Does Healthy Food Access Matter in a French Urban Setting?

The Role of Food Retail Structure

France Caillavet
French National Institute for Agricultural Research,
INRA-UR1303, ALISS Ivry-sur-Seine

Gayaneh Kyureghian*
Department of Food and Resource Economics, Korea University

Rudy Nayga
Department of Agricultural Economics and Agribusiness, University of Arkansas,
Norwegian Agricultural Economics Research Institute
Department of Food and Resource Economics, Korea University

Coline Ferrant
Sciences Po, Observatoire Sociologique du Changement
INRA-UR1303, ALISS Ivry-sur-Seine

Pierre Chauvin
Inserm, Sorbonne Universités UPMC Paris 06, UMRS 1136,
IPLESP, ERES, F-75012 Paris, France

Selected Paper prepared for presentation at the Allied Social Science Associations 2015 ASSA Annual Meeting, Boston, MA, January 3-5, 2015.

This research has received funding from the French National Research Agency (ANR), the French Institute of Public Health Research (IRESP) and the French Interministerial Committee of Urban Affairs. It received also some support from the Mutualité Française.
SIRS survey received the approval of the CNIL (Commission Nationale Informatique et Liberté) via the authorization n°05; 1024 following demand n° 904851 of the 17.06.2004)

Does Healthy Food Access Matter in a French Urban Setting?

The Role of Food Retail Structure

It is maintained that limited access to healthy food and relatively easy access to less healthy food, among other economic and environmental factors, are accountable for poor dietary choices and are ultimately associated with major public health concerns (Walker *et al.* 2010; Economic Research Service Report to Congress 2009). The linkage between the food environment and food choices and spending patterns, and, more fundamentally, food security has been a subject of interest in academic and policy debates. Although much discussed and widely researched, the linkage between food retail availability and dietary choices or health status is still largely unclear (Walker *et al.* 2010; Kyureghian, Nayga and Bhattacharya 2013; Kyureghian and Nayga 2013).

The literature on this subject is predominantly correlational and the findings are mixed. In a comprehensive review of literature on disparities in access to healthy food, Larson, Story and Nelson (2009) report that the majority of studies suggest that there is a relationship between the availability of supermarkets and healthy diets, and that this varies by socio-demographic groups. In another systematic review, Beaulac, Krisjansson and Cummins (2009) report mixed results concerning the effect of availability and quality of healthy foods in disadvantaged areas. Bitler and Haider (2011), on the other hand, provide a comprehensive analysis and discussion of the empirical literature on food deserts.

There are both economic and empirical problems in estimating the impact of the food environment on dietary choices and health. On the economic front, two problems stand out. First, food unavailability or restricted availability is hard to factor-in in the demand analysis. In other words, while availability is a necessary condition for purchase, unavailability is neither a necessary nor sufficient condition for non-purchase. Second, there is no logically compelling or theoretically established causality path linking retail access to dietary choices.

On the empirical front, the choice of the area of residence, which has certain characteristics, is unlikely to be uncorrelated with other behavioral and lifestyle choices that might give rise to dietary choice as well. Consequently, any approach that disregards this relationship is likely to produce biased results. A number of other difficulties associated with the complexity, inherent lack of precise definition, and measurement of food availability, accessibility and affordability have made this task increasingly difficult in the past (Kyureghian, Nayga and Bhattacharya 2013; Kyureghian and Nayga 2013).

Despite the factual relevance of the commonly accepted belief that the food retail environment shapes dietary choices, there have been no empirical studies on the causal dependency between the two. In this article, we set to explore the impact of retail food availability on a staple healthful food consumption – fruits and vegetables. We do so by addressing the aforementioned empirical and economic problems. We employ fruit and vegetable (FV) consumption data for Paris and its suburbs in France.

Several works point out to the role of FV as a preventive factor in many health problems, such as obesity and cardiovascular diseases, among others. Nowadays, it is the most recommended food group in governmental dietary guidelines, including in France. For example, the Food and Health Programs, carried out in France since 2001advocate a specific goal of consuming five servings of FV daily, as an essential component of a healthy diet.

Spatial or territorial disparities of environmental determinants have been a source of concern in France (Cadot and Chaix 2009; Leal *et al.* 2011; Chaix *et al.* 2011; Chaix *et al.* 2012;

Casey *et al.* 2012). These French studies confirm the influence of retail environment on obesity in urban areas, but none of them take into account the endogeneity issues discussed above. In particular, we find no study that examined the direct relationship between spatial inequalities in food access and healthy food patterns, such as FV consumption in France. This is an important issue in France considering its rising incidence of obesity and poor dietary habits. For example, the last National Nutritional Survey (ENNS 2007), which collected data on food consumption in France from 2006-2007, shows that most French consumers do not meet the recommended dietary guidelines for FV (potatoes excluded). Specifically, the majority of the respondents – 57%, did not achieve the daily recommended five servings of FV, and approximately 35% of respondents consumed less than the threshold considered for low consumption (i.e., 3.5 servings per day). Although some improvement is observed compared to the preceding survey in 1998-99, (60% of consumers not reaching guidelines and 41% low consumers), promotion of the consumption of FV remains a key objective of French food policy. Furthermore, there is an inequality issue since higher consumption of FV is related to increasing income and education levels. For instance, social inequalities of purchases have been found for fresh products (Caillavet *et al.* 2009, Plessz and Gojard 2012).

