Imperial College London

National Heart and Lung Institute

Department of Respiratory Epidemiology and Public Health

The use of dietary patterns empirically derived from Principal Components Analysis and alternative strategies to identify associations between diet and disease.

by

Ioannis Bakolis

A thesis submitted for the degree of

Doctor of Philosophy of Imperial College London

March 2013

Declaration of Originality

I declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at Imperial College London or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at Imperial College London or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged. All other work mentioned in the thesis is appropriately referenced.

Copyright Declaration

Abstract

Dietary patterns derived empirically using principal components analysis (PCA) are widely employed for investigating diet-disease relationships. The aim of the study was to investigate whether PCA performed better at identifying associations between diet and disease than analysing each food on the FFQ separately, a process we refer to as exhaustive single food analysis (ESFA).

A systematic review of nutritional epidemiology literature relating to the use of PCA in identifying dietary patterns in observational and cohort studies from 2004-2009 was employed. Furthermore, we simulated diet and disease data using real food frequency questionnaire data and assuming that a number of foods or dietary pattern intakes were causally associated with disease. In each simulation, ESFA and PCA were employed to identify foods associated with disease using logistic regression, allowing for multiple testing and adjusting for energy intake. ESFA was further adjusted for principal components, foods which were significant in unadjusted ESFA, and propensity scores. For each method, we investigated the power, with which we could identify an association between diet and disease, and the power and false discovery rate (FDR) for identifying associations with specific food intakes. We apply our innovative methodology to a real dietary dataset (GA^2LEN survey).

ESFA had greater power to detect an association of diet with disease than PCA, and greater power and lower FDR for identifying associations with specific foods. FDR increased with increasing sample size using both methods. However, when ESFA was adjusted for foods that were significant in unadjusted ESFA, FDRs were controlled successfully at the desired level of 20%.

Our results raise questions about the use of PCA in nutritional epidemiology. Adjusted ESFA identifies foods that are causally linked to disease with a low rate of false discoveries, and surprisingly good power. These findings were not fully supported from the analysis of the GA^2LEN data-set.

Acknowledgements

First and foremost, I would like to extend my most sincere appreciation and gratitude equally to both of my supervisors Prof. Peter Burney and Dr.Richard Hooper. Without their help, inspiration, guidance and patience, the completion of this thesis would not have been possible. Peter gave me the freedom and the environment to think and explore every possible scenario of my research, providing constructive criticism to every stage of this work. Richard, as a supervisor was brilliant in every aspect, always there to answer to my questions and to challenge my ideas, showing a great commitment to the project even after departing from Imperial College London. Their support to me, both professionally and personally, was invaluable, and helped me become a better academic researcher.

I would like to thank also my friend and colleague, Stephanie McNeil, for spending so much time listening to my complaints, suggesting me good music and always trying to make me feel better; Deborah Jarvis and Seif Shaheen, for growing an epidemiologist out of me; James Potts for doing the hard work of handling the data and having numerous canteen lunches with me; Ischa Kummeling for helping me blend in at the first months in the group and for being a wonderful desk mate; Vanessa Garcia-Larsen for introducing me to nutritional epidemiology; and Hilary Burton for always being helpful and supportive throughout these four years. Finally all the rest of my colleagues at Respiratory Epidemiology and Public health group for making it a great place to work at.

I am grateful to my examiners Prof. John Thompson and Dr. Teresa Norat for constructive and very useful comments and suggestions during my Viva.

I am also very grateful to my friends, Kostas, Lykourgos, Alexandros, Tasos, Spyros, Myrto, Dhan, Lucia, Aecio and Lorenzo, for the talks, travels, laughs and drinks that we shared together and who have helped me with one way or another to finish this thesis during the last four years.

This thesis is dedicated to my family, to Gerasimos, Kiriaki, Andreas, Gerasimos jr, and Thodoris for their unconditional presence and love in my life **(Το διδακτορικό είναι αφιερωμένο στην οικογένεια μου, στον Γεράσιμο, στην Κυριακή, στον Ανδρέα, στον Γεράσιμο τον νεότερο, στον Θοδωρή για την χωρίς όρους αγάπη και παρουσία τους στην ζωή μου)** and to Lucia for making my life brighter especially in the last gloomy period of this work.

Table of Contents

Index of Figures

Figures in Chapter 2

Figures in Chapter 3

Figures in Chapter 4

Figures in Chapter 6

Index of Tables

Tables in Chapter 4

Tables in Chapter 5

Tables in Chapter 6

Tables of Systematic Review

1 Introduction

"In the Pre-Socratic Era, mathematical knowledge was thought to supply an ideal, from which every-day empirical knowledge fell short. Specifically, if the world of observation doesn't fit the world of mathematics so much the worse for world of observation."

Bertrand Russell

1.1 Dietary pattern analysis and nutritional epidemiology

Epidemiology has been defined as the study of the patterns of disease occurrence in populations and of the factors that influence those patterns (Rothman & Greenland, 1998). Nutritional epidemiology is therefore the study of the nutritional determinants of disease and is concerned with effects of diet on chronic diseases. These effects are multifactorial in origin and may take years, if not decades, to develop (Willett, 1998).

The public have an increasing interest in the role of diet in the aetiology of certain chronic diseases and has been often confused by the contradictory results of the empirical findings of nutritional epidemiology.

These contradictory results in observational studies are due to the complex nature of diet and the vast number of potentially relevant nutrients and foods. Because the prevailing method of analysis has been to study the relation of outcomes to intake of single foods or nutrients, this complexity of diet may be difficult to take into account.

An alternative and increasingly popular approach, which claims to resolve the limitations of single food analysis, is to explore associations by identifying dietary "patterns" from Food Frequency Data with the use of Principal Component Analysis (PCA) (Schwerin et al., 1981, Kant, 2004, Newby & Tucker, 2004, Hu, 2002). PCA groups food items according to the degree that they are correlated with each other, and aggregates them to distinct dietary patterns. The concept of PCA lies in the explanation of dietary behaviours of the population and the way that foods are consumed in combination with each other. However, PCA as a method applied in nutritional epidemiology raises conceptual issues and statistical problems which call for a subjective judgment by the researcher (Newby & Tucker, 2004).

Slattery (2008) claimed that eating patterns derived from PCA characterized the diet associated disease risk better than anyone food or nutrient. However, in order for this statement to be true in disease prevention we need PCA to identify all of the foods (and only those foods) which, in combination, increase or decrease the risk of disease, and this food combination has to be easily translatable into an intervention.

1.2 Premises

The primary premises of this research are that

1. There is no comprehensive critical evaluation that establishes the methodological superiority of PCA despite its wide application in nutritional epidemiological studies.

2. Simpler methods than PCA could be equally or more effective for detecting diet and disease associations.

1.3 Aims and objectives

We aim

1. To give a historical overview of nutritional epidemiological methods for detecting diet-disease associations with a particular focus on the application of PCA.

2. To undertake a systematic review of PCA as employed in dietary pattern analysis for observational studies.

3. To compare the results of analysing each individual food on the Food Frequency Questionnaires (FFQ) separately in relation to disease risk, a process we refer to as an exhaustive single food analysis (ESFA), with empirically derived dietary patterns with the use of Principal Component Analysis (PCA) to identify diet–disease associations in diverse populations and comparing the performance of these two methods. For this we used Monte Carlo Simulations of a dietary data-set with a realistic correlation structure.

4. To compare the use of three different methods of adjustment to cope with confounding in an unadjusted EFSA.

5. To apply our innovative methodology to a real dietary dataset derived from the GA^{2}LEN survey and follow-up.

1.4 Data sources

Our two real dietary datasets and source of the food correlation matrices for our simulations were comprised of 856 adults aged 16-50 years old living in Greenwich and questioned as part of the F.L.A.G survey (Shaheen et al., 2001) and from 200 adults 29-54 years old living in Ipswich and Norwich interviewed as part of the UK ECRHS II diet survey (Hooper et al., 2010). For more detailed information see paragraph 4.8.

Our real dataset for the application of our conclusions from our Monte Carlo simulation results was comprised of 3057 adults aged 21-75 years old living in 17 centres across Europe; Belgium (Ghent), Denmark (Odense), Poland (Lodz, Katowice), Germany (Berlin, Duisburg), Portugal (Coimbra), Italy (Palermo), Sweden (Gothenburg, Stockholm, Umea, Uppsala), Netherlands (Amsterdam), FYROM (Skopje), United Kingdom (Southampton, London) and Finland (Helsinki) as part of the Global Allergy and Asthma Network of Excellence (GA^2LEN) survey. For more detailed information see paragraph 6.2.

1.5 Outline of thesis

- **Introduction.** A description of the problem studied, of the primary premises, aims and objectives is presented.

- **Background.** A literature review of the different methods used for dietary patterns analysis as applied in observational studies of diet with an emphasis on PCA.

- **Systematic Review.** A systematic review of nutritional epidemiology literature relating to the use of PCA in identifying dietary patterns in observational studies from 2004-2009.

- **Methods.** A detailed description of our Monte Carlo simulations, where we created a hypothetical population in which we tested which one of the two approaches, ESFA or PCA had greater power and lower false discovery rate, and which method of adjustment was more appropriate for dealing with confounding in ESFA.

- **Results.** Average estimates of percentages of power and false discovery rates for different sample sizes and different number of principal components. Average percentages of false discovery rates for different ways of adjustment of ESFA method for different sample sizes are presented.

- **Analysis of GALEN data-set: effect of diet on asthma.** This section presents

    1. An innovative two-step analysis with the use of generalized linear models. As a first step we explore associations of each individual food in the FFQ with respiratory and allergic outcomes. As a second step, we re-run the same analysis additionally adjusting for foods that were identified as statistically significant at the first step controlling the false discovery rate at 20%.

    2. Dietary patterns analysis with the use of Principal Component Analysis (PCA).

All of the statistical models were adjusted for potential demographic and environmental risk factors and multiple testing. Where necessary, multilevel and meta-analytical techniques were employed within this framework of analysis to account for between-centre and within-centre variation.

- **Discussion and Conclusion.** Interpretation of the study results and limitations of the study are presented along with ideas for future research.

2 Background

2.1 From single food and nutrient analysis to dietary patterns

2.1.1 A historical overview of the complexity of diet

Nutritional epidemiology is a scientific field with a history of over 200 years. Lind (1753) conducted one of the first clinical trials in 1776 to test the effect of citrus fruit in preventing scurvy and concluded that lemons and oranges have a protective effect on scurvy (Carpenter, 1986, Willet, 1998). In the late nineteenth century, during the era of industrialisation it was hypothesised that the inclusion of milk and vegetables in the human diet could eliminate beriberi and pellagra (Messina et al., 2001). Later on, it was proved that world epidemics of these diseases were due to vitamin and nutrient deficiency; vitamin C for scurvy, thiamine for beriberi and nicotinic acid for pellagra (Jacobs & Steffen, 2003). At the start of the twentieth century, this fundamental evidence and the discovery of most of the vitamins and minerals led to the belief that nutrient deficiency was the primary cause of disease symptoms (Jacobs & Steffen, 2003). So, the attention of nutritional science moved towards the investigation of specific foods and nutrients (vitamins, lipids, amino acids) with protective effects; the attention of industry to the production of nutrient supplements; and the public health policy towards prevention strategies (Mertz, 1984).

However, in the second half of the twentieth century and with the advent of westernized chronic diseases, nutrient deficiency has not been the only explanation for diet-disease associations. Although single food and nutrient analysis has continued to make important steps forward in identifying nutrient-disease associations (Shekelle et al., 1981, Knert et al., 2004, Giovannucci et al., 2006), recent results have raised questions about the aetiology of diet in certain chronic diseases. Specifically, a lack of empirical evidence from randomised controlled trials based on observational findings, has been observed in chronic diseases like asthma (Shaheen et al., 2007, Pearson et al., 2004, Fogarty et al., 2003), cancer (Greenberg et al., 1994, Schatzkin et al., 2000) and cardiovascular disease (Hennekens et al., 1996). This may be due to the moment of intervention in life, duration of the intervention, follow-up period of the intervention as well to the various conceptual and methodological limitations of single food and nutrient analysis in taking into account the complex biological and behavioural effects of diet on disease.

2.1.2 Limitations of single food and nutrient analysis and the dietary pattern analysis solution

Occurrence of chronic diseases in the second half of the twentieth century has been linked with multiple factors and not only with vitamin or nutrient deficiency. As Willet pointed out in the second chapter of his book *Nutritional epidemiology* major energy sources (proteins, carbohydrates, fats, alcohol), food additives, chemical contaminants (pesticides, herbicides), microbial toxic contaminants, inorganic contaminants, chemicals formed in the cooking or processing of food and natural toxins have played a role in understanding the relation of diet to disease. Moreover, dietary causes of disease have posed a significant challenge to the science of epidemiology, since the diet of an individual represents a large and complex set of exposures that are difficult to measure (Willet, 1998).

Independent effects of a dietary component of interest may be difficult to identify and can be partly confounded by other dietary exposures and an individual's socio-demographic (Gex-Fabry, Raymond & Jeanneret, 1988, Northstone, Emmett & Rogers, 2008) and behavioural patterns. In addition, collinear associations between food or nutrient variables can increase the uncertainty of the estimated models (Jacques & Tucker, 2001, Michels & Schulze, 2005, Randall et al., 1990, Kant et al., 1991). Individuals who try to eat a healthy diet are likely to lead a healthy lifestyle in general and it is not always possible to measure all important markers of a healthy lifestyle (Michels, 2003). A good example for potential confounding and lifestyle factors has been given in relation to per capita meat intake and colon cancer (Armstrong & Dol, 1975). However, in rich countries, people can afford to eat fat rather than starchier grain products. Some aspects of the diet in these countries, or other factors in the life-style, probably do cause certain kinds of cancer and protect against other kinds. So far, epidemiologists can identify only a few of these factors with any real confidence and fat is not among them (Willet et al. 1998).

Furthermore, there are unmeasured additive and interactive effects when foods are consumed in combination, which are difficult to take into account by single food or nutrient analysis (Sacks et al., 1995, Newby et al., 2006b). Studies have claimed that dietary constituents may interact with each other biologically in complex ways, for instance one nutrient may modify the absorption, metabolism or requirement for another nutrient (Willet, 1998, Sacks et al., 1995) and these interactions may have an effect on health. For example, vitamin D acts

synergistically with interleukins to inhibit the proliferation of MC-7 breast cancer cells and that calcium supplementation decreases the risk of adenoma but only in patients who consume a low fat diet (Messina et al., 2001).

Finally, since there is a large number of foods that are consumed by any population, when we try to identify associations between food or nutrient intake and disease, we are testing a large number of hypotheses and indiscriminate multiple testing can result in chance findings (Teo & Chong, 2006, Benjamini & Hochberg, 1995).

Over the last thirty years, based on the hypothesis that the complexity of diet-disease associations could not adequately be addressed using only traditional approaches, there has been an explosion in the use of dietary pattern analysis in nutritional epidemiological studies (Kant, 2004, Newby & Tucker, 2004, Hu, 2002). The main idea behind the use of dietary patterns is that diet of an individual may usefully be described in terms of a limited number of continuous variables, each representing a selection of different foods either eaten together of mutually excluded from the diet. So, since traditional approaches could not always take into account the complexity of the diet, and the use of a large number of different food intakes creates problems of multicollinearity and multiple testing, an alternative is to look at a small number of dietary dimensions each made up of a combination of foods. This alternative method was supported by two randomised controlled clinical trials where a dietary pattern approach seemed to be effective in lowering blood pressure (Dietary Approaches to Stop Hypertension (DASH) trial) and be protective in the recurrence of a number of potential outcomes after a first myocardial infarction (Lyon Heart Study).

In the first trial, 459 adults were randomly assigned and received a control diet, a diet that was high in fruits and vegetables and a combination diet that was high in fruits, vegetables and low-fat dairy products with reduced saturated and total fat for eight weeks. There was a decrease in the systolic and diastolic blood pressure of the individuals for the combination diet (5.5 and 3.0 mm Hg reduction of systolic and diastolic blood pressure), and for the diet rich in fruits and vegetables (2.8 and 1.1 mm Hg reduction of systolic and diastolic blood pressure) compared to the people who received a control diet (Appel et al., 1997).

In the second trial, a Mediterranean type diet (patients assigned to the experimental group were asked to comply with a Mediterranean-type diet) compared to a prudent Western-type diet (patients of the control group received no dietary advice from the investigators but

nonetheless were advised to follow a prudent diet by their attending physicians) reduced the number of adverse events of 423 patients after their first myocardial infarction (de Lorgeril et al., 1999). Specifically, all-cause and cardiovascular mortality and the combination of recurrent myocardial infarction and cardiac death (RR: 0.28; 95% CI:0.15-0.53), or the preceding plus major secondary end points (unstable angina, stroke, heart failure, pulmonary or peripheral embolism) (RR: 0.33; 95% CI:0.21-0.52), or the preceding plus minor events requiring hospital admission (RR:0.52; 95% CI:0.38-0.74) were reduced in the Mediterranean type diet group compared to the prudent Western-type diet group.

Dietary patterns (also referred to as food or eating patterns) can be derived either *a priori* or empirically from a set of data.

2.2 *A priori* dietary patterns

A priori dietary patterns use current nutritional knowledge from empirical research or theory based on prevailing hypotheses and guidance about the role of food items and nutrients in disease prevention. Diet is assessed by a dietary index that a research group has created in order to rank the presence or absence of certain food or nutrient characteristics, and the resulting score is used as an overall measure of dietary quality. As reported in a review by Kant et al. 2004 (Kant, 2004) the reported dietary indexes/scores as dietary patterns can be grouped into three major categories:

i) Dietary variety-based scores, such as the Dietary variety score based on a cumulative number of food items consumed on 15 consecutive days (Drewnowski et al., 1997), or the CARDIA dietary questionnaire in which diet variety was defined as the number of unique food items reported (Slattery et al., 1997).

ii) Scores derived from food related dietary guidance, such as the Diet Quality Index (DQI) (Haines et al., 1999) and the Health Eating Index (HEI) (Hann et al., 2001) which are analytic scoring tools (in a scale of 100 points) being used to measure compliance with dietary recommendations and guidelines.

iii) Mediterranean dietary scores which assess the conformity to the traditional Mediterranean diet with a 10 unit scale which relies on nine dietary components that capture the essence of the traditional Mediterranean diet. In the Mediterranean dietary score food items such as vegetables, legumes, fruits and nuts, fish and seafood and

cereals are presumed to be beneficial for health, whereas meat and dairy products are presumed to be harmful (Trichopoulou et al., 2003, Chatzi et al., 2007).

2.3 Empirically derived dietary patterns

Empirically derived eating patterns are not defined *a priori* but are data driven. Statistical methods are used to generate patterns from collected dietary data. Dietary assessment methods that are commonly used to collect the dietary data are food frequency questionnaires (FFQ), diet recalls or dietary records. These three dietary assessment methods are different in terms of how costly they are and what measurements of diet they provide (Willet, 1998). In brief, FFQ's are usually measure habitual consumption (portions of a food item per day, or units per week) of an individual over a long period (usually one year), while diet recall's and dietary records collect data over a small period (usually one week or less). In nutritional epidemiology data-driven methods that are commonly used to derive dietary patterns are presented below with a detailed description on Principal Component Analysis.

2.3.1 Cluster Analysis

Cluster analysis is the second most popular method after Principal Component Analysis for identifying dietary patterns of the population (Kant, 2004, Newby & Tucker, 2004, Hu, 2002). The basic aim of cluster analysis in nutritional studies is to find natural groupings, if any, of a set of individuals according to their dietary intake of specific food items. Newby et al.'s 2004 review indicated 2 commonly reported methods used in nutritional epidemiology for cluster analysis , which are going to be described in brief; Ward's and K-means (Chatfield & Collins, 1980, Anderson, 2003). Studies usually used both approaches in order to decide on the number of clusters (dietary patterns) that should be derived.

Ward's method is a hierarchical clustering method and is designed to optimize the minimum variance of individual's dietary intakes within clusters, which represent our dietary patterns. It uses analysis of variance at each merging step, considering all possible pairs of clusters and retaining the one with the smallest increase in the error sum of squares (Ward, 1963).

The K-means algorithm is a non-hierarchical simple, iterative procedure which is based on the definition of a point (centroid) in the space of records which represents an average location of the particular cluster (using the squared Euclidian distances between observations to determine cluster position). Thus, the coordinates of this point are averages of dietary intakes of all subjects who belong to the cluster. Each individual is positioned in space on the basis of

intake of numerous foods. Food choices common to all contribute less to cluster formation than those choices made by some and not by others. Clusters are named according to food items that on average contributed relatively more to total energy intake. The reason for this standardization was to account for differences in total energy needs due to demographic and lifestyle factors (Anderson et al., 2011). Labelling of the clusters could also be performed on the basis of the percentage of people in each cluster consuming lower or higher-than-median value of the food items stratified for.

After deciding on the number of clusters and the number of foods that constitute them, diet and disease associations are examined with clusters being the categorical exposure variables (Kant, 2004, Reedy et al., 2010, Engeset et al., 2005, Newby, Muller & Tucker, 2004b, Costacou et al., 2003).

