EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE   **Open Access**

On the interplay between educational attainment and nutrition: a spatially-aware perspective

Narges Azizi Fard[1], Gianmarco De Francisci Morales[2], Yelena Mejova[2] and Rossano Schifanella[1,2*]

*Correspondence:
rossano.schifanella@unito.it
[1] Computer Science Department,
University of Turin, Corso Svizzera
185, Turin, Italy
[2] ISI Foundation, Via Chisola 5, Turin,
Italy

Abstract

Food choices are an integral part of wellbeing and longevity, yet poor nutrition is responsible for millions of deaths every year. Among the complex mosaic of determinants of food choices are demographic, socioeconomic, physiological, and also cultural. In this work, we explore the connection between educational attainment, as a proxy for cultural capital, and food purchases, as a proxy for food consumption. Unlike existing studies, which use diaries and surveys, we use a large-scale dataset of food-related products purchased from a major retailer in London over the course of one year. By using this high-resolution dataset, we are able to explore the spatial dependence of the various factors impacting food choices, and estimate their direct and indirect spatial effects. We characterize food consumption across two complementary dimensions of (1) diet composition, and (2) diet variety. By building spatial auto-regressive models on these variables, we obtain an improved fit compared to traditional regression, and illustrate the importance of spillover effects. Our results consistently confirm the association between a higher educational attainment and a healthier diet, even when controlling for spatial correlation. First, a low educational level is connected to diets high in carbohydrates and low in fibers. Second, it is also associated with higher consumption of sweets and red meats, while high educational level is linked to a greater consumption of fruits, vegetables, and fish. Third, highly-educated areas show an increased nutritional diversity, together with a lower caloric intake. Finally, we show the presence of spillover effects within the neighboring communities, which would need to be taken in consideration when designing public health policies and interventions.

Keywords: Food choices; Cultural capital; Educational attainment; Spatial analysis

1 Introduction

Globally, in 2017 approximately 11 million deaths and a loss of 255 million disability-adjusted life-years were attributable to poor diet [1]. In the UK, 63% of the adult population are overweight and 27% are obese.[1] Given that nutrition is a key determinant for

---

Springer

health and wellbeing, understanding food choices is of paramount importance. The process that leads to food choices is a complex one, which has been studied from several points of view, ranging from the biological [2, 3], to the demographic [4–6] and socioeconomic [7], focusing predominantly on developed Western countries. However, only recently has the cultural aspect started to garner attention [8–10]. Culture can be seen as a sort of collective memory that influences individual behaviors [11]. Food choices can be reinforced via the existence of "lifestyle enclaves" [12], where small preferences get amplified by elective affinities. For instance, occupation might influence food intake through work-related social circles [13, 14].

In this work, we take inspiration from the theory of the French sociologist Pierre Bourdieu, which links health and lifestyle to social class identity [15]. This identity is then differentiated through "taste" in music, art, and—in our context—culinary preferences. The theory connects taste to the pursuing of cultural capital, a non-material resource that accumulates throughout the life course [16]. Bourdieu introduced three forms of cultural capital [17]: *incorporated*, e.g., skills, competencies, personal effort, and time investment, *objectified*, e.g., possession of books, dictionaries, instruments related to artistic expressions, and *institutionalized*, e.g., educational attainment at the level of individual or family.

We focus on the third form of cultural capital, operationalized as the level of educational attainment. In particular, we show how education can explain inequities in food choices that go beyond socioeconomic determination and the consequent barriers to food access due to cost. To this extent, there is general agreement among researchers [4, 18, 19] that education and income are distinct concepts that are likely to make separate and unique contributions to health outcomes [20]. Education has been widely studied in connection to health and food choices for several reasons. First, it may provide the tools to access and comprehend dietary information and its impact on health. Second, social diffusion theory suggests that highly educated people generally take up innovations sooner than less-educated [21], which might affect the success of health-related interventions. Third, is can affect time use and thus the opportunities to allocate time to food acquisition and preparation.

Existing studies around cultural capital and nutrition use small-scale, traditional methodologies [10, 22–25] such as food diaries, questionnaires [26, 27], and surveys [21, 28]. As such, existing datasets have not allowed the modeling of spatial clustering, which is an important confounder for both culture and health, as it has been shown in other fields [29]. The presence of this confounder may invalidate the associations discovered by conventional methods such as OLS regression, and produce estimates that are biased and inconsistent. This confounder occurs when individual data points are not independent and identically distributed, rather they are spatially correlated. Observed variation in the dependent variable may thus arise from latent influences related to culture, infrastructure, recreational amenities, and a host of other factors not present in the data. Modeling these latent influences is important especially because many health interventions work at a community level, such as "farm to school" programs [30], food literacy initiatives [31], and community gardens [32].

In this work, we use a unique fine-grained dataset that allows us to reveal the role of space and geography in modeling the relationship between food consumption and culture. Unlike traditional approaches, we use a large-scale log of the food-related purchases of 1.6M customers of a major retail chain in London across the time span of a year [33].

This resource allows us to observe—in an indirect way—the daily food choices at an unprecedented granularity, with the assumption that supermarket purchases largely represent the dietary intake of a household.[2] This granularity enables the exploration of two complementary dimensions that capture multiple facets of food consumption: (1) diet composition, along the dimensions of macronutrients and product categories, and (2) diet diversity.

Through the use of spatially-aware regression models, which include multiple environmental features of an areal unit, we illustrate the effects of spatial clustering in the emergence of localized communities with homogeneous behavior, and the advantages over standard regression approaches on fitting performance and biases in the estimates. We further show the presence of spillover effects within neighboring communities, all of which need to be considered when designing policies and interventions, or informing decision support systems for public health.

2 Related work

A large body of literature explores the interplay between socioeconomic status and food consumption, showing how inequities in the access to resources, privilege, or power, play a role in shaping people's dietary habits. The dimension of wealth, often modeled with the average household income or employment rate, has been connected to the ability to afford certain categories of food products. Cost has been mentioned as one of the main obstacles to a widespread consumption of fruits, vegetables [34, 35], or lean meats [36], and connected to the lack of a healthy diet [37]. The individual and household disadvantages are compounded by residential segregation. For example, disadvantaged communities often face spatio-structural barriers to access to food [38], as well as to physical activities, fitness clubs, and weight loss programs [36], all of which have effects on health outcomes. As a result, disadvantaged communities have a higher incidence of obesity than wealthy ones [39]. In this direction, a vast body of literature studies the relationship between food environment and diet to capture the degree of food access using both respondent-based perceived measures and quantitative approaches that leverage Geographic Information System (GIS) technology [40]. The latter commonly use store density (using buffer distances), or proximity to the nearest food store to operationalize food access [41]. Another common method involves store audits, in which researchers estimate the shelf-space occupied by certain foods in a store, or assess product variety and food prices within stores using measures such as the Nutrition Environment Measure Survey (NEMS) [42]. Several food environment conceptualizations have been proposed, mainly divided into community food environment and consumer food environment [43] that draw attention to the distribution of food sources within a community or within a local retailer, respectively. It has been suggested that a largely accepted food access conceptualization involves 5 dimensions relevant in the healthcare setting: availability, accessibility, affordability, acceptability, and accommodation [41]. Refer to [44] for a literature review. In addition, spatial modeling and GIS have been adopted to characterize additional dimensions of the food ecosystem. For example, Khushi et al. [45] investigated spatial inequality in food consumption, nutrient consumption, and production-consumption gaps at the sub-national levels

---

[2]The Family Food module of the UK Living Costs and Food Survey 2018/2019 (https://www.gov.uk/government/publications/family-food-201819/family-food-201819) that characterizes the household shopping and eating habits through questionnaires, indicates how the average expenditure on food and drink consumed at home, per person per week, represents 69% of the total spending.

in the Punjab province of Pakistan. Dohyeong et al. [46] characterized the spatial patterns of unhealthy food consumption in South Korea and they modeled the presence of areas with constrained access to fresh and nutritious foods, providing guidelines for targeted nutrition and public health programs. Moreover, methods for mapping provincial spatial food consumption data by accounting for spatial variability in population structure (age and gender) have been proposed in [47] with the intent to inform policy makers interested in promoting the consumption of locally produced food, as assessing localized nutritional demand. However, in contrast with our analysis, all these studies are based on corse geographical units such as administrative districts or regions, making it hard to address the variability in consumption within localized communities, e.g., neighborhoods, that often show a pronounced diversity especially in multi-cultural and multi-ethnic megacities.

Along with the socioeconomic dimensions, it is broadly acknowledged how the cultural group to which one belongs is of great importance when it comes to food preferences [48]. In the 1980s, the French sociologist Bourdieu [15] proposed a theory on the relationship between material and non-material capital to explain social inequalities, stratification and the distribution of power. Bourdieu connected taste, a multidimensional concept involving attributes, such as musical, artistic and culinary preferences, to the pursuing of cultural capital, a non-material resource that accumulates throughout one's life course [15].

Even though several studies attempted to quantitatively characterize the different forms of cultural capital and their relation to food choices [10, 22–25], they were mainly based on interviews and questionnaires on small samples of the population and potentially affected by common biases [49] related to the way a question is designed or administered. Kamphuis et al. [16] performed a systematic review of cultural capital indicators; they identify several indicators of family institutionalized (e.g. parents' education completed) and objectivized (e.g. possession of books, art) or incorporated cultural capital (e.g. cultural participation, skills). After designing a questionnaire to capture these dimensions along with food habits of the participants, they found evidence of a connection between cultural capital and healthy food choices. The link between healthy diet and cultural capital has been observed in several studies. For instance, in a study of a cohort of adolescents in Norway, Fismen et al. [50] identified cultural capital as a stronger predictor than material capital of disparities in consumption of fruit and vegetables (positively correlated), and it was the only significant predictor of consumption of sweets and sugared soft drinks (negatively correlated).