To address the empirical issues mentioned above, we use two identification strategies that exploit variation in two instruments: (i) the availability of metro or rail stations and (ii) business entry approval rates (i.e., number of business approvals conditional on number of applications). For the first source of variation, we capitalize on the empirical evidence that stores flock around metro or rail stations, possibly as a response to high traffic or elevated demand in those areas. Considering that the presence of a metro or rail station in a region might not be exogenous for the same reasons that the retail environment is not, we utilize the size of a metro or rail station, expressed by the daily number of commuters, as a source of exogenous variation.

The second variation – the share of approved applications to start or expand a retail business, is a measure of stringency of entry deterrence for retail entities. This is a supply side-variation, effectively regulated by the zoning committees that make the approval decision.

We directly address some of the technical difficulties related to the measurement of food retail availability. The lack of consensus in the literature concerning the area of reference, or the appropriate geographic unit, for the food retail environment has been extensively discussed and credited as one of the possible reasons for the mixed findings in the literature (Kyureghian, Nayga and Bhattacharya 2013; Bitler and Haider 2011). For example, some of the geographic units that have been popularly used in past studies, such as census tracts or census blocks or zip codes, have been criticized as too narrow since they can potentially underestimate the retail presence, while other geographic units, such as counties, or states, were deemed much too large and uninformative. To ascertain the robustness of our findings, we consider four different reference areas, ranging from the smallest (IRIS, equivalent to the US census tracts) to the largest (areas encircled by a perimeter equidistant from a core of either the residence of the respondent or a metro/rail station).

Similar criticism has been brought up against the actual measure of the food retail availability. Numerous definitions were entertained in the literature so far – the number of supermarkets, combined number of stores, the combined shelf length for FV in food stores, etc. In this study, we make use of eight different measures of availability, defined across four different areas of reference. Finally, we utilize an instrumental variable estimation technique to estimate the relationship between the frequency of consuming the recommended FV servings and food retail availability.

Data

To deal with the complex and multifaceted nature of our proposed analysis, we draw data from several sources. Health, Inequalities and Social Ruptures (Santé, Inégalités et Ruptures Sociales or SIRS) is the source we use for data on dietary behavior and demographic profile of the respondents. This data set also provides the linear distance from each respondent's dwelling to the nearest metro/rail station. The second data source we use is the Permanent Database of Facilities dataset, compiled by the National Institute of Statistics and Economic Studies (Institut National de la Statistique et des Etudes Economiques or INSEE), which includes information on the commercial retail structure and the presence of food outlets. We also use the TDLinx data, which are provided by Trade Dimensions (the Nielsen company), and provide information on numbers of stores by store type and total areas (square meters) along with areas designated to food sales in each store (most of the store types), complete with the longitude and latitude information of each outlet. The data from the Commercial Zoning Boards (Commissions Départementales d'Aménagement Commercial or CDAC) provides information about the retail entry applications and application approval rates in French départements. The Metropolitain transport system (RATP) and the suburban railway system operating in Ile de France (SNCF Transilien) were the source of the information on the number of daily commuters.

The data on food patterns and health collected in the Parisian region come from SIRS for 2010. The SIRS cohort study is a longitudinal socio-epidemiological, population-based survey of the French-speaking adult population in the Paris metropolitan area (Paris and its suburban départements, a region with a population of 6.5 million), conducted since 2005 in the framework of a collaborative research project between the French National Institute for Health and Medical Research (INSERM) and the National Centre for Scientific Research (CNRS).

This survey was based on a three-stage cluster random sample of approximately 3,000 adults (areas, households, household adult members) stratified according to the socioeconomic status of the neighborhood. The primary sampling units were census blocks (with about 2000 inhabitants each): 50 were randomly selected (over-representing the poorer neighborhoods) from the 2595 eligible ones in Paris and its suburbs. Subsequently, 60 households were randomly chosen from a complete list of dwellings in each surveyed block. Lastly, one adult was randomly selected from each household. Data were collected through at-home, face-to-face interviews during the third wave of data collection in 2010 (for instance, see Vallée *et al.* 2011, or Martin-Fernandez *et al.* 2012 for an extensive description of the methodology).

The variable of interest in this survey is the likelihood that a respondent is frequently consuming the recommended five FV servings per day. The respondents choose from among the alternatives 'every day', 'very frequently', 'frequently', 'occasionally, or 'never'. The dependent variable in our models is a dichotomous variable that equals to unity if respondent is consuming frequently the recommended five servings per day and zero if only consuming the recommended FV servings occasionally or not at all. On average, 63% of the individuals in our sample reported consuming the recommended five daily servings of FV frequently. The average age of the sample is 50 years, with women being the majority (approximately 60 percent). The average monthly income from all sources is 2,624 Euros. Households with children comprise approximately 35% of the sample. Approximately 52-53% of the sample has college level or higher education and works either full-or part-time. The respondents are evenly distributed over the city of Paris and the three suburbs (about three quarters of the respondents live in the 3 suburbs of Paris). While only a few are dieting (for any reason), a vast majority are aware of the 5-a-day recommendation. The variable names, description, and summary statistics are presented in table 1.