2.3.2 Reduced rank regression

Reduced Rank Regression (RRR) applied in nutritional epidemiology is described in detail in Hoffman et al.'s 2004 paper (Hoffmann et al., 2004). In brief, RRR is a statistical data reduction technique which defines linear combinations of food intakes that maximally explain intermediate markers of disease. RRR requires two sets of variables for the identification of dietary patterns; *predictors* which could be the dietary intakes derived from a food frequency questionnaire and *responses* which could be nutrient intakes (Schulz et al., 2005, Hoffmann et al., 2005, Nothlings et al., 2008) or biomarkers that are in the pathways between the foods and health outcomes (McNaughton, Mishra & Brunner, 2008, Liu et al., 2009). RRR determines linear functions of *predictors* by explaining as much variation in a group of *response* variables. Finally, factor scores of the predictors are used as the exposure variables and their association with disease is investigated (Kroke, 2004).

2.3.3 Other data-driven methods

In brief, other less commonly data driven methods applied in food frequency data are:

i) Conditional Gaussian mixture modelling. A latent variable solution is proposed for dietary pattern analysis by using a finite mixture model to identify mutually exclusive subgroups of individuals with different dietary profiles. The main assumption of this method is that if subgroups exist in a sample of individuals who are distinguished by their dietary profiles, these subgroups would be expected to have different food intake

probability distributions which could be explained by a conditional Gaussian mixture model (Fahey et al., 2007).

ii) Cluster analysis of principal component scores (see a detailed description of principal component scores at paragraph 2.4.5) instead of food items. In Brief, principal component scores are calculated by applying principal components analysis on a Food Frequency Questionnaire, and then these obtained principal component scores entered in the cluster analysis procedure (He et al., 2009).

iii) Treelet transform (TT) which can be viewed as an amalgamation of PCA and hierarchical clustering methods (Gorst-Rasmussen et al., 2011). TT algorithm locates the two variables in the dataset with the largest correlation, performs PCA on them and creates a corresponding score. A merge is indicated in the cluster tree. This scheme continues until all of the variables have joined the cluster tree. TT procedure results in the production of pattern scores by aggregating dietary intake values according to correlation. Furthermore, TT singles out a smaller number of interrelated dietary variables than PCA by introducing sparsity to the principal component loadings; that is making a many loadings exactly zero (for a detailed description of principal component loadings see paragraph 2.4.7).

iv) In some studies researchers have employed factor analysis. Although factor analysis shares aims with PCA, it is not recommended for the analysis of nutritional data. The reason is that PCA is commonly used to define dietary patterns because the principal components are certain mathematical functions of the observed variables, whereas common factors are not expressible by the combination of the observed variables. Even when people say that they are employing factor analysis in nutritional studies they may be employing PCA, and this misconception depends on the statistical package that these studies have used (Agur-Collins et al., 2009, Hughes et al. 2009, Yuna et al., 2009).

2.4 Principal Component Analysis (PCA) and nutritional epidemiology

2.4.1 A brief history of PCA and its application in nutritional epidemiology

PCA is one of the oldest techniques of multivariate analysis. It was introduced by Pearson (1901) and developed by Hotelling in 1933. Over the last 80 years, as Jolliffe points out there has been a wide application of the Principal Component Analysis (PCA) method in the fields of psychology, agriculture, genetics, biology, chemistry , physics, meteorology quality control (Jolliffe, 2010). The use of PCA in nutritional epidemiology as a method for identifying dietary patterns is the focus of this thesis.

PCA as a method for reduction of measurements used to assess nutritional status dates back to a study by Drion published in 1961. In this study fourteen nutrients entered the PCA, and four principal components were identified which expressed (i) level of total consumption, (ii) relative importance of animal versus vegetable protein, (iii) quantity of fruits and vegetables (other than potatoes) consumed and (iv) the quantity of butter and margarine consumed (Drion, 1961). Another study in the 1970's employed PCA to reflect underlying processes that had created the correlation among thirty-two variables (demographic, socio-economic, anthropometric, dietary, biochemical and urinary) which were used to assess nutritional status (Gurthie et al., 1973). It wasn't until 1981 when the association between 7 eating patterns derived from PCA and health was examined from the TEN-STATE and HANES I survey (Schwerin et al., 1981). Although the PCA approach dates back half a century, there has been an explosion in the use of the method in recent years.

Specifically, our systematic review (see Chapter 3) identified 163 papers employing PCA in observational studies between 2004-2009 and other systematic reviews identified 41 papers between 1998-2004, and 13 between 1981-1997 (Kant, 2004, Newby & Tucker, 2004). The main reason why the PCA method applied in nutritional epidemiology gained so much attention after 1998 was a paper published in the American Journal of Epidemiology by Slattery et al., which associated a **"Western"** dietary pattern with increased risk of colon cancer and identified a **"prudent"** dietary pattern as protective against colon cancer in a population-based study conducted in Northern California, Utah, and Minnesota (Slattery et al. 1998). These two patterns of diet identified by Slattery were also identified in two other large US cohorts, where a "prudent" pattern characterised by intake of vegetables, fruit, legumes, whole grain, fish and poultry, and a "Western" pattern characterised by intake of red meat, processed meat, refined grains, sweets and dessert, French fries, and high-fat dairy products

were identified and subsequently linked to coronary heart disease (CHD) (Hu et al., 2000, Fung et al., 2001), colon cancer (Wu et al., 2004, Fung et al., 2003) and chronic obstructive pulmonary disease (Varraso et al., 2007a, Varraso et al., 2007b). This accumulating evidence along with the methodological problems that PCA claims to solve boosted the application of the method in nutritional studies (see Chapter 3; Figure 3.2.1). However, before we provide the methodological reasons why PCA became such a popular method in nutritional epidemiology, we will give a more detailed description of the method.

2.4.2 Description and general purpose of PCA in nutritional epidemiology

PCA, as applied in nutritional epidemiology, is a multivariate statistical method which uses food frequency questionnaire data (Willett, 1998, Hu, 2002), dietary records (Perrin et al., 2005) or dietary history questionnaire (Yannakoulia et al., 2008, Robinson et al., 2009) to aggregate information, reflect underlying processes and explain the variance-covariance structure of a set of correlated food intake variables. Specifically, PCA examines the relationships among a set of k correlated dietary exposures by transforming them to a new smaller set of uncorrelated overall dimensions of diet, while aiming to explain as much of the variation present in the original dietary exposures; these overall dimensions of diet are called principal components, or in our case, dietary patterns.

Although k components (dietary patterns) are required to reproduce the total system variability of the k original food intake variables, often much of this variability could be accounted from a smaller number of p components (dietary patterns). So, the original dataset of n measurements of k food intake variables could be reduced to a data set of n measurements of p components. In the next paragraph we will try to describe PCA more formally.

2.4.3 Derivation of population dietary patterns (principal components)

Let's assume that we have a random vector of *k* correlated food intake variables

$X^T=[X_1, X_2...,X_k]$ that have a covariance matrix k_xk

$$V[X^{\mathrm{T}}]=\Sigma=\begin{bmatrix}\sigma_{11} & \sigma_{21} & & \sigma_{k1}\\ \sigma_{12} & \sigma_{22} & & \sigma_{k2}\\ \vdots & & & \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_{kk}\end{bmatrix} \quad (2.4.3.1)$$

PCA solves the problem of finding a new set of k variables $Y^T=[Y_1, Y_2,..,Y_k]$ called dietary patterns (principal components) which are uncorrelated linear combinations of the original food intake variables $X_1,X_2,...,X_k$ and whose variances are large as possible and decrease from first to last. The first principal component is the linear combination of food intakes with maximum variance. The second principal component is the linear combination of food intakes with maximum variance subject to being uncorrelated with the first component. The k^{th} principal component is the linear combination of food intakes with maximum variance subject to being uncorrelated with all of the previous components.

More formally, we consider the linear combinations as

$$Y_1 = a_1^{\mathrm{T}} X = a_{11}X_1 + a_{21}X_2 + \ldots + a_{k1}X_k$$

$$Y_2 = a_2^{\mathrm{T}} X = a_{12}X_1 + a_{22}X_2 + \ldots + a_{k2}X_k \quad (2.4.3.2)$$

.

.

$$Y_k = a_k^{\mathrm{T}} X = a_{1k}X_1 + a_{2k}X_2 + \ldots + a_{kk}X_k$$

where

$$a_j^{\mathrm{T}} = [a_{1j},..., a_{kj}] = \begin{bmatrix} a_{11} & a_{21} & & a_{k1} \\ a_{12} & a_{22} & & a_{k2} \\ \vdots & & & \\ a_{1k} & a_{2k} & \ldots & a_{kk} \end{bmatrix} (j=1, 2,..., k) \quad (2.4.3.3)$$

is a vector of constants (coefficients).

The general linear form of equation (2.4.3.2) is given by

$$Y = \alpha^T X \quad (2.4.3.4)$$

Each dietary pattern Y_j $(j=1,2,..,k)$ is a weighted sum of the food intake variables X_i's $(i=1,2,...,k)$, and a_{ij}'s $(i=1,2,..,k\ j=1,2,..,k)$ are the weights of the food intake variable i on pattern j. According to Chatfield and Collins (Chatfield & Collins, 1980; page 24) the mean of vector Y as described by equation (2.4.3.4) is given by the formula

$E(Y) = a^T E(X)$ (2.4.3.5)

Moreover, variance and covariance of dietary patterns Y_j's ($j=1,2,..,k$) are given from the formulas

$Var(Y_j) = a_j^T \Sigma a_j$ (2.4.3.6)

$Cov(Y_j, Y_w) = a_j^T \Sigma a_w$ *(j#w; j=1,..., k; w=1,...,k)* (2.4.3.7)

From equation (2.4.3.6), we can maximize arbitrarily the variance $Var(Y_j)$ of dietary pattern Y_j ($j=1,2,..,k$) by multiplying the factor a_j^T ($j=1,2,..,k$) with a constant factor. So in order to have maximum variance and a deterministic value, a_j^Ts should be subject to the specific constraints of algebraic orthogonality.

$a_j^T a_j = 1$ ($j=1,..,k$) (2.4.3.8)

$Cov(Y_j, Y_w) = a_j^T \Sigma a_w = 0$ *(j#w; j=1,...k; w=1,...,k)* (2.4.3.9)

Dietary patterns $Y_1, Y_2, \ldots, Y_k$ are derived in decreasing order of importance in terms of explaining the largest proportion of total variance of the dietary intakes of the population. In order to do so, we need to use a mathematical optimization procedure for finding the minima and maxima of a function of several variables subject to a specific constraint through the method of Lagrange multipliers (Chatfield & Collins, 1980, Johnson & Wichern, 1982). Particularly, in order to maximize the variance of a **principal component (dietary pattern)** we need to introduce a new variable λ, which is called a Lagrange multiplier and it is the stationary point for the Lagrange function (for this point the derivative of the Lagrange function is zero and is the point when the function stops to increase or decrease). The mathematical problem that this procedure poses could be easily solved with the help of matrix algebra. So, a non-null solution for determining the a_j*'s* in order for the dietary patterns to have the required properties of orthogonality is by finding the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k \geq 0$ of the covariance matrix Σ of the food intake variables which are the roots of the equation

$|\Sigma - \lambda I| = 0$ (2.4.3.10)

To each eigenvalue λ_i *(i=1,..,k)* corresponds a vector c_i, which is called an eigenvector such that

$$\Sigma c_i = \lambda_i c_i \quad (2.4.3.11)$$

Principal components (dietary patterns) are the eigenvectors defined by equation (2.4.3.11).

Also using equation (2.4.3.10) and (2.4.3.6) we have that

$$Var(Y_1) = a_1^{\mathrm{T}} \Sigma a_1 = a_1^{\mathrm{T}} \lambda_1 I a_1 = \lambda_1 \quad (2.4.3.12)$$

Thus, since we want to maximize the variance $Var(Y_1)$ of the ***1st*** **principal component (dietary pattern)** subject to the constraint that $a_1^{\mathrm{T}} a_1 = 1$ we want to choose the largest eigenvalue λ_1 and the corresponding eigenvector $a_1^{\mathrm{T}} = [a_{11},..., a_{k1}]$ for this eigenvalue (Anderson, 2003, Bartholomew & Steele, 2002). In addition, in order to maximize the variance $Var(Y_2)$ of the ***2nd*** **principal component (dietary pattern)** Y_2 subject to the following constraints

$$a_2^{\mathrm{T}} a_2 = 1 \quad (2.4.3.13)$$

$$Cov(Y_1, Y_2) = a_1^{\mathrm{T}} \Sigma a_2 = 0 \quad (2.4.3.14)$$

we will choose λ_2 to be the second largest eigenvalue for the corresponding eigenvector $a_2^{\mathrm{T}} = [a_{12},..., a_{k2}]$. So accordingly, in order to maximize variance $Var(Y_j)$ of the ***jth*** **principal component (dietary pattern)** Y_j we will choose λ_j to be the j^{th} largest eigenvalue corresponding to the eigenvector $a_j^{\mathrm{T}} = [a_{1j},..., a_j]$.

So as a general framework

$$Var(Y_j) = \lambda_j \ , j=1,...,k \quad (2.4.3.15).$$

In the case that some of the eigenvalues of the covariance matrix Σ are equal, there is not a unique way to choose the corresponding eigenvectors but they should always be chosen to be orthogonal.

To sum up, this derivation of the Principal Components (dietary patterns), weights (coefficients) and variances as eigenvectors and eigenvalues of a covariance matrix is standard and appears in most text books. All equations presented above are taken from Chatfield & Collins and Johnson & Wichern (Chatfield & Collins, 1980, Johnson & Wichern, 1982).

Furthermore, we would like to express the total variance being explained by the principal components (dietary patterns) of the original food intake variables. So, let's assume again that (k_xk) matrix of eigenvectors

$$A = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{k1} \\ a_{12} & a_{22} & & a_{k2} \\ \vdots & & & \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} \quad (2.4.3.16)$$

And the (k_x1) vector of principal components

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} \quad (2.4.3.17)$$

with the covariance matrix of Y given by

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & & \\ 0 & & \dots & \lambda_k \end{bmatrix} \quad (2.4.3.18)$$

From equations (2.4.3.1), (2.4.3.6) and (2.4.3.18) we can derive that

$$\Lambda = Var(Y) = A^{\mathrm{T}} \Sigma A \quad (2.4.3.19)$$

The spectral decomposition of a square matrix could be given by

$$\Sigma = A \Lambda A^T \quad (2.4.3.20)$$

since A is an orthogonal matrix and

$$AA^T = I \quad (2.4.3.21)$$

Furthermore in linear algebra the trace of two square $k_x k$ matrices A and B is the sum of their diagonal elements

$$trace(A) = a_{11} + a_{22} + ... + a_{kk} = \sum_{i=1}^{k} a_{ii} \quad (2.4.3.22)$$

$$trace(B) = \beta_{11} + \beta_{22} + ... + \beta_{kk} = \sum_{i=1}^{k} \beta_{ii} \quad (2.4.3.23)$$

with the properties

$$trace(AB) = trace(BA) \quad (2.4.24)$$

so in our case

$$trace(\Sigma) = \sigma_{11} + \sigma_{22} + ... + \sigma_{kk} \quad (2.4.3.25)$$

$$trace(\Lambda) = \lambda_1 + \lambda_2 + ... + \lambda_k \quad (2.4.3.26)$$

So combining (2.4.24), (2.4.3.25), (2.4.3.26), (2.4.3.15), (2.4.3.21) we have that

$$\sum_{i=1}^{k} Var(X_i) = trace(\Sigma) = trace(A^T \Lambda A) = trace(A^T A \Lambda) = trace(\Lambda) = \sum_{j=1}^{k} \lambda_j = \sum_{j=1}^{k} Var(Y_j)$$

(2.4.3.27)

So the total population variance of food intake variables X_i's could be expressed by the sum of the variance of the dietary patterns Y_j's in terms of their eigenvalues and the proportion of total variance of the food items explained by the dietary patterns is given by

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \ldots + \lambda_k} \quad (2.4.3.28)$$

Consequently, the proportion explained by the first *p* dietary patterns together is

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_p}{\lambda_1 + \lambda_2 + \ldots + \lambda_k} \quad (2.4.3.39)$$

2.4.4 Principal Component Analysis with the use of a correlation matrix

In nutritional epidemiology, food intake variables X_i's are assumed to be continuous and usually standardised to have a mean zero and variance 1. So, when food intake variables are standardized before entering the PCA procedure, instead of using the covariance matrix Σ we effectively use the correlation matrix P. Let's assume that r_{ij} is the correlation coefficient between food item *i* and food item *j*, then

$$r_{ij} = \frac{Cov(X_i, X_j)}{Var(X_i)Var(X_j)} \; i=1, 2,...,k, j=1,..,k \quad (2.4.4.1)$$

$$\mathrm{P} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & & r_{2k} \\ \vdots & & & \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix} \quad (2.4.4.2)$$

All the properties for finding the principal components of a covariance matrix of our food intake variables are valid when we use the correlation matrix. As we can observe from the correlation matrix P all the diagonal terms are 1. So in this case, and according to equations (2.4.3.22) (2.4.3.23) and (2.4.3.27)

$$trace(P) = r_{11} + r_{22} + ... + r_{kk} = \sum_{i=1}^{k} r_{ii} = k \quad (2.4.4.3)$$

and the proportion of the total variance for j^{th} component (dietary pattern) will be given by

$$\frac{\lambda_j}{k} \quad (2.4.4.3)$$

Standardization and use of the correlation matrix needs to be applied in dietary data because of their widely different scales (e.g. bread could be measured in grams/per day and beer in pints /per day). If they are not standardized, PCA will simply pick up the variable that has the highest variance as the direction of greatest variability (Jolliffe, 2010). Another argument for using the correlation matrix is that we need to make food intake variables directly comparable between each other. This approach will also help to make principal components (dietary

patterns) more comparable across different studies and populations (Chatfield & Collins, 1980, Jolliffe, 2010).

Although we follow the exact same mathematical procedure for deriving the principal components, eigenvalues and eigenvectors provided from the correlation matrix *P* are different from that of the covariance matrix. Furthermore, principal components derived from the correlation matrix don't give the same information with the covariance matrix. However, variables that tend to have large weights to one component tend to be highly correlated with that component (Johnson & Wichern, 1982, Jolliffe, 2010). Finally, there are appropriate methods discussed in detail by Chatfield & Collins and Jolliffe when eigenvalues of the covariance or correlation matrix are small, equal or zero (Chatfield & Collins, 1980; page 62, Jolliffe, 2010; page 27).

2.4.5 Derivation of sample dietary patterns (principal components)

In nutritional epidemiology, we don't usually have information about the population under investigation but for a specific sample. So, using similar notations with that of paragraphs 2.4.3 and 2.4.4, we can assume that we have n independent drawings $x_1, x_2, ..., x_n$ from some k dimensional population with covariance matrix *S*, correlation matrix *R,* eigenvalues $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq ... \geq \tilde{\lambda}_k \geq 0$, vector of constants $\tilde{a}_j^{\mathrm{T}} = [\tilde{a}_{1j}, ..., \tilde{a}_{kj}]$ and p empirically derived principal components $y^T=[y_1, y_2,..,y_p]$.

PCA is a linear procedure and was originally developed for the normal multivariate distribution and samples from it, so the variables will need to be continuous or approximately normal in order for the method to provide accurate estimates. However is a very narrow view to apply PCA only when data are approximately normal (Jolliffe, 2010). As nutritional data are highly skewed no assumptions could be satisfied for the underlying population and without these assumptions is very difficult to derive the sampling properties of the above estimates. Hence, PCA should be used in nutritional epidemiology with caution and mainly as a descriptive and not as an inferential tool.

2.4.6 The identification of important dietary patterns (How many dietary patterns?)

In this paragraph we present a number of rules on how to derive an appropriate number of dietary patterns. Continuing from the previous paragraph, after calculating the eigenvalues of a population or sample correlation matrix, the next step is to determine which dietary patterns are accounting for a large proportion of total variance of the standardized food items; or in other words which eigenvalues are "large" and which are "small". There are three prevailing criteria in the majority of PCA studies to determine the number of dietary patterns.

Firstly, the number of patterns could be determined by examining the percentage of total variance that the dietary patterns explain in the original dataset and is derived from the equation (2.4.4.3). Most statistical textbooks of PCA provide a rule of thumb, which determines an acceptable number of principal components when they explain around 70-90% of total variance of the original variables (Chatfield & Collins, 1980, Bartholomew & Steele, 2002, Anderson, 2003, Jolliffe, 2010). However, in nutritional studies, the percentage of total variance reported is rather smaller with commonly reported percentages for dietary patterns being between 10% and 30% (median: 24%, IQR: 19.9-31.3; see paragraph 3.2.5 and the review by Newby (2004)).

Secondly, as we mentioned above, it is more appropriate for nutritional data to derive dietary patterns from the correlation matrix of food intake variables. In this case, another rule of thumb is to retain those components with eigenvalues >1. This rule is known as the Kaiser criterion (Kaiser, 1960) and the main idea behind it is that components with eigenvalues (variances) below one explain less variation than any one of the original variables. However, in nutritional epidemiology, a number of studies who decided the number of patterns based on the cut-off point for eigenvalues used a different cut-off point (median value:1.6, IQR:1.25-2) in order to retain a small and more easily interpretable number of dietary patterns (See paragraph 3.2.5 and the review by Newby (2004)).

Thirdly, a more objective way on deciding on the number of components is with the scree plot which was named by Catell (1966) and inspired by the scree at the bottom of a mountain slope. With the eigenvalues ordered from largest to smallest, this is a plot of the magnitude of an eigenvalue versus the dietary pattern number. The aim is to identify an elbow which

corresponds to the point after which the addition of more dietary patterns explains relatively little more of the variance.

