**Worldwide dietary patterns and their association with socioeconomic data: an
ecological exploratory study**

Authors

Gabriel Gonçalves da Costa^a

Bruno Francisco Teixeira Simões^b

Alessandra da Silva Pereira^c

^aSchool of Nutrition, Federal University of The State of Rio de Janeiro (UNIRIO), Urca,
296, Rio de Janeiro, Rio de Janeiro

^bFederal University of the State of Rio de Janeiro (UNIRIO), Quantitative Methods,
Avenue Pasteur, Urca, 458, Rio de Janeiro, Rio de Janeiro

^cFederal University of the State of Rio de Janeiro (UNIRIO), School of Fundamental
Nutrition, Urca, 296, Rio de Janeiro, Rio de Janeiro

Correspondance:

Gabriel Gonçalves da Costa (gabrielgc@ufrj.br)

Number of pages 19
Number of Tables 4
Number of Words 4857
Abstract 230
Introduction 710
Methodology 1004
Results 690
Discussion 2341
Conclusion 108

Conflicts of interest: The authors declare no conflicts of interest.

Author Contributions:

Conceptualization: GC and BFT
Methodology: GC and BFT
Project Administration: All authors
Investigation (data collection): GC and BFT
Writing – Original Draft: All authors
Writing – Review and Editing: All authors
Supervision: BFT and ASP

Abstract

Objective: To identify dietary patterns in food availability data at the global level using multivariate statistical methodology, to associate the identified dietary patterns with socioeconomic data and to analyze the adequacy of the applied statistical methods in data of food availability. **Methods:** Principal Component Analysis was applied to food availability data of 172 UN registered countries available at FAOSTAT database in Food Balance Sheets section. The Principal Components were correlated with socioeconomic data available from the World Bank database. **Results:** Five principal components were identified, each characterizing a dietary pattern. The first one, a westernized dietary pattern, was composed of energy-dense and processed foods, foods of animal origin, alcoholic beverages, but also, albeit less, by vegetables, fruits and nuts, being correlated with income, urbanization and trade liberalization. This westernized pattern was characterized more by western, animal origin and processed foods, yet preserving unprocessed and regional foods. The other dietary patterns were three agricultural patterns characterized more by regional foods, especially starchy staples, and one coastal dietary pattern composed of fish and seafoods, being associated with GINI index, poverty, and female labor force. **Conclusions:** Principal Component Analysis was adequate to identify dietary patterns in food availability data. A westernized dietary pattern was identified, being associated with income, urbanization and trade liberalization. This association did not occur for the remain of the dietary patterns identified, these being less driven by economic development.

Keywords: Multivariate Analysis, Western Diet, Western Dietary Pattern, *Food Supply*, Food Balance Sheets.

Introduction

According to the Food and Agriculture Organization of the United Nations (FAO), Food Balance Sheets (FBS) presents the quantity of food items available for human consumption, being also named food availability (or food disappearance) in the literature. It is calculated from food supply (food production plus imports minus raw quantity produced in the country that is exported) after factors such as food waste in the supply chain and industrial usage are subtracted from the country's domestic supply. Per capita food availability for human consumption (that is, the quantity of food available for human consumption), calculated at the country level, is associated with household food availability and individual food consumption, this latter assessed by food frequency questionnaires and the 24-hour dietary recall in studies with representative samples of the country's population. Food availability presents a comprehensive picture of the patterns of food supply (associated with food consumption) of a given country during a specified reference period ^{1, 2}.

In recent decades, food availability has increased in many countries, not only in total raw quantity but also in variety. There has been a transition from a scenario characterized by a low variety of starchy staples to one with a greater variety of foods, with an increase in the availability of animal origin foods and energy-rich foods (vegetable oils, animal fats and sugars), as well as the level of food processing. This process has resulted in changes in dietary patterns of populations around the world and is part of the nutrition transition, as well as the process of westernization and globalization of diets ³⁻⁶.

Today a western dietary pattern characterizes the diet of a considerably part of the global population. Currently, this dietary pattern is associated with health and environmental problems, and this association will become stronger especially if historical trends of patterns of food availability of the past continue in the future. In the health dimension, there is for example its association with noncommunicable chronic diseases. For environmental sustainability, there is the discussion about the environmental impact of food of animal origin and of processed food ⁷⁻¹¹.

Food availability and consumption are known to depend on a myriad of factors, such as economic, political, cultural, social, or even biophysical, which eventually interact with each other ¹². It has regional and cultural variations ¹³, having also socioeconomic drivers. For example, drivers of the availability of western and

processed food are income, urbanization, trade liberalization and economic development in general ^{3, 5, 11, 14}. It is known that food availability trends differed historically according to region, being this affected by socioeconomic, regional and cultural differences. Some regions preserved traditional foods more than others in response to the westernization process and to nutrition transition ^{3, 14, 15}.