Institutionalized cultural capital in the form of educational attainment has been widely adopted by the studies that focused on modeling cultural inequities and food choices, since level of education arguably affects what type of social milieu people inhabit, and consequently it affects what type of food one is exposed to [15]. Moreover, the availability of aggregated data at a fined-grained geographical scale, usually from the census, is another driving reason of this choice, one that we also embrace in this work. A low educational level has been connected to diets higher in fat density [51–53], ultra-processed and ready made foods [54], sugar-rich [55] products, meat products (especially red meat) [21], and to a lower food group variety [8, 21]. On the contrary, highly educated people tend to consume more fruits and vegetables [55], fish [56], and to follow a more diverse diet. At last, social diffusion theory suggests that highly educated people generally take up innovations sooner than less-educated people. For example, in the UK, foods and diets low in saturated fat were adopted by the tertiary-educated before others [57]. Social epidemiologists also

suggest that education enables people to rise up the social class hierarchy, thus allowing them greater power over outcomes in their lives, for example through higher incomes. In this study, we aim to re-examine some of these trends using a new high-granularity, large dataset of nutritional behavior.

3 Methods

3.1 Data

In this work, we characterize food consumption by using the Tesco Grocery 1.0 dataset [33] that contains an anonymized record of 420M food items purchased by 1.6M fidelity card owners who shopped in one of the 411 Tesco stores in Greater London during 2015. Tesco is the largest food retailer in UK with around 30% market share and a solid geographical coverage in the area of study. The dataset contains aggregated and privacy-preserving data views that combine individual purchases at different spatial granularities by using the home location field from the loyalty card application as the way to geolocate customers. The fine-grained geographical information included in Tesco Grocery 1.0 is the key to link food consumption data to any attribute that can be measured at the level of statistical census areas, e.g., demographic, socioeconomic, and health determinants. In [33], the authors provide an analysis of the representativeness of the Tesco consumers base by comparing the number of unique customers to the general population, and report a solid match. Moreover, they prove the ecological validity of the dataset by comparing the grocery purchases with metabolic syndrome conditions that are strongly linked to food consumption habits. An in-depth discussion on sample bias is provided in the Limitations section of the current work.

To model food consumption, we focus on three groups of variables of interest: *macronutrients*, *product categories*, as a proxy for diet composition, and *diet variety*. The first captures the nutritional properties of a food product and it is connected to the concept of energy intake that we measure in calories. In fact, a food item contains different types of nutrients in different proportions, which are transformed by the human body into energy and structural material for its growth and maintenance. We consider the following nutrients that have been connected to diet and culture in the literature: *fats*, *carbohydrates*, *proteins*, and *fibers*. A few studies distinguish between different types of fats, e.g., saturated fats that are fat molecules that have no double bonds between carbon molecules because they are saturated with hydrogen molecules. However, in our work, we model the lipids intake within a single macro category, this is justified also by the high degree of correlation observed between the variables *fats* and *saturated fats* in the Tesco dataset ($\rho_{\text{fats}}^{\text{saturated fats}} = 0.8$, p-value < 0.001). A similar observation holds for *sugar* that loosely refers to a number of carbohydrates, such as monosaccharides, disaccharides, or oligosaccharides, as its excessive consumption has been implicated in the onset of obesity, diabetes, cardiovascular diseases, dementia, and tooth decay [58–60]. Accordingly, we do not consider sugar separately in the experimental setting due to the strong correlation with the macronutrient carbohydrates ($\rho_{\text{carbohydrates}}^{\text{sugar}} = 0.85$, p-value < 0.001).

The second group of variables is related to a classification of the food products into categories. Even though several food taxonomies have been proposed in the literature [61], there is no consensus on how to group foods [62]. In this work, we adopt the classification in [33] which focuses on the following categories: *oils*, *fish*, *produce*, *red meats*, *readymade*, and *sweets*. Among the food categories, we observe two pairs of highly correlated variables, *poultry* and *red meats* ($\rho_{\text{poultry}}^{\text{red meats}} = 0.77$, p-value < 0.001), and *sweets* and

grains ($\rho_{\text{sweets}}^{\text{grains}}$ = 0.79, `p-value` < 0.001). As such, in the rest of the experiments, we use *red meats* to characterize food products containing animal flesh, and *sweets* to include baked products using grain flours and, for example, candies or chocolate. The *produce* category includes fresh vegetables and fruits, while *readymade* contains pre-cooked meals that are usually available in a specific area of the store and that need to be open, often warmed-up, and eaten. At last, to test the hypothesis that eating a wide variety of foods improves dietary adequacy, we adopt the normalized entropy of the distributions of nutrients (*h_nutrients*), and food groups (*h_products*) as proxies for variety. Moreover, we consider the *weight*, and the calories intake (*energy*) of the average product sold in an area as a measure of quantity. At this stage, we have for each spatial unit a characterization of the diet quality and variety in that area that could be potentially linked to census variables. It is worth noting that we represent the nutritional features of the hypothetical average product consumed in an area, since we cannot characterize individual or group behaviors.

We characterize the educational attainment variables by using the 2011 Census data and, in particular, the table *Highest level of qualification by sex*[3] in the *Local Characteristics* series. The highest level of qualification is derived from the question asking people to indicate the types of qualifications held. The following levels are available: *no qualifications*, *level 1*, *level 2*, *apprenticeship*, *level 3*, *level 4 and above*, and *other qualifications*. In this work, we restrict our analysis to a representative class for the low-educated (level 1) and high-educated (level 4+) population strata. This design choice implies a linear approximation of the effect of education, which might hide more complex effects of the distribution of education levels. However, we think that the interpretability of the model, and its applicability, offer a good tradeoff for this simplifying assumption. To account for confounding variables that could influence both education and than food choices, we explore the dimensions of gender, age, and income, which have been extensively linked to dietary habits in previous work. Gender and age, in the form of average age, are extracted from the census, while economic status is quantified via a model-based estimate of the equivalized net income per household after the deduction of housing costs.[4] All the variables are standardized with zero mean and unit variance. A summary of all the variables used in this study can be found in Table 1, and a characterization of the spatial distribution and cross-correlation is presented in the Additional file 1.

Data is aggregated at the geographical level of administrative areas in UK, thus implementing a privacy-preserving methodology. In particular, we adopt as a reference the spatial units of Middle Super Output Areas (MSOA), that have an average population of 7200 [63] and an average surface of 1.6 square kilometers within the Greater London region. In contrast with previous work, MSOA provide a considerably finer granularity that enables community-level observations. We refer to the variable `representativeness` (mean = 0.37, σ = 0.16) defined in [8] as the min-max normalized ratio between the number of unique customers and the number of residents to characterize how much the user base captures the census statistics. We limit our analysis to the 774 out of 983 geographical units that have a `representativeness` $\geq$ 0.15 (see Fig. 1 for more details on the

---

[3]https://www.nomisweb.co.uk/census/2011/local_characteristics

[4]Income estimates for small areas, England and Wales. Additional details on the methodology could be found here.

Table 1 Description and source of the variables used for characterizing food consumption, educational attainment, and socioeconomic confounding factors

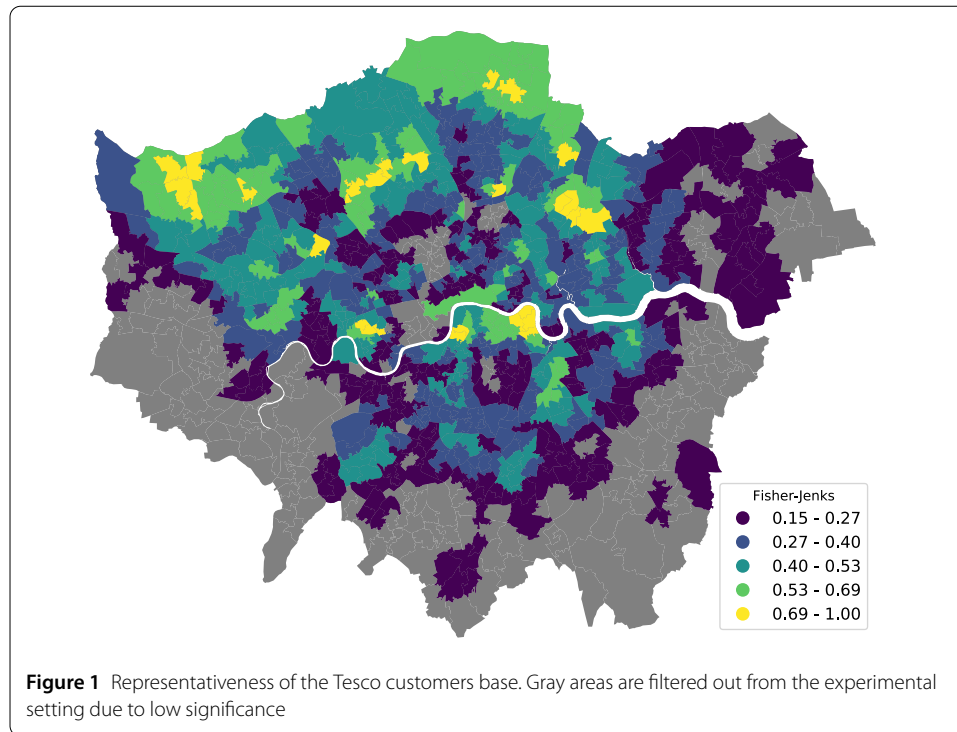| Variable	Source	Description
Nutrients		
fats	[33]	grams of fats contained in the typical product
proteins		grams of proteins contained in the typical product
carbs		grams of carbohydrates contained in the typical product
fibers		grams of fibers contained in the typical product
Product Categories		
oils	[33]	fraction of products in the fats and oils category
fish		fraction of products in the fish category
produce		fraction of products in the fruits and vegetables category
sweets		fraction of products in the sweets category
red meats		fraction of products in the red meats category
readymade		fraction of products in the readymade category
Variety		
h_products	[33]	Shannon entropy of the distribution of the items belongings to the different products categories
h_nutrients		Shannon entropy of the distribution of the calories given by all the nutrients
Misc		
weight	[33]	average food product's weight (not including drinks)
energy		average number of calories in a food product
Education		
low (level1)	census	1-4 O Levels/CSE/GCSEs (any grades), Entry Level, Foundation Diploma, NVQ Level 1, Foundation GNVQ, Basic/Essential Skills
high (level4+)	census	Degree (e.g., BA, BSc), Higher Degree (e.g., MA, PhD, PGCE), NVQ Level 45, HNC, HND, RSA Higher Diploma, BTEC Higher level, Foundation degree (NI), Professional qualifications (e.g., teaching, nursing, accountancy)
Socioeconomic		
female	census	fraction of the female population
income	census	estimate of the equivalised net income per household after the deduction of housing costs
age	census	average age