As previously mentioned, to address the issues associated with different reference areas, we use four different areas of reference – IRIS, TRIRIS, a circle with the respondent at the center, and a circle with the metro/rail station at the center, with the distance of the respondent to the nearest metro/rail station as the radius. IRIS (an acronym of ‘aggregated units for statistical information’) are geographical units defined by INSEE. They are comparable to U.S. census tracts. IRIS typically contain between 1,800 and 5,000 inhabitants. TRIRIS areas, on the other hand, are combinations of three adjacent IRIS’s. Since IRIS’s are the finest area delineations, typically all the respondents in the same IRIS have the same metro/rail station as the closest station. Consequently, we define the radius as the maximum distance from the residents in an IRIS to the metro/rail station.

With the reference areas defined, we then proceed with building the metrics for food availability – (i) number of stores in the area per 1 square kilometer, (ii) total food retail area per 1 square kilometer, and (iii) a diffusion metric reflecting the availability from different types of stores.

The availability measures we employ are the total number of stores – hypermarkets, supermarkets, superettes and frozen markets. Readers interested in details concerning the French retail system are referred to Appendix A. The total food areas are the summation of areas designated to food sales in all types of stores in the reference area. The numbers of stores are available from INSEE data (for IRIS and TRIRIS areas only) and from the TDLinx data (for resident- or metro-centered circles only). The food designated areas are available from TDLinx (for resident- or metro-centered circles only). Since we have TDLinx data for only 2013, we added the numbers of stores that closed after 2010, but were open as of 2010. Unfortunately, we cannot identify the stores that were open in 2013 but not in 2010 (i.e., new stores that opened after 2010). However, we believe that the number of new openings, if any, is insignificant and proceed hereafter with this assumption.

We would like to capture not only the store availability, but also the variety of store types in the area. Basically, we conjecture that a decentralized availability is preferred to centralized availability. In other words, two different types of stores with half the area are preferred to one store with twice the area. To capture both the availability and variety of availability by store type and the amount of area devoted to food sales from the stores in the reference area, we use a variation of the Berry index (Thiele and Weiss, 2003). We define it as: $BI_{ri} = 1 - H_{rj} = 1 - w_j \sum_i^n s_{ij}^2$, where BI_{ri} is the diversity index of store type j, in reference area r, H_{rj} is the Herfindahl index of concentration for store type j, in area r, s_j are the shares of food area of store type j in the food area of all store types, and w_j are the shares of store type j in all stores. Basically, this variation of the Berry index is measuring the dispersion of stores by type and by area designated to food. It is bound between 0 and 1. For example, a $BI_{rj} = 0$ indicates that there is one store available in the area, and so the higher index means more diversified store types with smaller surface areas. The different types of availability measures, reference areas, data sources, and years used in our analysis are exhibited in table 2.

As expected, the average number of stores (all types combined) is lowest in IRIS areas (the smallest areas) at 0.4337 stores on average. Slightly higher averages are observed in TRIRIS areas, and the largest numbers of stores are in the respondent-centered circles (6.4580 stores on average). Naturally, the latter have the largest average food surface as well. The Berry indices are close, indicating that there is slightly more diversification in metro-centered circles compared to the respondent-centered ones.

Finally, the excluded instruments – the shares of retail outlet applications approved and the number of daily commuters per metro/rail station, come from the Loi Royer (see the appendix for a detailed discussion) and RATP,/SNCF Translinien, respectively. For the individuals living inside Paris area, we obtained the data from the Metropolitain transport system (RATP), which also provides the data on annual number of commuters. For the individuals living outside the Paris area, data on the number of commuters were obtained from the railway company operating in Ile de France (SNCF Transilien). It provides a range of the daily number of commuters in each station: less than 300 commuters/day, between 300 and 1000 commuters /day, etc. We obtained commuter data for 2011 and 2013, but the data for the year of interest – 2010, were not available. The commuter numbers by stations from 2011 to 2013 were remarkably similar, so we considered it reasonable to use these to extrapolate the 2010 values. As the summary statistics reflect in table 1, the average number of daily commuters is over 12,000.

Commercial zoning boards (Commissions Départementales d'Aménagement Commercial) which are charged of controlling the development of large stores (over 1000m^2) exist at the département level, which is the main administrative unit in France. Listings of all applications submitted, along with the main characteristics of the store, project (brand name of the applicant, size in square meters, nature (creation or extension), specialty (food or non-food), and address) and the positive or negative decision (i.e., approvals/disapprovals) of the board are available for 2010. Although the zoning committee decisions are made at the département level, we have information on the applications and approvals at the arrondissement level and, therefore, can perform one-to –one matching with the main consumption data. To ward off concerns of self-selection bias due to application fees and other costs being associated with applying, and the lack of variability in the approval/application ratio, we expressed the approval rate in each arrondissement not as the ratio of approvals to applications in each arrondissement, but as the ratio of the approvals per arrondissement to applications in the whole of a département. This also reflects the relative standing (or variation) of the different arrondissements in the same département as far as entry stringency is concerned. The numbers of applications and approvals are exhibited in table 4. As the summary statistics reflect in table 1, the average percentage distribution of approvals per arrondissement is almost 8.