Socioeconomic variables and food availability have been already previously studied at the national or continental level, with multiple countries and/or continents being analyzed simultaneously with substantial historical coverage ^{4, 16}. Even if food availability as calculated from FBS was already used to identify both a priori and a posteriori dietary patterns, analyses with multivariate statistical tools to study a posteriori dietary patterns in these data at the global and regional level may still be lacking in the literature ^{9, 17}. Additionally, it is known that diets across the globe are becoming increasingly homogeneous, a direct feature of the westernization process, and there is evidence that the nutrition transition is growing at a faster pace in many regions of the globe ¹⁸.

In food availability data (both food supply and household food availability), dietary patterns can also be called dietary availability patterns ^{19, 20}. These nomenclatures can also be used when analyzing food supply data derived from Food Balance Sheets, even if food supply is not able to measure food consumption and diet, as the 24-hour dietary recall and the food frequency questionnaire ^{1, 2, 9, 19, 20}. The nomenclature “dietary pattern” is mostly referred to statistical patterns derived a posteriori, from multivariate statistical methodologies from food data, i.e., food supply (or food availability), household food availability or food consumption ^{19, 20}.

The objective of this exploratory study was to identify dietary patterns in food availability data at the global level, to associate these patterns with socioeconomic data and to analyze the adequacy of the multivariate statistical methods applied.

Methodology

Food availability data in kg/capita/year (item: food supply quantity kg/capita/yr) between 1961 to 2013 (representing time series with 53 data observations units) from Food Balance Sheets of each of the 172 countries registered at United Nations available at FAOSTAT database were downloaded for this ecological study. This temporal gap

represents all years that food balance sheets data was available at the time. For the present analysis, all data was used with the aim of obtaining a representative sample.

In Food Balance Sheets (FBS) from FAO database (FAOSTAT), food availability is calculated from food supply (food production plus imports minus raw quantity produced in the country that is exported) after factors such as food waste in the supply chain and industrial usage of the commodity are subtracted from the domestic supply. A distinction is made between the quantities exported, fed to livestock, used for seed, put to manufacture for food use and non-food uses, losses during storage and transportation, and food supplies available for human consumption. The per caput supply of each such food item available for human consumption (food availability per capita) is then obtained by dividing the respective quantity by the related data on the population partaking on a given country ^{1,2}.

Data were downloaded by 18 food groups: Meat, Milk – Excluding Butter, Eggs, Animal Fats, Offals, Alcoholic Beverages, Sugars & Sweeteners, Vegetable Oils, Pulses, Starchy Roots, Cereals – Excluding Beer, Stimulants, Fish – Seafood, Vegetables, Oilcrops, Treenuts, Fruits – Excluding Wine and Spices, with one dataset for each food group. These food groups represent almost all food groups available in FAOSTAT. Sugar Crops, Aquatic Products, and Miscellaneous were not analyzed due to the high frequency of missing values.

Given the non-gaussian distribution of many of the time series available and for obtaining one single value of food availability by food group for each country, medians for each time series were obtained, since middle values represents a robust estimate of asymmetric frequency distributions. This process was done for each food group in separate datasets, and the medians obtained in each dataset were inserted in a dataset, this latter containing all food groups, one food group per column, and one country per line. In the dataset composed of only medians of time series of food groups for the given countries, all raw data in kg/capita/day was standardized to z-scores.

Multivariate normality of standardized data was analyzed by the Shapiro-Wilk Multivariate Normality Test. Principal Component Analysis (PCA) by Spearman correlation method was used to identify dietary patterns, using Kaiser's Rule as criteria for the number of principal components to be retained (then retaining all principal components with eigenvalues higher than one unit). For the adequacy of PCA, Kaiser-Meyer-Olkin (KMO) and Bartlett's Tests were performed.

Each principal component retained by Kaiser's Rule in the standardized data was then correlated with the medians of the 18 food groups for each country in raw data (kg/capita/year) to characterize each dietary pattern. Food groups with significant ($p < 0.05$) and positive correlation with one principal component were considered to contribute to its construct, characterizing one dietary pattern.

In PCA, high-dimensional and linear dependent data are reduced to fewer variables, the Principal Components. In this transformation, there is a minimum loss of variance of the data. The first principal component explains the greater share of variance and the second explains the greater share of the reminiscent variance unexplained by the first. The number of principal components to be retained is usually analyzed by the Kaiser's Rule when the correlation matrix is used¹⁹⁻²².

Each principal component was then statistically correlated using the Spearman method with socioeconomic data downloaded at the World Bank database, given the non-parametric distribution of the principal components (Shapiro-Wilk Normality Test, $p < 0.05$). Given this asymmetric distribution, quantile regressions with the middle value as conditional quantile ($\tau = 0.5$) between the first principal component and socioeconomic data were performed.

