent variable is nutrient ratio $ht^{scaled}$, the adjusted nutrient ratio. This measure is 0 when the migrant household's nutrient ratio is identical to the average in the origin, and 1 when the migrant household's nutrient ratio is identical to the average in the destination. Standard errors are clustered at the household level.
## A Supplementary Tables and Figures

### A.1 Representativeness of Nielsen and Nutrition Data

#### Table A.1: Representativeness Nielsen data

<table>
<thead>
<tr>
<th></th>
<th>Nielsen Mean</th>
<th>US Census Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH Head Age</td>
<td>47.4</td>
<td>48.9</td>
</tr>
<tr>
<td></td>
<td>(10.3)</td>
<td></td>
</tr>
<tr>
<td>HH Head Years of Education</td>
<td>14.4</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td>(3.3)</td>
<td></td>
</tr>
<tr>
<td>HH Income</td>
<td>65,932</td>
<td>68,918</td>
</tr>
<tr>
<td></td>
<td>(45,690)</td>
<td></td>
</tr>
<tr>
<td>HH Size</td>
<td>2.59</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td></td>
</tr>
<tr>
<td>White (0/1)</td>
<td>0.82</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td></td>
</tr>
<tr>
<td>Moved Out of Birth State</td>
<td>0.32</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td></td>
</tr>
</tbody>
</table>

#### Table A.2: Share of Sales and UPCs Covered in Nutrition Data

<table>
<thead>
<tr>
<th></th>
<th>UPCs</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>45%</td>
<td>74%</td>
</tr>
<tr>
<td>2014</td>
<td>41%</td>
<td>73%</td>
</tr>
<tr>
<td>2013</td>
<td>39%</td>
<td>72%</td>
</tr>
<tr>
<td>2012</td>
<td>37%</td>
<td>72%</td>
</tr>
<tr>
<td>2011</td>
<td>37%</td>
<td>72%</td>
</tr>
<tr>
<td>2010</td>
<td>37%</td>
<td>72%</td>
</tr>
<tr>
<td>2009</td>
<td>35%</td>
<td>71%</td>
</tr>
<tr>
<td>2008</td>
<td>33%</td>
<td>70%</td>
</tr>
<tr>
<td>2007</td>
<td>29%</td>
<td>65%</td>
</tr>
<tr>
<td>2006</td>
<td>32%</td>
<td>64%</td>
</tr>
<tr>
<td>2005</td>
<td>31%</td>
<td>63%</td>
</tr>
<tr>
<td>2004</td>
<td>29%</td>
<td>63%</td>
</tr>
</tbody>
</table>
A.2 Area level healthfulness

Figure A.1: Regional nutrient ratio: controlling for demographic characteristics

Notes: The unit of observation is a household month. The figure plots the coefficients from a regression of the nutrient ratio on region dummies, both with and without controls for education, income, household size, and race. Coefficients are relative to the Middle Atlantic region. Standard errors are clustered at the household level.

Table A.3: Variation in area-level nutrient ratio explained by income

<table>
<thead>
<tr>
<th>(1)</th>
<th>Area-Level Nutrient Ratio</th>
<th>b/se</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median Income in 1000s of USD</td>
<td>0.008634***</td>
</tr>
<tr>
<td></td>
<td>Median Income in 1000s of USD squared</td>
<td>-0.000051***</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.288586***</td>
</tr>
<tr>
<td>Observations</td>
<td>799</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.082</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table shows coefficients on a regression of the area-level expenditure score on a quadratic in median household income.
Table A.4: Area-level nutrient ratio and healthfulness of marketing

<table>
<thead>
<tr>
<th></th>
<th>(1) On Display</th>
<th>(2) Featured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutrient Ratio (std. dev.)</td>
<td>0.00270* (2.26)</td>
<td>0.00168 (1.03)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0886*** (88.47)</td>
<td>0.140*** (102.39)</td>
</tr>
<tr>
<td>Observations</td>
<td>687</td>
<td>693</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.007</td>
<td>0.002</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows coefficients from a regression of the ratio of items on display or featured that are considered healthful by the Thrifty Food Plan groups on the nutrient ratio of grocery purchases at the 3 digit ZIP code level. Both measures are derived from the Nielsen Scantrack data.