Throughout the construction of the final dataset, where missingness was observed, regional means, where available, were imputed. Otherwise, the mean values of the observed observations were imputed for the missing values. The final data set has 2,963 observations.

Empirical Methodology

Our basic empirical approach is to explain variation in the frequency of consumption of FV by observed variation in the food retail availability. We estimate the following models:

$$FV_{ir} = \beta No\_Stores\_IRIS_r + a_d + \gamma X_i + \varepsilon_{ir} \quad (1)$$

$$FV_{il} = \beta No\_Stores\_TrIRIS_l + a_d + \gamma X_i + \varepsilon_{il} \quad (2)$$

$$FV_{in} = \beta No\_Stores\_R_n + a_d + \gamma X_i + \varepsilon_{in} \quad (3)$$

$$FV_{in} = \beta FoodArea\_R_n + a_d + \gamma X_i + \varepsilon_{in} \quad (4)$$

$$FV_{in} = \beta BI\_R_n + a_d + \gamma X_i + \varepsilon_{in} \quad (5)$$

$$FV_{im} = \beta No\_Stores\_M_m + a_d + \gamma X_i + \varepsilon_{im} \quad (6)$$

$$FV_{im} = \beta FoodArea\_M_m + a_d + \gamma X_i + \varepsilon_{im} \quad (7)$$

$$FV_{im} = \beta BI\_M_m + a_d + \gamma X_i + \varepsilon_{im} \quad (8)$$

where i's index individual respondents; r, l, n and m index the four areas of reference – IRIS, TRIRIS, Resid-centered and Metro-centered areas, respectively; No_Stores_IRIS, No_Stores_TrIRIS, No_Stores_R, $FoodArea\_R$, BI_R, No_Stores_M, and $FoodArea\_M$ are the availability measure variables as described in table 2; a_d is a regional dummy indicating whether the respondent lives in the city of Paris or in suburbs; X_i is a vector of demographic and dietary variables, listed under the *Demographic Controls* in table 1; β and γ are parameters to be estimated; and ε_{im} is the error term. The parameter of interest – β, indicates the effect of the availability variables on the frequency of consuming the recommended number of FV. The results of the estimations of models (1) through (8) are reported in table 4.

Four different model specifications were estimated. The marginal effects from regressing food availability variables only on the likelihood of frequently consuming the recommended amount of FV are reported in column (A). As can be seen, none of the availability measures explains the variations in the dependent variable. We then investigate the robustness of our results to the inclusion of various area and individual characteristics. Column (B) displays the results when a regional dummy is added to the models, and columns (C) and (D) display the results from adding a full set of demographics and health knowledge variables, respectively. None of these specifications reveals a statistically significant relationship between retail availability and the likelihood of frequently consuming the recommended FV servings, consistent with some previous findings reported in the literature (Kyureghian, Nayga and Bhattacharya 2013; Kyureghian and Nayga 2013).

Endogeneity

While the empirical approach discussed above is simple and intuitive and the results are robust across different model specifications, the validity of this approach rests upon the assumption that the variation in the availability variables is exogenous. As discussed above, it seems plausible that the choice of the area of residence is correlated with other behavioral and lifestyle choices that might give rise to dietary choice as well. If this is the case, then the marginal effects we derived above are biased. An alternative empirical approach is to resort to instrumental variable estimation techniques. For identification, we need a source of exogenous variation. One source of exogenous variation is the stringency measures of business entry. This measure basically indicates how easy or likely are new stores (including food) to open or expand in the reference area. The approval decisions are made at the level of départements (four of them in our sample), but the numbers of application and approval are available on more refined geographical levels (see table 3). We maintain that even if it is remotely possible that the approval decisions themselves might be endogenous, the variation within the départements is exogenous. This, combined with the fact that our data on applications and approvals cover all businesses, not just food retailing, bolsters our confidence in the exogeneity of this source of variation. Hence, the approval rates are expressed as the distribution of approvals of the entire département over the arrondissements, rather than as the ratio of the approvals on arrondissement level and the applications on the arrondissement level. In other words, this structure reflects the internal variation in each département.

The second source of exogenous variation we entertain is the existence of metro/rail stations in the reference area. Due to limited parking spaces in urban areas, it is natural to assume that retail stores would be motivated to be located close to public transportation such as in metro/rail stations. Indeed, when investigating the relative locations of metro/rail stations and food retail outlets for the entire greater Paris area (not only the IRIS's in our consumption data), it is clearly observable that there is a strong interdependence. The latter is even stronger in the city of Paris (Figure 1). A closer consideration, however, reveals that an obvious criticism of this source could be that the metro/rail station could itself be endogenous for the same reason that the number of food stores or any other retail or zoning feature could be. That is, the respondents with specific perceptions of health and diet could self-select to reside in areas with a particular level of availability of public transportation and retail outlets. To ward off such doubts about our instrument, we opt for a feature of metro/rail stations rather than rely on a measure about their existence – i.e., we are utilizing information on how big the station is. That is, we are using the average number of daily commuters as the source of exogenous variation. Given these instruments, we then re-estimate models (1) through (8) using instrumental variable regressions.