Socioeconomic data for GNI per capita in constant 2010 dollars (ID: NY.GNP.PCAP.KD), Poverty Headcount Ratio (ID: SI.POV.DDAY), Urban Population (ID: SP.URB.TOTL.IN.ZS), Trade – % of GDP (ID: NE.TRD.GNFS.ZS), Labor force – female (ID: SL.TLF.TOTL.FE.ZS) and GINI index (ID: SI.POV.GINI) were downloaded. Each socioeconomic variable was also available in years, and medians of each variable (also from time series of 1961 to 2013 – the same temporal gap that of food availability data) were also obtained for each country. These data were then correlated using the Spearman method with the principal components.

A cutoff point of 1,90 dollars per capita per day with 2011 Purchasing Power Parity (PPP) values was utilized for poverty. Trade – % of GDP was transformed into raw data using GDP per capita (2010 dollars) as a parameter (ID: NY.GDP.MKTP.KD). For more information about these socioeconomic data, the ID of the data in the World Bank database is provided in parenthesis.

Countries were also classified by income level, using the World Bank Country classification by income level 2019-2020. 23 countries were classified as low-income, 45 as lower middle income, 49 as upper middle income and 54 as high income.

Download and analysis of FAOSTAT and World Bank data were performed in 2018. Country classifications data were downloaded and subgroups of countries by income level classification were analyzed in 2020. More recent data from food balance sheets from 2014 to 2017 are already currently available but are still in a preliminary form, utilizing different methodology than of the previous years.

To classify the magnitude of correlation, it was considered 0.0 to 0.3 (or 0.0 to -0.3) for weak correlation, 0.3 to 0.7 (or -0.3 to -0.7) for moderate correlation and 0.7 to 1.0 (or -0.7 to -1.0) for strong correlation. All data and analyses code can be assessed at <https://osf.io/br4vu/> and can also be provided by contact by e-mail with the first author.

Software R version 3.4.4 was used for data analysis, along with the packages: stats, ggplot2, and factoextra. For the multivariate normality test, the package mvnrmtest was used. For PCA analysis, the prcomp function was used, being the correlation matrix of the original data considered for performing the PCA.

Results

Given the non-gaussian distribution of the data (Shapiro-Wilk Multivariate Normality Test, $p < 0.05$), and KMO (Overall MSA = 0.77) and Bartlett's Test ($p < 0.01$), PCA was the adequate technique to be applied.

The standard deviation (square root of the eigenvalue of each principal component) and the proportion of variance of each principal component can be seen in **Table 1**, being the proportion of variance of each component is the eigenvalue divided by total variance (sum of eigenvalues). In the present analysis, 69.55% of the variance-covariance of the data (that was standardized in z-scores) can be explained by the retained principal components and five principal components with eigenvalues more than one unit were retained by Kaiser's rule ²¹.

The correlation between the principal components obtained and the food availability raw data can be seen in **Table 2**. Food groups with positive and significant correlations characterize the dietary pattern, expressed as a principal component. In **Table 3**, correlations of the dietary patterns with socioeconomic data can be visualized.

The first principal component (PC1) identified a westernized dietary pattern, with a positive correlation with GDP per capita, % of urban population, and Trade - % of GDP, all socioeconomic variables that are proxy of socioeconomic development ²³.

This principal component explained 35.48% of the variance of the data, with a standard deviation of 2.53.

In the analysis stratified by income country classifications, the correlation between PC1 and income (GNI per capita) was 0.34 (95%CI: -0.08-0.66, p-value = 0.14), 0.62 (95%CI: 0.40-0.78, p-value < 0.01), 0.41 (95%CI: 0.14-0.62, p-value < 0.01) and 0.65 (95%CI: 0.46-0.78, p-value < 0.01) in low-income, lower middle income, upper middle income, and high-income country classifications, respectively. The correlation between PC1 and urbanization (% of urban population) was 0.40 (95%CI: -0.008-0.70, p-value = 0.062), 0.68 (95%CI: 0.48-0.81, p-value < 0.05), 0.59 (95%CI: 0.37-0.75, p-value < 0.01) and 0.43 (95%CI: 0.19-0.63, p-value < 0.01) in low-income, lower middle income, upper middle income, and high-income country classifications, respectively. The correlation between PC1 and trade liberalization (Trade - % of GDP) was -0.1 (95%CI: -0.49-0.33, p-value = 0.67), 0.20 (95%CI: -0.1-0.47, p-value = 0.20), 0.23 (95%CI: -0.06-0.48, p-value = 0.13) and 0.58 (95%CI: 0.36-0.73, p-value < 0.01) in low-income, lower middle income, upper middle income, and high-income country classifications, respectively.