A further criterion for deciding on the number which is described in more detail by Jolliffe (2010) is the Bartlett test which tests the hypothesis that all the eigenvalues are equal, assuming that our data are following a multivariate normal distribution

2.4.7 Principal component interpretability

Another way to decide on the number of dietary patterns could be done on the basis that the groupings of the food items suggested by the pattern have a realistic interpretation. Interpretation of the dietary pattern could be aided with the use of component loadings, correlation coefficients and the appropriate method of rotation (see below).

2.4.7.1 Component loadings/correlation coefficients and their contribution to the label of a dietary pattern and its interpretability

Component loading is the numeric size of a food intake variable within a principal component (dietary pattern). When the covariance matrix of food items is analyzed, component loadings are rescaled coefficients a_{ij}'s which we can calculate them as

$$a_{ij}^{*} = \lambda_j^{1/2} a_{ij} \quad \text{(i=1,.., k; j=1,...,k)} \quad (2.4.7.1.1)$$

The basic idea is to rescale the coefficients, in order that those coefficients for the most important components are larger than those of the less important components. In nutritional data, where the correlation matrix of the food items is analyzed, a_{ij}^{*} may be interpreted as the correlation coefficient (or component correlation) between food item *i* and dietary pattern *j* (a detailed proof is provided by Chatfield & Collins at page 63).

A positive correlation coefficient means that the food item is positively associated with the dietary pattern whereas a negative correlation coefficient reflects an inverse association with the dietary pattern. Correlation coefficients values of ≥ 0.30 and < - 0.30 between food items and the dietary pattern usually determine the label of the pattern (median value: 0.3, IQR: 0.3-0.4; see section 3.2.5 and review from Newby & Tucker (2004)). The rationale for ignoring near-zero correlation coefficients is that these will correspond to only minor displacements in the direction of the variables which they multiply, so can be safely disregarded (Cadima & Jolliffe, 1995).

2.4.7.2 Method of rotation

Interpretability of the patterns may increase with the appropriate method of rotation. Let's assume that we derive *p* components from *k* food items. By rotating the *p* components we could find a new set of components in the p-dimensional space which are more easily interpretable. There are two families of rotations, the orthogonal and the oblique. Orthogonal family rotates the derived *p* principal components in a way that the *p* components or component loadings maintain the orthogonality between them and are uncorrelated. Oblique families relax the orthogonality constraint in order to gain a simpler structure for the components. Although, Catell (1978) and Richman (1986) give non-exhaustive lists of 11 and 19 such criteria, there have been two methods of rotation, varimax(orthogonal) and promax(oblique) being used in nutritional epidemiology (See paragraph 3.2.5 and the review by Newby (2004)).

Varimax rotation is being used in the majority of the studies (see section 3.2.5 and review form Newby & Tucker (2004)) and the choice of this method over other methods of rotation is arbitrarily, that is no justification is given by the nutritional literature. This may be happening on the basis that the dietary patterns should be as uncorrelated as possible even after rotation. Fortunately, as Jolliffe (2010) points out, different choices of criteria, at least within orthogonal rotation, often make little difference to the results.

The method was suggested by Kaiser in 1958 (Kaiser, 1958) and attempts to find an orthogonal rotation that is close to a simple structure by finding *p* principal components with few large (towards the maximum possible value) component loadings or component correlations and as many near-zero one. The total variance being explained by the components remains unchanged. So, let's assume that *R* is a $k_x k$ orthonormal rotation matrix such as

$$RR^T = I \quad (4.2.7.2.1)$$

and R_{ij} is the element of the matrix in the i^{th} row and j^{th} column, and *L* is a $p_x k$ orthonormal matrix of column eigenvectors then varimax method maximizes the squared component loadings in each component according to the following definition by Stegman (2006).

$$R_{varimax} = \arg\max_R \left(\sum_{j=1}^{k}\sum_{i=1}^{p}(LR)_{ij}^4 - \frac{1}{p}\sum_{j=1}^{k}\left(\sum_{i=1}^{p}(LR)_{ij}^2\right)\right)^2 \quad (2.4.7.2.1)$$

As Jolliffe (1995) suggests, by using the normalization constraint at equation (4.2.7.2.1) the rotated varimax loadings are orthogonal but the rotated components could be correlated.

There have been 5 studies employing a promax rotation to the derived principal components in nutritional epidemiology (three of them as indicated by Newby (2004) review and two studies from our systematic review -see paragraph 3.2.5). A detailed description of the method is provided by Hendrickson and White (1964).

When the number of dietary patterns is determined and labelled then the next step is to include them as explanatory variables in a regression model.

2.4.8 Principal Components Regression (PCR)

Principal Components Regression (PCR) as applied in nutritional epidemiology is a regression analysis technique in which instead of using all the food intake variables as predictors, it uses the principal component scores of the derived dietary patterns which measure the conformity of an individual's diet to the given pattern.

More formally, we apply PCA in a nutritional dataset and the dietary patterns are empirically derived with the criteria mentioned in paragraphs 2.4.6 and 2.4.7. Principal component scores $[Y_1^*, Y_2^*, \ldots Y_p^*]$ are calculated for each pattern by summing the observed consumption from all standardized food items $(X_1, X_2, \ldots, X_k)$, weighted by the principal component loadings or correlations a^*_{ij}'s according to the equation

$$Y^*_j = a^*_{1j} X_1 + a^*_{2j} X_2 + \ldots + a^*_{kj} X_k \quad , j=1,\ldots, p \quad (2.4.8.1)$$

Then, we run an ordinary least square regression with use of a generalized linear model for exploring the association between a potential outcome and the p selected derived dietary patterns.

$$y_j = \beta_o + \beta_1 Y^*_{1j} + \beta_2 Y^*_{2j} + \ldots + \beta_p Y^*_{pj} \quad , j=1,\ldots,k \quad (2.4.8.2)$$

This technique has the advantage that, in the presence of confounding and mulitcollinearity, the uncorrelated dietary pattern scores may be more easily interpretable than the correlated food intakes. Furthermore, in the regression model, with the use of the uncorrelated principal component scores instead of all the correlated food intakes the estimated regression coefficients may remain unaffected from the presence of mulitcollinearity and more stable

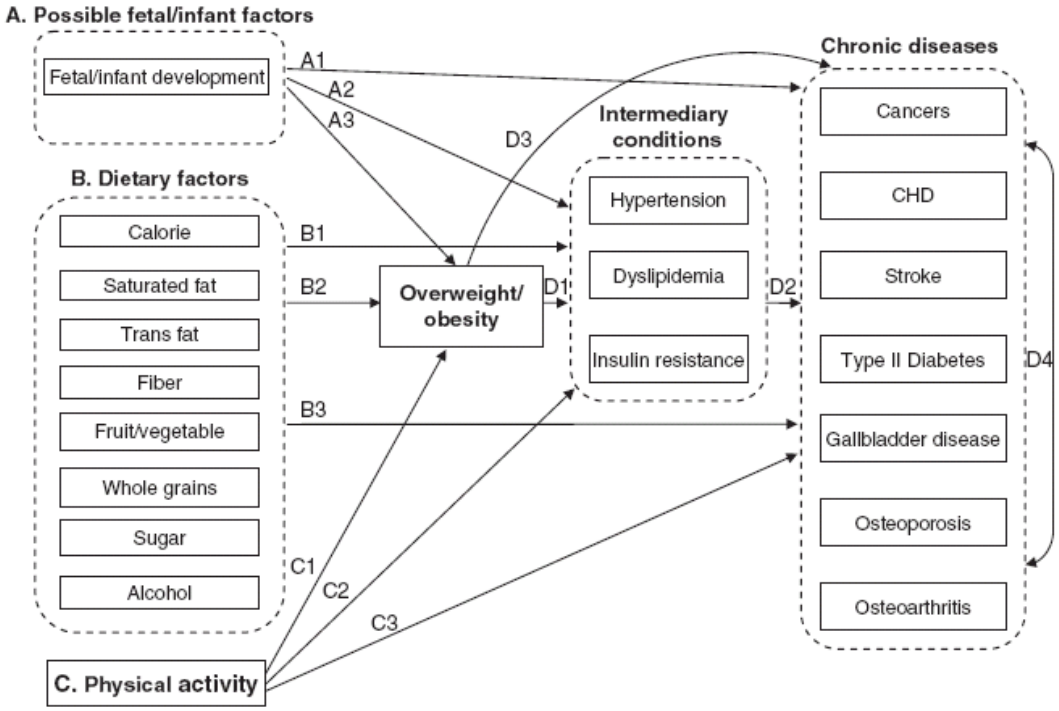
estimates can be obtained (Jolliffe, 2010). Because of the presence of correlation even after the rotation of a dietary pattern, usually patterns are included in the equation one at a time. Further adjustment of the dietary patterns for other socio-demographic and lifestyle confounders is usually employed.

2.5 Methodological justifications of Principal Component Analysis (PCA) and Principal Components Regression (PCR) and causal pathways of diet and disease at the dietary pattern, food and nutrient level.

As mentioned above (paragraph 2.1), although dietary causes of diet on disease could be logically addressed by epidemiology, the complex nature of diet has posed an unusually difficult challenge to this discipline. One of the most well studied exposures for epidemiologists is cigarette smoking and this is due to the accurate and easily obtainable quantitative information on the assessment of smoking status of an individual. On the other hand, diet represents a large set of dietary intake variables (exposures) which are highly correlated with each other. According to Willet (1998) everyone eat fat, fiber, and vitamin A and exposures cannot be characterized as present or absent; rather, they are continuous variables, often with a rather limited range of variation. As we can observe from Figure 2.5.1 (taken from page 61 (Kim & Popkin , 2010) paper) , there is numerous dietary factors which could act independently or interact with each other and potentially lead to the development of a chronic disease through different causal pathways. These pathways could be confounded by lifestyle, behavioural and early life patterns.

There are a number of methodological arguments on the rationalizing the use of PCA and PCR and how successfully they deal with the complexity of diet-disease associations (a list of them is presented for years 2004-2009 in our systematic review in Chapter 3 and Table A at the end of the thesis). Claims made for PCR include (i) that it captures additive effects of foods and nutrients on the investigated health outcome, which are too small to be detected when they are examined independently (Hu, 2002, Newby et al., 2006); (ii) that it captures the interactive effects of different nutrients and foods on the investigated health outcome (Hu, 2002, Varraso et al., 2007b); (iii) that it resolves aspects of confounding and mulitcollinearity between food intakes, so allowing more accurate estimates of the effect of diet on disease (Jacques & Tucker, 2001, Randall et al., 1990); (iv) that it reduces the risk of chance findings arising from multiple statistical testing (Hu, 2002, Teo & Chong, 2006).

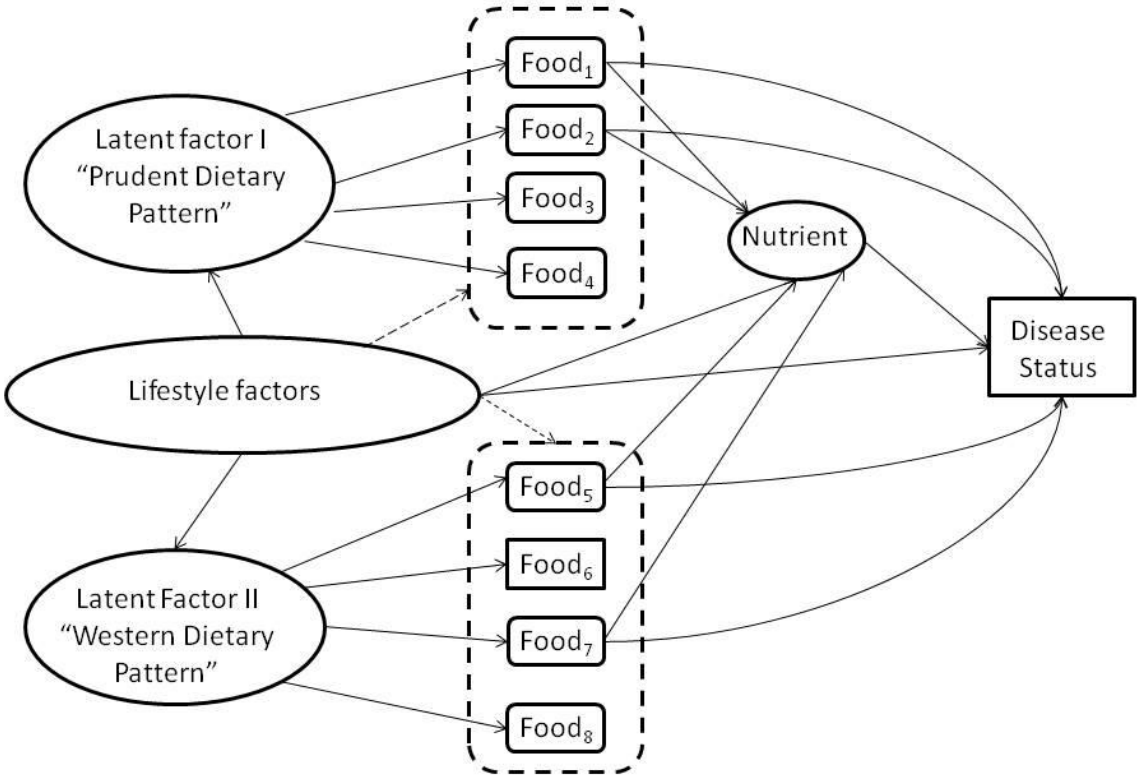
Figure 2.5.1 Key pathways for diet, physical activity, and obesity on nutrition-related non-communicable diseases (Kim & Popkin , 2010).



The argument for (iv) is that PCA reduces the number of food items into fewer dietary patterns and consequently with PCR the number of hypotheses to be tested. However the first three arguments are highly questionable, since there is a lack of quantitative evidence to support them.

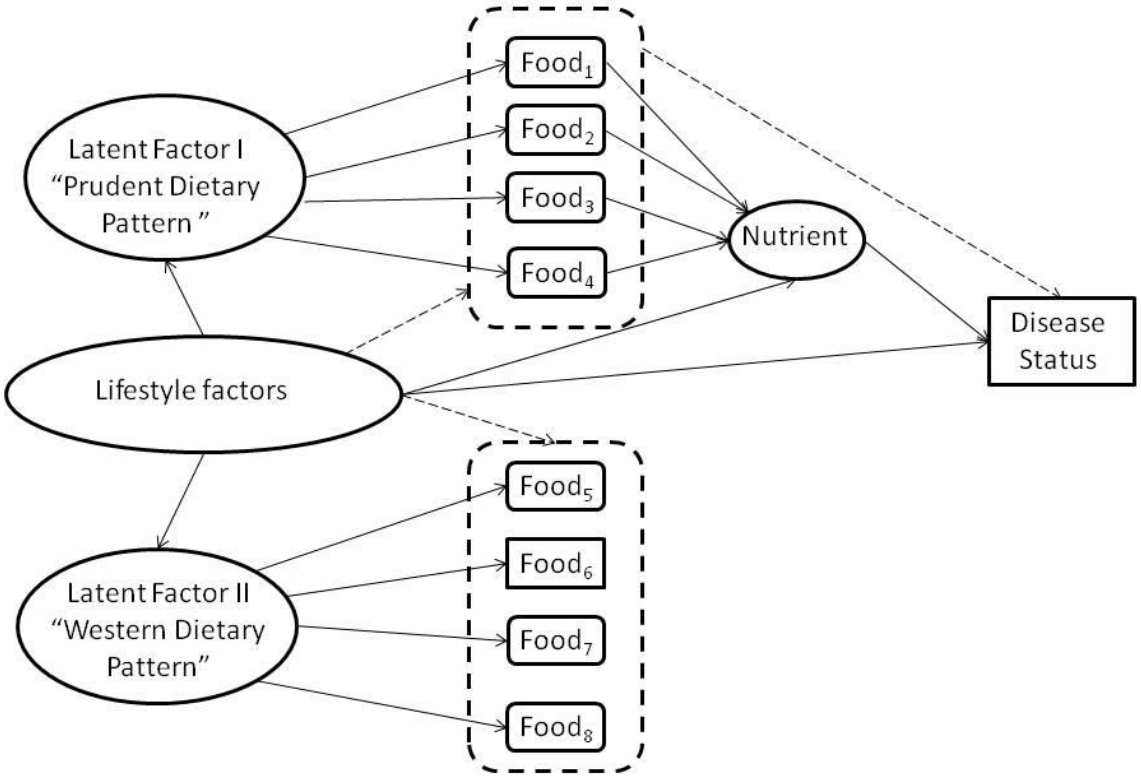
More formally, we can infer the presence of latent (i.e. unobserved) factors underlying dietary food intakes from the fact that we identify dietary patterns with the use of PCA; these identified dietary patterns are strongly correlated with a number of individual food intake variables. However, we are interested in finding causal effects of food intakes; thus, if we design a clinical trial to change food (see fruit/vegetable box in Figure 2.5.1) or nutrient (derived from foods; see saturated fat box in Figure 2.5.1) intakes of individuals, we could directly change the status of their disease risk (Figures 2.5.2, 2.5.3 and 2.5.4). Disease risk could be altered with a mechanism which could be different for each food, or these might all be foods that contain some special nutrient. (Figure 2.5.2). Associations of dietary pattern, food and nutrient intakes with disease risk could be confounded by lifestyle factors.

Figure 2.5.2 Potential pathways for diet when food intakes are not expected to be identified by PCA.



For example, observational studies have provided strong evidence that selenium is associated with the risk of asthma. Foods that are rich in selenium are Brazil nuts, sunflower seeds, fish (tuna, halibut, sardines, flounder, salmon), shellfish (oysters, mussels, shrimp, clams, scallops), meat (beef, liver, lamb, pork), poultry (chicken, turkey), eggs, mushrooms (button, crimini, shiitake), grains (wheat germ, barley, brown rice, oats) and onions. These foods are not expected to be found as foods that are highly correlated with a dietary pattern arising from PCA. However, if their standardised intakes are added together a "selenium" pattern score could be identified which is closely related to disease risk. Nevertheless, if PCA could identify this selenium pattern then we would have a different causal diagram (Figure 2.5.3). We try to address both of the causal pathways observed in the causal diagrams in figures 2.5.2 and 2.5.3 in our simulation study in Chapters 4 and 5.

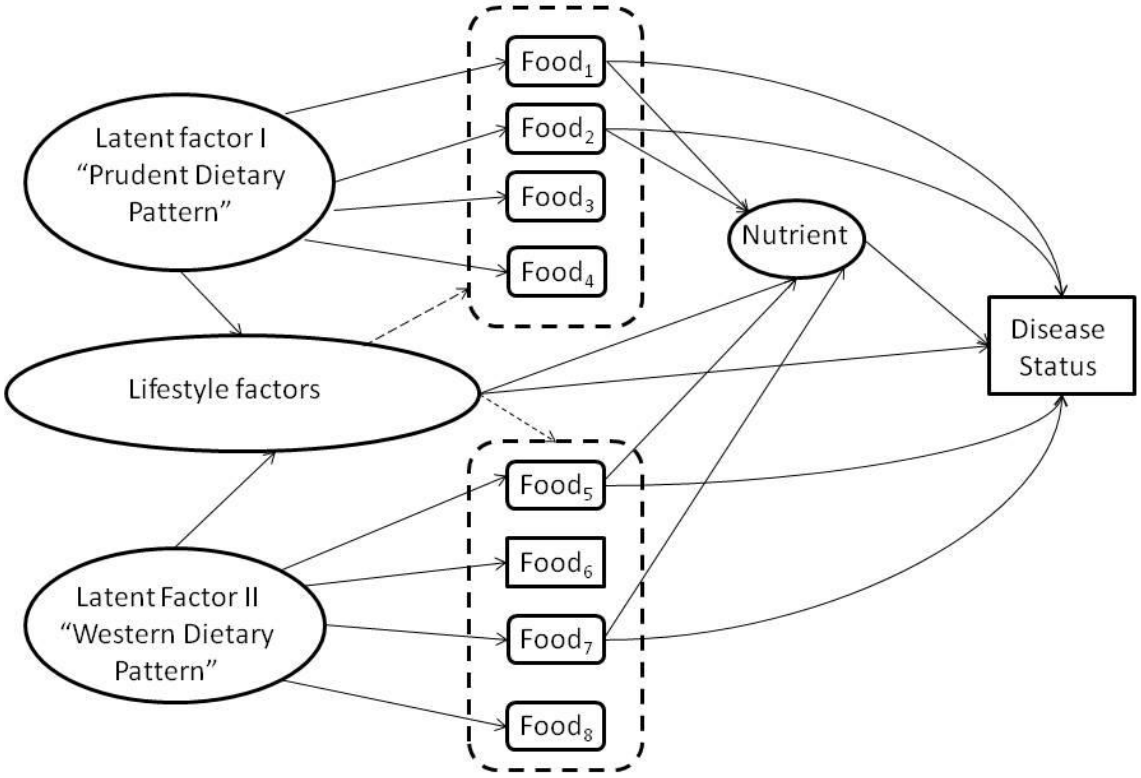
Figure 2.5.3 Potential pathways for diet when food intakes are expected to be identified by PCA.