A.3 Short-Run movers

A.3.1 Descriptives

Table A.5: Panelist 3 digit ZIP code movers: move year

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>568</td>
</tr>
<tr>
<td>2006</td>
<td>412</td>
</tr>
<tr>
<td>2007</td>
<td>423</td>
</tr>
<tr>
<td>2008</td>
<td>640</td>
</tr>
<tr>
<td>2009</td>
<td>553</td>
</tr>
<tr>
<td>2010</td>
<td>508</td>
</tr>
<tr>
<td>2011</td>
<td>564</td>
</tr>
<tr>
<td>2012</td>
<td>581</td>
</tr>
<tr>
<td>2013</td>
<td>653</td>
</tr>
<tr>
<td>2014</td>
<td>655</td>
</tr>
<tr>
<td>2015</td>
<td>714</td>
</tr>
<tr>
<td>Total</td>
<td>6271</td>
</tr>
</tbody>
</table>

Notes: The table shows the move year for each of the movers identified in the Nielsen HomeScan panel (N=6,271).
Table A.6: Household Income and Size Changes and Area-Level Healthfulness Change

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ZIP Nut Ratio Diff</td>
<td>ZIP Nut Ratio Diff</td>
<td>ZIP Nut Ratio Diff</td>
</tr>
<tr>
<td></td>
<td>b/se</td>
<td>b/se</td>
<td>b/se</td>
</tr>
<tr>
<td>Income change (1000s of USD)</td>
<td>0.000066</td>
<td>0.000071</td>
<td>0.000071</td>
</tr>
<tr>
<td></td>
<td>(0.00010)</td>
<td>(0.00010)</td>
<td>(0.00010)</td>
</tr>
<tr>
<td>Household size change</td>
<td>-0.001155</td>
<td>-0.001449</td>
<td>-0.001449</td>
</tr>
<tr>
<td></td>
<td>(0.00363)</td>
<td>(0.00364)</td>
<td>(0.00364)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.006014**</td>
<td>-0.006054**</td>
<td>-0.006051**</td>
</tr>
<tr>
<td></td>
<td>(0.00242)</td>
<td>(0.00242)</td>
<td>(0.00242)</td>
</tr>
<tr>
<td>Observations</td>
<td>219141</td>
<td>219141</td>
<td>219141</td>
</tr>
</tbody>
</table>

Notes: The dependent variable in the regression is the destination-origin expenditure score experienced by a given household upon moving. Standard errors are clustered at the household level.

Table A.7: Life Events Coinciding with Moves

<table>
<thead>
<tr>
<th></th>
<th>Movers</th>
<th>Random Year</th>
<th>Non-Movers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household Composition Change:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage</td>
<td>.024</td>
<td>.007</td>
<td></td>
</tr>
<tr>
<td>Divorce</td>
<td>.009</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>Child birth</td>
<td>.007</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>Empty nest</td>
<td>.002</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td><strong>Employment Change:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male HH Head Retires</td>
<td>.016</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Female HH Head Retires</td>
<td>.012</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Household Head Loses Job</td>
<td>.064</td>
<td>.0188</td>
<td></td>
</tr>
<tr>
<td>Income increase</td>
<td>.058</td>
<td>.034</td>
<td></td>
</tr>
<tr>
<td>Income decrease</td>
<td>.138</td>
<td>.059</td>
<td></td>
</tr>
<tr>
<td>Any event</td>
<td>.3313</td>
<td>.135</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table shows the occurrence of life events. Column (1) shows the share of move years that coincides with each event among short-run movers identified in the Nielsen panel. Column (2) shows the share of randomly picked years coinciding with such events for non-movers in the Nielsen panel.
A.3.2 Robustness and Alternative Measures

Figure A.2: Income Effects: Total Food Spending

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{r(h,t)}$ on indicators for time relative to move interacted with the ZIP-code nutrient ratio change associated with the move (equation 6). The coefficient for relative period -1 is normalized to zero. The dependent variable is total food expenditure per month. The regressions contain period x year dummies and household fixed effects. Standard errors are clustered at the household level.