In **Table 4**, the results of the quantile regressions demonstrate that income was statistically significant associated with PC1 even after adjustment for urbanization and trade liberalization.

The second principal component (PC2) identified an agricultural dietary pattern, with a strong and positive correlation with the Cereals food group, a moderate and positive correlation with the group of Vegetables, a positive and weak correlation with the group of Treenuts, and with no significant and positive correlation with any of the socioeconomic variables under study. Given its moderate and positive correlation with the group of Vegetables, this dietary pattern was classified as a transitional agricultural dietary pattern. This principal component explained 12.52% of the variance of the data, having a standard deviation of 1.50.

The third principal component (PC3) also identified an agricultural dietary pattern, with positive and weak or moderate correlations with the food groups of Offals and Pulses, respectively. This principal component had positive and significant correlation with Poverty Headcount Ratio and GINI index. This principal component explained 0.08% of the variance of the data, with a standard deviation of 1.21.

The fourth principal component (PC4) identified a coastal dietary pattern, with a positive and weak correlation with the food group of Fish – Seafood and with no

significant and positive correlations with any socioeconomic variable. This principal component explained 0.07% of the variance of the data, with a standard deviation of 1.31.

The fifth and last principal component (PC5) identified another agricultural dietary pattern, with a positive and moderate correlation with the food group of Starchy Roots. This principal component had a positive and significant correlation with the female labor force variable. This principal component explained 0.06% of the variance of the data, with a standard deviation of 1.05, being the last principal component with standard deviation and eigenvalue higher than one unit, and hence the latter retained after applying the PCA.

Discussion

Principal Component Analysis was the chosen statistical multivariate method to identify dietary patterns in this study given that the analyzed data was not normally distributed, even after z-scores standardization of the medians, and given its adequacy analyzed by the KMO and Bartlett's Tests. Factor analysis, another multivariate statistical technique with similarities with PCA, can only be applied when the data are normally distributed^{21, 22}. In this sense, the multivariate normality was assessed by the Shapiro-Wilk Multivariate Normality Test, with a $p < 0.05$ indicating that the data were not normally distributed (the null hypothesis of this statistical test is that the data have a multivariate normal distribution).

The westernized dietary pattern is associated to socioeconomic variables that are all proxy of economic development^{3, 23}. Previously, with different methodology, similar results were found in dietary patterns derived by multivariate statistical techniques using food availability data⁹. This first pattern identified in the analysis (PC1) is the one that most explains the variability of the data.

The second principal component, identifying a transitional agricultural dietary pattern, is characterized as transitional accordingly to previous studies,⁹ capturing a pattern characterized by the expansion stage of the nutrition transition, where vegetables also appear alongside with starchy staples in the dietary pattern, occurring before the substitution stage where the dietary pattern switches for a mainly western one^{3, 24, 25}.

Income influence food availability as stated by Bennett's Law. With rising incomes, there is a shift in diets, from a starchy staple mainly scenario to a more varied

one, with the presence of more calorie-rich and western foods ^{12, 14, 26}. Also, Engel's Law establishes that not only income needs to be accounted for when assessing drivers of dietary patterns, but also income distribution ²⁷. For example, growing demand for meat increases in a U-shaped curve, and after some amount of income, the demand for meat stagnates or even declines, this association being also dependent on geographic and cultural factors ^{3, 24}. That is, there is a saturation point where demand for food availability no longer responds to growing incomes ^{3, 24, 27}.

Urbanization has a similar role in driving dietary patterns. Food availability changes occur first in urban regions, only occurring in rural regions if there is income increase ⁵. Although more debate is needed in the literature, there is some evidence that income and urbanization have different roles in diet, the first acting increasing variety of foods in the menu, the latter acting changing regional traditions of diets and causing westernization ¹⁵. There is also evidence that the nutrition transition, as for dietary patterns change, is occurring at lower levels of income and urbanization than previously thought ¹⁶.

It is important to note that further exploratory analyses were performed to assess the correlation between the principal components (dietary patterns) and the socioeconomic variables by country income classification (World Bank Country Classification by income level 2019-2020) to assess the confounding effect of income in the association between the dietary patterns and other socioeconomic variables such as trade liberalization and urbanization. These analyses could only control but not remove entirely the confounding effect of income, because country income classifications are not continuous variables but categorical ones, and each income classification comprises a range of income values. Hence, the correlation between PC1 (the westernized dietary pattern) and income was not entirely removed in low-income, lower middle income, upper middle income, and high-income country classifications, respectively, suggesting that the influence of income on the dietary patterns was not entirely removed after stratifying for income classifications.