While it is tempting to use both instruments at the same time in the first stage regressions, it is perhaps wise to avoid this as the two are likely to be correlated, which would result in spurious results. Consequently, the choice of the relevant excluded instruments is guided by the choice reference areas. For example, while the share of approvals can be a logical choice for all reference areas, the number of commuters seems to be a reasonable choice if the reference area is defined based on or around a metro/rail station. We performed the first stage estimations for each of the instruments.

Results

The marginal effects of the IV estimations, along with the first stage regression parameters and F-values of excluded instruments are reported in table 5. The first stage results generally indicate that the instruments are strong as all first stage parameter estimates for the excluded instruments are highly significant. The F statistics of excluded instruments confirm the validity of the instruments as they mostly exceed the Stock-Yogo critical values (Stock and Yogo 2005). The Wald test of exogeneity of the suspect variable (the availability measures) mostly rejects the null of exogeneity, confirming the appropriateness of the choice of IV estimation method.

In contrast to the probit estimates, the IV estimates are uniformly significant, albeit of different signs. As mentioned previously, IRIS's and TRIRIS's are very small geographical areas and hence it is possible that they could give a rather distorted view of the retail environment since retail presence in these geographic areas can be almost invisible. The marginal effects of retail availability, therefore, should be taken with a grain of salt. The interesting and rather remarkable result is the negative effects of the number of stores, both in resident- or metro-centered circles. This tendency seems to feature throughout the literature (Walker *et al.* 2010; Kyureghian, Nayga and Bhattacharya 2013; Kyureghian and Nayga 2013). The marginal effects indicate that a one-unit increase in the number of stores in resident- or metro-centered areas would decrease the probability of frequent FV consumption by 0.0187 and 0.0327, or by 1.87% and 3.27%, respectively. While the reason for this finding is unclear, it is possible that an increase in the accessibility of store triggers substitution patterns that are otherwise dormant. More attuned to our earlier expectations, our results also indicate that an increase in the food designated area has positive and small, although steady effect on the likelihood of frequent consumption of recommended number of FV servings. Specifically, an increase of 1 m^2 of food area can trigger a 0.01% increase in the probability of frequently consuming the recommended number of FV servings.

Interestingly, the dispersion indices have conflicting signs, depending on the area of reference. The marginal effect for the resident-centered area measure of dispersion indicates a positive effect on FV consumption frequency. That is, 1/100 increase in the dispersion index will improve the chance of being a 5-a-day consumer by 0.0098. The marginal effect for the metro-centered area measure of dispersion indicates a negative effect on FV consumption frequency, or a 1/100 increase in the dispersion index will reduce the probability of being a 5-a-day consumer by 0.0116.

To explain this phenomenon, we turn to differences in these two definitions of reference areas. The average number of stores (not adjusted to the surface area) is larger in metro-centered area compared to resident-centered area (numbers not reported here). This order is reversed, however, when comparing the average number of stores per $1km^2$. This indicates that in the metro-centered areas with a large number of stores also have large surface areas or, equivalently have longer radii. Recall that radii are equal to the distance to the closest metro/rail station. This clearly indicates on metro-centered areas with large number of stores are located in suburbs, where the distance to the nearest metro/rail stations is observably higher. This situation is reversed for resident-centered areas: areas with high number of stores are located in Paris, and the ones with smaller number of stores are located in suburbs. In summary, the concentration of different types and large number of stores is away from metro/rail stations and close to the residential areas in Paris, and the concentration of many and various types of stores is away from residents and around the metro/rail stations in suburbs.

With this in mind, our results indicate that an increase in dispersion in Paris has positive effect on improving the odds of FV consumption, and more dispersion in suburbs actually decreases these odds. This finding actually resonates with those of Michimi and Wimberly (2010), who use national level cross sectional data for seven years to demonstrate an inverse association between the odds of consuming F&V five times a day and the distance to supermarket in metropolitan areas, but not in non-metropolitan areas.

Concluding Remarks

There has been plenty of interest from individual consumers and the media in the US and elsewhere about the "French Paradox". For example, it has been reported that the French people eat high caloric and fatty foods and yet have low incidence of cardiovascular related diseases. The bad news for the French, however, is that their obesity rate and incidence of poor dietary habits have been increasing in the last decade. This has become of great concern in France. For example, a vast majority of the French population now do not follow the recommended guidelines of consuming five servings of fruits and vegetables per day. Hence, the increase of fruit and vegetable intake has become a public health priority in France. While there are worries that availability of food stores could be blamed for this phenomenon, there is scant information on the effect of food access and availability on food choice in France. Past research are also mired by a myriad of empirical and data issues. We address some of these issues in this article using French data from different sources and an instrumental variable estimation. We also employ different measures of food availability, different geographic units or reference areas, and different model specifications to test the robustness of our results.

Our results generally suggest that, in accordance with some previous research, the number of food retail outlets is negatively related with the likelihood of frequently consuming the recommended number of fruit and vegetable servings per day. However, our results also indicate that the effect of total area of the stores devoted to food sales is positively related with the likelihood of frequently consuming the recommended number of fruit and vegetable servings per day.