spatial coverage). The hierarchical structure and the shapefiles of the various census units are provided by the Open Geography Portal[5] of the Office for National Statistics (ONS).[6]

3.2 Spatial modeling

Since a large extent of socioeconomic and cultural phenomena are driven by spatially-aware data generation processes, and the modeling of food consumption has been rarely addressed in space, we approach our problem with spatial econometrics tools. To formalize the neighboring relation between areal units we refer to spatial weights [64] and we adopt a contiguity approach based on the binary *queen* criterion where $w_{h,k} = 1$ if the areas h and k share at least one vertex, 0 otherwise. To test the sensitivity of the results to the choice of a different spatial arrangements, we explore alternative weights structures based on distance, in particular the *k-nearest neighbors (knn)*, where each area has a fixed number of k closest neighbors, and approaches based on kernel functions with adaptive

---

[5]http://geoportal.statistics.gov.uk

[6]https://www.ons.gov.uk/

Figure 1 Representativeness of the Tesco customers base. Gray areas are filtered out from the experimental setting due to low significance

bandwidth. To diagnose the presence of spatial dependence in the outcome variables or in the residuals of the regressors, we rely on the global measure of spatial autocorrelation Moran's I [65] that tests how a variable in space differs significantly from the expected value under the null hypothesis of spatial randomness.

In this work, we start from the following vector notation of the general linear spatial econometric model for cross-sectional data:

$$Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + u, \quad \text{where } u = \lambda Wu + \varepsilon. \tag{1}$$

Y denotes an $N \times 1$ vector with the observations of the dependent variable for every spatial unit in the sample ($i = 1, \ldots, N$), ι_N is a $N \times 1$ vector of ones associated with the constant parameter α, X contains a $N \times K$ matrix of exogenous explanatory variables, with the associated parameters β represented in a $K \times 1$ vector, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^T$ is a vector of disturbance terms, where the ε_i are independently and identically distributed error terms with zero mean and variance δ^2. W denotes an $N \times N$ nonnegative matrix describing the spatial arrangement of the units in the sample. The model specifies three main terms: (1) an *endogenous* interaction effect ρWY, (2) an *exogenous* interaction effect $WX\theta$, and (3) an interaction effect amongst *error terms* λWu. They model, respectively, the interplay between the value of the dependent, independent, and error terms in a spatial unit i and the values of the other spatial units. From the general nested model, the configuration of the parameters ρ, θ, λ leads to different spatial model specifications. For example, $\lambda = 0$, while removing the lagged errors, leads to the definition of the Spatial Durbin Model [66, 67] (SDM), that resolves in a Spatial Lag Model [68] (SLM) when $\theta = 0$. In a similar way, nullifying the lagged dependent variable component ($\rho = 0$) defines the Spatial Error Durbin Model [69] (SEDM) and the simpler Spatial Error Model [68] (SEM) when

also $\theta = 0$ (the spatial dependence is modeled via the error term alone). When all the ρ, θ, λ parameters are null, the specification traces back to a standard linear regression. In this settings, the interpretation of results involves two main components: a *direct* impact that links the characteristics of spatial unit with the value of the dependent variable in the same unit, and an *indirect* impact that models the spillover effects. The spillover effects can be further categorized in *local* or *global*. In local spillovers ($\rho = 0$) the indirect effects are measurable only in the neighboring units, e.g., areas where $W_{ij} \neq 0$. This leads to the adoption of a SEDM or SEM specifications. In contrast, in global spillover effects ($\rho \neq 0$) the indirect influence of a spatial unit falls on the entire set of locations, producing high-order effects observable even in spatial units that are not directly connected. This is compatible with the SDM or the SLM specifications.
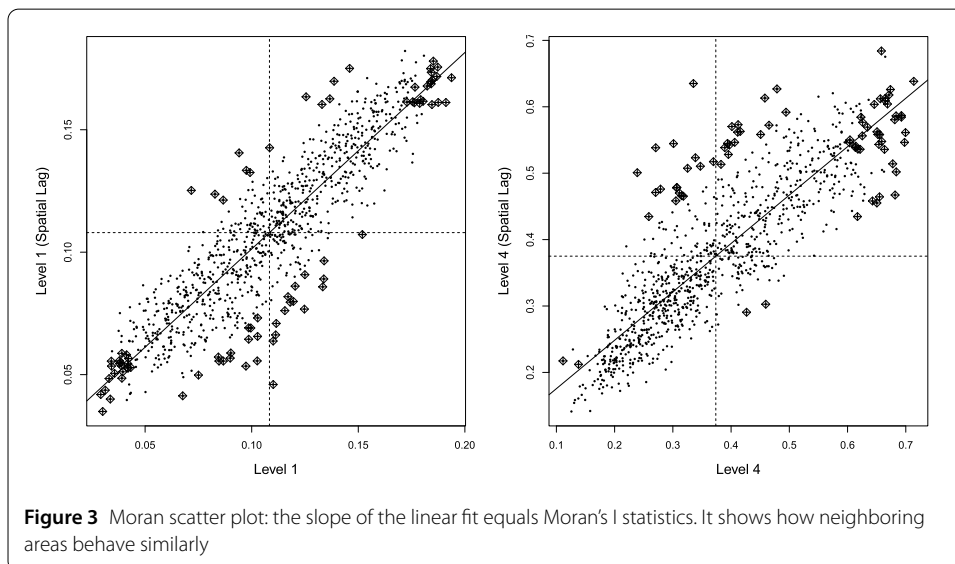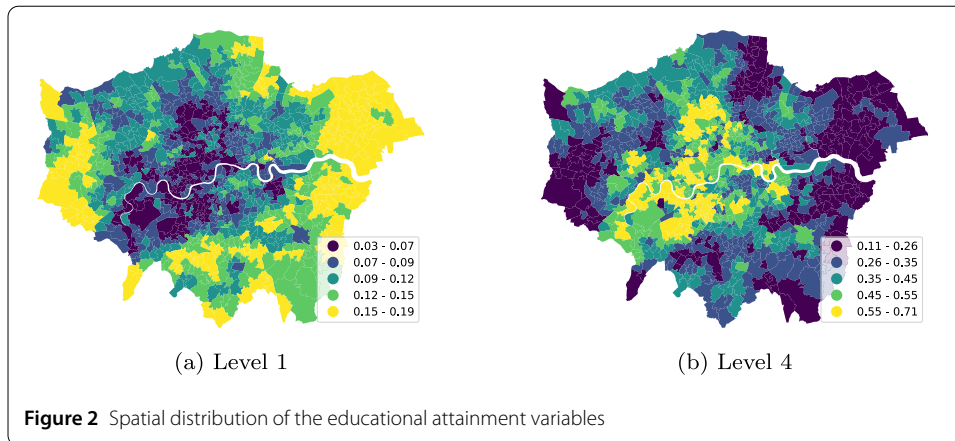
In the literature, different approaches have been adopted to find the most appropriate spatial econometric model specification given an empirical use case. In this work, following the discussion in [70], we use a theoretical rather than data-driven approach to guide the decision. In the case of food choices, it is difficult to form a reasonable argument to include endogenous interaction effects even though they are found statistically. Including endogenous interaction effects would imply that the consumption of a particular food in an area has effects on the entire city, which is difficult to justify. The scenario points to a local spillover specification and, accordingly, we focus our analysis on the SDEM and SEM models. For completeness, we also run the Lagrange Multiplier [71] (LM) diagnostics based on the OLS residuals in their standard and robust forms that are at the core of the methodology described in [72].

For the experiments we use the functions `errorsarlm` and `lagsarlm` of the `spatialreg`[7] package in R.

4 Results

To assess the presence of spatial dependence in the outcome, we run a permutation test for the Moran's I statistic in the case of the educational attainment variables `level1` and `level4` with the number of random permutations $n = 10{,}000$. We observe a strong positive autocorrelation in both cases $I_{level1} = 0.7202$, `p-value` < 0.0001 and $I_{level4} = 0.684$, `p-value` < 0.0001. The presence of clusters of areas with similar behavior is confirmed by the visual inspection of the spatial distribution of values shown in Fig. 2 and from the Moran I scatterplot in Fig. 3. It is worth noting how the two variables show complementary spatial patterns: in fact, the areas with a high prevalence of a low-educated population (dark areas in the left map) correspond to the areas with a low presence of high-educated population (yellow areas in the right map) and vice versa. In this direction, we mainly focus on modeling the high education outcome (`level4`), and we show how the complementary variable for low education (`level1`) performs when space allows. To test if the observed spatial autocorrelation could be explained by the spatial structure of the covariates alone, we first run a linear regression analysis and we check for the presence of autocorrelation in the residuals [72]. Fitting performance is evaluated with the Akaike Information Criterion [73] (AIC), the Bayesian Information Criterion [74] (BIC), and the Nagelkerke's pseudo R^2 [75] when appropriate.