In this sense, one can argue that urbanization and income can fundamentally indicate one same phenomenon, that is, of economic development. Their correlation was positive and strong ($r = 0.8$, 95%CI: 0.74-0.85, $p < 0.05$). However, urbanization had positive and moderate/strong correlations with the westernized dietary pattern in different subgroups of country income classifications, suggesting an independent association of this variable. Both urbanization and income maintained statistically

significant associations with the westernized dietary patterns after adjusting for trade liberalization in the quantile regression. Additionally, urbanization maintained association with the westernized dietary pattern after adjusting for income only in **Table 4** (model 2).

Trade liberalization, represented (as a proxy) in this study by the variable Trade – % of GDP, can change the commodities market, where countries with greater insertion in the global trade market present more dependence on foods of the western dietary pattern. Also, there is the influence of marketing and the market of foods ²⁸. This variable also was strong and positively correlated with income ($r = 0.63$, 95%CI: 0.53-0.71, $p < 0.05$). This variable lost statistical significance with the westernized dietary pattern after adjusting for urbanization and income. In the analysis of income strata, trade liberalization was only statistically significant correlated with this dietary pattern in high income countries.

Even if correlation is not equal to causation, trade liberalization had a stronger correlation with the westernized dietary pattern in high income countries, suggesting that the play of income comes first as of trade liberalization in the westernization process. The correlation between the westernized dietary pattern and income is higher in the transition between low income to lower middle income. Also, urbanization has stronger correlations comparing middle income to low income countries, suggesting that the play of the urbanization process in westernization of dietary patterns comes in lower levels of income when compared to trade liberalization.

These socioeconomic variables analyzed above (trade liberalization, income, and urbanization) did not presented positive and significant correlations with the agricultural dietary patterns identified in this study in a similar manner as other studies as well ^{9, 25}, since these patterns are mainly composed by starchy staple foods, and these foods are not correlated with economic development when analyzed individually using food availability data ^{3, 6, 16}. Even if the inferences realized in the present work reproduce in other studies, the associations outlined above between the principal components and the socioeconomic variables are also dependent on the unit of the latter (i.e., percentages and not raw values for trade liberalization and urban population).

Fish and seafood are present both in PC1 and PC4. This food group can be driven both by economic development and by geographical factors. In PC1, fish and seafood are driven by economic development, being this food group present in a westernized dietary pattern, a pattern with higher amount of protein and foods of animal

origin^{13, 16}. The rise in income is a main driver for the growth of food availability of fish worldwide. In PC4, it is driven by geographical factors since this pattern is not driven by income and urbanization and PC4 does not have significant correlation with any socioeconomic variable under study. Fish and seafood comprise an important food group for food security and human nutrition. In low-income populations that depend heavily on a narrow range of calorie-dense staple foods and can consume fish locally available, fish can represent a much-needed means of nutritional diversification that is relatively cheap. Within countries, consumption is higher in coastal marine and inland water areas^{6, 13, 16}.

Agricultural Patterns – regional and cultural differences

Given these results, none of the agricultural dietary patterns identified were characterized by more than one starchy staple (cereals, pulses, or starchy roots food groups). Other factors than the socioeconomic ones can explain differences between PC2, PC3, and PC5 (all these three being characterized as agricultural dietary patterns), as for example, cultural and regional factors. In Africa, there is a great variety of food spending patterns, especially starch staples, and this can explain why the food groups of cereals, starchy roots, and pulses did not share a principal component. Regions where agricultural dietary patterns prevail have low income and high food insecurity levels⁶, and in the present study PC3 had positive and significant correlation with poverty (analyzed by Poverty Headcount Ratio by a cutoff point of 1,90 dollars per capita per day of 2011 Purchasing Power Parity) and GINI, where a higher GINI equals to higher inequality.

The positive and moderate correlation between PC5 and the female labor force can be at least partly explained by a phenomenon called agricultural feminization. The percentage of women working in agriculture is rising in many countries in Africa, Asia, and Latin America, because of greater migration rates of men compared to women. In developing countries, employment opportunities increase in non-farm sectors, and when men move out of agriculture to other activities, women tend to remain on the farm or move out more slowly. Their responsibilities in agriculture may increase in response to this phenomenon. Even in Latin America, where farming has traditionally been a male occupation, the share of women in agricultural employment is increasing when analyzing historical trends¹⁶. The correlation between PC5 and the female labor force

variable in this study was positive in all subgroups of income classification, but was higher in countries of middle income (countries classified as lower and upper middle income), being only positive and weak in low and high income countries. The correlation between PC5 and female labor force was 0.10 (95%CI: -0.32-0.50, p-value = 0.65), 0.49 (95%CI: 0.23-0.68, p-value < 0.01), 0.46 (95%CI: 0.20-0.66, p-value <0.01) and 0.11 (95%CI: -0.16-0.37, p-value = 0.45) in low-income, lower middle income, upper middle income, and high-income country classifications, respectively. Without stratification by income country classification, the correlation between PC5 and female labor force is 0.43, as reported in **Table 3**.