Furthermore, latent factors such as a "prudent" dietary pattern that is observed in the majority of the PCA studies of diet (see Chapter 3) could be an unobservable measure of the "desire to be healthy" which makes individuals to eat healthier, but also to exercise more and follow a prudent lifestyle. If this is true, then intervening to change food intakes will not alter disease risk of the individuals, and lifestyle factors are on the causal pathway between dietary patterns and disease (Figure 2.5.4).

Finally, there are a number of methodological considerations on the application of PCA and PCR in nutritional epidemiology. We discuss all these issues in the next paragraph.

Figure 2.5.4 Potential pathways for diet when lifestyle factors are on the causal pathway between diet and disease.



2.6 Methodological considerations of Principal Component Analysis (PCA) and Principal Components Regression (PCR)

To the best of our knowledge, the greater ability of PCR to capture additive effects of foods and nutrients on the investigated health outcome when compared with the analysis of single foods or nutrients has not been critically evaluated. This is one of the goals of our simulation study described in Chapters 4 and 5.

PCA, as a statistical technique and its extension PCR do not take into account statistical interactions between food or nutrient exposures, since dietary patterns are merely linear combinations of food intake variables and nutrients are linear combinations of specific food variables as well.

PCR does not solve the problem of collinearity between dietary exposures, but rather ignores the problem. If a dietary pattern is found to be associated with disease, the question still

remains as to which of the correlated or confounded food intakes that make up the dietary pattern are implicated in the association. PCR cannot be specific about which particular foods or nutrients form the dietary pattern are responsible for having a protective or positive effect with associated disease and being informative about biological relationships between dietary constituents and disease risk (McCann et al., 2001b). We are trying to address extensively this argument with our simulations.

Furthermore, in PCR, if principal components are not easily interpretable, conclusions from the regression equation may not have a clear meaning. Even when dietary patterns that are derived according to the high-variance criterion of paragraph 2.4.6 and 2.4.7 (eigenvalues >1, scree plot, total variance being explained by the dietary pattern) have a meaningful interpretation, the omitted dietary patterns with low variance are not necessary unimportant for the regression model (Jolliffe, 2010). In addition, as mentioned in paragraph 2.4.7, the interpretation of dietary patterns which are associated with the outcome is usually based on the correlation coefficients ≥ 0.30 and < -0.30 between the food item and the derived dietary pattern, which can also suggest a subset of variables that can be used implicitly or explicitly, in a simplified interpretation (Jeffers, 1967). The rationale for ignoring near-zero loadings or correlation coefficients is that these will correspond to only minor displacements in the direction of the variables which they multiply, so can be safely discarded (Cadima & Jolliffe, 1995). However this approach of selecting a subset of variables according to non-zero or large correlation coefficients could provide misleading results, since this approach doesn't take into account correlations between the variables (Cadima & Jolliffe, 1995).

PCR, when one component alone contributes to the fit of the model, could fail to the account for the variation in the response variable. This could happen even in the case when the remaining principal components account for a large proportion of variation in the original variables . In addition, omitting low variance principal components could be problematic in the regression model, because they may be strongly associated with the dependent variable (Hadi & Ling, 1998).

Finally, PCA raises both conceptual and statistical issues. Conceptual issues include the appropriate number of patterns derived from PCA, how patterns should be named, whether food items should be aggregated before entering the PCA, how input variables should be correctly quantified, whether or not input variables should be adjusted for total energy intake, and whether dietary patterns should be derived separately for men and women. Statistical

issues which call for a subjective judgment by the researcher include choice of the correct cut off points for eigenvalues, choice of the cut-off points for principal component loadings or component correlations in order to label a pattern, appropriate method of rotation of PCA, total variance being explained by the dietary patterns, and the interpretation of dietary patterns (Kant, 2004, Newby et al., 2004, Michels & Schulze, 2005, McCann et al., 2001b). In addition, dietary patterns empirically derived with the use of PCA are not reproducible (Kant et al. 2004) and considered to express measures of lifestyle (Slattery et al., 1999, Maskarinec, Novotny & Tasaki, 2000). All these issues are discussed in detail on the discussion and conclusion section of the thesis in paragraph 7.3.1.3 with the help of our systematic review of Chapter 3 and two other reviews done previously (Kant, 2004, Newby et al., 2004).

In addition, in the next paragraph we will try to investigate why PCA is different from the other miultivariate techniques and why is preferred over them in nutritional analysis.

2.7 A Comparison of PCA with other multivariate techniques for dietary pattern analysis

2.7.1 PCA and distinguishing between cases and controls

An important issue in nutritional epidemiology is the characterisation of dietary intake in relating diet to chronic disease risk. Disease risk could be estimated correctly if distinction between cases and controls is achieved according to their nutritional status. As McCann et al. (2001b) suggested, if the method used to characterise dietary intake is inaccurate with regard to measurement of the characteristics that distinguish cases from controls, diet of the cases may appear more like the diet of the controls; thus, reducing our ability to identify important dietary intake risk factors.

PCA lacks some essential features for investigating the disease structure of particular populations, as its main purpose is data reduction. Principal Component Analysis does not provide a group assessment, and would require a priori definition to distinguish between healthily and unhealthy groups of individuals according to their specific dietary intakes. But even then, PCA does not aim to obtain a clear picture of disease variation but to summarize the overall dietary intake variation of the population.

As we described in Chapter 2.4, dimensionality of a large number of dietary intake variables is reduced into two or more high-variance principal components or so-called dietary patterns. Then with the use of PCR the effect of dietary patterns on disease is estimated. As Jolliffe (2010) described in his textbook, a common problematic assumption when PCA is being employed for distinguishing between healthy and unhealthy individuals is that the correlation matrix is the same for all groups. The second problem encountered in using principal components based on a common within-group correlation matrix to distinguish between groups is that there is no guarantee that the separation between groups will be in the direction of the high-variance PCs.

For example, if the first two principal components account for a high proportion of the variance, they can also be used to provide a two-dimensional graphical representation of the data showing how good is the separation between our cases and our controls (Figures 1 and 2; pages 202 & 203 from Jolliffe (2010)). In both figures the two groups are well separated, but in the first the separation is in the direction of the first principal component, whereas in the second the separation is orthogonal to this direction. Thus, in the second case, PCA searches for the direction showing the largest total variance and fails to distinguish between the two groups; PCA will only be useful for distinguishing between groups in the case where within- and between-group variation have the same dominant directions. If this does not occur (and in general there is no reason to expect to do so) then omitting the low-variance principal components may actually throw away most of the information concerning between-group variation (Jolliffe, 2010). Thus, 134/163=82% of the papers (Table E) could potentially have misused PCA in this context.

Discriminant analysis is a method which is concerned with identifying well defined groups or populations of individuals. Assumptions are made about the structure of the populations, and the main objective is to construct rules for assigning individuals to populations of cases or controls according to their dietary intake. There are different forms of Discriminant analysis which could be potentially be used in nutritional epidemiological studies. Linear Discriminant Analysis (LDA) is a statistical technique similar to regression analysis except that the dependent variable is categorical rather than continuous (Jolliffe, 2010) and its aim is to predict class membership to be a case or a control based on a set of predictor dietary intake variables (McCann et al., 2001b). However, LDA is not suitable in this context because dietary intake variables are highly correlated with each other, so the assumption of independence is violated. In this case, Discriminant Analysis of Principal Components (DAPC) could provide an

alternative. In brief, DAPC relies on data transformation using PCA as a prior step to LDA, which ensures that variables submitted to LDA are perfectly uncorrelated (Jombart et AL. , 2010).

Figure 2.7.1.1. Two data sets whose direction of separation is the same as that of the first (within-group) PC.

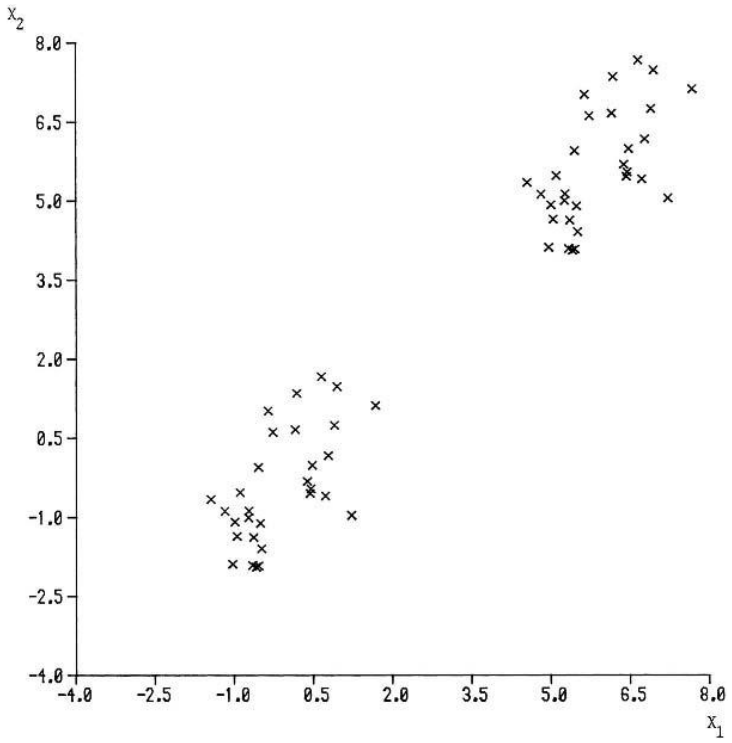
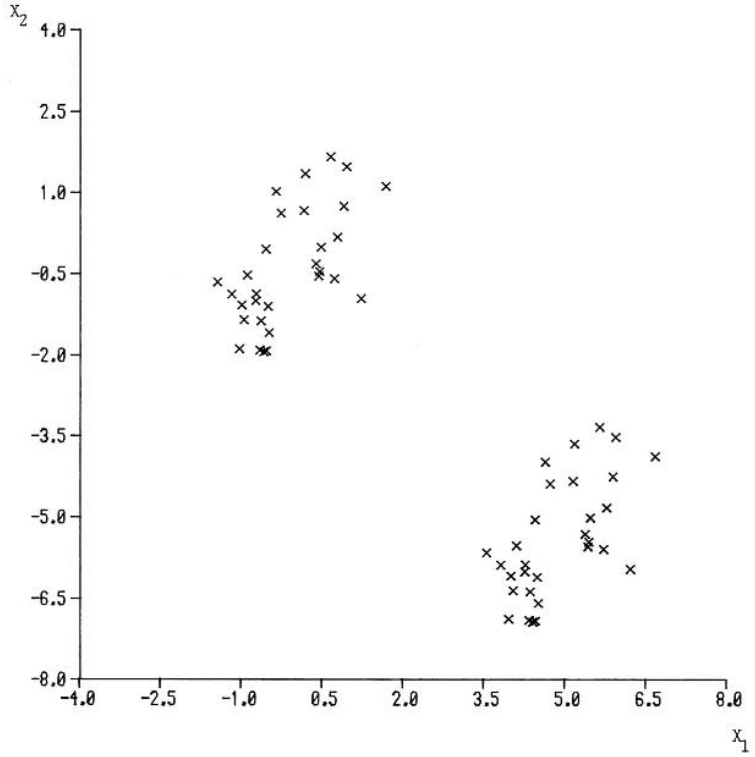


Figure 2.7.2.2. Two data sets whose direction of separation is orthogonal to that ofthe first (within-group) PC



2.7.2 Other multivariate techniques

Differences in the results given by PCA with a-priori methods could be explained simply on the basis that dietary patterns derived depend solely on how well individuals are scoring on a predefined notion of diet. In this paragraph we will focus on the conceptual and methodological differences between PCA and the two other most commonly used data-driven techniques, Cluster analysis and Reduced Rank Regression. Other methods mentioned from the literature, such as cluster analysis of principal component components scores and Treelet transform are methods that do not have a wide application and they are an amalgamation of Cluster analysis and PCA, so by reviewing Cluster analysis and PCA we are partially reviewing them as well.

Main fundamental differences between PCA, Cluster Analysis and RRR rely on how dietary pattern variables are constructed and what they represent when they are associated with disease risk. In brief, PCA of data from food frequency questionnaires (FFQs) groups food items and creates a number of dietary pattern variables dependent on the degree with which their reported intakes are correlated. Principal component scores are continuous summary measures which are used as exposures in our analysis (usually the highest quintile of the score is compared with the lowest one). Cluster analysis classifies individuals into naturally existing, mutually

exclusive groups on the basis of intake of numerous foods. After clusters are defined, there is no consideration for the variation of intake within a cluster among the individuals who comprise the cluster. Clusters are used as a categorical exposure by using the largest cluster as the reference category (Reedy et al., 2010). RRR groups food intake variables and constructs dietary patterns according to the covariance matrix of specific biomarkers (taken by blood samples), nutrients, (defined by a dietary questionnaire) or other biological and epidemiological evidence which are associated with the food intake variables and are assumed to be linked with the investigated disease outcome.

So, in a simpler conceptual framework, PCA aims to explain whether there are underlying factors that explain people's diet variation; cluster analysis aims to identify different groups of individuals in the population that are consuming on average different diets; and RRR aims to find out what variation in diet is important for the development of disease with the use of an *a priori* established biological hypothesis.

The main advantage of cluster analysis over PCA is that cluster analysis creates mutually exclusive groups which can easily be used in the analysis. On the contrary, in PCA, empirically derived patterns after a varimax rotation could be correlated with each other (see paragraph 2.4.7). Furthermore, because individuals are assigned to specific clusters according to their diets, results could be more interpretable compared to PCA.

On the other hand, PCA has the advantage of offering a technique for constructing a continuous score for individuals over a large number of foods. PCA does not only compare one group with a reference category as cluster analysis does. For example, let's assume that we derived two identically labelled dietary patterns with the use of cluster analysis and PCA; with cluster analysis we are testing if individuals who are sharing a similar "fish and vegetables" diet compared to those that sharing a "meat and potatoes" one have different disease distributions; with PCA we investigate the effect that has on disease the tendency of a population to follow a "fish and vegetables" or a "meat and potatoes" pattern. However, studies that compared PCA and cluster analysis have provided similar evidence in the nutritional literature (Bamia et al. 2007, Crozier et al, 2006, Kant et al., 2004)

The main advantage of the RRR method compared to PCA is that the pattern that is identified is associated with disease for a specific biological reason. However, when there is not a clear underlying relationship between specific markers and disease, or there are factors which are related with food intake but not with the specific markers, PCA is still claimed to be an

appropriate method to use (Kroke, 2004). Another potential advantage of RRR over PCA is that factor scores derived from RRR do not represent combinations of foods that characterize a pattern of the population diet, but combination of foods that describe a specific biomarker in the causal path with disease for that population. Finally, the RRR method as distinct from PCA and Cluster Analysis doesn't involve arbitrarily decisions on the number of factors to be derived, since this is determined by the number of biomarkers that are used as responses. Food items derived from RRR could have more plausible explanation from the foods indicated by PCA However both methods have poor repeatability of results when they used in different populations (Hoffman et al., 2004).

In conclusion, the wide use of PCA over the two other data-driven methods (Cluster Analysis and RRR) in nutritional epidemiology relies on the fact that it is simple to apply, creates a continuous numeric score and it doesn't require any plausible biological research hypothesis beforehand. PCA's long history and successful application as a data-reduction technique to different fields of research, led to the explosion of the method in the last decade in nutritional field; as indicated by our systematic review in Chapter 3 and the systematic review of Newby (2004).

In order to strengthened our arguments from our literature review in Chapter 2 and before describing our simulation study and present the results from it, we provide in Chapter 3 systematic evidence on how researchers employed PCA and PCR in years 2004-2009, and what were the key findings of the methods on the investigated health outcomes in nutritional epidemiology.

3 Systematic Review: Dietary patterns derived empirically from food frequency questionnaire data using PCA

“Suppose everyone had a box with something in it: we call it a "beetle." No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.”

Ludwig Wittgenstein

3.1 Systematic Review

This chapter systematically review the literature on nutritional epidemiology as it relates to the use of Principal Component Analysis (PCA) for identifying dietary patterns in observational studies.

3.1.1 Aim and objectives

The main aim of the systematic review is to provide a presentation and discussion (see paragraph 7.3 in Discussion and Conclusion section of the thesis) of systematically collected evidence of the methodological and conceptual issues of PCA and PCR as employed in dietary pattern analysis in nutritional epidemiology. Objectives of the systematic review are:

- To identify all the relevant literature between 2004 and 2009 regarding observational studies **(Tables A, B, C and D)**.

- To report, in each study, on how other researchers justify the use of PCA **(see paragraph 3.2.1 and Table A)**; on the different population settings in which these studies took place and on the sample size being used **(see paragraph 3.2.2 and Table B)**; on the type and design of questionnaires being used **(see paragraph 3.2.3 and Table B)**; on how researchers prepared the data before entering the PCA procedure **(see paragraph 3.2.4 and Table B)**; on the methodological decisions and numerical thresholds undertaken for deciding the number and the labelling of dietary patterns **(see paragraph 3.2.5 and Table B)**; on the number and the total percentage of variance of the food intake variables being explained by dietary patterns **(see paragraph 3.2.6 and Table B)**; and on the validation methods for PCA that have being used **(see paragraph 3.2.8 and Table B)**.

- To report on the foods and food groups that were correlated highly with a dietary pattern and how these influenced the labelling of the pattern **(see paragraph 3.2.7 and Table C)** ; on their corresponding principal component loadings or correlations that have being used for the labelling of the **"Western"** and **"prudent"** dietary pattern **(see paragraph 3.2.7 and Table D)**; on the level of consistency of these dietary patterns across the studies **(see paragraph 3.2.7 and Tables C and D)**; on the major findings from studies that have explored specific effects of dietary patterns on socio-demographic characteristics **(see paragraph 3.2.9 and Table C)** and disease **(see paragraph 3.2.10 and Table C)**.

- To discuss methodological issues of PCA in nutritional epidemiology **(see paragraph 7.3 in Discussion and Conclusion section)** and conclude on how PCA and PCR are being used in this specific context **(see paragraph 7.3 in Discussion and Conclusion section)**.

3.1.2 Inclusion and exclusion criteria

The inclusion criteria that were relevant to the systematic review are defined formally below:

1. Empirical papers deriving dietary patterns with the use of PCA in nutritional epidemiology.
2. Empirical papers with the objective of illustrating, testing, criticizing or appraising PCA compared to other methods being employed for dietary pattern analysis.

The exclusion criteria for the systematic review are defined formally below:

1. Conceptual papers on methodological issues of the dietary patterns approach using PCA including think pieces and reviews of methods papers without original data (this is covered from our literature review in Chapter 2).
2. Conceptual papers comparing methods of identifying dietary patterns without original data (this is covered from our literature review in Chapter 2).
3. Studies that were using *a priori* dietary patters such as quality index, Mediterranean diet score or healthy eating index.
4. Studies that were using data-driven methods other than PCA.

Four further exclusion criteria were also employed:

1. Papers prior to 2004. The reason for this is that there is already a systematic review published by Newby (2004) covering previous years.
2. Articles which provided only abstracts.
3. Articles presented only to conferences.
4. Non-English language papers (there were 4 non-English language papers that were excluded).

3.1.3 Search strategy

The search strategy used a variety of approaches. The main part was undertaken using electronic databases, but references were also obtained by other methods, such as hand searching journals and citation searches.

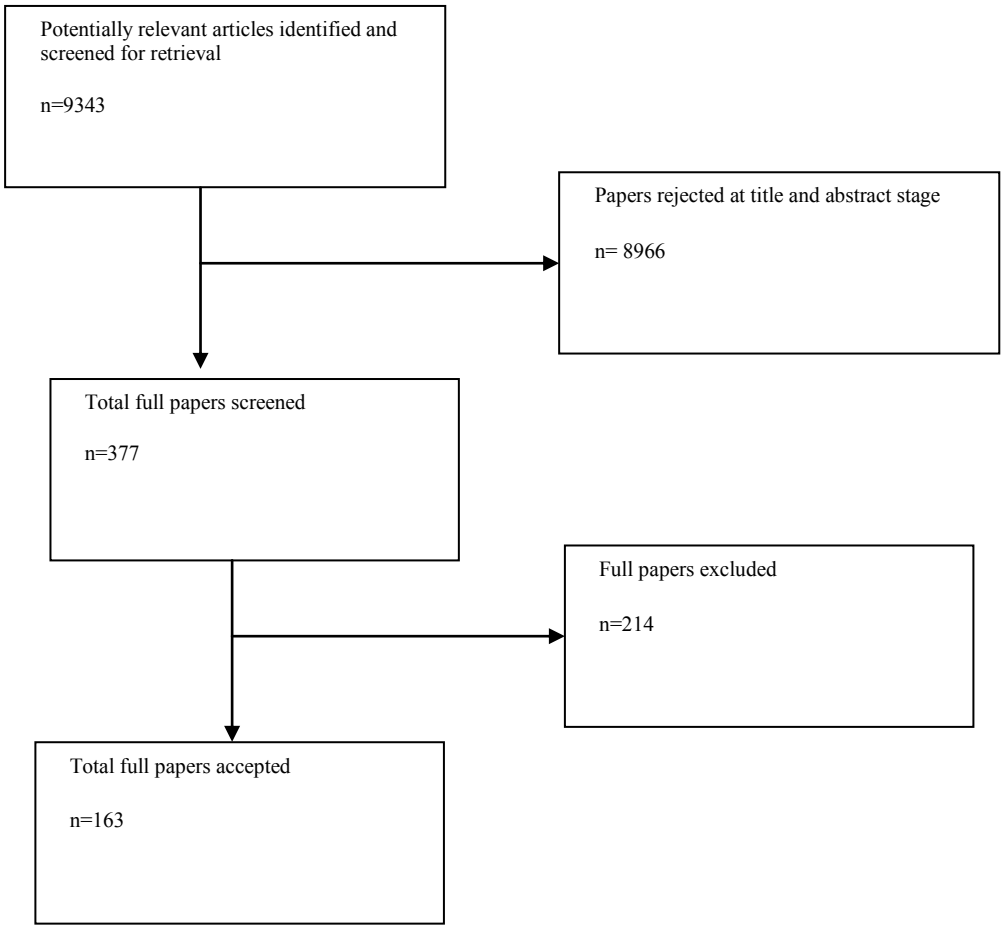
Specifically, from electronic databases, to identify dietary studies that have used PCA in dietary pattern analysis, OVID was searched using the terms "(factor analysis and (diet or nutr*)) OR (principal component analysis and (diet or nutr*)) OR dietary pattern* OR ((food pattern* or eating pattern*) OR ((PCA and (diet or nutr*)). In addition, Pubmed using the terms (factor analysis OR PCA OR principal component analysis AND (diet OR nutr*)) OR ((PCA and (diet or nutr*)) was also searched.