The results in this study provide motivation for future research to conduct a comprehensive analysis in France of the effect of retail environment on consumption of not only fruits and vegetables, but also related food groups given recent increases in French obesity rates. Such an analysis could also provide more insights that could be used in analyzing the intended and unintended consequences of policy actions aimed at creating incentives to increase food availability in specific areas of the country (e.g., urban areas). With data availability, future studies should also analyze more definitive measures of consumption such as random weight purchases of fruits and vegetables.

References

Beaulac, J., E. Kristjansson, and S. Cummins. 2009. "A Systematic Review of Food Deserts, 1966-2007." *Preventing Chronic Disease* 6(3):A105-A115.

Bitler, M., and S.J. Haider. 2011. "An Economic View of Food Deserts in the United States." *Journal of Policy Analysis and Management* 30(1):153-176.

Cadot, E., and B. Chaix. 2009. "Quartier de Résidence: La Presence de Fast-Food Augmente le Risque d'Obésité." SIRS info n°7, April.

Caillavet, F., C. Lecogne, and V. Nichèle. 2009. "La Consommation Alimentaire : Des Inégalités Persistantes Mais Qui Se Réduisent." *La Consommation, INSEE Références*: 49-62.

Caillavet, F., and S.Y. Chen. 2012. "Les Modes d'Approvisionnement Alimentaire." In Chauvin P. (dir). ALISIRS, Les Inégalités Sociales et Territoriales Des Pratiques Alimentaires, de l'Obésité et de l'Insécurité Alimentaire dans l'Agglomération Parisienne en 2010. Rapport 2012.

Casey, R., B. Chaix, C. Weber, *et al.* 2012. "Spatial Accessibility to Physical Activity Facilities and to Food Outlets and Overweight in French Youth." *International Journal of Obesity* 36: 914-919.

Chaix, B., Y. Kestens, K. Bean, C. Leal, N. Karusisi, K. Meghiref, J. Urban, M. Fon Sing, C. Perchoux, F. Thomas, J. Merlo, and B. Pannier. 2011. "Cohort Profile: Residential and non-residential environments, individual activity spaces and cardiovascular risk factors and diseases—The 5 RECORD Cohort Study." *International Journal of Epidemiology*: 1-10.

Chaix, B., K. Bean, M. Daniel, S.N. Zek, Y. Kestens, H. Charreire, *et al.* 2012. "Associations of Supermarket Characteristics with Weight Status and Body Fat: A Multilevel Analysis of Individuals within Supermarkets (RECORD Study). Plos One 7(4), April.

Kyureghian, G., and R.M. Nayga Jr. 2013. "Food Store Access, Availability and Choice When Purchasing Fruits and Vegetables." American Journal of Agricultural Economics 95(5): 1280-86.

Kyureghian, G., R.M. Nayga Jr., and S. Bhattacharya. 2013. "The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis." *Applied Economic Perspectives and Policy* 35(1): 69-88.

Larson, N.I., M.T. Story, and M.C. Nelson. 2009. "Neighborhood Environments: Disparities in Access to Healthy Foods in the U.S." *American Journal of Preventative Medicine* 36(1):74-81.e10.

Leal C., K. Bean, F. Thomas, and B. Chaix. 2011. "Are Associations between Neighbourhood Socioeconomic Characteristics and Body Mass Index or Waist Circumference Based on Model Extrapolations?" *Epidemiology:* 22(5).

Stock, J.H., and Yogo, M. 2005. "Testing for Weak Instruments in Linear IV Regression." In D.W.K. Andrews and J.H. Stock, eds. Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge: Cambridge University Press: 80-108.

Thiele, S., and C. Weiss. 2003. "Consumer Demand for Food Diversity: Evidence for Germany." *Food Policy* 28: 99-115.

Walker, E., C. R. Keane, and J. G. Burke. 2010. "Disparities and Access to Healthy Food in the United States: A Review of Food Deserts Literature." *Health & Place* 16:876–84.

Table 1. Summary Statistics

Variable Name	Variable Description	Mean (Standard Deviation)
Dependent Variable		
FV	Binary variable = 1 if eating the recommended 5 servings of FV frequently; = 0 if eating occasionally or never	0.6363 (0.4782)
Food Availability Measures		
No_stores_IRIS	Number of all food stores in the IRIS area	0.4337 (0.6634)
No_stores_TRIRIS	Number of all food stores in the TRIRIS area	1.4536 (1.0749)
No_stores_R	Number of all food stores in the circle around residence	6.4580 (8.7454)
No_stores_M	Number of all food stores in the circle around metro/rail station	4.8212 (5.6402)
Food Area_R	Combined area (in m^2) devoted to food sales in all food retail stores in the circle around residence	3111.07 (3877.96)
Food Area_M	Combined area (in m^2) devoted to food sales in all food retail stores in the circle around metro/rail station	2252.43 (2443.66)
B_R	Berry Index is the diversity index of FV availability in the circle around residence	0.5671 (0.1904)
B_M	Berry Index is the diversity index of FV availability in the circle around metro/rail station	0.5771 (0.1972)
Excluded Instruments		
Approved_Arr	Share of approved applications in an arrondissement in all applications in each department	7.9924 (6.1538)
Commuter	Daily number of commuters by metro/rail station	12014.82 (7839.31)