There is evidence of the nutrition transition occurring in sub-Saharan African countries and India. However, low food diversity can still be a prevailing characteristic of the dietary patterns of these regions ^{16, 29}. In this sense, the analysis of medians of time series may be useful for the purpose of this work, which is to identify dietary patterns in worldwide data and their association with socioeconomic data. However, this approach may not identify more recent trends.

Food balance sheets have limitations, for example, they do not reflect the actual food consumption of the population. Also, this FAO methodology does not consider losses in food preparation and food preparation methods, leftovers after food consumption, representing the food that is available for human consumption in each region. Comparing with methods used for assessing food consumption, like the 24-hour dietary recall, food balance sheets overestimate the mean regional values for a variety of food groups ^{2, 3, 30-32}.

Low- and middle-income countries at different stages of the nutrition transition are more likely to have a lower quality of food availability data, as they have a marked presence of subsistence agriculture, this one not included in the calculation ³³. In developing countries there may be an underestimation, while in developed countries overestimation of per capita food availability ^{2, 32}. Also, the method does not capture regional and cultural variations within the country ³.

However, all methods for measuring diet have limitations ³⁴. The choice of the appropriate method depends on the purpose of the study. In this sense, per capita food availability method by FAO food balance sheets generates important data for trend assessment over time, being the only one available for a large number of countries in large time ranges and so may be suitable for an ecological study analyzing multiple countries at the same time, given the absence to date of other methods for a large

number of countries ^{3, 35}. Global food consumption databases with data from national country samples are still under development until this date.

Another possible limitation of the study is the fact that the analysis were performed in medians for each time series of food availability data. Longitudinal data from 1961 to 2013 (all available years) were used to gain information in this exploratory analysis. The median values, both for food availability and for socioeconomic data were calculated with the aim to summarize with a robust estimator each time series (longitudinal data, from 1961 to 2013) into one single value for each country and food group, creating a new dataset where statistical analyses (PCA and correlations) were performed in order to analyze the data in an exploratory manner. In this sense, each median value of this dataset can be from a different year since this value depend on each time series. However, the aim of using medians was only to summarize into one single value the frequency distributions of each time series, to capture general trends in this food availability data available at FAOSTAT, given the exploratory nature of this study. Multivariate statistical analyses in nutrition can be exploratory, and hence dietary patterns identified by these techniques are often called a posteriori dietary patterns ¹⁹. It is of knowledge of the authors that few studies analyzed dietary patterns in an a posteriori manner in food availability data derived from Food Balance Sheets ^{9, 17, 25}, and hence the exploratory nature of this study.

It is of knowledge that when performing the statistical analyses performed herein in median values, the effect of time cannot be analyzed properly, and hence the generalization to the nutrition transition is limited. However, the a posteriori dietary patterns identified in this study can be partly explained by the nutrition transition ⁴ and the westernization of dietary patterns ^{3, 15} that occurred in the time frame that the data is available. In this sense, the a posteriori dietary patterns identified herein can be explained by the literature.

Additionally, some correlations between dietary patterns and socioeconomic variables performed in subgroups of country income level may lack statistical power due to low sample sizes. Low income countries had the smallest sample size of all strata of income levels (n = 23). However, these subgroup analyses were performed with the aim of assessing confounding of income in the association between the identified dietary patterns and socioeconomic variables only.

Lastly, there is the fact that this study analyzes publicly available data and use open code statistical software, making available the data and analysis code in an open

access platform ³⁶. This makes it useful for replication initiatives, reproducibility assessment of analysis and for statistical learning, both current growing needs in science ^{37, 38}.

Conclusion

Dietary patterns were identified in food availability per capita data at the global level after Principal Component Analysis was applied. Five dietary patterns were identified, a westernized dietary pattern, a transitional agricultural dietary pattern, two agricultural dietary patterns and a coastal dietary pattern. The westernized dietary pattern was associated with socioeconomic data that are proxy of economic development, more specifically, income, urbanization, and trade liberalization. This association did not occurred for the other dietary patterns identified, these being less driven by economic development and more by regional characteristics, having significant correlations with GINI index, poverty, and female labor force. Principal Component Analysis was adequate to identify dietary patterns.

References

1. Serra-Majem L, MacLean D, Ribas L, Brule D, Sekula W, Prattala R, et al. Comparative analysis of nutrition data from national, household, and individual levels: results from a WHO-CINDI collaborative project in Canada, Finland, Poland, and Spain. *J Epidemiol Community Health* 2003; 57(1):74-80.
2. Serra-Majem L. Food availability and consumption at national, household and individual levels: implications for food-based dietary guidelines development. *Public Health Nutr* 2001; 4(2b):673-6.
3. Kearney J. Food consumption trends and drivers. *Philos Trans R Soc Lond B Biol Sci* 2010; 365(1554):2793-807.
4. Popkin BM. The nutrition transition and its health implications in lower-income countries. *Public Health Nutr* 1998; 1(1):5-21.
5. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev* 2012; 70(1):3-21.
6. Alexandratos N, Bruinsma J. *World agriculture towards 2030/2050: the 2012 revision*. Rome: FAO; 2012.