Titles and/or abstracts from all articles retrieved from these searches were reviewed to determine whether they should be included. Reference lists from selected articles were also reviewed to locate additional papers that were not retrieved in the OVID and Pubmed search. The search was restricted to human studies reported in the English language and published up to December 2009. Data abstraction questions of the systematic review are reported in Appendix I. All the data abstracted from the papers were stored in several excel sheets by year of publication.

3.1.4 Papers identified

The electronic search in OVID generated 11891 references in total which, after duplicates, amounted to 9343 unique papers. Of these, 377 articles were retrieved following examination of the title and abstract, of which 163 were considered to be consistent with the inclusion and exclusion criteria and were formally included in this review (Figure 3.1.4.1). Pubmed database was further explored for papers that were not identified by the OVID search but didn't provide any additional papers.

Figure 3.1.4.1. Summary of study selection and exclusion: electronic literature searches 2004-2009 from OVID.



3.1.5 Methods of Analysis

Summary tables and graphs are used to to help explain the patterns in the data. For each data abstracted question we provide percentages on the number of papers that include information, and summary statistics (median and interquatile range). Descriptive statistics of the systematic review are presented in Table E.

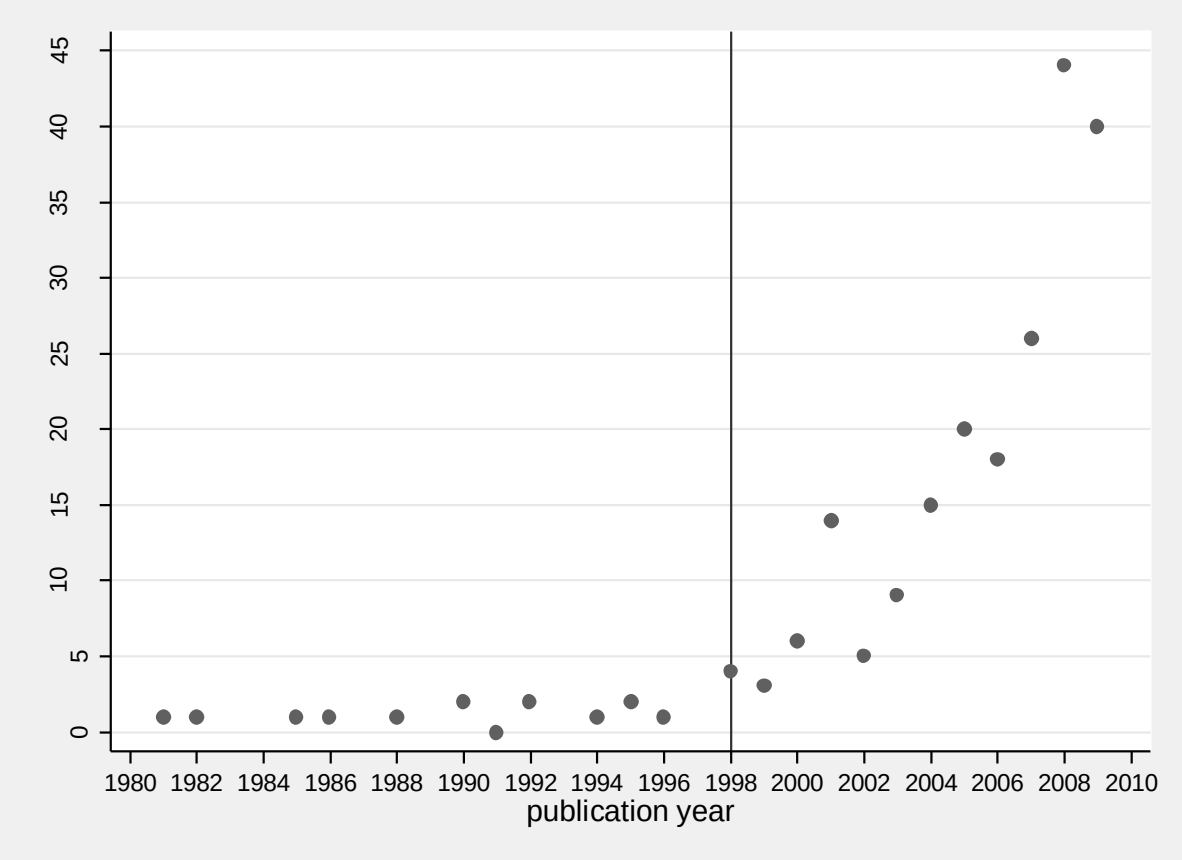
3.2 Summary findings

Our systematic review identified 163 papers employing PCA in observational studies between 2004-2009. Figure 3.2.1 presents number of these papers by these years along with data from two other systematic reviews (41 papers between 1998-2004, and 13 between 1981-1997) (Kant, 2004, Newby & Tucker, 2004). Articles from each table are summarized in their own section below. All the tables summarize why PCA is employed, how PCA is applied and how associations between diet and disease and socio-demographic population characteristics are explored through the method in each abstracted paper. Papers identified within the same cohort but reporting different research hypothesis are summarized in the same row of the table; in table A studies are organized by year of publication and alphabetical list of the author; in table B, studies are organized alphabetically by type of design and study cohort; in Table C, studies are organized alphabetically by health outcome; and in table D, studies are organized by year of publication, cohort study and alphabetical list of the author. Papers appear to be published at 2010 in our systematic review have been published electronically in year 2009, and that's why they are included. Although the term "principal component analysis" is used in this thesis, other terms, such as "factor analysis" (Agur-Collins et al., 2009), "principal components factor analysis" (Bertuccio et al., 2009) and "principal factor analysis" (Muller et al., 2009) were encountered for the same technique in papers in our systematic review. The main reason is that certain textbooks treat PCA as a special case of factor analysis, as do certain statistical computer packages (e.g. the command PROC FACTOR is SAS (SAS Institute, Inc, Cary, NC)).

Table A describes the main methodological justifications being used for the application of PCA in the abstracted dietary pattern studies; Table B summarizes the use of PCA on the populations and sample sizes being used in the articles, the type and design of questionnaires being used, how the data derived from these questionnaires are managed before entering the analysis, the criteria being used for employing PCA and articles which tried to validate the empirically derived patterns. Table C shows the number of patterns empirically derived with

the use of PCA in each article, and their associations with disease outcomes and socio-demographic characteristics. Finally, Table D presents the food elements that constitute a **"prudent"** and a **"Western"** dietary pattern and how they are labelled. For ease of reading all the tables are displayed together at the end of the thesis. A broader discussion of study findings on how PCA is commonly applied in nutritional epidemiology appears in the next ten paragraphs. Percentages and summary statistics of reasons why and how PCA is applied in the specific in nutritional epidemiology are presented in Table E at the end of the thesis.

Figure 3.2.1: Number of publications per year of empirically derived dietary patterns with the use of PCA



3.2.1 Main methodological justifications for the use of PCA in dietary pattern analysis

Methodological justifications for use of PCA (as described as well in paragraph 2.5) in the specific context as stated by the abstracted literature are that it is assumed that it capture additive effects of foods and nutrients on the investigated health outcome which are too small to be detected when they are examined independently (8.5% of the studies). PCA is also sometimes assumed to capture the effect of interaction and synergy between the different foods

and nutrients (15.3% of the studies). It has been claimed that PCA resolves aspects of confounding and mulitcollinearity, so allowing more accurate estimates of the effect of diet on disease (31.9% of the studies). It is also claimed that it can reduce the risk of chance findings arising from multiple statistical testing (4.2% of the studies). Public health policy is argued to be better understood in terms of dietary patterns than in terms of individual foods (8.5% of the studies). Finally, studies have claimed more general justifications on the basis that PCA could explain the complexity and provide a better assessment of the overall diet (18.4% of the studies) **(Table A and E)**.

3.2.2 Populations and sample size being used

Principal Component Analysis was employed in a variety of observational studies for diverse populations and settings, in every continent, including several countries in Europe , North America including Canada, Latin America (Argentina, Brazil, Mexico, Costa Rica , Uruguay), Middle East (Iran) and Asia (China, Japan, Singapore, Korea, Bangladesh), Caribbean (Jamaica, Puerto Rico), Australia, Canada and Africa (Botswana, Mauritius). Sample size of the populations being used varied from 115 (Custodio das Dores et al., 2007) to 492382 (Flood et al., 2008) (**Table B)**.

3.2.3 Details on type and design of questionnaire being used

Quantitative food-frequency questionnaires (FFQs) and semi quantitative food frequency questionnaires (SFFQs) (designed for measuring food consumption quantitatively and qualitatively) were employed as the primary assessment method in 93.2% of the studies. Some studies used diet history questionnaire (1.8%) (EPIC, DHQ) (Masala et al., 2007, Cottet et al., 2009, Robinson et al., 2009, Yannakoulia et al., 2008, Okubo et al., 2008, Waijers et al., 2006) , 24 hour recall (2.4%) (Cui et al., 2007, Kesse-Guyot et al., 2009, Hamer & Mishra, 2010, Kim et al., 2007), 48 hour recall (Mikkila et al., 2005) and dietary records (1.8%) (Balder, Goldbohm & van den Brandt, 2005, Cuco et al., 2006, McNaughton et al., 2007, Perrin et al., 2005, Newby, Muller & Tucker, 2004b) to assess diet. Individual food and nutrient intakes were derived from these methods and most studies collapsed the original measured dietary items into a smaller number of input variables, usually food groups, for entry into the principal component analysis (28.4% of the studies). Food items in the questionnaire varied from 11 (Takaoka & Norback, 2008) to 255 (Bakolis et al., 2010, Pala et al., 2006) (median value: 92, IQR: 21-204) and food groups derived varied from 13 (Romaguera et al., 2008) to 74 (Ambrosini et al., 2008) (median value: 38, IQR: 19-69). Scale of the FFQ being used varied

from 5 (Wiles et al., 2009, Sadakane et al., 2008, Keskitalo et al., 2008, Shimazu et al., 2007) to 10 (Bakolis et al., 2010, Flood et al., 2008, Campbell, Sloan & Kreiger, 2008, Yang, Kerver & Song, 2005) point scales (median value: 7; IQR:5-10). Four studies used macronutrient and/or micronutrient intakes in the analysis rather than foods or food groups (Bertuccio et al., 2009, Edefonti et al., 2008, De Stefani et al., 2008a, Corrao et al., 2004) **(Table B and E)**. Finally, 64% of the studies were cross-sectional, 19% were cohort, 17.1% were case-control studies and 2.4% were clinical trials **(Table E)**.

3.2.4 Preparation of data before the application of PCA

Food intake variables may be measured in frequency (servings), weight (grams), or daily percent energy contribution. Conversions to grams/d or grams/week and standardization of the food intake variables was reported in 16% of the studies. Moreover, in 10 studies food intake variables were adjusted for energy intake by the residual method before entering the PCA procedure (Willett, 1998) **(Table B and E)**.

3.2.5 Empirical derivation and labelling of the dietary patterns

Decisions for retaining the number of principal components were based on the following criteria. Cut - off points for eigenvalues to decide the number of principal components ranged from 1 (Kaiser Criterion) to 3 (Kim et al., 2008) (median value: 1.6, IQR: 1-2). In addition, 49.1% of papers used a scree plot and 42.9 % of the studies derived patterns according to their principal component interpretability. In one paper Van de Voet's test (DiBello et al., 2008) was also used for deciding the number of factors.

Decisions for retaining the label of principal components were based on the following criteria; Cut off points for component loadings or component correlation coefficients for deciding which foods constituted the dietary pattern ranged from 0.15 (Hughes et al., 2009) to 0.6 (Park et al., 2005) (median value: 0.3, IQR: 0.3-0.4). When principal component analysis was employed on a number of nutrients (3 studies), cut off points for component loadings ranged from 0.39 (De Stefani et al., 2007) to 0.63 (Edefonti et al., 2008). Principal components were rotated with the use of varimax (orthogonal) rotation for better interpretability of the patterns in 52% of the studies. Only four studies used promax (oblique) rotation (Kim et al., 2008, He et al., 2009, Lau et al., 2008, De Stefani et al., 2007) **(Table B and E)**.

3.2.6 Number of dietary patterns and percentage of total variance being explained by them

The number of empirically derived dietary patterns ranged from 2 to 10 (median: 3, IQR: 2-4) and the total percentage of variance being explained by the dietary patterns ranged from 11.2% (Hooper et al., 2010) to 88% (Romaguera et al., 2008) (median: 24%, IQR: 19.9-31.3). Not all the papers provided information of the total variance being explained by the dietary patterns.

3.2.7 Foods and food groups that were deemed to constitute a dietary pattern and how these influenced the labelling of the pattern

Principal Components were labelled quantitatively according to:

- Food items with the highest principal component loading such as **"coffee"**, **"bread"**.
- Food groups with the highest principal component loading such as **"vegetable"**, **"sweets"**, **"meats"**, **"alcohol"**, **"fruit", "bread eaters"**, **"less fish", "confectionary"**, **"plant-based"** (3 studies) , **"animal food"**, **"salad vegetables"**, **"vegetable-soy"**, **"meat-sweet"**, **"stew"**.

- Combination of food items and food groups such as **"vegetables, fruit and milk"**, **"fruit, salad, cereals, and fish"**, **"meat and fast food"**, **"fruit and milk"**, **"fish and sauce"**, **"potato and fish"**, **"cereals and legumes"**, **"alcohol and butter**", **"fruits and vegetables"**, **"whole food"**, **"potato and fish"**, **"cakes and sweets"**, **"fruits"**, **"coffee and dairy"**, **"meat-starch"**, **"sweets and soft drinks"**, **"British meat and two vegetables"**, **"fruit-rich"**, **"meat-rich"**, **"refined-grains"**, **"meet dim-sum"**, **"sweetened beverages and sugars"**, **"pasta and meat"**, **"olive oil and salad"**, **"sweet and dairy"**, **"pork processed meat and potatoes"**, **" cooked vegetables"** and **"green"**.

- Descriptions of dietary composition of the food items that were highly correlated to the pattern such as **"carbohydrate"**, **"antioxidants"**, **"low-fat/low-sugar"**, **"high-fat"**, **"high-protein/high-fat"**, **"phytoestrogen-rich"**, **"animal proteins"**, **"vitamins and fibre"** and **" unsaturated fats".**

- Descriptions of the way that foods were cooked such as **"processed"** (11 studies) with high positive principal component loadings or correlation coefficents between the

pattern and processed meat products, white bread, French fries, salty snacks, and sugar-sweetened drinks and high negative principal component loadings or correlation coefficients between the pattern and oily fish, high-fibre breakfast cereals, and lean fish.

Furthermore, dietary patterns were labelled qualitatively according to foods or food groups which were assumed to provide a degree of benefit. The majority of the studies decided to label their empirically derived dietary patterns in this way, as **"prudent"** (43 studies). The foods comprising a **"prudent"** diet varied from one study to another but the following characteristics appeared as part of a **"prudent"** diet in more than two studies: high intakes of vegetables, cheese, olive oil, fruit, wholemeal bread, non-fried fish, poultry, rice/pasta, beans, eggs, seafood, yoghurt and breakfast cereals and low intakes of white bread, roast potatoes/ chips, red/processed meat, full-fat milk, full-fat spread, crisps, confectionery, sugar, tea/coffee and Yorkshire puddings/ pancakes, tinned vegetables, cakes and biscuits and soft drinks **(Table D)**. Other dietary pattern names which were highly positively or negatively correlated with the same food items as the **"prudent"** pattern were entitled **"healthy"** (30 studies)**, "health conscious"** (5 studies), **"health aware"** and **"heart healthy"**. Under the same framework, patterns were labelled according to how unhealthy they were assumed to be, such as **"junk food"** and **"junk"** (3 studies)**, "fast food"** (2 studies) **"avoidance"** and **"unhealthy" (Table C)**.

Another way of labelling a pattern was according to the degree which was positively correlated with foods or food groups which were linked with a specific lifestyle. Specifically, patterns were labelled **"Western"**, **"Western-like"**, **"Western-type"**, **"Macho"** in 65 studies when they were strongly positevely correlated with cured and red meats, white bread and rolls, chocolate, margarine, butter desserts, potatoes, sweets, pizza, soft drinks, French fries, coffee, alcohol, high-fat dairy products, hamburger, eggs, bacon, mayonnaise, doughnuts, and negatively correlated with rice, vegetables and low fat milk **(Table D)**. In addition patterns were labelled **"vegetarian"** (6 studies) if they included negative loadings for food items that were avoided by vegetarians, principally meats; **"diet",** if they included foods items that were consumed mainly by people who were on a diet; **"sweet tooth"** and **"drinker"** (6 studies) **(Table C)**.

Other qualitative labels referred to cultural or geographic descriptions of dietary intake, such as **"traditional"** (39 studies)**, "modern"**, **"Iranian"**, **"Andean-like", "Mediterranean"**,

“ethnic”, **“yellow earth”**, **“green water”**, **“adopter”** , **“new affluence”**, **“oriental”**, **“native Mexican”**, **“urban/rural”**, **”seasonal”**, **“Korean”**, **“Swedish”**, **“healthy American”**, **“Japanese”**, **“traditional southern”**, **“traditional Finnish”**, **“traditional Dutch”** and **“canteen”**.

Finally, some articles chose to name the principal components as **“Component 1 and 2”** and **“Factor 1 and 2” (Table C).**

(Note: all counts refer to separate study populations and not to separate reports from the same study population. We counted only patterns that were observed in more than one population)**.**

3.2.8 Validation methods being used after PCA for confirmation of the derived patterns

Very few validation studies of eating patterns have been performed. Five of the validation studies considered the stability of patterns over time, an indicator of reproducibility, by comparing factor solutions from an FFQ at different time points (Northstone, Emmett & Rogers, 2008, Lau et al., 2008, Fung et al., 2005, Khani et al., 2004, Togo et al., 2004). Confirmatory factor and maximum likelihood analysis for principal components were used to test the internal validity of a pattern in 8 of the studies. Internal validation of the empirically derived principal components was also performed by randomly splitting the study sample and repeating the principal component analysis to the two sub-samples in 8 of the studies; by deriving patterns separately for men and women and comparing the results in 3 of the studies; by using Cronbach's alpha reliability coefficient which evaluates to what extent food items measure the same underlying content when the food items are combined into a scale in 5 of the studies; by using Barlet's test of sphericity (1 study) which tests the null hypothesis that our sample was randomly drawn from a population in which the correlation matrix was an identity matrix (a matrix full of zeros, except for ones on the main diagonal); by employing the Kaiser - Meyen-Ollkin test (2 studies) which tests whether the partial correlation coefficients among our food intake variables are small; by constructing simplified dietary patterns according to Schulze method (1 study) (see paragraph 4.9); and by employing a different method of rotation (2 studies) (**Table B)**.

3.2.9 Associations between Principal Components and socio-demographic characteristics

Table C presents results from 60 studies (36.8% of the overall number fo studies) that have examined associations between empirically derived dietary patterns with the use of PCA and

socio-demographic characteristics. Studies reported a high **"fruits, vegetables and oily fish"** pattern consumption (labelled as **"prudent"** or **"health conscious"** or **"vegetable** or **"fruit and vegetable"** or **" Mediterranean"** or **"healthy"** or **"low-fat"** or **"vegetable-soy"** or " **fruit, salad, cereals, and fish"** and were positively associated with women, older age, vitamin and supplement use, height, income, social class, strenuous exercise, education, urban living, partnership status, non-smoking status, non-obese status. A high **"red meat, processed foods, sugary items and soft drinks"** pattern of consumption (labelled as **"Western"** or **"processed"** or **"processed meat"** or **"organ meat and fast food"** or **"meat and French fries"** or **"alcohol/meat"** or **"unhealthy"** or **"high energy"** or **"meat-dim sum"** or **"high-meat"** or **"fat-reduced and diet-foods"** or **"fat and sugar"** or **"junk"** or **"vegetables and fruit and milk"**) was inversely associated with age, exercise, education, social class, height and positively associated with being male, BMI, smoking and drinking, height urban and rural residence**.** A **"convenience foods"** pattern consumption was inversely related to age and positively related to higher education in both genders and a **"starch, sauces, and vegetables"** pattern consumption was associated with high education and an urban residence (Kesse-Guyot et al., 2009). People consuming a **"carbohydrate"** pattern showed lower use of vitamins and were less likely to be overweight or obese. An **"alcohol"** pattern was inversely associated with weight gain (Jackson et al., 2009). Men in the highest tertile of the "**sugary foods and sweet baked goods"** pattern were more likely to be current smokers and were less likely to have tertiary education (Jackson et al., 2009, De Stefani et al., 2009). A **"traditional"** pattern consumption based on a UK population was more common in men, in couples and was associated with higher alcohol use (Robinson et al., 2009). People consuming a **"traditional / Iranian"** dietary pattern score were older, slightly more physically active (Esmaillzadeh & Azadbakht, 2008b). People scoring high in a **"Japanese"** pattern were less likely to be smokers and drinkers (Hirose et al., 2007). Subjects with a high score for the "**drinker"** pattern were younger and were more frequently current smokers. **"Stew"** pattern was inversely correlated with education (Mizoue et al., 2006). Higher scores for the "**traditional Dutch dinner**" dietary pattern were associated with women who had a lower level of education, were current smokers, and were more overweight (Waijers et al., 2006). The **"traditional Korean"** (Yang, Kerver & Song, 2005) dietary pattern was negatively associated with length of residence in the U.S. for both men and women. Older women and with no children had high scores for the **"canteen"** pattern (Sieri et al., 2004) **(Table C)**.