Table 1. – Continued

Variable Name	Variable Description	Mean (Standard Deviation)
Demographic Controls		
Age	Age of respondent in years	49.9585 (17.1542)
Education	Binary variable = 1 if respondent has college or higher education; = 0 if high school or less	0.5245 (0.4995)
Gender	Binary variable = 1 if respondent is male; = 0 otherwise	0.3945 (0.4888)
Income	Monthly household income, in Euros	2624.06 (3122.71)
Child	Binary variable = 1 if there are children in the household; = 0 otherwise	0.3530 (0.4780)
Employed	Binary variable = 1 if respondent employed full- or part-time; = 0 otherwise	0.5299 (0.4992)
Suburb	Binary variable = 1 if respondent resides in suburbs; = 0 if respondent resides in Paris	0.7418 (0.4377)
Diet	Binary variable = 1 if respondent follows any diet; = 0 otherwise	0.1147 (0.3188)
Health knowledge	Binary variable = 1 if respondent knows about the dietary guidelines for FV consumption; = 0 otherwise	0.9690 (0.1735)

Table 2. Food Retail Availability and Excluded Instrument Variables: Reference Area Definition, Coverage, Source and Year

Availability Variable	Reference Area Center	Reference Area	Availability Data Year	Availability Data Source	Excluded Instrument Variable	Reference Area Center	Reference Area	Instrument Data Year	Instrument Data Source
No of retail stores	Resident	IRIS	2010	INSEE	Approved_Arr[1]	Resident	Arrondissement	2010	Loi Royer
No of retail stores	Resident	TRIRIS	2010	INSEE	Approved_Arr[1]	Resident	Arrondissement	2010	Loi Royer
No of retail stores	Resident	Circle	2013[2]	TDLinx	Approved_Arr[1]	Resident	Arrondissement	2010	Loi Royer
Total Food Area	Resident	Circle	2013[2]	TDLinx	Approved_Arr[1]	Resident	Arrondissement	2010	Loi Royer
Berry Index	Resident	Circle	2013[2]	TDLinx	Approved_Arr[1]	Resident	Arrondissement	2010	Loi Royer
No of retail stores	Metro/rail station	Circle	2013[2]	TDLinx	Commuter	Metro/rail station	Metro/rail station	2010[3]	RATP, SNCF Translinien
Total Food Area	Metro/rail station	Circle	2013[2]	TDLinx	Commuter	Metro/rail station	Metro/rail station	2010[3]	RATP, SNCF Translinien
Berry Index	Metro/rail station	Circle	2013[2]	TDLinx	Commuter	Metro/rail station	Metro/rail station	2010[3]	RATP, SNCF Translinien

[1] Approved_Arr is the percentage of approved applications in an arrondissement in all applications in each department.
[2] Combined number of stores open in 2013 and stores closed after 2010.
[3] The numbers of daily commuters in 2010 are extrapolated from 2013 and 2011 daily commuter data.

Table 3. Applications to the Commercial Zoning Boards and Approvals in 2010

Department	SIRS Arrondissements	Number of applications	Number of approvals	Rate of approval
75- Paris (total)		**5**	**5**	**100.0%**
	13	2	2	40.0%
	14	1	1	20.0%
	17	1	1	20.0%
	20	1	1	20.0%
92- Hauts de Seine (total)		**14**	**14**	**100.0%**
	1	4	4	28.6%
	2	4	4	28.6%
	3	6	6	42.9%
93- Seine St-Denis (total)		**8**	**8**	**100.0%**
	1	3	3	37.5%
	2	3	3	37.5%
	3	2	2	25.0%
94- Val de Marne (total)		**6**	**6**	**100.0%**
	1	2	2	33.3%
	2	0	0	0.0%
	3	4	4	66.7%

Table 4. Effect of Availability on Fruit and Vegetable Consumption: Probit Regressions

Availability Measures		Dependent Variable: FV No Additional Controls	Region Control	Region + Demographic Controls	Region + Demographic Controls + Health Knowledge
		(A)	(B)	(C)	(D)
No of stores IRIS	(1)	-0.0111 (0.0132)	-0.0122 (0.0131)	-0.0103 (0.0128)	-0.0094 (0.0128)
No of stores TRIRIS	(2)	-0.0103 (0.0082)	-0.0091 (0.0081)	-0.0086 (0.0080)	-0.0081 (0.0079)
No of stores Resid. Circle	(3)	0.0002 (0.0010)	-0.0004 (0.0010)	0.0001 (0.0010)	0.0002 (0.0010)
Food Area Resid. Circle	(4)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Berry Index Resid. Circle	(5)	0.0377 (0.0465)	0.0599 (0.0469)	0.0540 (0.0457)	0.0607 (0.0456)
No of stores Metro Circle	(6)	0.0003 (0.0016)	-0.0008 (0.0016)	-0.0004 (0.0015)	-0.0003 (0.0015)
Food Area Metro Circle	(7)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Berry Index Metro Circle	(8)	0.0790* (0.0450)	0.1139** (0.0457)	0.0755* (0.0447)	0.0839* (0.0446)

Each coefficient corresponds to a separate regression. Demographic controls include age, gender, education, employment, income, presence of children in the households and diet. The regional control is an indicator variable equal to 0 if Paris, and 1 otherwise. Health knowledge control is represented by diet and the knowledge of dietary guidelines. Sample size is 2963 for models (1) – (4), (6) and (7); 2903 for (5) and 2843 for (8). Robust standard errors are in parentheses.