7. Borges CA, Rinaldi AE, Conde WL, Mainardi GM, Behar D, Slater B. Dietary patterns: a literature review of the methodological characteristics of the main step of the multivariate analyzes. *Rev Bras Epidemiol* 2015; 18(4):837-57.
8. Eng JY, Moy FM, Bulgiba A, Rampal S. Consistency and Generalizability of Dietary Patterns in a Multiethnic Working Population. *J Acad Nutr Diet* 2018; 118(7):1249-62.
9. Oggioni C, Lara J, Wells JC, Soroka K, Siervo M. Shifts in population dietary patterns and physical inactivity as determinants of global trends in the prevalence of diabetes: an ecological analysis. *Nutr Metab Cardiovasc Dis* 2014; 24(10):1105-11.
10. Springmann M, Godfray HC, Rayner M, Scarborough P. Analysis and valuation of the health and climate change cobenefits of dietary change. *Proc Natl Acad Sci USA* 2016; 113(15):4146-51.
11. Popkin BM, Reardon T. Obesity and the food system transformation in Latin America. *Obes Rev* 2018; 19(8):1028-64.
12. Godfray HC. Food for thought. *Proc Natl Acad Sci USA* 2011; 108(50):19845-6.
13. Food and Agriculture Organization of the United Nations (FAO). *The future of food and agriculture – Trends and challenges*. Rome: FAO; 2017.
14. Cirera X, Masset E. Income distribution trends and future food demand. *Philos Trans R Soc Lond B Biol Sci* 2010; 365(1554):2821-34.
15. Pingali P. Westernization of Asian diets and the transformation of food systems: Implications for research and policy. *Food Policy* 2007; 32(3):281-98.
16. Food and Agriculture Organization of the United Nations (FAO). *The State of Food and Agriculture: Leveraging Food Systems for Inclusive Rural Transformation*. Rome: FAO; 2017.
17. Porkka M, Kummu M, Siebert S, Varis O. From food insufficiency towards trade dependency: a historical analysis of global food availability. *PloS One* 2013; 8(12):e82714.
18. Popkin BM. Relationship between shifts in food system dynamics and acceleration of the global nutrition transition. *Nutrition Reviews* 2017; 75(2):73-82.
19. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002; 13(1):3-9.
20. Nascimento S, Barbosa FS, Sichieri R, Pereira RA. Dietary availability patterns of the Brazilian macro-regions. *Nutrition Journal* 2011; 10(1):79.

21. Mardia K, Kent JT, Bibby JM. *Multivariate Analysis*. London: Academic Press; 1980.
22. Johnson RA, Wichern DW. *Applied multivariate statistical analysis*. New Jersey: Prentice Hall Upper Saddle River; 2002.
23. Abreu ES, Viana IC, Moreno RB, Torres EAFS. Alimentação mundial: uma reflexão sobre a história. *Saúde e Sociedade* 2001; 10:3-14.
24. Vranken L, Avermaete T, Petalios D, Mathijs E. Curbing global meat consumption: Emerging evidence of a second nutrition transition. *Environmental Science & Policy* 2014; 39:95-106.
25. Oggioni C, Cena H, Wells JC, Lara J, Celis-Morales C, Siervo M. Association between worldwide dietary and lifestyle patterns with total cholesterol concentrations and DALYs for infectious and cardiovascular diseases: an ecological analysis. *J Epidemiol Glob Health* 2015; 5(4):315-25.
26. Bennett MK. *The world's food*. New York: Harper & Bros. Publ.; 1954.
27. Godfray HC, Crute IR, Haddad L, Lawrence D, Muir JF, Nisbett N, et al. The future of the global food system. *Philos Trans R Soc Lond B Biol Sci* 2010; 365(1554):2769-77.
28. Hawkes C. Globalization and the nutrition transition: A case study. In: Pinstrup-Andersen P, Cheng F, editors. *Food Policy for Developing Countries: Case Studies*. Ithaca, NY: CUL Initiatives in Publishing (CIP); 2007.
29. Golzarand M, Mirmiran P, Jessri M, Toolabi K, Mojarad M, Azizi F. Dietary trends in the Middle East and North Africa: an ecological study (1961 to 2007). *Public Health Nutr* 2012; 15(10):1835-44.
30. Hawkesworth S, Dangour AD, Johnston D, Lock K, Poole N, Rushton J, et al. Feeding the world healthily: the challenge of measuring the effects of agriculture on health. *Philos Trans R Soc Lond B Biol Sci* 2010; 365(1554):3083-97.
31. Popkin BM. Nutrition in transition: the changing global nutrition challenge. *Asia Pac J Clin Nutr* 2001; 10:S13-S18.
32. Del Gobbo LC, Khatibzadeh S, Imamura F, Micha R, Shi P, Smith M, et al. Assessing global dietary habits: a comparison of national estimates from the FAO and the Global Dietary Database. *Am J Clin Nutr* 2015; 101(5):1038-46.
33. Vandevijvere S, Chow CC, Hall KD, Umali E, Swinburn BA. Increased food energy supply as a major driver of the obesity epidemic: a global analysis. *Bull World Health Organ* 2015; 93(7):446-56.