3.2.10 Major findings between Principal Components and health outcomes

Table C presents results from 120 studies (73.6% of the total number of studies) that have examined potential associations between empirically derived dietary patterns with the use of PCA and disease outcome. Dietary patterns to major health outcomes which are examined will be presented briefly as evidence of the potential predictive consistency of the PCA method only.

Derived patterns were examined in relation to many different outcomes, including indicators of cardiovascular or coronary heart disease; different type of cancers (gastric, breast, prostate, colorectal, bladder, gastric, lung, pancreatic, laryngeal, endometrial); overweight and obesity; metabolic syndrome and its components, including hypertension, blood pressure, cholesterol, blood glucose, and blood insulin measures; type 2 diabetes; all-cause mortality; depression; respiratory conditions including asthma, chronic bronchitis, wheeze, cough with phlegm and COPD; Preeclampsia; VTE (venous thromboembolism); Behavioural scores; spina bifida; myocardial infarction; Barrett's oesophagus; school attainment; dyslipidemia; Crohn's Disease, constipation; overall health status; and stroke.

Overall mortality

An increase in the pattern score which measures the adherence to the **"plant-based"**, (Bamia et al., 2007), **"fruit-rich"** (Cai et al., 2007) and a **"healthy traditional"** (Waijers et al., 2006) diet was associated with a lower overall mortality. An association of the **"pasta & meat"** (Bamia et al., 2007), **"meat-rich"** (Cai et al., 2007) pattern with increased overall mortality was suggested by the literature. Furthermore, "**prudent"** (Masala et al., 2007), "**Mediterranean"** (Waijers et al., 2006)**, "meat, potatoes, legumes and bread"** and **"vegetables, fruits and dairy products"** (Hoffmann et al., 2005) patterns were not associated with overall mortality.

Cardiovascular Disease (CVD)

A "**whole grains and fruit"** (Nettleton et al., 2009) and a **"prudent"** pattern were associated with a lower risk of cardiovascular disease (Heidemann et al., 2008). On the other hand, a **"Western"** (Heidemann et al., 2008), **"animal food"** (Shimazu et al., 2007) and **"healthy" (**Esmaillzadeh & Azadbakht, 2008a) patterns were associated with a greater risk of incident CVD. In addition, a **"component 1"** pattern which loaded mainly on low-fat products like fish,

vegetables, legumes, greens, and salads, as well as olive oil was associated with lower likelihood of having increased burden of CVD ((et al., 2007a).

Myocardial Infarction (MI)

A **"vegetable"** (DiBello et al., 2008), and **"prudent"** (Iqbal et al., 2008) patterns were associated with a significantly decreased risk of myocardial infarction. .A **"western"** pattern showed a U-shaped association with Acute Myocardial Infarction (Iqbal et al., 2008). **"Healthy"** and **"alcohol"** patterns were positively associated with increased risk of MI (Akesson et al., 2007).

Type 2 Diabetes mellitus including Type II diabetes

Dietary patterns characterized by high intake of **"whole grains, fruit, nuts/seeds"**, **"green leafy vegetables", "low-fat dairy"** (Nettleton et al., 2008b) and **"prudent"** (Montonen et al., 2005) characteristics of diet were associated with lower diabetes risk. Food patterns characterized by high intake in "**high fat and sugar"** (McNaughton, Mishra & Brunner, 2008) and **"tomatoes, beans, refined grains, high-fat dairy, and red meat"** (Nettleton et al., 2008) were associated with increased risk of type 2 diabetes. **"Western"** and **"prudent"** (Imamura et al., 2009, Fung et al., 2004a) pattern scores were not associated with Type 2 diabetes. Finally, a **"Westernized breakfast"** was inversely and a "**seafood"** dietary pattern was positively associated with A1C concentrations related to type 2 diabetes (Nanri et al., 2008b).

Metabolic Syndrome

Subjects in the highest quintile of a **"healthy"** dietary pattern score had a lower odds ratio for the metabolic syndrome. Consumption of a **"Western"** (Esmaillzadeh & Azadbakht, 2008a) and **"sweets"** (Noel et al., 2009) dietary pattern was adversely associated with incident metabolic syndrome. **"Korean traditional"**, "**Western"** (Kim et al., 2007) and **"prudent"** (Lutsey, Steffen & Stevens, 2008) patterns were not statistically significant associated with metabolic syndrome.

Types of Cancer

Breast

A **"prudent"** diet was inversely (Hirose et al., 2007, Agurs-Collins et al., 2009) not (Fung et al., 2005, Kroenke et al., 2005, Robinson et al., 2004), and positively associated (Murtaugh et al., 2008) with breast cancer. A **"Western"** pattern was positively (Cottet et al., 2009, Murtaugh et al., 2008, Wu et al., 2009, Ronco et al., 2006) and not associated (Fung et al., 2005, Agurs-Collins et al., 2009, Kroenke et al., 2005, Robinson et al., 2004) with breast cancer. **"Traditional"**, **"healthy"**, **"stew"** (Ronco et al., 2006)**, "vegetables"** (Takata et al., 2007)**, "salad vegetables"** (Sant et al., 2007), "**salad–sauce–pasta/grain"** (Tseng et al., 2008)**, "vegetables"** (Takata et al., 2007) diets were significantly protective against breast cancer.

Colon

An increased risk of colon cancer and distal adenoma was suggested with higher **"Western"** (Wu et al., 2004a, Meyerhardt et al., 2007) and **"pork, processed meats and potatoes"** (Dixon et al., 2004) pattern scores. Higher **"prudent"** pattern scores were only weakly and non-significantly associated with decreased risk of colon cancer or distal colon adenoma (Wu et al., 2004a).

Colorectal

A "**Mediterranean"** pattern significantly reduced colorectal cancer (Cottet et al., 2005). High scores of "**meat-eaters"** (Kesse, Clavel-Chapelon & Boutron-Ruault, 2006), **"pork, processed meats and potatoes"** (Dixon et al., 2004), **"meat and potatoes"** (Reedy et al., 2010) and **"red meat"** (Flood et al., 2008) patterns had higher risk of developing colorectal cancer risk. **"Healthy"**, **"fruits and vegetables"**, **"drinker"**, **"snacks"** (Kesse, Clavel-Chapelon & Boutron-Ruault, 2006, Cottet et al., 2005), **"meat–dim sum"**, **"vegetable–fruit–soy"** (Butler et al., 2008), **"healthy"**, **"Japanese"** and **"animal food"** (Mizoue et al., 2005) patterns were not associated with colorectal cancer risk.

Gastric

"Healthy", "prudent" (Kim et al., 2004, De Stefani et al., 2004) **"fruit, salads, vegetables, dairy products, fish and meat"** (Bastos et al., 2010) **"vitamins and fiber"** , **"vegetable", "mixed"** (De Stefani et al., 2004) patterns were associated with decreased risk of gastric cancer. A positive association between gastric cancer risk and the "**animal products"** ,"**starch-rich"** (Bertuccio et al., 2009) ,"**Western"** (Campbell, Sloan & Kreiger, 2008) and

"**starchy"** (De Stefani et al., 2004) dietary pattern was observed. A "**Western"** (Kim et al., 2004) dietary pattern was not associated with risk of gastric cancer.

Lung

"**Antioxidants"** (De Stefani et al., 2008a) **''salad vegetables''** and **''sweet foods''** (Balder, Goldbohm & van den Brandt, 2005) pattern were all inversely associated with risk of lung cancer. A **"high-meat"** pattern was associated with a strong increase in risk of lung cancer (Dixon et al., 2004). The **"carbohydrates"** pattern was not associated with risk of lung cancer (De Stefani et al., 2008a).

Prostate

An increased risk for prostate cancer was observed with a higher intake of a **"Western"** (Ambrosini et al., 2008, Arkkola et al., 2008) and a **"Southern"** pattern (Tseng et al., 2004). There were no associations between **"meat & potatoes"** (Muller et al., 2009) **"red meat-starch"** (Tseng et al., 2004) **"Western"** (Wu et al., 2006), **"vegetable", "health-conscious"** (Ambrosini et al., 2008), **"Mediterranean", "fruit & salad"** (Muller et al., 2009), **"prudent"** (Wu et al., 2006), **"vegetable-fruit"** (Tseng et al., 2004) and **"healthy and carbohydrate"** (Jackson et al., 2009) patterns with overall prostate cancer risk.

Other

The **"prudent"** pattern was directly associated with risk of bladder cancer (De Stefani et al., 2008a). **"Pattern 5"** (drinker) and **"pattern 6"** (western) were directly associated with risk of laryngeal cancer whereas the **"pattern 2"** (healthy) was protective. (De Stefani et al., 2007) . No associations were observed between the **"prudent"** and **"Western"** pattern and the risk of pancreatic cancer (Michaud et al., 2005). Intakes of **"fruits and vegetables"** pattern were associated with a reduction in risk of pancreatic cancer (Nkondjock et al., 2005). A **"healthy"** pattern was not significantly associated with decreased risk of Renal Cell Carcinoma (all cancer and kidney cases) (Rashidkhani et al., 2005). A **"plant-based"** diet had higher ovarian cancer risk (Chang et al., 2008).

Respiratory and allergic symptoms

A **"vegetarian"** dietary pattern was positively associated with asthma (Bakolis et al., 2010). There were no evidence that **"fish, fruits and vegetables"** (Hooper et al., 2010b), **"health**

conscious" (Shaheen et al., 2009) **"western"** and **"prudent"** (Bakolis et al., 2010) patterns were associated with asthma. A **"Western"** pattern was associated with an increased risk of reporting frequent asthma attacks and a **"nuts and wine"** with decreased risk (Varraso et al., 2009). A **"prudent"** pattern was positively associated with chronic bronchitis (Bakolis et al., 2010). An **"urban"** component of diet was strongly associated with positive skin tests after adjusting for place of residence (Hooper et al., 2008). The **"prudent"** pattern was inversely associated and a **"Western"** pattern was positively associated with the risk of newly diagnosed COPD (Varraso et al., 2007a, Varraso et al., 2007b). The **"meat–dim sum"** pattern was positively associated with new-onset cough with phlegm (Butler et al., 2008). A pattern including **"fast food, juice and soft drinks"** was related to wheeze and respiratory infections (Takaoka & Norback, 2008). Univariately, a **"health conscious"** pattern was positively associated with eczema and a **"processed"** pattern was negatively associated with atopy (Shaheen et al. 2009).

Obesity

"Healthy" (Okubo et al., 2008, Newby et al., 2004a), **"prudent"** (Murtaugh et al., 2007, Paradis et al., 2009, Newby et al., 2006) patterns were significantly associated with a lower risk of obesity. Individuals with a high consumption of **"Western"** (Murtaugh et al., 2007, Paradis et al., 2009), "**Japanese traditional"** (Okubo et al., 2008), "**pasta & meat"** (Pala et al., 2006), **"animal foods"** (Shin, Oh & Park, 2007) and **"fish and sauce"** (Craig et al., 2010) pattern were more likely to be obese. No significant associations between **"green"**, **"sweet"** and **"traditional"** patterns and obesity were found (Togo et al., 2004).

High-density lipoprotein (HDL) Cholesterol

Serum HDL cholesterol was inversely associated with **"health aware"** (Hamer & Mishra, 2010), **"sweets"** (Newby et al., 2004b) and **"Western"** (Deshmukh-Taskar et al., 2009) dietary pattern. **"Meat"**, **"Western", "vegetable"** and **"protein and alcohol"** (Hamer & Mishra, 2010, Sadakane et al., 2008, Newby et al., 2004b) patterns were associated with higher total HDL cholesterol.

Mental and Behavioural Health

Improved behavioral scores were significantly associated with a **"healthy"** pattern (Oddy et al., 2009). Participants in the highest tertile of the **"whole food"** pattern had lower odds of

depressive symptoms than those in the lowest tertile. Patterns labeled as **"sweets"** and **"meat and products"** were positively associated with anxiety score in females (Yannakoulia et al., 2008). Child Behaviour Checklist for mental health scores was significantly associated with the **"Western"** dietary pattern. A **"junk"** food pattern was negatively associated with school attainment (Feinstein et al., 2008). In addition, high consumption of **"processed food"** was associated with increased odds of depressive symptoms (Akbaraly et al., 2009). There was little evidence to support an association between **"junk food"** intake and overall behavioral difficulties or other sub-scales of the childhood behavioral problems (Wiles et al., 2009).

Other health outcomes

A **"balanced"** pattern was associated with decreased risk and an **"animal protein"** with increased risk of hypertension (Chen et al., 2006)**.** Strong adherence to the **"health-conscious"** dietary pattern was inversely associated with Barrett's esophagus (Kubo et al., 2008). A **"traditional Western"** in girls was positively associated and a **"prudent"** was inversely associated with Chron's Disease (D'Souza et al., 2008). There were no consistent association of **"vegetable-fruit"**, **"potato-sweats and meat"** and **"alcohol-snacks"** patterns with Actinic Keratoses acquisition (Hughes et al., 2009). A pattern characterized by high consumption of **"fish and olive oil"** and low intake of **"red meat"** was positively associated with lumbar spine bone mineral density (Kontogianni et al., 2009). Individuals who had a high consumption of a **"health conscious"** diet had lower sperm DNA damage. Furthermore, sperm concentrations were much higher in men who strongly adhere to the **"traditional Dutch"** dietary pattern (Vujkovic et al., 2009a). A significantly increased risk of spina bifida was observed for offspring in mothers with a weak use of the **"Mediterranean"** dietary pattern (Vujkovic et al., 2009b). A dietary pattern characterized by high consumption of **"whole grains, fruit, vegetables"**, and **"low-fat dairy foods"** was associated with lower spot urine collection (Nettleton et al., 2008b). The **"prudent"** pattern was inversely and the **"Western"** pattern was positively associated with plasma concentrations of CRP (CRP are markers of endothelial dysfunction) (Lopez-Garcia et al., 2004). A **"Vegetables, plant foods and vegetable oils"** pattern associated with lower risk and a **"processed meat, salty snacks, and sweet drinks"** pattern with increased risk of Pre-eclampsia (Brantsaeter et al., 2009). A **"high-dairy, high-fruit and vegetable, high-starch, low-alcohol"** pattern was significantly and inversely associated with glucose tolerance abnormality. Additionally, a **"dairy products and fruits**

and vegetables" pattern was associated with decreased risk of developing a glucose tolerance abnormality ($P<0.05$) (Mizoue et al., 2006).

A detailed discussion and overall conclusions of these findings are given at paragraph 7.3. Moreover, after being informed with the help of our literature (Chapter 2) and systematic (Chapter 3) review on the application of PCA procedure in nutritional epidemiology we proceed to Chapter 4, where PCA is critically evaluated and compared with an Exhaustive Single food Analysis with the use of a Monte Carlo simulation study.

4 Methods

"The truth content of our theories, even the best of them cannot be verified by scientific testing, but can only be falsified."

Karl Popper

4.1 Overview

Our aim of this chapter is to describe the methodology behind our simulation study. A simulation study as Maldonado and Greenland (1997) suggested is a study which repeatedly answers a "What if?" question. In our case, this "What if?" question asks about the performance of two statistical methods, an exhaustive single food analysis (ESFA) and a Principal Components Regression (PCR).

Our main premise is that in order PCR to characterize the diet associated disease risk better than anyone food, then PCR should identify combinations of foods in the population which are causally linked with disease. We set out to investigate whether PCR really performs better in these respects than an analysis of individual food intakes.

In order to test this hypothesis, we created a hypothetical population in which food intake derived from a food frequency questionnaire could be related to disease under the conditions that we controlled for and in a way that we could specify in terms of relative risks. In our simulation model, a number of randomly selected foods that are not highly correlated with each other and a "Western" dietary pattern (a selected number of foods that are correlated with each other) are assumed to be in the causal path with disease.

Our objective is to investigate whether analysing each individual food on the FFQ separately, a process we refer to as an exhaustive single food analysis (ESFA) is as good as a method as constructing posterior dietary patterns and effectively identifying combination of foods that are causally linked with disease in this hypothetical population with the use of PCR.

4.2 Simulation study

4.2.1 Principal of Bootstrapping

In literature, the term bootstrap is attributed to Rudolph Erich Raspe and its story on the adventures of Baron Munchausen. The Baron in order to prevent himself from drowning in a lake pulled himself out of his bootstraps. This term is used when someone wants to describe the impossible. In statistics, bootstrap is a special case of the Monte Carlo simulation method, where re-sampling techniques are used to assign measures of accuracy to statistical estimates and was originated by Efron in 1979 and developed by Efron and Tibshirani (1993), as a way to tackle inaccurate and complex mathematical procedures. We will give a detailed description of the method as employed in our simulation study.

4.2.2 Simulating diet

Individuals who took part in the F.L.A.G survey (Data-set 1-see paragraph 4.8.1) and ECHRS UK study (Data-set 2-see paragraph 4.8.2) are two representative random samples of the populations living in Greenwich and Ipswich/Norwich. In each study, we treat the sample we have as if it were itself a reference population. We sample from the sample we have with replacement. Sampling with replacement means that some of the individuals will be part of our new sample more than once and some others will not be part at all. This imitates the way we would have drawn a new sample if we wanted to conduct another study for the same reference population. These samples of our sample are called bootstrap samples.

More formally, because PCA, as applied in nutritional epidemiology, depends on the correlation matrix of all dietary intake variables in particular we wanted to ensure realistic correlation structure in ours bootstrap simulations. For this reason we used real data sets. The real data-set that we used is referred to here as the reference data-set. For each individual, dietary consumption was recorded with the use of a food frequency questionnaire (FFQ) over a number of foods and beverages and dietary intake variables were constructed for our two reference data-sets. Food intake variables in the dataset were standardised to have a mean 0 and a standard deviation of 1; whereby a negative value corresponds to a smaller-than-average value of dietary intake; a zero value corresponds to an average value; and a positive value corresponds to a larger-than-average value. To simulate a dietary data-set (corresponding to the bootstrap sample of individuals as described above) with a realistic correlation structure between food intake variables we sampled with replacement from the reference data-set. When we sample with replacement, sample values of food intake are independent and it is like sampling from an infinite population.

4.2.3 Simulating disease from randomly selected food items and a simplified dietary pattern

We then simulated the presence or absence of disease assuming a logistic model for the disease risk. This model has been widely used in previous simulation studies in epidemiology (Fewell et al., 2007, Pastor & Guallar, 2001, Peduzzi et al., 1996). In each simulation we assumed that disease risk depended on a linear combination of *m* food intakes from the FFQ. Suppose these foods are indexed $i_1, i_2, \ldots, i_m$, and absolute food intakes $x_{i_1}, x_{i_2}, \ldots, x_{i_m}$ are standardised to have zero mean and unit standard deviation. We assumed a logistic model for the disease risk, *p* that is

$$\log\frac{p}{1-p} = a + b_1 + b_2 x_2 + \ldots + b_m x_m \quad (4.2.3.1)$$

or solving (4.2.3.1) for *p,* we have that

$$p = \frac{1}{1+e^{-(a+b_1x_1+b_2x_2+\ldots b_mx_m)}} \quad (4.2.3.2)$$

We chose the constant *a* so that the baseline risk at the average intake of all foods was 0.15, *i.e. a*=ln(0.15/(1-0.15)). Constants b_1, b_2, …, b_m were chosen so that the odds ratio per standard deviation of food intake was 1.5 or 1/1.5 depending on whether the food was assumed to increase or decrease the risk of disease, *i.e.* $b_j = \pm \ln(1.5)$. Combing data from all the dietary pattern studies in our systematic review and from Newby (2004) we observed that the median value of statistical significant odds ratio was 1.38 (IQR: 0.77-1.99) (see table C at the end of the thesis and Newby (2004)). We decided to use a slighter higher odds ratio, in order to increase the power of our simulation study.

The value of *m* was chosen to take two different values for each of the two data-sets; around one in seven of the total number of foods on the two FFQs, that is *m*=30 (of 217) in Data-set 1, and *m*=10 (of 74) in Data-set 2; and one in twenty of the total number of foods on the two FFQs, that is *m*=10 (of 217) in Data-set 1, and *m*=4 (of 74) in Data-set 2. The *m* foods were chosen in two different ways.