---

[7]https://r-spatial.github.io/spatialreg/index.html

(a) Level 1　　　　　　　　　　　　(b) Level 4

Figure 2 Spatial distribution of the educational attainment variables

Figure 3 Moran scatter plot: the slope of the linear fit equals Moran's I statistics. It shows how neighboring areas behave similarly
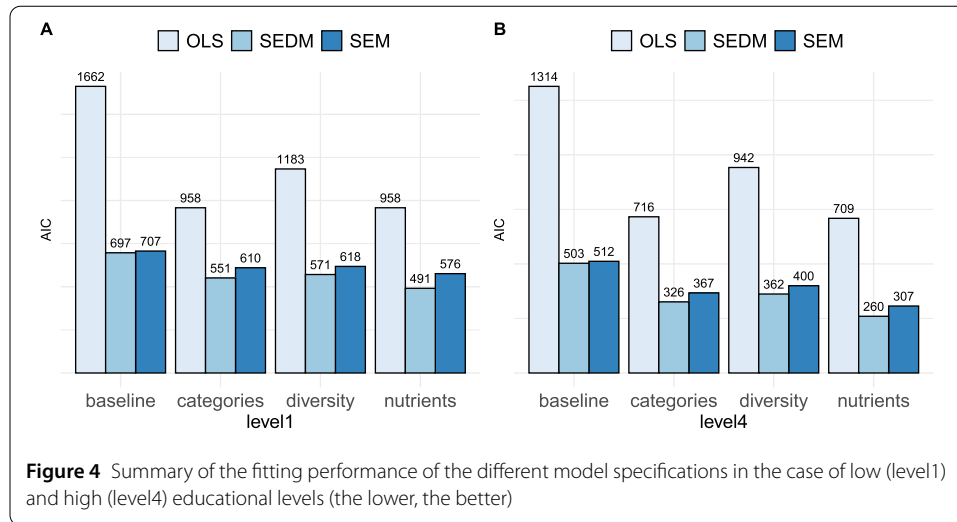
We organize the rest of this section in two main parts that follow the same methodological pipeline and that capture the complementary dimensions of food consumption: (1) diet composition, along the dimensions of macronutrients and product categories, and (2) diet diversity.
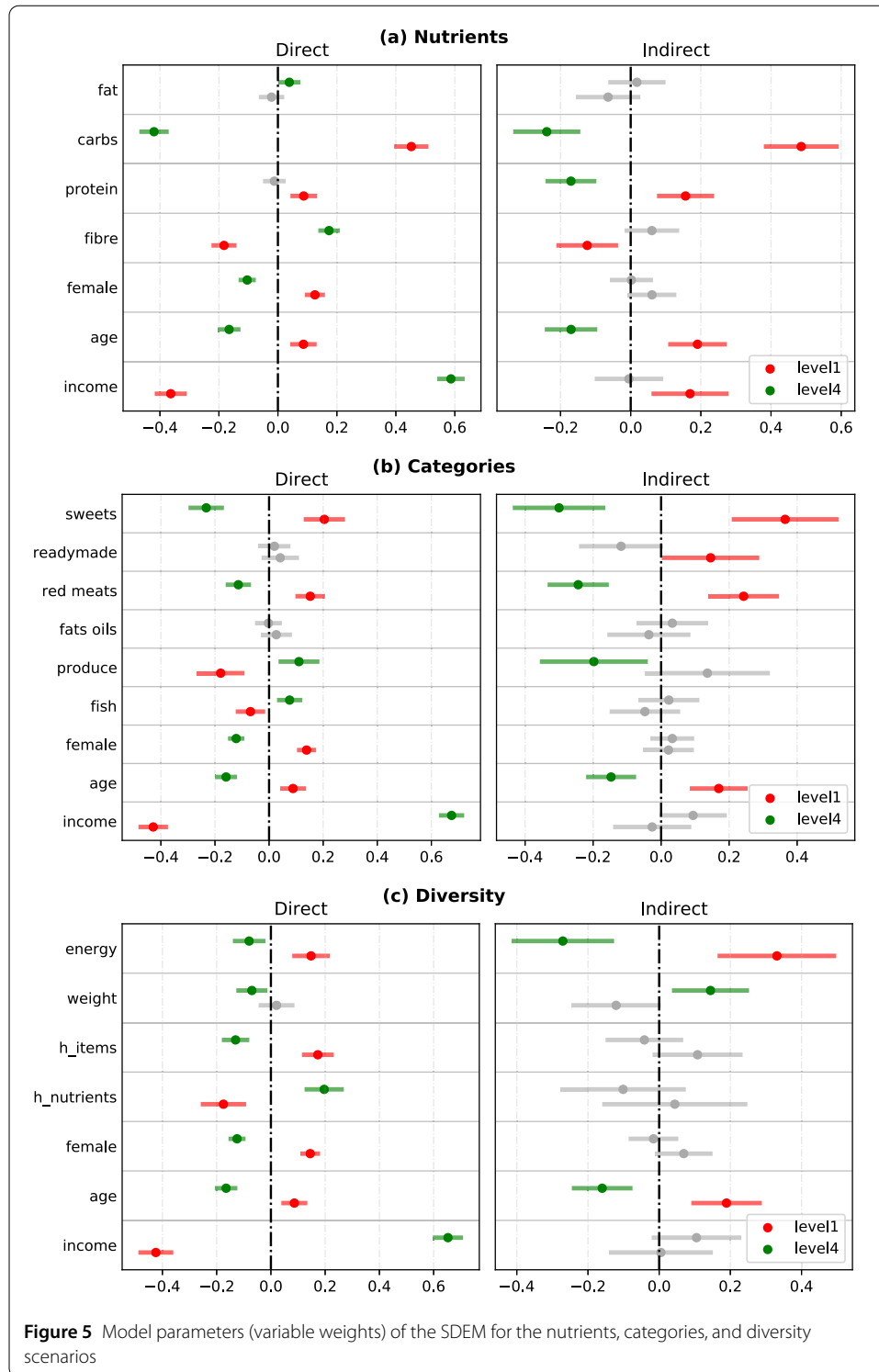
4.1 Diet composition

Macronutrients In this section, we explore the interplay between educational attainment and the consumption of different macronutrients. We first compute a linear regression model observing the presence of significant spatial autocorrelation in the residuals (I_{level4} = 0.4409, `p-value` < 0.0001) (refer to Additional file 2 for more details). To account for local spillover effects, we estimate the SEM and SDEM models. The spatial specification choice is coherent with the results of the robust Lagrange Multiplier tests, $\text{LM}^{\text{error}}_{\text{level4}}$ = 154.32, `p-value` < 0.0001 and $\text{LM}^{\text{lag}}_{\text{level4}}$ = 39.549, `p-value` < 0.0001 that suggests the adoption of a lagged error specification due to the higher value of the corresponding statistics. Figure 4 summarizes the fitting performance of the alternatives tested, thus identifying SDEM as the best performing model across measures (AIC = 490). A likeli-

Figure 4 Summary of the fitting performance of the different model specifications in the case of low (level1) and high (level4) educational levels (the lower, the better)

hood ratio test using the function `LR.sarlm` in the `spatialreg` package confirms the goodness of the choice of SDEM over SEM (LR = 61.077, p-value < 0.0001).

In Fig. 5(a) we present the SEDM regression results for the target variables `level1` and `level4`. In both cases, the parameter λ is highly significant (λ_{level1} = 0.67 and λ_{level4} = 0.71, p-value < 0.0001), thus confirming the presence of a strong spatial error lag in the empirical data. The group of variables on the left side defines the direct impact while the variables on the right side estimate the indirect effect of the neighboring spatial units as defined in Sect. 3. Moreover, we test for the presence of spatial autocorrelation in the residuals, observing respectively a Moran's I_{level1} = −0.02, p-value = 0.8 and I_{level4} = −0.024, p-value = 0.83 that show how the SEDM, unlike a standard linear regressor, produces an uncorrelated spatial structure in the residual in accordance with the hypothesis of independence. Finally, we observe the presence of heteroscedasticity via a studentized Breusch–Pagan [76] test (BP$_{\text{level4}}$ = 69.6, p-value < 0.0001). While heteroscedasticity does not affect the estimation of the coefficients, it biases the estimation of the significance; however, since the observed p-values have, in the most cases, values below 10^{-3}, the net effect is not substantial.

Product categories   We explore the dimension of the food categories following the same pipeline described in the previous section. The residuals of a linear regressor shows a significant spatial autocorrelation (I_{level4} = 0.41, p-value < 0.0001), and the LM tests are coherent with the current choice of spatial specification (LM$_{\text{level4}}^{\text{error}}$ = 112.5, p-value < 0.0001 and LM$_{\text{level4}}^{\text{lag}}$ = 60.7, p-value < 0.0001). SDEM shows the best fitting performance also for the case of food categories (see Fig. 4), in accordance with a likelihood ratio test (LR = 59.7, p-value < 0.0001). Figure 5(b) summarizes the direct and indirect impacts for the fitted model, confirming significant spatial effects in the error terms (λ_{level1} = 0.69 and λ_{level4} = 0.7, p-value < 0.0001). The application of SDEM produces residuals free from spatial autocorrelation (Moran's I_{level1} = −0.015, p-value = 0.71 and I_{level4} = −0.02, p-value = 0.77). At last, we observe heteroscedasticity (BP$_{\text{level4}}$ = 58, p-value < 0.0001) similarly to the macronutrients case.