Figure A.3: Moves not coinciding with life events

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{r(h,t)}$ on indicators for time relative to move interacted with the ZIP-code nutrient ratio change associated with the move (equation 6). The coefficient for relative period -1 is normalized to zero. The dependent variable is total food expenditure per month. The regressions contain period x year dummies and household fixed effects. Standard errors are clustered at the household level.
Notes: The unit of observation is a household month. The figure plots coefficients $\theta_{r(h,t)}$ on indicators for time relative to move (equation 5). The coefficient for relative period -1 is normalized to zero. The dependent variable is $nutrient_{ratio}^{scaled}_{ht}$, the adjusted nutrient ratio in relative period t. This measure is 0 when the migrant household’s nutrient ratio is identical to the average in the origin, and 1 when the migrant household’s nutrient ratio is identical to the average in the destination. Panel (a) shows the monthly change in the adjusted nutrient ratio and panel (b) shows the same plot but for annual frequency. The regressions contain period x year dummies and household fixed effects. Standard errors are clustered at the household level.
Figure A.5: Alternative diet measures

(a) Expenditure score

(b) Healthy foods: whole grain products, fruits and vegetables

(c) Unhealthy foods: soft drinks, soda, fruit drinks, sugar, sweets, and candy

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{r(h,t)}$ on indicators for quarters relative to move (equation 5). The coefficient for relative quarter -1 is normalized to zero. The dependent variable in panel (a) is expenditure score $\text{expenditure score}_{h,t}^{scaled}$, the adjusted expenditure score in relative quarter t. This measure is 0 when the migrant household’s expenditure score is identical to the average in the origin, and 1 when the migrant household’s expenditure score is identical to the average in the destination. Panels (b) and (c) show this for more specific food groups, including: (a) healthy foods: whole grain products, fruits and vegetables; (b) unhealthy foods: soft drinks, soda, fruit drinks, and ades, sugar, sweets, and candy. shows the same result for the expenditure score. The regressions contain quarter x year dummies and household fixed effects. Standard errors are clustered at the household level.
Figure A.6: Interacted with change in area-level obesity rates

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{c(h,t)}$ from equation (6) on the interaction term between quarter relative to move dummies and the change in area-level obesity rates upon moving. The coefficient for relative quarter -1 is normalized to zero. The dependent variable is the household's nutrient ratio in relative quarter $t$. The regression contains quarter x year dummies and household fixed effects. Standard errors are clustered at the household level.

Figure A.7: Interacted with dummy for move to healthier area

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{c(h,t)}$ from equation (6) on the interaction term between quarter relative to move dummies and an indicator for moving to a higher nutrient ratio ZIP code. The coefficient for relative quarter -1 is normalized to zero. The dependent variable is the household’s nutrient ratio in relative quarter $t$. The regression contains quarter x year dummies and household fixed effects. Standard errors are clustered at the household level.
Figure A.8: Moves to and from USDA food deserts

(a) Moves to USDA food deserts

(b) Moves from USDA food deserts

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{r(h,t)}$ from equation (6) on the interaction term between quarter relative to move dummies and an indicator variable for moving to or from a ZIP code defined by the USDA as a food desert. The coefficient for relative quarter -1 is normalized to zero. The dependent variables are the nutrient ratio for household $h$ in quarter $t$. The regression contains quarter x year dummies and household fixed effects. Standard errors are clustered at the household level.