First, they were randomly chosen in each simulation. We considered three models of this kind: in Model 1, all *m* foods were assumed to be protective; in Model 2, *m*/2 foods were assumed to increase the risk of disease, and the other *m*/2 were assumed to have protective effects; in Model 3, all *m* foods were assumed to increase the risk of disease. Results of these simulations tell us about the average performance of different methods when we do not restrict *a priori* the combinations of foods that might be important for disease risk. Furthermore, in our simulations, we don't assume *a priori* that foods must be highly correlated with each other in order to be associated with disease risk.

Second, they were predetermined in each simulation to be foods making up a simplified "Western" dietary pattern (see paragraph 4.9), which was assumed to be positively associated (Model 4) and negatively associated (Model 5) with disease risk. These foods are listed in Table 4.9.2.1, and were chosen as the food intakes with the highest positive component correlations on a "Western" dietary pattern obtained using PCA from the original, reference

data-sets (30 foods for Data-set 1 paragraph 4.9.2, 10 foods for Data-set 2-see paragraph 4.9.3). Results from models 4 and 5 might be expected to favour PCA as a means of identifying dietary associations, since the model is based on a principal component in the populations under investigation.

4.2.4 Bernoulli trial

Hence, having simulated the food intake data for a new individual for the sample, we then calculated the probability of the outcome, *p,* for each individual using equation 4.2.3.2. We determined whether or not the individual has the disease by generating a uniform random number between 0 and 1, and observing whether it was less than *p*. If the uniform generated number was less than *p* then the individual had the disease, but otherwise not. In this way, a Bernoulli trial for each individual in the data-set was randomly simulated and a hypothetical binary variable (our hypothetical disease) was generated for our data.

4.3 Principal Components Regression (PCR) and Exhaustive Single Food Analysis (ESFA) of the simulated diet-disease dataset

4.3.1 Description of PCR and ESFA procedures

The simulated dietary data were subjected to a PCA with a fixed number of principal components, with the use of the correlation matrix of the bootstrap sample of the reference dataset (see Chapter 2), according to the way that was described in the nutritional literature (see Chapter 3); majority of studies that employed PCA derived 2 to 10 principal components (median value was 3, IQR: 2-5) which were varimax rotated. A principal component score was calculated for each varimax rotated dietary pattern for each individual in the sample to represent the individual's level of intake for the pattern. Resulting rotated principal component scores (dietary patterns intake) were investigated for their associations with disease using a logistic regression adjusted for total energy intake. Hence, let's assume that Y_i denote the disease outcome, $(total.energy)_i$ the total energy intake, and Y_{ij}^* the j^{th} pattern for the i^{th} individual of our sample of n individuals, then

$$\log\left[\frac{E(Y_i / Y_{ij}^*)}{1-E(Y_i / Y_{ij}^*)}\right] = \beta_{oj} + \beta_{1j}Y_{ij}^* + \beta_{2j}(totalenergy)_i \quad , i=1,...,n;\ j=1, ..., p \quad (4.3.1.1)$$

where $\beta_{0j}, \beta_{1j}, \beta_{2j}$ are the regression coefficients of the j^{th} pattern intake variable.

For comparison, we considered the results of analysing each individual food on the FFQ separately in relation to disease risk, a process we refer to as an exhaustive single food analysis (ESFA), adjusted for total energy intake. This is a univariate or independent screening approach as we know from, for example, SNP screening in statistical genetics (Laird, 2011; page 111). Hence, let's assume that Y_i denote the disease outcome, $(\text{total.energy})_i$ the total energy intake and X_{ij} the j^{th} food intake variable for the i^{th} individual of our sample of n individuals, then

$$\log\left[\frac{E(Y_i / X_{ij})}{1-E(Y_i / X_{ij})}\right] = \beta_{oj} + \beta_{1j} X_{ij} + \beta_{2j}(total.energy)_i \quad i=1...n,\ j=1...k \quad (4.3.1.2)$$

where $\beta_{0j}, \beta_{1j}, \beta_{2j}$ are the regression coefficients of the j^{th} food intake variable.

Observational nutritional studies are strongly advised to be adjusted for energy intake (Willet, 1998, Jakes et al, 2004). As pointed out by Willet, adjustment for total energy intake should be considered because the level of intake might be a risk factor, might distort the effect of a food or a nutrient on the potential outcome, and the variation of nutrient intake between individuals might reflect variations of individual's energy intake levels (Willett et al., 1997). Total energy intake was calculated in our reference data-set from food frequency questionnaire using data from the British food composition tables (McCance and Widdowson, 1991).

ESFA was in the first instance unadjusted for effects of other food intakes, but in order to cope with confounding we also carried out an ESFA adjusting for effects of foods other than the index food.

So, let's assume that for the i^{th} individual V_{ij} is a vector of covariates which represent our confounders and ζ_j is a vector of regression coefficients for these covariates, our logistic model in equation 4.3.1.2 could take the general form,

$$\log\left[\frac{E(Y_i / X_{ij})}{1-E(Y_i / X_{ij})}\right] = \beta_{oj} + \beta_{1j} X_{ij} + \beta_{2j}(total.energy)_i + \sum_{n=1}^{m} \beta_m V_{im} \quad i=1...n \quad (4.3.1.3)$$

The confounders V_{ij} are actually considering here other foods. Because in practice we don't want to have to include all the foods in the regression equation (4.3.1.2), we need practical strategies to adjust for all the foods we're not interested in. Hence three different strategies of adjustment were employed in our logistic model in (4.3.1.3) as presented below.

4.3.2 Description of ESFA procedure adjusted for the first five principal components of diet

Firstly, the simulated dietary data were subjected to a PCA using a varimax rotation. Five principal components were identified and their principal components scores were calculated for each individual (see paragraph 2.4). An ESFA procedure adjusted for the resulting five rotated principal component scores (covariates) was carried out with the use of the logistic model in equation 4.3.1.3. Because net confounding by correlated foods and patterns could result to biased associations between diet and disease, adjustment for dietary patterns has been suggested by the literature (Imamura et al., 2009). We choose five components of diet because that was the number of dietary patterns of diet that were identified in the original population (see paragraphs 4.9.2 and 4.9.3).

4.3.3 Description of ESFA procedure adjusted for all the food intakes which were significant in the unadjusted ESFA) controlling the false discovery rate at 20% level

In this case, as a first step we run an ESFA procedure adjusting for total energy intake as described in paragraph 4.3.1, keeping the food variables (covariates) with those regression coefficient estimates that had a P-value lower than a specific threshold. The threshold was determined by the Benjamini and Hochberg procedure controlling the rate of our false discoveries at 20% (see paragraph 4.4). Then we re-run (second step) an ESFA procedure adjusting for energy intake and for all these foods (covariates) that were significant in the first round of analysis (unadjusted ESFA) with the use of the logistic model in equation 4.3.1.3. This method is conceptually similar with the Iterative sure independence screening method (ISIS) proposed by Fan and Lv (2008). However in our case we choose our covariates based on a multiple test procedure (Benjamini and Hochberg) and not on a penalized likelihood method. Furthermore we aim to control the false discovery rate, whereas sure screening method focuses on missed discoveries (Fan & Lv, 2008).

4.3.4 Description of ESFA procedure adjusted for a propensity score predicting the index food intake from the other food intakes

Propensity scores where first established by Rosenbaum and Rubin (1983) (Rosenbaum & Rubin D. P., 1983) and can be defined as the probability of exposure to a specific treatment for randomised controlled trials or to a potential risk factor for observational studies given

observed covariates. The development of the score was first developed for binary exposures (Rubin, 1997) but it was extended and generalized for continuous and ordinal exposures (Kosuke Imai & David A van Dyk., 2004). A generalized linear model could be employed for calculating the propensity score given an observed set of covariates. Once the propensity score is estimated it could be used as a confounder in a conventional multivariate outcome model, (Sturmer et al., 2006).

Specifically, in our simulation study we defined the propensity score ps_{ij} as the predicted value of an individual's intake of a specific food *i* given all the other intakes of foods that an individual consumed in our study with the use of a linear regression model. So let's assume that X_{ij} is the j^{th} food item for the i^{th} individual of our sample of *n* individuals, and β_j are the regression coefficient for the j^{th} food item then

$$ps_{ij} = \beta_o + \sum_{\substack{i=1 \\ i \# j}}^{k} \beta_j X_{ij} \quad (4.3.4.1)$$

The resulting regression model produces for each food item *j* a corresponding propensity score according to 4.3.4.1. These propensity scores were considered as a confounder (covariate) in our ESFA procedure for each food item *j* with the use of equation 4.3.1.3.

Before discussing about the "environment" that we want to evaluate our procedures (PCR, ESFA and adjusted ESFA) we will give a description of the multiple test procedures which are a vital part of our simulation study.

4.4 Multiple Test Procedures

4.4.1 Overview

In hypothesis testing, when we want to test the association between two variables in our sample, we are testing the null hypothesis (that is the two variables are unrelated and that any apparent difference in our sample is due to chance) over the alternative. A p-value is the probability of obtaining a test statistic for exploring this association at least as large as the one observed in our sample, assuming that the null hypothesis in our population is true. For example, when we want to test if the effect of a food intake variable on disease in our sample is zero, we observe a value of a z-test. P-value gives the probability that this z-value value to be as large as the one we observe, assuming that the population of z-values is following a standard

normal distribution. If this p-value is lower than an arbitrary cut-off value of significance α, which is our type I error, then our effect estimate is not zero.

However, as Rothman (1990) points out "if *n* independent associations are examined for statistical significance, the probability that at least one of them will be found statistically significant is $1-(1-\alpha)^n$, if all *n* of the individual hypotheses are true". So for example, in our case, when an ESFA is applied in our data-set 1 (see paragraph 4.8.1) 217 separate logistic regression models are employed to explore the association between each separated food item and disease, and consequently 217 hypotheses are tested. So if we assume that our type I error is 0.05 for the 217 independent hypothesis, then the probability that at least one of them will be found statistically significant is $1-(1-\alpha)^{217}= 0.999$, assuming that all of 217 of the null hypotheses are true. So we have a probability of almost 1 to report a false positive result, if all of the null hypotheses are true.

Multiple test procedures are used for tackling the issue of multiple comparisons. Two well-established methods are presented to this thesis as a response to this problem; Bonferroni inequality and the Benjamin and Hochberg method of the rate of false discoveries which we will describe below.

4.4.2 Bonferroni inequality and family-wise error rate (FWER)

Let's assume that $\{P_1, P_2 \dots P_m\}$ is a set of our observed p-values for our m null hypotheses $\{H_1, H_2 \dots H_m\}$ and α is our acceptable type I error**.** Bonferroni inequality is defined by the formula

$$\Pr[\min(P_j: 1\leq j \leq m) \leq \alpha /m=q] \leq \alpha \quad (4.4.2.1)$$

or if P-values are ordered in ascending order $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$, then

let *k* to be the largest *(i)* for which $P_{(i)} < \frac{a}{m}$ (4.4.2.2)

then reject all the corresponding $H_{(i)}$, i=1,2,...,k .

In the Bonferroni procedure this upper bound q=α /m is denoted as the family-wise error rate (FWER) and defined as the probability that at least one of the null hypotheses in our set will be rejected. More formally, If the upper bound q ($0 < q < 1$) is a FWER then we can be 100(1- α) % confident that all null hypothesis in the subset of p-values that are below *q* are false

(Newson, 2003). However, controlling the FWER with the Bonferroni inequality is a conservative method, which could omit important associations between diet and disease, and could lead to false negative findings because of loss of power in our study. Furthermore, in some situations this could be unnecessary. For example, FWER is important when a conclusion from the various individual inferences is likely to be erroneous when at least one of them is.

In our case, if we want to test the hypothesis if each one of individual food item on the FFQ is separately associated with disease, we don't want to falsely accept a specific null hypothesis (that there is no association between a specific food item and disease) because some of the null hypothesis (for the other food items) are falsely rejected (Benjamini and Hochberg 1995). In order to tackle with these problems Benjamini and Hochberg suggested a less conservative procedure the maximum permissible false discovery rate (FDR).

4.4.3 Benjamini and Hochberg procedure and false discovery rate (FDR)

As defined by Benjamini and Hochberg (1995), false discovery rate (FDR) is the expected proportion of true null hypothesis that declared significant (false positives) among all the hypotheses that declared significant and controls the proportion of the rejected null hypotheses which are erroneously rejected.

Let's assume that $\{P_1, P_2 \ldots P_m\}$ is a set of our observed p-values for the corresponding m null hypotheses $\{H_1, H_2 \ldots H_m\}$ and q* is the maximum permissible FDR. P-values are ordered in ascending order $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$.

let *k* to be the largest *i* for which $P_{(i)}$ for which $P_{(i)} \leq \frac{i}{m} q^*$ (4.4.3.1)

then reject all the corresponding $H_{(i)}$, i=1,2,...,k .

Benjamini & Hochberg (1995) proved under general conditions that the above procedure for rejecting hypotheses leads to a false discovery rate of no more than q*. If all of null hypothesis are true then the FDR is FWER.

4.5 Evaluation of performance of ESFA and PCR procedures in each simulation experiment

First, we investigated the statistical **power** with which ESFA and PCA could detect whether there was *any* association between diet and disease. For ESFA, using the methodology described in Chapter 4.3.1 we considered that an association had been found if any of the food intakes were statistically significantly associated with disease after applying a Bonferroni correction (see chapter 4.4.2) for the number of foods (family-wise $P<0.05$) (Miller, 1981). For PCR, using the methodology described in paragraph 4.3.1, we considered an association had been found if any of the dietary patterns were significantly associated with disease after applying a Bonferroni correction (see Chapter 4.4.2) for the number of patterns identified . In both situations, we control our results for the FWER because we wanted to be 95% confident that *all* of statistical significant associations are real.

We also wanted to see how well the two procedures identified the specific combinations of foods that were causally linked with disease. We compared the power and the false discovery rate (FDR) of ESFA and PCA for detecting these associations. In this context we extend the concept of "power" to mean the proportion of foods included in the model which were identified as significant. The FDR is the proportion of discoveries, or significant findings, which are false (Figure 4.5.2.1). More formally, **power** is defined as the number of true significant results identified by the method (True Positives) divided by the number of foods that are causally linked with disease (TP/ (FN+TP)), and **false discovery rate (FDR)** is defined as the number of false significant results (False Positives) identified by the method divided by the total number of significant results identified by the method (FP / (FP + TP)), or 0 if FP + TP = 0).

Figure 4.5.2.1. How the results of dietary analyses can be broken down

	Foods not causally linked with disease	*Foods causally linked with disease*
Foods declared non-significant	True Negatives (TN)	False Negatives (FN)
Foods declared significant	False Positives (FP)	True Positives (TP)

Power = *TP / (FN + TP)*

False Discovery Rate = FP / (FP + TP) *if* FP + TP > 0
= 0 *if* FP + TP = 0

For ESFA, we considered there was a "significant" effect of a food if it was identified as such using the multiple testing procedure of Benjamini and Hochberg, with a nominal false discovery rate set to 20% (Benjamini & Hochberg, 1995) (see chapter 4.4.3). For PCA, we considered there was a "significant" effect of a food if it had correlation >0.3 or <-0.3 with a dietary pattern that was significantly associated with disease (P<.0.05) – this being the way in which individual foods tend to be highlighted in a PCA (median value of all the studies in our systematic review was 0.3 – see paragraph 3.2.5). Furthermore, we control the rate of our false discoveries at 20% since we want to be 80% confident that *some* of the statistical significant associations that we observe are real, or that 80% of these associations are expected to be real. Note that the Benjamini-Hochberg procedure is designed to control the FDR at no more than the nominal level, but here false discoveries (of foods) occur not just as random errors, but also because of confounding with other foods, so the nominal rate may be exceeded.

4.6 Standards error of our Monte Carlo simulations

In each simulation, we are sampling with replacement N individuals from our real-reference data-set (Data-set 1-see paragraph 4.8.1, Data-set 2-see paragraph 4.8.2) and create a sample. In each sample we construct our diet (X_b^*) and disease (Y_b^*) variables according to paragraph 4.2, and we estimate our three parameters of interest according to paragraphs 4.3, 4.4 and 4.5; (i) **power (**$\hat{\theta}_1^*$) with which ESFA and PCA could detect whether there was *any* association between diet and disease; (ii) **power** ($\hat{\theta}_2^*$) with which ESFA and PCA could detect specific combination of foods that are causally linked to disease and (iii) and **false discovery rate** ($\hat{\theta}_3^*$)

of ESFA and PCA for detecting these combinations of foods. We want to calculate the corresponding standard errors for these estimation results.

Thus, in case (i) our estimate of **power (**$\hat{\theta}_1^*$) is the proportion of statistical significant associations that were declared significant out of B replications (in our case B=10000). Standard error of this proportion is calculated by the formula

$$\hat{se}(\hat{\theta}_1) = \sqrt{\frac{1}{B} p(1-p)} \quad (4.6.1)$$

In addition, we construct our standard errors for our estimates (ii) **power** ($\hat{\theta}_2^*$) and (iii) **false discovery rate** ($\hat{\theta}_3^*$) according to the following algorithm (Efron & Tibshirani, 1993).

1. Select B independent samples (X_b^*, Y_b^*) of size n from our reference data-set. In our case B=10000 replications
2. We estimate our two parameters of interest for each sample

   $\hat{\theta}_{cb}^*$, for $b=1,...,B$ and c=2,3
3. we estimate the standard error $\hat{se}(\hat{\theta}_c)$ by the sample standard deviation and according to the formula

$$\hat{se}(\hat{\theta}_c) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_{cb}^* - \hat{\theta}_c^*)^2} \text{ , c=2, 3} \quad (4.6.2)$$

where $\hat{\theta}_c^* = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_{cb}^*)$ (4.6.3)

4.7 Sample size calculation of our simulation experiments

Using the "powerlog" sample size calculation routine in Stata (Ender, 2002), we determined that a sample size of 330 would achieve 80% power at the 5% significance level to detect an odds ratio of 1.5 per standard deviation, using an unadjusted logistic regression with no allowance for multiple testing(Table 4.7.1).

Table 4.7.1.Power Analysis

Power	N
0.60	194
0.65	221
0.70	252
0.75	287
0.80	328
0.85	380
0.90	451

4.8 Reference data sets

Our two real dietary data-sets and source of the food correlation matrices for our simulations were comprised of adults living in Greenwich as part of the F.L.A.G survey and Ipswich and Norwich as part of the UK ECRHS II diet survey. We used only controls from both data-sets in order to have a more representative reference correlation matrix for our simulation process. Food frequencies were converted to intakes in g/d by multiplying frequency of consumption by the weight of standard portion sizes using British food composition tables (Paul, Southgate & Buss, 1986) and standardised to have a mean 0 and a standard deviation of 1.

4.8.1 Food, Lifestyle & Asthma in Greenwich Survey (F.L.A.G)

The original dataset being used in our study was based on 856 adults aged 16-50 years without asthma who responded to an asthma survey in a random sample of adults 16-50, registered with 40 general practices in Greenwich, South London, UK, in autumn 1996 (Marks et al., 1997, Premaratne et al., 1999). Individuals were mailed a dietary questionnaire in September 1997. Usual diet was assessed (previous 12 months) using a food frequency questionnaire (FFQ) based on one used previously (Calvert et al., 1997). Food frequency questionnaires (FFQ) recorded a consumption as frequencies of 217 different foods (from never to 6d a week) and drinks.

4.8.2 European Community Respiratory Health Survey II (ECRHS II UK)

ECRHS-I ran from 1990 to 1995. At each centre, a random sample of at least 3000 adults aged 20–44 years was selected using a local sampling frame. From those who responded, a random sample of at least 600 adults was selected to undergo a detailed clinical examination. Eight to ten years later, these subjects were contacted to take part in a follow-up study

(ECRHS-II) and invited to a local clinic for further assessments, including an interviewer-administered questionnaire (European Community Respiratory Health Survey II Steering Committee, 2002).

Dietary assessments were included in ECRHS-II at some centres, though the method and protocol differed between countries. In the present study, we report results from 2 centres in UK, where FFQ were administered: Ipswich and Norwich. Three thousand three hundred and eighty-seven adults at these centres were contacted to take part in ECRHS-II. The UK FFQ was adapted from one developed for EPIC-UK (Bohlscheid-Thomas et al., 1997). It recorded a consumption of 198 different foods over the last 12 months as frequencies (from never to 7 d a week) and number of portions consumed on each of these days (portions being defined on the questionnaire). Some aggregation of food items into food groups was performed and this process led to a list of 74 food groups whose intake in g/d. In our study, we included 201 adults aged 29-54 years living in Ipswich and Norwich.

4.9 Construction of a simplified "Western" dietary pattern

4.9.1 Overview

Schulze et al 2003 (Schulze et al., 2003) proposed a method to construct a simpler form of a dietary pattern variable applied previously in the field of psychology (Comrey, 1988) and described extensively by Jolliffe (Jolliffe, 2010). The main idea of this method is to associate food items with each of the first few dietary patterns and then retaining those food items which are more strongly associated with the first dietary patterns. The choice of which food items to retain for each pattern should be determined by looking at the strength of the relationship between the food item and the dietary pattern So, in our case, a simplified **"Western"** pattern was constructed by selecting 30 items from data-set 1 (see paragraph 4.9.2), and the 10 items from data-set 2 (see paragraph 4.9.3), that were most strongly positively correlated with the Western principal component of the population.