Figure 5 Model parameters (variable weights) of the SDEM for the nutrients, categories, and diversity scenarios

4.2 Diet variety

Spatial patterns are observed in the residuals of a linear regressor (Moran's $I_{\text{level4}} = 0.53$, p-value < 0.0001), and the LM tests are coherent with the current choice of spatial specification ($LM_{\text{level4}}^{\text{error}} = 197.4$, p-value < 0.0001 and $LM_{\text{level4}}^{\text{lag}} = 42.3$, p-value < 0.0001). Figure 5(c) summarizes the direct and indirect effects in the case of the best performing

model SEDM (see Fig. 4 in accordance with the likelihood ratio test LR = 52.4, p-value < 0.0001). The parameters λ_{level1} = 0.78 and λ_{level4} = 0.78, p-value < 0.0001 confirm the presence of significant spatial patterns. We acknowledge the presence of heteroscedasticity (BP_{level4} = 30.05, p-value = 0.008) and missing spatial autocorrelation in the residual of SEDM (Moran's I_{level1} = −0.03, p-value = 0.88 and I_{level4} = −0.03, p-value = 0.93).

4.3 Sensitivity analysis

In this section, we explore the extent to which the observed results are sensitive to changes in the experimental design. First, we focus on a comparative analysis between the baseline model with only the socioeconomic confounds and the complete model including the food choices variables. The interplay between educational attainment and socioeconomic determinants has been extensively studied in previous work and, consistently, we observe how they play a primary role in the predictive framework. However, as shown in Fig. 4, adding the food consumption dimensions does provide a significant improvement in the fitting performance. We measure a 30%, 21%, and 18% reduction of the AIC for the nutrients, food categories, and diet variety cases (level1) and, specularly, a 48%, 35%, and 28% reduction in the case of the high education outcome (level4). A consistent behavior is observed in relation to alternative performance metrics, e.g., BIC and Nagelkerke R^2 as summarized in Additional file 2. Moreover, it is worth noting that the spatial-aware models consistently outperform the standard linear regression framework by a large extent (on average we observe an improvements in AIC greater or equal to 75%) underscoring the benefits of taking into account the geographical structure of the determinants.

Second, we focus on the choice of the weighting scheme that has a central role in a spatial econometric framework [64]. We extend the initial experimental settings based on a contiguity approach with different distance-based spatial arrangements methods: the *k*-nearest neighbors (k-nn), in which each spatial unit is connected to a fixed number of *k* closest neighbors, and a class of kernel functions with adaptive bandwidth (gaussian, quadratic, triangular, quartic, and uniform). For simplicity, we present the nutrients and level4 case, similar results apply to the other scenarios. In the case of the nearest neighbors approach, we explore the range $k \in [3, 15]$ obtaining the best performing model with $k = 6$ (see Fig. 6(A)) and an overall performance that is slightly lower than the contiguity case (AIC = 278). We present the full results for the best performing model with $k = 6$ in Additional file 2 showing how the learned relations are stable and change only partially in strength. Switching to the kernel weights approach, we study the behavior of different classes of kernel functions exploring a bandwidth size within the same range of the k-nn scenario. Figure 6(B)–(F) summarize the observed performance curves showing similar results across methods. An extensive comparison between kernel functions is out of the scope of the paper, however, the best performing kernel settings is the triangular function with bandwidth size equals to 9 (AIC = 257) which reaches a very similar output to the contiguity case (AIC = 260). These results confirm the stability of the learned relations across a wide range of spatial arrangements.

5 Discussion

The first dimension of food choices that we explore is related to nutrients consumption. First, we focus on the direct impacts that model the effect of the intrinsic characteristics of a spatial unit on the educational attainment variable. As shown in Fig. 5(a), we

Figure 6 AIC fitting performance under different weighting schemes
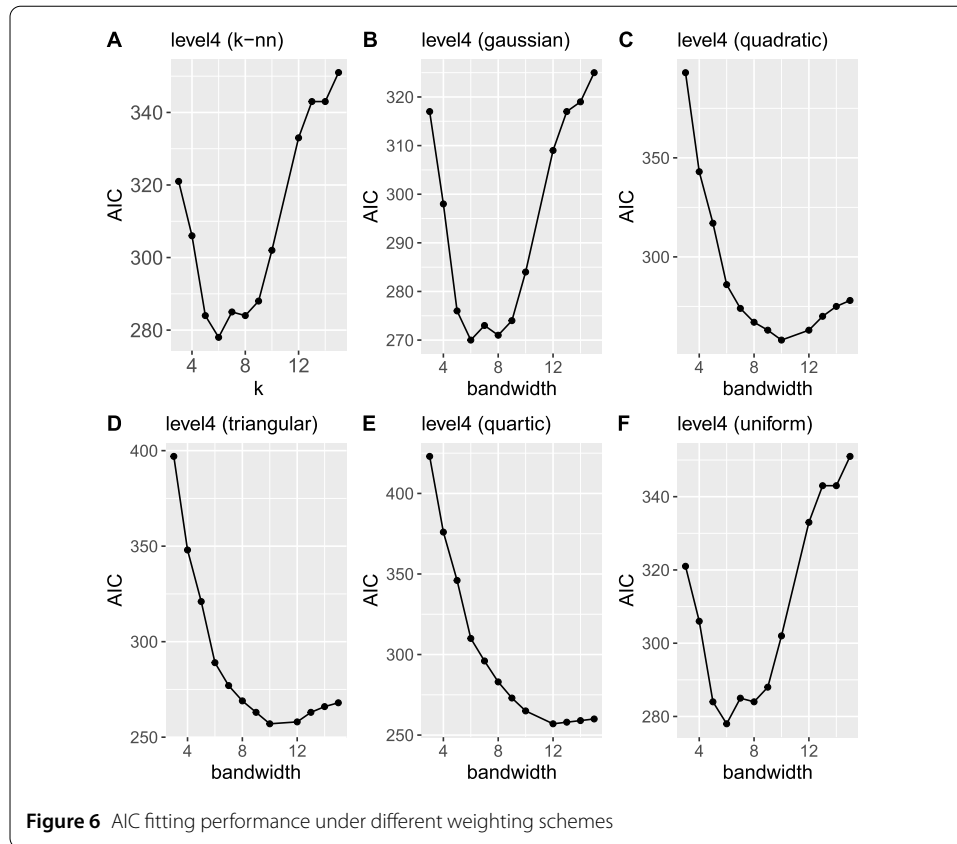
observe that a low educational level is connected to diets high in carbohydrates [55], including sugar. Conversely, areas with a predominance of highly educated residents show a higher consumption of fibers, which provides a range of important health benefits, particularly in preventing heart and cardiovascular diseases, stroke, hypertension, diabetes, obesity, and some gastrointestinal pathologies [77–79]. Diets higher in fat density have been associated to lower education in several studies [51–53], which is consistent with the empirical measure of rank correlation observed in our scenario (ρ_{level4}^{fats} = −0.36 and ρ_{level1}^{fats} = 0.37, `p-value` < 0.001). However, in a multivariate settings and discounting for the presence of the other predictors, the fat variable—due to its strong interplay with the carbohydrates variable (ρ_{fats}^{carbs} = 0.57)—appears to have a weak positive effect on the high education outcome that could be due to the presence of multicollinearity or misspecification of the model [80]. At last, we observe a non statistically significant direct association with protein consumption, this could be potentially related to the diverse set of sources of proteins and that have been associated in literature with different heath outcomes, socioeconomic factors, as well as impact on the environment. In fact, protein supply comes from both plant, e.g., legumes, soya, nuts and seeds, and animal sources, e.g., fish and seafood, poultry, pork and beef, and derivatives of milk such as dairy products. These products have been associated with low and high educational levels depending on their relation with a healthy diet; in this scenario, a definitive association is hard to pinpoint. When focusing on the spatial spillover effects, we do not observe significant neighboring effects for fiber and fat, while carbohydrates shows also in this case the stronger effect, that underlines how the neighboring units affect the model's decision in the same direction as

the observed direct effects. For the case of protein, the results show an indirect negative impact on the `level4` variable. We remark that the spatial autoregressive model does not necessarily capture a causal model, i.e., we do not assume that the food choices have a causal effect on the education attainment. Therefore, the interpretation of the indirect effect needs some care, as they represent a spatial spillover of the covariates and the dependent variable, rather than actual influence of the neighboring units. The direction of these spillover effects is in almost all cases in accordance with the direct effects, which is a confirmation of the robustness of the results. These spillovers may be caused by several modeling factors, including the granularity of the spatial discretization, the specific choice of neighborhood function, and the arbitrariness of the spatial borders.