Figure A.9: Short-Run Cross-State Moves

Notes: The unit of observation is a household quarter. The figure plots coefficients $\theta_{r(h,t)}$ on indicators for time relative to move (equation 5), restricted to movers that move across State lines. The coefficient for relative period -1 is normalized to zero. The dependent variable is $nutrient\ ratio_{nt}^{relative}$, the adjusted nutrient ratio in relative period $t$. This measure is 0 when the migrant household’s nutrient ratio is identical to the average in the origin, and 1 when the migrant household’s expenditure score is identical to the average in the destination. The regression contains quarter x year dummies and household fixed effects. Standard errors are clustered at the household level.
A.4 Long-run movers

A.4.1 Descriptives

Table A.8: Number of primary shoppers selected by stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number of Primary Shoppers Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Primary shopper indicated in survey</td>
<td>35,159</td>
</tr>
<tr>
<td>2. Household head</td>
<td>770</td>
</tr>
<tr>
<td>3. Female household head</td>
<td>1,277</td>
</tr>
<tr>
<td>4. Oldest individual</td>
<td>307</td>
</tr>
<tr>
<td>5. Drop duplicate individuals</td>
<td>10</td>
</tr>
<tr>
<td>6. Select random respondent</td>
<td>536</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>38,059</strong></td>
</tr>
</tbody>
</table>

*Notes:* The table shows the number of primary shoppers identified in the life history survey, by selection stage. See section 3.2.2.
A.4.2 Robustness and Alternative Measures

Figure A.10: Alternative diet measures

(a) Expenditure score

(b) Healthy food share

(c) Unhealthy food share

Notes: The unit of observation is a household year. Subfigure (a) plots coefficients $\theta_{r(h,t)}$ on indicators for years since move groups pooling across age at move categories, subfigure (b) plots coefficients $\beta_{a(h)}$ on indicators for age at move categories pooling across years since move categories (equation 8), and subfigure (c) shows within-household changes by plotting coefficients of a flexible regression on years since move dummies including household fixed effects (equation 9). In all cases the dependent variable is $expenditure \ score_{ht}^{scaled}$, the adjusted expenditure score. This measure is 0 when the migrant household’s expenditure score is identical to the average in the origin, and 1 when the migrant household’s expenditure score is identical to the average in the destination. Standard errors are clustered at the household level.
Notes: The unit of observation is a household month. This plots the coefficients $\theta_{t(h,t)}$ on indicators for years since move in equation (9) including household fixed effects. The dependent variable is $nutrient\ ratio_{ht}^{scaled}$, the adjusted nutrient ratio, but where the origin and destination state nutrient ratios are computed using national sales weights rather than local sales weights. As a result, this plot depicts whether households converge to their destination area’s healthfulness of food access. The measure is 0 when the migrant household’s nutrient ratio is identical to the average in the origin, and 1 when the migrant household’s nutrient ratio is identical to the average in the destination. Standard errors are clustered at the household level.
B TFP categories and the expenditure score

B.1 Formal Definition

The expenditure score measures the extent to which a household’s grocery purchases deviate from the expenditure shares recommended by the USDA Center for Nutrition Policy and Promotion (CNPP)’s “Thrifty Food Plan” (TFP). The expenditure score follows the measure used by Volpe et al. (2013), and is defined as:

$$Expenditure\ Score_{ht} = \left[ \sum_{c \in C_{\text{Healthful}}} (sh_{cht} - sh_{ch}^{TFP})^2 | sh_{cht} < sh_{ch}^{TFP} \right] \right. + \left. \sum_{c \in C_{\text{Unhealthful}}} (sh_{cht} - sh_{ch}^{TFP})^2 | sh_{cht} > sh_{ch}^{TFP} \right]^{-1}$$

where $c$ indexes TFP food categories, $sh_{cht}$ denotes the percent of household $h$’s grocery expenditures in month $t$ spent on products in category $c$, and $sh_{ch}^{TFP}$ is the category $c$ expenditure share, also in percent units, that the TFP recommends for a household with the same gender-age profile as household $h$.

As there are no clear guidelines for which food categories are most important for health, the index construction gives equal weight to all categories. Following Volpe et al. (2013) I take the inverse of the squared loss function so that higher scores are indicative of healthfulness.