Table 5. Effect of Availability on Fruit and Vegetable Consumption: IV Regressions

			1st Stage			2nd Stage
Endogenous Variable	**Model**	**Excluded Instrument**	**Instrument Parameter Estimate (Standard Errors)**	**F-test of Excluded Instrument (Standard Errors)**	**Wald Test of Exogeneity Chi-Sq (Standard Errors)**	**Marginal Effects (Standard Errors)**
No of stores IRIS	(1)	Approved_Arr	-0.0062*** (0.0017)	13.13 (0.0003)	6.96 (0.0083)	-0.3321*** (0.0542)
No of stores TRIRIS	(2)	Approved_Arr	0.0091*** (0.0028)	10.49 (0.0012)	6.92 (0.0085)	0.2234*** (0.0387)
No of stores Resid. Circle	(3)	Approved_Arr	-0.1559*** (0.0231)	45.28 (0.0000)	6.79 (0.0092)	-0.0187*** (0.0057)
Food Area Resid. Circle	(4)	Approved_Arr	30.4806*** (10.3813)	8.59 (0.0034)	6.61 (0.0101)	0.0001*** (0.0000)
Berry Index Resid. Circle	(5)	Approved_Arr	0.0025*** (0.0005)	23.50 (0.0000)	5.06 (0.0245)	0.9806*** (0.2636)
No of stores Metro Circle	(6)	Approved_Arr / Commuter	-0.0828*** (0.0165)	24.95 (0.0000)	6.23 (0.0126)	-0.0327*** (0.0090)
Food Area Metro Circle	(7)	Approved_Arr	37.6907*** (7.3265)	26.38 (0.0000)	6.59 (0.0103)	0.0001*** (0.0000)
Berry Index Metro Circle	(8)	Commuter / Approved_Arr	0.0000*** (0.0000)	6.87 (0.0088)	2.10 (0.1478)	-1.1607** (0.5204)

*** Indicates significance at 1% level; ** Indicates significance at 5% level. In models (6) and (8) both instruments render almost identical marginal effects.

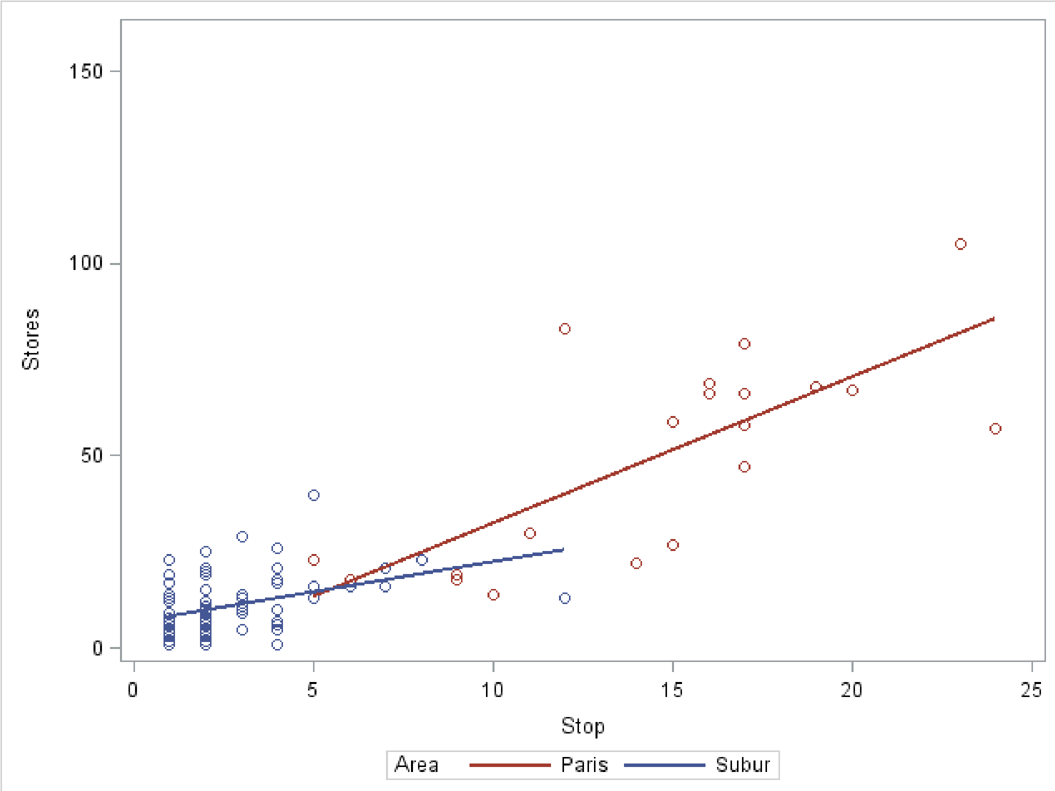

Figure 1. Metro/rail stations and number of stores by region.