Switching now to food categories, there is a vast literature discussing the interplay between food choices and socioeconomic factors, and to a lesser extent, cultural capital. A high intake of fruit and vegetables is one of the cornerstones of a healthy diet, and has been recommended to the general public to reduce the risk of cardiovascular, coronary hearth diseases, and stroke [81]. Consistently with previous work, we observe a positive association between high educational attainment and consumption of vegetables and fruits [50, 55, 82, 83]. The prevalence of sweets and, in general, products high in density of sugar, is evident in communities with a lower educational level [50], an effect that, similarly to the case of carbohydrates, is the strongest in intensity in our experimental scenario. Focusing on animal-based products, it is interesting to note the different behavior between the variables `fish` and `red meats`. While high-educated people are less likely to be regular consumers of several meat products [21] ($\beta = -0.114$), in particular for the case of processed meats, they tend to consume more `fish` [56] ($\beta = 0.076$), including seafood. Moreover, fish [84] and seafood [85] are an excellent source of protein and provide a range of benefits for major health outcomes among adults, even when taking into account that the presence of contaminants such as mercury in polluted natural environment could pose potential risks[84]. We observe non significant relations for the `readymade` and `oils` categories. In the first case, consumption of readymade and ultra-processed food products have been associated to the lower educational strata [54], however, even if the pairwise correlation follows this tendency ($\rho_{\text{level4}}^{\text{readymade}} = -0.31$ and $\rho_{\text{level1}}^{\text{readymade}} = 0.45$, `p-value` < 0.001), when adjusting for confounders the relation becomes not significant. Moreover, the tendency of creating healthy versions of readymade food products to target singles, small households, or professionals that tend to have a higher level of education is a factor to take into in consideration when exploring this dimension in its full extent. Consistently with the difficulty to characterize `fats` consumption in the case of nutrients, the heterogeneity of the `oils` category does not allow to draw a significant picture. In the case of `level4`, the contribution of the spatial spillover effects are relevant for the dimensions of `sweets`, `red meats`, and `produce`, thus indicating the importance to consider the influence of neighboring areas and the community effects. Indirect impacts are moderately significant for the `readymade` category in the expected direction.

Education is a valuable variable to consider as a proxy for consumers' dietary knowledge and ability to process nutritional information. As a consequence, a solid body of research associates more educated subjects to the awareness and the importance of a balanced diet [21, 86, 87]. Consistently, we observe that the variety of nutrients `h_nutrients` is positively ($\beta_{\text{level4}} = 0.2$) associated to a higher level of educational attainment. It is worth noting that the variety of products `h_items` that reflects the number of unique purchased

products follows an opposite trend (β_{level4} = −0.13). This result means that, even though it seems that low-educated customers select amongst a wider range of products, the nutritional variety is limited. Moreover, high-educated people show the tendency to have a smaller caloric footprint (β_{level4} = −0.08) and to consume a smaller quantity of food products in weight (β_{level4} = −0.07) that is consistent with previous studies linking this observation to a lower incidence to obesity and a lower average Body Index Mass (BMI) [88]. The variables `energy` and `weight` show a significant or moderately significant spatial spillovers.

Further exploring the causal pathways of the relationships will help in designing effective interventions for improving health outcomes and dietary behavior in particular. For instance, Chandola et al. [89] examined six hypothetical pathways linking education and health via structural equation modeling. Since they found a combination of mechanisms being involved, they conclude that "improvements in population educational attainment may not automatically lead to improvements in population health". Such pathways may also involve access to outside resources including nutritional and health services. When examining the connection between education and health outcomes of an older Japanese cohort, Oshio [90] finds regular health check-ups to be one of the primary mediators. Further, previous studies have shown [91, 92] that health literacy in particular may be an important factor mediating the relationship between educational attainment and many health behaviors, such as being physically inactive, making diet choices, and being obese. Ongoing policy efforts are already attempting to incorporate health education and related services into educational environments, such as the "Whole School, Whole Community, Whole Child" (WSCC) approach developed by the U.S. Centers for Disease Control and Prevention (CDC) [93] and the "Skilled for Health" initiative lead by U.K.'s National Health Service (NHS) [94]. The long-term impact of such interventions within communities will be made clearer by using data-driven, anonymous monitoring of nutritional behaviors of large cohorts, such as one presented in this work.

Limitations   There are few limitations and open points that should be mentioned:
- Our study is based upon the dataset provided in [33] that aggregates the purchasing history of customers of a specific retailer and who have opted for a loyalty card. Even though the authors provide a representativeness score and some empirical evaluation of the biases introduced in their study, the user base is not exempt from sample bias. It might occur that the average customer is more likely to represent a specific level of educational attainment, and, by reflection, specific age and income profiles. For example, a student that is in the process of obtaining a degree, might be less willing to sign in for a loyalty card and, therefore, be accounted for in the study. This could lead to a biased estimation of the interplay between education and food choices; however, the large-scale nature of the dataset and the extensive adoption of the Tesco loyalty program [95] by its customers might reduce this effect. Moreover, food choices are aggregated at the level of administrative units that does not enable the characterization of the dietary habits of individuals, and models the average behavior in a geographical area instead.
- There is no consensus in food and nutrition research on how to group food products in coherent categories [62] leaving the choice to the specific use case. However, the heterogeneity of food products in a group can be very high, smoothing out the

intrinsic differences in health outcomes, affordability, or sociodemographic adoption determinants such as education or gender. For example, proteins are not all the same: there is a wide variety of foods that provide a protein intake that have a very different source, organoleptic properties, connection with health outcomes, impact of the environment and sustainability, or ethical concerns. Not being able to control the aggregation schema limits the applicability of an hypothesis-driven approach where the groups formation is guided by the research question under study.

- Education attainment is only one aspect of the broader concept of cultural capital (institutionalized cultural capital) and, in general, of the cultural substrate that has been arguably identified as crucial when it comes to model food choices. For example, to better capture demographic variations, other measures should be considered. For instance, Rohit et al. [96] show that, for older African Americans, reading level may be a better predictor of baseline neurocognitive status than the years of schooling (possibly due to different quality of schooling available to students of different races). Taking into consideration other measures of cultural capital in general, and cognitive development in particular, may reduce observed racial disparities [97].

- The tight interplay between food variables in characterizing dietary habits and the complex construct of socioeconomic determinants give rise to multicollinearity effects in the explanatory variables. Even though some researchers seem to assume that different socioeconomic indicators reflect the same underlying information and can therefore be used interchangeably [18] (typically, correlations between education, occupation, and income are weak to moderate with magnitude in the range 0.3–0.6 in developed countries [98]), we embrace the evidence from several previous studies that shows the unique contribution of each indicator. Moreover, the relevance of a specific indicator might differ between subgroups of the population, such as between adults and adolescents.

- As pointed out in previous work, the process that underlies food choices is multifaceted and it involves a broad range of dimensions that are not fully captured in our study. These omitted variables are left out partially due to the unavailability of location-aware data with compatible spatial and temporal scales. Moreover, we aim at testing a specific set of hypotheses rather than exploring a wider spectrum of determinants. It is worth noting that these unobserved confounding factors could potentially explain away the link between education and diet; however, we rely on the extensive literature that explores the extent of this relation to corroborate our findings. We speculate that ethnicity, religion, or the dimensions being part of the Index of Multiple Deprivation[8] (IMD) could play a role in this direction.

6 Conclusion

The interplay between educational attainment and food choices has been the subject of a wide body of literature especially for the important ramifications in the public health domain. In this work, we explored the interplay between institutionalized cultural capital, a form of cultural capital theorized by the sociologist Pierre Bordieu, and dietary choices showing how education plays an important role beyond socioeconomic determination. To this extent, we adopted an anonymized large-scale record of food purchases in a major

---

[8]https://www.gov.uk/government/collections/english-indices-of-deprivation

grocery store chain in the Greater London area to quantitatively model food consumption across the three complementary dimensions of macronutrients, food categories, and diet variety. Purchases were geographically aggregated at the level of fine-grained administrative areas (MSOA) and, unlikely most of previous work, we explored this relation in space with the adoption of spatial autoregressive models that aim at capturing the direct and indirect impacts that the spatial dependence induces. We observed that highly-educated areas tend to follow a healthier and more diverse diet, characterized by a higher consumption of fibers, fruits, vegetables, and fish products, along with a more balanced and diversified nutritional intake. On the contrary, a low educational attainment is generally connected to diets high in carbohydrates, sweets and red meats, as well as to a higher caloric intake and average portion size. These relations are consistent with the findings emerging from literature and they allow to map with an unprecedented spatial granularity the behavior of localized communities enabling the design of health policies and interventions that better adhere to the social, economic, and cultural contexts of a place.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-021-00273-y.

Additional file 1. Explorative spatial data analsysis (PDF 11.7 MB)
Additional file 2. Complement to the spatial regression analysis (PDF 126 kB)

Availability of data and materials
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
RS conceived, designed, and supervised the project. NAF contributed to the design of the study, writing the protocol, data preparation, and analysis. GDFM, and YM contributed to the interpretation and impact. NAF drafted the manuscript. RS, GDFM, and YM performed the quality assessment and revised the manuscript. All authors have read and approved the submitted version.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, Mullany EC, Abate KH, Abbafati C, Abebe Z et al (2019) Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet 393(10184):1958–1972
2. Goldberg LR, Strycker LA (2002) Personality traits and eating habits: the assessment of food preferences in a large community sample. Pers Individ Differ. https://doi.org/10.1016/S0191-8869(01)00005-8
3. Johansen SB, Næs T, Hersleth M (2011) Motivation for choice and healthiness perception of calorie-reduced dairy products. A cross-cultural study. Appetite. https://doi.org/10.1016/j.appet.2010.11.137
4. Krieger N, Williams DR, Moss NE (1997) Measuring social class in us public health research: concepts, methodologies and guidelines. Annu Rev Public Health. https://doi.org/10.1146/annurev.publhealth.18.1.341
5. Cooke LJ, Wardle J (2005) Age and gender differences in children's food preferences. Br J Nutr. https://doi.org/10.1079/bjn20051389
6. Wadołowska L, Babicz-Zielińska E, Czarnocińska J (2008) Food choice models and their relation with food preferences and eating frequency in the Polish population: POFPRES study. Food Policy. https://doi.org/10.1016/j.foodpol.2007.08.001

7. Martikainen P, Brunner E, Marmot M (2003) Socioeconomic differences in dietary patterns among middle-aged men and women. Soc Sci Med 56(7):1397–1410. https://doi.org/10.1016/S0277-9536(02)00137-5

8. Aiello LM, Schifanella R, Quercia D, Del Prete L (2019) Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Sci. arXiv:1905.00140. https://doi.org/10.1140/epjds/s13688-019-0191-y.