34. Archer E, Lavie CJ, Hill JO. The Failure to Measure Dietary Intake Engendered a Fictional Discourse on Diet-Disease Relations. *Front Nutr* 2018; 5:105.
35. Barnard ND. Trends in food availability, 1909–2007. *The American Journal of Clinical Nutrition* 2010; 91(5):1530S-1536S.
36. Open Science Framework [Internet]. [cited 2019 Dezember 26]. Available from: <https://osf.io/dashboard>
37. Goodman, SN; Fanelli, D; Ioannidis, JPA. What does research reproducibility mean?. *Science translational medicine* 2016; 8(341):341-341.
38. Peels, R; Bouter, L. The possibility and desirability of replication in the humanities. *Palgrave Communications* 2018; 4(1):95.

Table 1

Standard deviation, individual proportion and cumulative proportion of explained variance of each retained principal component

Principal Components Statistics	PC1	PC2	PC3	PC4	PC5
Standard Deviation	2,5271	1,5013	1,21352	1,13411	1,05841
Proportion of Variance	0,3548	0,1252	0,08181	0,07146	0,06223
Cumulative Proportion	0,3548	0,48	0,56184	0,63329	0,69553

Table 2

Statistical significant correlations between food availability data (middle values of time series for each country) and principal components

Food groups	PC1	PC2	PC3	PC4	PC5
Alcoholic Beverages	0,68	ns	ns	ns	ns
Animal Fats	0,84	ns	ns	ns	ns
Cereals - Excluding Beer	ns	0,91	ns	ns	ns
Eggs	0,87	ns	ns	ns	ns
Fish - Seafood	0,34	-0,49	-0,46	0,23	ns
Fruits - Excluding Wine	0,56	-0,44	ns	-0,42	-0,26
Meats	0,88	ns	ns	ns	ns
Milk - Excluding Butter	0,88	ns	ns	ns	ns
Offals	0,65	ns	0,3	ns	0,22
Oilcrops	ns	-0,38	-0,35	ns	ns
Pulses	-0,4	ns	0,27	-0,5	-0,3
Spices	ns	ns	-0,35	ns	-0,49
Starchy Roots	ns	-0,57	ns	-0,12	0,5
Stimulants	0,87	ns	ns	ns	ns
Sugars & Sweeteners	0,8	ns	ns	ns	-0,36
Treenuts	0,53	0,25	-0,4	-0,29	0,22
Vegetable Oils	0,65	ns	-0,21	-0,27	ns
Vegetables	0,62	0,38	-0,38	-0,23	ns
*ns equals to non-significant					

Table 3

Correlations between socioeconomic data and principal components

Socioeconomic Variables	PC1	PC2	PC3	PC4	PC5
GINI Index	-0,45	-0,22	0,29	ns	ns
GNI per capita in constant 2010 dollars	0,89	ns	ns	ns	ns
Labor force – female	ns	-0,27	ns	ns	0,43
Poverty Headcount Ratio	-0,86	ns	0,2	ns	ns
Urban Population	0,88	ns	ns	ns	ns
Trade (% of GDP)	0,68	ns	ns	ns	ns
*ns equals to non significant					

Table 4				
Quantile regressions between PC1 and socioeconomic variables				
Model 1				
Factors	Beta	Standard error	t value	p-value
(Intercept)	-3.91440	0.34327	-11.40326	< 0.001
GNI per capita in constant 2010 dollars (US\$)	0.00007	0.00002	2.98405	0.00337
Urban Population (% of total population)	0.06356	0.00959	6.63113	< 0.001
Trade (% of GDP in US\$)	0.00000	0.00000	1.18707	0.23727
Model 2				
Factors	Beta	Standard error	t value	p-value
(Intercept)	-3.82618	0.35792	-10.69018	< 0.001
GNI per capita in constant 2010 dollars (US\$)	0.00009	0.00003	3.56881	< 0.001
Urban Population (% of total population)	0.05938	0.01042	5.69800	< 0.001