The recommended individual expenditure shares from the TFP are outlined in Carlson et al. (2007). The individual recommended category $c$ expenditure share for each household member $i$, denoted $sh_{ci}^{TFP}$ is determined by the person’s age and gender. Then weights are assigned to each household member following the OECD equivalence scale:

$$w_{\text{adult}} = \frac{1+(n_{\text{adult}}-1) \times 0.5}{1 + (n_{\text{adult}} - 1) \times 0.5 + n_{\text{children}} \times 0.3}$$

$$w_{\text{child}} = \frac{0.3}{1 + (n_{\text{adult}} - 1) \times 0.5 + n_{\text{children}} \times 0.3}$$

The recommended category $c$ expenditure shares for each household member is then:

$$sh_{ch}^{TFP} = \sum_i w_i sh_{ci}^{TFP}$$
B.2 Healthful and unhealthful food categories

The healthful and unhealthful food categories are:

<table>
<thead>
<tr>
<th>Healthful</th>
<th>Unhealthful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole grain products</td>
<td>Non-whole grain breads, cereals, rice, pasta, pies, pastries, snacks, and flours</td>
</tr>
<tr>
<td>Potato products</td>
<td>Whole milk products</td>
</tr>
<tr>
<td>Dark green vegetables</td>
<td>Cheese</td>
</tr>
<tr>
<td>Orange vegetables</td>
<td>Beef, pork, veal, lamb and game</td>
</tr>
<tr>
<td>Canned and dry beans, lentils, and peas</td>
<td>Bacon, sausage, and luncheon meats</td>
</tr>
<tr>
<td>Other vegetables</td>
<td>Fats and condiments</td>
</tr>
<tr>
<td>Whole fruits</td>
<td>Soft drinks, sodas, fruit drinks, and ades</td>
</tr>
<tr>
<td>Fruit juices</td>
<td>Sugars, sweets, and candies</td>
</tr>
<tr>
<td>Reduced fat, skim milk, and low-fat yoghurt</td>
<td>Soups</td>
</tr>
<tr>
<td>Chicken, turkey, and game birds</td>
<td>Frozen or refrigerated entrees</td>
</tr>
<tr>
<td>Eggs and egg mixtures</td>
<td></td>
</tr>
<tr>
<td>Fish and fish products</td>
<td></td>
</tr>
<tr>
<td>Nuts, nut butters, and seeds</td>
<td></td>
</tr>
</tbody>
</table>

Following Todd et al. (2010) I match the TFP food groups with Nielsen products using the Quarterly Food-at-Home Price Database (QFAHPD). Only a share of households report purchases of random-weight items, and for this reason I leave out all such purchases and use only UPC-coded purchases. Handbury et al. (2015) follow the same strategy and state that many of the random weight categories are too broad to infer meaningful nutritional information for random weight items.

I aggregate the 52 QFAHPD food groups to the 24 TFP food categories using the correspondence created by Volpe and Okrent (2013). Two TFP food categories, cheese and meat, contain both healthful and unhealthful food groups. Following Handbury et al. (2015), I assume that the aggregate TFP cheese and meat categories are unhealthful.
B.3 Correlation diet quality measures

To give a better sense of what the measures of diet quality are capturing, I present correlations of the two summary measures with the spending shares on food categories that are easier to interpret. Table B.2 shows the correlation between the nutrient ratio, expenditure score and spending shares on two unhealthful food categories and two healthful food categories.

Table B.2: Correlations between diet quality measures

<table>
<thead>
<tr>
<th></th>
<th>Nutrient Ratio</th>
<th>Expenditure Score</th>
<th>Healthy Share</th>
<th>Unhealthy Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutrient Ratio</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure Score</td>
<td>0.2517</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy Share</td>
<td>0.3797</td>
<td>0.5827</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Unhealthy Share</td>
<td>-0.1377</td>
<td>-0.4739</td>
<td>-0.3019</td>
<td>1</td>
</tr>
</tbody>
</table>