9. Khawaja M, Mowafi M (2006) Cultural capital and self-rated health in low income women: evidence from the urban health study, Beirut, Lebanon. J Urban Health. https://doi.org/10.1007/s11524-006-9051-8

10. Abel T (2007) Cultural capital in health promotion. In: Health and modernity: the role of theory in health promotion. https://doi.org/10.1007/978-0-387-37759-9_5

11. Franchi M (2012) Food choice: beyond the chemical content. Int J Food Sci Nutr. https://doi.org/10.3109/09637486.2011.632403

12. DellaPosta D, Shi Y, Macy M (2015) Why do liberals drink lattes? Am J Sociol 120(5):1473–1511

13. Robinson N, Caraher M, Lang T (2000) Access to shops: the views of low-income shoppers. Health Educ J. https://doi.org/10.1177/001789690005900202

14. Barratt J (1997) The cost and availability of healthy food choices in southern Derbyshire. J Hum Nutr Diet. https://doi.org/10.1046/j.1365-277X.1997.00487.x

15. Bourdieu P (1984) Distinction: a social critique of the judgement of taste. Routledge & Kegan Paul, London

16. Kamphuis CBM, Jansen T, Mackenbach JP, Van Lenthe FJ (2015) Bourdieu's cultural capital in relation to food choices: a systematic review of cultural capital indicators and an empirical proof of concept. PLoS ONE. https://doi.org/10.1371/journal.pone.0130695.

17. Bourdieu P (1986) The forms of capital. In: Richardson J (ed) Handbook of theory and research for the sociology of education. Greenwood, New York, pp 241–258

18. Turrell G, Hewitt B, Patterson C, Oldenburg B (2003) Measuring socio-economic position in dietary research: is choice of socio-economic indicator important? Public Health Nutr. https://doi.org/10.1079/phn2002416

19. Moreira PA, Padra PD (2004) Educational and economic determinants of food intake in Portuguese adults: a cross-sectional survey. BMC Public Health. https://doi.org/10.1186/1471-2458-4-58

20. Bellisle F (2003) Why should we study human food intake behaviour? NMCD, Nutr Metab Cardiovasc Dis. https://doi.org/10.1016/S0939-4753(03)80010-8

21. Worsley A, Blaschea R, Ball K, Crawford D (2004) The relationship between education and food consumption in the 1995 Australian national nutrition survey. Public Health Nutr 7(5):649–663. https://doi.org/10.1079/PHN2003577

22. Abel T (2008) Cultural capital and social inequality in health. J Epidemiol Community Health 62(7):13. https://jech.bmj.com/content/62/7/e13.full.pdf. https://doi.org/10.1136/jech.2007.066159

23. Christensen VT (2011) Does parental capital influence the prevalence of child overweight and parental perceptions of child weight-level? Soc Sci Med. https://doi.org/10.1016/j.socscimed.2010.11.037

24. Shim JK (2010) Cultural health capital: a theoretical approach to understanding health care interactions and the dynamics of unequal treatment. J Health Soc Behav. https://doi.org/10.1177/0022146509361185

25. Veenstra G (2007) Social space, social class and Bourdieu: health inequalities in British Columbia, Canada. Health Place. https://doi.org/10.1016/j.healthplace.2005.09.011

26. Cade JE, Burley VJ, Warm DL, Thompson RL, Margetts BM (2004) Food-frequency questionnaires: a review of their design, validation and utilisation. Nutr Res Rev. https://doi.org/10.1079/nrr200370

27. Steptoe A, Pollard TM, Wardle J (1995) Development of a measure of the motives underlying the selection of food: the food choice questionnaire. Appetite. https://doi.org/10.1006/appe.1995.0061

28. Skuland SE (2015) Healthy eating and barriers related to social class. The case of vegetable and fish consumption in Norway. Appetite. https://doi.org/10.1016/j.appet.2015.05.008

29. LeSage JP (2008) An introduction to spatial econometrics. Rev Écon Ind 123:19–44

30. Powell LJ, Wittman H (2018) Farm to school in British Columbia: mobilizing food literacy for food sovereignty. Agric Human Values 35(1):193–206

31. Cullerton K, Vidgen HA, Gallegos D (2012) A review of food literacy interventions targeting disadvantaged young people

32. Siewell N, Thomas M (2015) Building sustainable neighborhoods through community gardens: enhancing residents' well-being through university–community engagement initiative. Metrop Univ 26(1):173–190

33. Aiello LM, Quercia D, Schifanella R, Del Prete L (2020) Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London. Sci Data. https://doi.org/10.1038/s41597-020-0397-7

34. Cummins SCJ (2007) Neighbourhood food environment and diet: time for improved conceptual models? Prev Med 44(3):196–197

35. Bolton-Smith C, Brown CA, Tunstall-Pedoe H (1991) Nutrient sources in non-manual and manual occupational groups. Results from the Scottish heart health study (SHHS). J Hum Nutr Diet 4(5):291–306. https://doi.org/10.1111/j.1365-277X.1991.tb00111.x

36. Pampel FC, Krueger PM, Denney JT (2010) Socioeconomic disparities in health behaviors. Annu Rev Sociol 36(1):349–370. PMID: 21909182. https://doi.org/10.1146/annurev.soc.012809.102529

37. Alkon AH, Block D, Moore K, Gillis C, DiNuccio N, Chavez N (2013) Foodways of the urban poor. Geoforum 48:126–135. https://doi.org/10.1016/j.geoforum.2013.04.021

38. Bower KM, Thorpe RJ Jr, Rohde C, Gaskin DJ (2014) The intersection of neighborhood racial segregation, poverty, and urbanicity and its impact on food store availability in the United States. Prev Med 58:33–39

39. Giskes K, Avendaňo M, Brug J, Kunst AE (2010) A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. Obes Rev 11(6):413–429. https://doi.org/10.1111/j.1467-789X.2009.00658.x

40. McKinnon RA, Reedy J, Morrissette MA, Lytle LA, Yaroch AL (2009) Measures of the food environment. A compilation of the literature, 1990–2007. Am J Prev Med. https://doi.org/10.1016/j.amepre.2009.01.012

41. Charreire H, Casey R, Salze P, Simon C, Chaix B, Banos A, Badariotti D, Weber C, Oppert JM (2010) Measuring the food environment using geographical information systems: a methodological review. Public Health Nutr. https://doi.org/10.1017/S1368980010000753

42. Glanz K, Sallis JF, Saelens BE, Frank LD (2007) Nutrition environment measures survey in stores (NEMS-S). Development and evaluation. Am J Prev Med. https://doi.org/10.1016/j.amepre.2006.12.019

43. Glanz K, Sallis JF, Saelens BE, Frank LD (2005) Healthy nutrition environments: concepts and measures. Am J Health Promot. https://doi.org/10.4278/0890-1171-19.5.330

44. Caspi CE, Sorensen G, Subramanian SV, Kawachi I (2012) The local food environment and diet: a systematic review. Health Place. https://doi.org/10.1016/j.healthplace.2012.05.006

45. Khushi S, Ahmad SR, Ashraf A, Imran M (2020) Spatially analyzing food consumption inequalities using gis with disaggregated data from Punjab, Pakistan. Food Secur. https://doi.org/10.1007/s12571-020-01057-4

46. Kim D, Lee CK, Seo DY (2016) Food deserts in Korea? A GIS analysis of food consumption patterns at sub-district level in Seoul using the KNHANES 2008–2012 data. Nutr Res Pract 10(5):530–536

47. Morrison KT, Nelson TA, Ostry AS (2011) Mapping spatial variation in food consumption. Appl Geogr. https://doi.org/10.1016/j.apgeog.2010.11.020

48. Vabø M, Hansen H (2014) The relationship between food preferences and food choice: a theoretical discussion. Int J Bus Soc Sci 5(7):145–157

49. Choi BCK, Pak AWP (2005) A catalog of biases in questionnaires. Prev Chronic Dis 2:A13

50. Fismen A-S, Samdal O, Torsheim T (2012) Family affluence and cultural capital as indicators of social inequalities in adolescent's eating behaviours: a population-based survey. BMC Public Health 12(1):1036

51. Milligan RAK, Burke V, Beilin LJ, Dunbar DL, Spencer MJ, Balde E, Gracey MP (1998) Influence of gender and socio-economic status on dietary patterns and nutrient intakes in 18-year-old Australians. Aust N Z J Public Health. https://doi.org/10.1111/j.1467-842X.1998.tb01419.x

52. Baghurst KI, Record SJ, Baghurst PA, Syrette JA, Crawford D, Worsley A (1990) Sociodemographic determinants in Australia of the intake of food and nutrients implicated in cancer aetiology. Med J Aust. https://doi.org/10.5694/j.1326-5377.1990.tb126148.x

53. Smith AM, Owen N (1992) Associations of social status and health-related beliefs with dietary fat and fiber densities. Prev Med. https://doi.org/10.1016/0091-7435(92)90080-2

54. Baraldi LG, Martinez Steele E, Canella DS, Monteiro CA (2018) Consumption of ultra-processed foods and associated sociodemographic factors in the usa between 2007 and 2012: evidence from a nationally representative cross-sectional study. BMJ Open 8(3):e020574. https://bmjopen.bmj.com/content/8/3/e020574.full.pdf. https://doi.org/10.1136/bmjopen-2017-020574

55. Fernández-Alvira JM, Mouratidou T, Bammann K, Hebestreit A, Barba G, Sieri S, Reisch L, Eiben G, Hadjigeorgiou C, Kovacs E et al (2013) Parental education and frequency of food consumption in European children: the idefics study. Public Health Nutr 16(3):487–498. https://doi.org/10.1017/S136898001200290X

56. Kamphuis CBM, Groeniger JO, van Lenthe FJ (2018) Does cultural capital contribute to educational inequalities in food consumption in the Netherlands? A cross-sectional analysis of the globe-2011 survey. Int J Equity Health. https://doi.org/10.1186/s12939-018-0884-z

57. Wardle J, Parmenter K, Waller J (2000) Nutrition knowledge and food intake. Appetite. https://doi.org/10.1006/appe.1999.0311

58. Malik VS, Popkin BM, Bray GA, Després JP, Willett WC, Hu FB (2010) Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. Diabetes Care. https://doi.org/10.2337/dc10-1079

59. Moynihan PJ, Kelly SAM (2014) Effect on caries of restricting sugars intake: systematic review to inform WHO guidelines. J Dent Res. https://doi.org/10.1177/0022034513508954

60. Amine EK, Baba NH, Belhadj M, Deurenberg-Yap M, Djazayery A, Forrestre T, Galuska DA, Herman S, James WPT, M'Buyamba Kabangu JR, Katan MB, Key TJ, Kumanyika S, Mann J, Moynihan PJ, Musaiger AO, Olwit GW, Petkeviciene J, Prentice A, Reddy KS, Schatzkin A, Seidell JC, Simopoulos AP, Srianujata S, Steyn N, Swinburn B, Uauy R, Wahlqvist M, Zhao-Su W, Yoshiike N, Rabenek S, Bagchi K, Cavalli-Sforza T, Clugston GA, Darnton-Hill I, Ferro-Luzzi A, Leowski J, Nishida C, Nyamwaya D, Ouedraogo A, Pietinen P, Puska P, Riboli E, Robertson A, Shetty P, Weisell R, Yach D (2003) Diet, nutrition and the prevention of chronic diseases. Am J Clin Nutr. https://doi.org/10.1093/ajcn/60.4.644a

61. EFSA (2015) The food classification and description system foodex 2 (revision 2). EFSA Support Publ 12(5):804. https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2015.EN-804. https://doi.org/10.2903/sp.efsa.2015.EN-804

62. Hodgson JM, Hsu-Hage BHH, Wahlqvist ML (1994) Food variety as a quantitative descriptor of food intake. Ecol Food Nutr. https://doi.org/10.1080/03670244.1994.9991395

63. ONS (2018) National statistics postcode lookup user guide (February 2018). http://geoportal.statistics.gov.uk/datasets/4ca06fae243147efb3df8a704653a99f

64. Rey SJ, Anselin L (2014) Modern spatial econometrics in practice: a guide to GeoDa, GeoDaSpace and PySAL. GeoDa Press LLC, United States

65. Moran PAP (1950) Notes on continuous stochastic phenomena. Institute of Statistics, Oxford University

66. Durbin J (1960) Estimation of parameters in time-series regression models. J R Stat Soc, Ser B, Methodol. https://doi.org/10.1111/j.2517-6161.1960.tb00361.x

67. Anselin L (1988) Spatial econometrics: methods and models. Springer, New York, p 284. https://doi.org/10.1007/978-94-015-7799-1

68. Darmofal D (2015) Spatial analysis for the social sciences. Analytical methods for social research. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139051293

69. LeSage J, Pace RK (2009) Introduction to spatial econometrics. J R Stat Soc, Ser A, Stat Soc. https://doi.org/10.1111/j.1467-985x.2010.00681_13.x

70. Pinkse J, Slade ME (2010) The future of spatial econometrics. J Reg Sci 50(1):103–117. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9787.2009.00645.x. https://doi.org/10.1111/j.1467-9787.2009.00645.x

71. Anselin L (1988) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geogr Anal. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

72. Anselin L (2005) Spatial regression analysis in R—a workbook. Urbana

73. Akaike H (1973) Information theory and an extension of the maximum likelihood principle. Springer, New York, pp 199–213

74. Schwarz G (1978) Estimating the dimension of a model. Ann Stat. https://doi.org/10.1214/aos/1176344136

75. Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. Biometrika 78(3):691–692. https://academic.oup.com/biomet/article-pdf/78/3/691/712023/78-3-691.pdf. https://doi.org/10.1093/biomet/78.3.691

76. Breusch TS, Pagan AR (1979) A simple test for heteroscedasticity and random coefficient variation. Econometrica. https://doi.org/10.2307/1911963

77. Pereira MA, O'Reilly E, Augustsson K, Fraser GE, Goldbourt U, Heitmann BL, Hallmans G, Knekt P, Liu S, Pietinen P, Spiegelman D, Stevens J, Virtamo J, Willett WC, Ascherio A (2004) Dietary fiber and risk of coronary heart disease: a pooled analysis of cohort studies. Arch Intern Med. https://doi.org/10.1001/archinte.164.4.370

78. McKeown NM, Meigs JB, Liu S, Wilson PWF, Jacques PF (2002) Whole-grain intake is favorably associated with metabolic risk factors for type 2 diabetes and cardiovascular disease in the Framingham offspring study. Am J Clin Nutr. https://doi.org/10.1093/ajcn/76.2.390

79. Anderson JW, Baird P, Davis RH, Ferreri S, Knudtson M, Koraym A, Waters V, Williams CL (2009) Health benefits of dietary fiber. Nutr Rev. https://doi.org/10.1111/j.1753-4887.2009.00189.x

80. Mosteller F, Tukey JW (1977) Data analysis and regression: a second course in statistics. Addison-Wesley series in behavioral science. Addison-Wesley, Reading. https://books.google.it/books?id=pGlHAAAAMAAJ

81. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum NN, Norat T, Greenwood DC, Riboli E, Vatten LJ, Tonstad S (2017) Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol. https://doi.org/10.1093/ije/dyw319

82. Flemmen M, Hjellbrekke J, Jarness V (2018) Class, culture and culinary tastes: cultural distinctions and social class divisions in contemporary Norway. Sociology. https://doi.org/10.1177/0038038516673528

83. Turrell G, Hewitt B, Patterson C, Oldenburg B, Gould T (2002) Socioeconomic differences in food purchasing behaviour and suggested implications for diet-related health promotion. J Hum Nutr Diet. https://doi.org/10.1046/j.1365-277X.2002.00384.x

84. Mozaffarian D, Rimm EB (2006) Fish intake, contaminants, and human health. JAMA. https://doi.org/10.1001/jama.296.15.1885

85. Hosomi R, Yoshida M, Fukunaga K (2012) Seafood consumption and components for health. Glob J Health Sci. https://doi.org/10.5539/gjhs.v4n3p72

86. Kant AK, Schatzkin A, Harris TB, Ziegler RG, Block G (1993) Dietary diversity and subsequent mortality in the first national health and nutrition examination survey epidemiologic follow-up study. Am J Clin Nutr. https://doi.org/10.1093/ajcn/57.3.434

87. Drescher LS (2007) Healthy food diversity as a concept of dietary quality: measurement, determinants of consumer demand, and willingness to pay. Cuvillier. https://books.google.it/books?id=FdCQppPAnQgC

88. Atella V, Kopinska J (2014) Body weight, eating patterns, and physical activity: the role of education. Demography. https://doi.org/10.1007/s13524-014-0311-z

89. Chandola T, Clarke P, Morris J, Blane D (2006) Pathways between education and health: a causal modelling approach. J R Stat Soc, Ser A, Stat Soc 169(2):337–359

90. Oshio T (2018) Widening disparities in health between educational levels and their determinants in later life: evidence from a nine-year cohort study. BMC Public Health 18(1):278

91. Friis K, Lasgaard M, Rowlands G, Osborne RH, Maindal HT (2016) Health literacy mediates the relationship between educational attainment and health behavior: a Danish population-based study. J Health Commun 21(sup2):54–60

92. Stormacq C, Van den Broucke S, Wosinski J (2019) Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. Health Promot Int 34(5):1–17

93. Lewallen TC, Hunt H, Potts-Datema W, Zaza S, Giles W (2015) The whole school, whole community, whole child model: a new approach for improving educational attainment and healthy development for students. J Sch Health 85(11):729–739

94. National Health Service: Enabling people to make informed health decisions. https://www.england.nhs.uk/ourwork/patient-participation/health-decisions/. Accessed: 2020-06-25

95. Stone M, Points S (2003) How Tesco is winning customer loyalty. J Database Mark Cust Strategy Manag. https://doi.org/10.1057/palgrave.dbm.3240219

96. Rohit M, Levine A, Hinkin C, Abramyan S, Saxton E, Valdes-Sueiras M, Singer E (2007) Education correction using years in school or reading grade-level equivalent? Comparing the accuracy of two methods in diagnosing HIV-associated neurocognitive impairment. J Int Neuropsychol Soc 13(3):462

97. Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y (2002) Reading level attenuates differences in neuropsychological test performance between African American and white elders. J Int Neuropsychol Soc 8(3):341

98. Abramson JH, Gofin R, Habib J, Pridan H, Gofin J (1982) Indicators of social class. A comparative appraisal of measures for use in epidemiological studies. Soc Sci Med. https://doi.org/10.1016/0277-9536(82)90267-2