4.9.2 Simplified "Western" Pattern derived from the F.L.A.G data-set

Dietary patterns were identified with the use of PCA from the F.L.A.G data-set. The principal components were rotated (varimax rotation) and the number of patterns were determined by examination of the scree plot of the eigenvalues. We extracted five components (dietary patterns), which explained 17% of the variance in the original 216 items. Individual foods that correlated >0.3 or <-0.3 with the varimax rotated principal components labelled the

dietary pattern. More detailed information of the analysis is presented elsewhere (Bakolis et al., 2010). From the five patterns that we identified we randomly choose the one that was labelled as **"Western"**. The following 30 food items or groups where highly positively correlated with the **"Western"** pattern and consisted our newly constructed simplified "Western" pattern are presented in table 4.9.2.1. This newly constructed simplified **"Western"** dietary pattern was strongly correlated with the original western pattern of the F.L.A.G dataset (r=0.84).

Table 4.9.2.1. List of foods comprising a "Western" pattern for each data-set [a]

Data-set 1	Data-set 2
roast potatoes	sausages
ham	donuts, pastries and tarts
ice cream	beer
pork - roast, chops	corned beef and luncheon meat
pork stew, casserole	hard cheeses
omelette/scrambled egg	tomato ketchup
fruit pies, tarts, crumbles	pizza
beef stew,casserole,mince,curry	beef burger
sponge cakes	fried egg, scrambled egg, omelette
fried fish in batter/breadcrumb	chips
baked beans	
chocolate biscuits	
sandwich/cream biscuits	
corned beef, spam, luncheon meat	
white bread and rolls	
fizzy soft drinks e.g. coke	
bacon	
fried egg	
milk chocolate	
bread crumbed chicken e.g. chicken nuggets	
crisps	
sponge puddings	
tomato ketchup	
chocolate snack bars	
meat pizza	
other fried snacks	
chips	
sausages - beef, pork	
beef burger, hamburger	
pies/pasties/sausage rolls	

[a] Data-set 1 is from the FLAG survey; Data-set 2 is from the UK ECRHS II survey

4.9.3 Simplified "Western" Pattern derived from the UK ECRHS II data-set

Similarly, dietary patterns were identified with the use of PCA to the UK ECRHS II data-set. We extracted five principal components from the examination of the scree plot, which were rotated orthogonally (varimax rotation) for better interpretation. Components explained 24.7% of the overall variation in the original 74 food items. Correlations between our 74 food items and our dietary patterns are shown in table 4.9.3.1. For consistency purposes, we constructed again a simplified **"Western"** dietary pattern from the following 10 food items that were positevely correlated highly with the **"Western"** pattern (Table 4.9.2.1). This newly constructed simplified **"Western"** dietary pattern was strongly correlated with the original **"Western"** pattern of the UK ECRHS II dataset (r=0.85).

Table 4.9.3.1. Correlations between food intakes and each of the five orthogonal rotated dimensions of diet, only correlations > 0.30 and <-0.30 are included in the table.

	pattern				
	vegetarian	**fruit and vegetables**	**meat and 2 vegetables**	**Western**	**deserts and cereals**
food items					
soy cheese, tofu, quern, grains	0.66	-	-	-	-
honey	0.62	-	-	-	-
lentils, dahl, mixed bean casserole	0.57	-	-	-	-
vegetarian paste	0.54	-	-	-	-
rice and rice dishes	0.53	-	-	-	-
kiwi, mango and pineapple	0.50	-	-	-0.40	-
garlic	0.48	-	-	-	-0.35
peppers	0.45	-	-	-	-
tomato	0.45	-	-	-	-
bean sprouts	0.38	-	-	-	-
apple	-	0.78	-	-	-
pear	-	0.75	-	-	-
orange	-	0.74	-	-	-
banana	-	0.64	-	-	-
peach and nectarine	-	0.62	-	-	-
grapes	-	0.57	-	-0.39	-
other fruit juice	0.31	-	-	-	-
sliced meat	-	-	0.58	-	-
beef steak	-	-	0.51	-	-
minced beef, meat stew and casserole	-	-	0.51	-	-
sausages	-	-	0.50	0.31	-
liver	-	-	0.49	-	-

pork chops	-	-	0.43	-	-
broccoli, cabbage and cauliflower	0.37	0.33	0.41	-	-
carrots	0.37	-	0.41	-	-
bacon	-	-	0.37	-	-
potato - boiled / mashed / baked	-	-	0.37	-	-
boiled egg	-	-	0.35	-	-
peas	-	-	0.35	-	-
green beans	0.36	0.34	0.32	-	-
poultry	-	-	0.32	-	-
pate	-	-	0.32	-	-
chips	-	-	-	0.55	-
fried egg, scrambled egg, omelette	-	-	-	0.50	-
beef burger	-	-	-	0.48	-
pizza	-	-	-	0.40	-
hard cheeses	-	-	-	0.37	0.34
tomato ketchup	-	-	-	0.37	-
corned beef and luncheon meat	-	-	-	0.36	-
beer	-	-	-	0.34	-
donuts, pastries and tarts	-	-	-	0.31	0.31
herbal tea	-	-	-	-0.32	-
yoghurt	-	-	-	-0.38	-
raspberries, red currants, blackcurrants	-	-	-	-0.46	-
cakes, puddings and desserts	-	-	-	-	0.55
breakfast cereals	-	-	-	-	0.44
chocolate	-	-	-	-	0.44
milk and milky drinks	-	-	-	-	0.44
choc bars and cereal bars	-	-	-	-	0.39
bread and rolls	-	-	-	-	0.37
ice cream	-	-	-	-	0.36
butter	-	-	-	-	0.35
wine	-	-	-	-	-0.35
jam and marmalade	-	-	-	-	-
peanut butter and choc spreads	-	-	-	-	-
biscuits	-	-	-	-	0.65
cream cheese	-	-	-	-	-
cottage cheese	-	-	-	-	-
soft cheeses	-	-	-	-	-
quiche	-	-	-	-	-
fish fillets / cakes / fingers	-	-	-	-	-
tinned fish	-	-	-	-	-
soup	-	-	-	-	-
strawberries	-	-	-	-	-
tinned or stewed fruit	-	-	-	-	-

nuts	-	-	-	-	-
orange juice	-	-	-	-	-
fizzy drinks	-	-	-	-	-
tea - black and green	-	-	-	-	-
coffee (not decaffeinated)	-	-	-	-	-
decaffeinated coffee	-	-	-	-	-
cider	-	-	-	-	-
fortified wine	-	-	-	-	-
liqueurs and spirits	-	-	-	-	-

4.10 Specification of simulation parameter values and analysis of simulation experiments

For each simulated experiment a number of 4, 10 (UK ECHRS II data-set) and 10, 30 (F.L.A.G data-set) food items with the same effect are chosen randomly and are causally associated with disease in 3 different ways. Additionally a simplified **"Western"** dietary pattern from our two datasets (F.L.A.G and E.C.H.R.S II) is causally associated with disease in 2 different ways. The number of individuals was set to be 100 (in only two case), 300, 600, 1200, 2400 and 4800 and the number of varimax rotated principal components was chosen to be 2, 5 and 10. Comparisons were made between our ESFA and PCA procedure for each different combination of our parameter values. ESFA procedure was further adjusted for effect of other foods in 3 different ways. Average percentages of power and false discovery rate were calculated, and after looking on the results of 10000 simulation trials, we observed that the standard errors for these percentages were < 0.5%.

4.11 Null simulations

Null simulations were conducted for testing the validity of our programming work. In this case no randomly selected number of foods or a simplified **"Western"** pattern was causally linked with disease (Model 0). So in this case we assumed a logistic model for the disease risk, p that is

$$p = \frac{1}{1+e^{-0.15}}, \quad (4.11.1)$$

Power for detecting *any* association between diet and disease was controlled at the desired level of 5% (Table 4.11.1).

Table 4.11.1. Power (%) of exhaustive single food analysis (ESFA) and principal components analysis (PCA) to detect any association between diet and disease (all estimates of power have standard error <0.5%).

Data-set and model[a]	Sample Size	ESFA	PCA		
			Number of Components		
			2	5	10
		Power	Power	Power	Power
Data-set 1					
Model 0	300	1.0	4.3	3.0	2.7
	600	2.0	4.9	3.7	2.4
	1200	3.3	4.1	4.0	3.7
	2400	3.7	4.8	4.6	4.2
	4800	4.2	4.6	4.9	4.8
Data-set 2					
Model 0	300	2.3	4.5	3.3	3.6
	600	3.0	4.7	4.6	4.1
	1200	3.8	4.2	4.1	4.6
	2400	4.9	4.6	4.5	4.6
	4800	4.5	4.6	4.9	5.0

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010).In Model 0 no foods or simplified **"Western"** pattern are selected at random in each replication from the foods on the FFQ.

The majority of null simulations in Table 4.11.1 don't reach their nominal level. Power of PCA is around 5% only for 2 and 5 principal components. Power of ESFA and PCA for 10 components reaches the nominal level only for large sample sizes (2400,4800). In addition, as the number of components increases power of PCA descreases and reaches a nominal level much lower than 5% for small sample sizes (300,600,1200).

One potential explanation is that the Bonferroni correction is a conservative method, meaning that the actual family-wise error rate will be smaller than the value it is fixed at when results are correlated . Suppose, for example, we are applying the Bonferroni correction to the results of 100 tests. For any result to be significant it must have a $P < 0.05/100$. If all the tests are independent, then the probability under the global null hypothesis that none of them are significant is $(1 - 0.05/100)^{100} \cong 0.95$. So, the probability that at least one result is significant under the null hypothesis is 0.05. However, in the extreme example of dependence, if we repeat the same test on the same data 100 times, the probability would have to be less than 0.05/100=0.0005 which is the probability that at least one result is significant under the null hypothesis. So, depending on how correlated the P-values are, the family-wise error rate might

be lower. So, due to highly correlated dietary intake variables in our simulation datasets, we don't expect the values of power under the null model in our simulations to be 5%, but lower than this value. This could be observed from Table 4.11.2, where moderate correlation coefficient values are observed between the components for five and ten principal components but not for two.

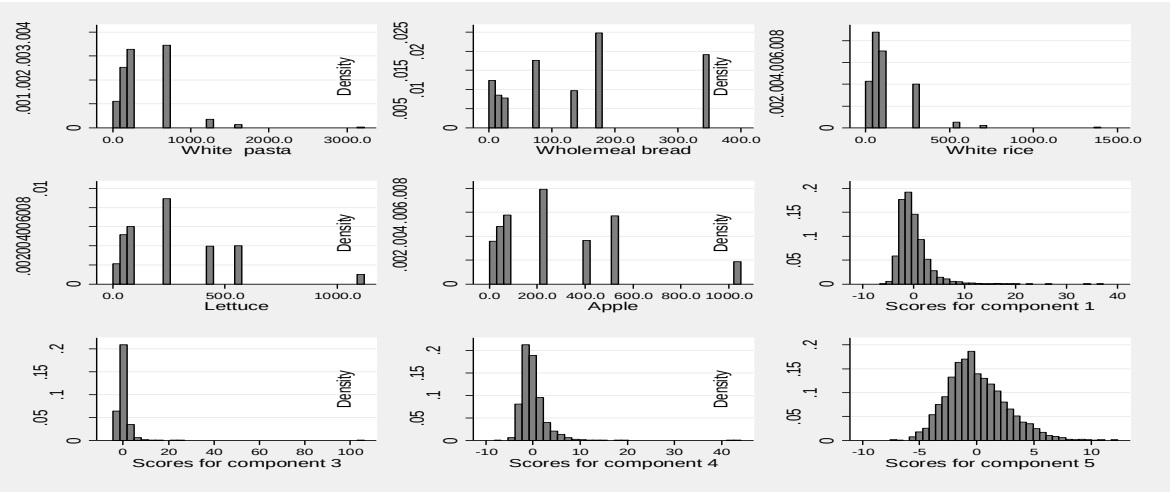
Furthermore, we observed for ESFA that power increased with sample size. However this pattern was not observed so clearly for PCA. One potential explanation that power increases with sample size is due to the effect of skewness of our dietary intake data. As we can see from Figure 4.11.1, dietary pattern intake variables derived from the F.L.A.G study are normally distributed but food intake variables are highly skewed to the right.

In the context of hypothesis testing this translates into true tail probabilities that are higher than nominal in the upper tail, resulting in fewer rejections, than there would be under a normal parent for one-tailed tests (Boos et al.,1998, Rieneke et al., 2003). In a simulation study (Rieneke et al., 2003) where the type I error rates were calculated, a strong dependence was shown between the error rates, skewness and sample size of the data according to the formula $\sqrt{\beta}(X)/\sqrt{n}$, where n is the size of the sample and $\sqrt{\beta}(X)$ is the measure of skewness. Particularly, when skewness was close to zero error rates were close to what we would have expected even for small sample sizes. However if the skewness is far from zero sample size must be large in order for error rates to be close to what we should have expected (Reineke et al., 2003). Thus, because of the skewness of our food intake data we have fewer rejections of the null hypothesis for small sample sizes, and as our sample size increases our power estimations are closer to the desirable error rate of 5%. The implications of lower significance level for ESFA in our null simulations could potentially lead to an underestimation of the average percentages of power of ESFA for small sample sizes in our monte carlo simulations reported in Chapter 5.

Table 4.11.2. Correlation coefficient values between randomly identified dietary patterns for sample size of 600 from the simulated F.L.A.G survey data-set. Different numbers of dietary patterns were derived with the use of PCA (2, 5 and 10).

2 Randomly simulated Principal Components										
Dietary pattern	I	II								
I	1.00									
II	-0.03	1.00								
5 Randomly simulated Principal Components										
Dietary pattern	1	2	3	4	5					
1	1.00									
2	-0.16	1.00								
3	0.21	0.12	1.00							
4	0.11	-0.11	0.07	1.00						
5	0.15	-0.03	0.10	0.03	1.00					
10 randomly simulated principal components										
Dietary pattern	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.35	1.00								
3	-0.21	-0.20	1.00							
4	0.16	0.04	0.04	1.00						
5	-0.06	-0.03	0.19	-0.05	1.00					
6	0.14	0.06	0.14	0.17	0.12	1.00				
7	0.19	0.09	-0.03	0.14	0.01	0.12	1.00			
8	0.14	0.16	-0.14	-0.03	-0.06	-0.04	0.03	1.00		
9	0.17	0.19	-0.04	0.04	0.05	0.08	0.00	0.03	1.00	
10	-0.19	-0.12	0.16	-0.07	0.12	0.03	-0.05	-0.10	-0.03	1.00

Figure 4.11.1. Histograms of random selected food intake and 4 randomly selected principal component analysis.



4.12 Programming

In this project, Monte Carlo simulations were developed and programmed using Stata 10 (Stata Corporation, College Station, Texas USA). Simulation programs were created from scratch; one for selecting randomly combinations of foods; two for generating a simplified **"Western"** dietary pattern; one for examining the **power** with which ESFA and PCR could detect whether there was *any* association between diet and disease; one for examining the **power** and **FDR** with which ESFA and PCR could detect a specific combination of foods; and three for examining for examining the **power** and **FDR** with which ESFA could detect a specific combination of foods under different ways of adjustment. Programs were designed to allow for a range of different number of

- bootstrap replications
- size of bootstrap samples
- size of baseline risks
- Effect sizes of the randomly selected number of food items and of the simplified "Western" patterns.
- size of the number of foods that are causally associated with disease
- ways that number of foods are causally associated with disease in the simulation model
- ways that a simplified "Western" pattern was causally associated with disease in the simulation model
- cut-off points of correlation coefficients of food items with the rotated dietary pattern
- cut-off points of family wise error rate (FWER) and false discovery rate (FDR)
- Number of rotated principal components identified.

Computer algorithms and commands being used are presented at the Appendix II

5 Results

5.1 Introduction

5.2 Power with which ESFA and PCA could detect whether there was *any* association between diet and disease

5.3 Power and FDR with which ESFA and PCA could identify specific combinations of foods between diet and disease

5.4 Power and FDR with which ESFA and PCA could identify specific combinations of foods between diet and disease adjusted in three different ways

5.1 Introduction

In this chapter results from our simulations are presented. Paragraph 5.2 present average estimates of percentages of power and paragraph 5.3 presents average estimates of percentages of power and FDR for different sample sizes and different number of principal component scenarios. Paragraph 5.4 shows average percentages of power and false discovery rates for different ways of adjustment of ESFA method for different sample sizes.

5.2 Power with which ESFA and PCA could detect whether there was *any* association between diet and disease

Tables 5.2.1, 5.2.2, 5.2.3 display average estimates of power of exhaustive single food analysis (ESFA) and principal component analysis (PCA), when 1 in 7 or 1 in 20 foods are randomly selected and causally linked with disease. Additionally, PCA and ESFA are evaluated when a simplified **"Western"** pattern is causally associated with disease.

Both methods had considerable power to detect any statistical effect between diet and disease. In the majority of the scenarios we investigated, ESFA had greater power than PCA to detect an association between diet and disease**,** when randomly selected foods where causally associated with disease Although PCA performed slightly better in some (21/240) simulations this was only when small sample sizes of 300 and 600 were used. The power of ESFA was increased with sample size. Average percentages of power estimates of ESFA didn't present a symmetrical pattern between model 1 (all selected food intakes are negatively associated with disease) and model 3 (all selected food intakes are positively associated with disease), and we get intermediate power estimates for Model 2 (half the selected food intakes are negatively associated and half are positively associated with disease) (Table 5.2.1 and 5.2.2). Similar qualitatively simulation results are observed in Table 5.2.3 for model 4 and 5 when a "Western" dietary pattern is causally associated with disease.

PCA outperforms ESFA in all scenarios when a "Western" pattern is causally associated with disease for sample sizes lower than 600 for all different odds ratio values (1.1, 1.3, and 1.5). The power of principal component analysis increases with the number of principal components and sometimes decreases. Furthermore, for PCA, averages percentages of power to detect any effect of diet on disease is roughly the same when all effects of foods are positive as when all effects are negative, and the power is less when there is a mixture of positive and negative effects (only for ECHRS II survey dataset at Table 5.2.2 we observe

intermediate power when there is a mixture of positive and negative effects of foods on disease).

Table 5.2.1. Power (%) of exhaustive single food analysis (ESFA) and principal components analysis (PCA) to detect <u>any</u> association between diet and disease (all estimates of power have standard error <0.5%). In models 1-3, *one in seven* foods (30 in data-set 1 and 10 in data-set 2) are selected at random in each replication from the foods on the FFQ. Family-wise error of Bonferroni correction at 5%.

Data-set and model[a]	Average Number of Cases	Sample Size	*ESFA*	PCA		
				Number of components		
				2	5	10
Data-set 1 (FLAG)						
Model 1 (All foods negatively associated with disease)	79.3	300	41.0	98.2[b]	98.2[b]	97.2[b]
	158.1	600	93.4	99.8[b]	99.8[b]	99.4[b]
	316.5	1200	100.0	100.0	100.0	100.0
	634.0	2400	100.0	100.0	100.0	100.0
	1266.8	4800	100.0	100.0	100.0	100.0
Model 2 (Half foods positively, half negatively associated with disease)	81.9	300	77.2	38.4	44.8	48.4
	163.8	600	99.2	56.0	69.7	80.0
	327.8	1200	100.0	78.6	87.9	93.2
	655.2	2400	100.0	86.8	95.2	100.0
	1311.5	4800	100.0	93.7	100.0	100.0
Model 3 (All foods positively associated with disease)	86.4	300	94.2	98.1[b]	99.9[b]	98.4[b]
	172.7	600	99.9	99.9	99.9	99.9
	345.7	1200	100.0	100.0	100.0	100.0
	690.6	2400	100.0	100.0	100.0	100.0
	1381.0	4800	100.0	100.0	100.0	100.0
Data-set 2 (ECRHS II)						
Model 1 (All foods negatively associated with disease)	59.8	300	30.2	44.5[b]	49.5[b]	49.4[b]
	119.4	600	82.1	66.2	77.8	82.0
	238.9	1200	99.7	82.8	92.5	96.6
	478.0	2400	100.0	92.2	97.9	99.7
	956.3	4800	100.0	96.1	99.5	99.9
Model 2 (Half foods positively, half negatively associated with disease)	62.3	300	65.1	42.3	53.6	58.9
	124.7	600	97.7	58.3	77.9	87.3
	249.2	1200	99.9	74.3	91.3	97.9
	498.4	2400	100	84.7	97.4	99.8
	997.3	4800	100.0	92.3	99.4	99.9
Model 3 (All foods positively associated with disease)	65.1	300	81.2	61.6	71.4	76.3
	129.8	600	99.3	78.9	89.6	95.6
	260.1	1200	100.0	90.0	97.6	99.5
	520.0	2400	100.0	95.2	99.3	99.9
	1040.3	4800	100.0	97.8	99.8	99.9

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 1 all selected food intakes are negatively associated with disease; in Model 2 half the selected food intakes are negatively associated and half are positively associated; and in Model 3 all selected food intakes are positively associated with disease. [b] Power of PCA exceeds that of ESFA.

Table 5.2.2. Power (%) of exhaustive single food analysis (ESFA) and principal components analysis (PCA) to detect <u>any</u> association between diet and disease (all estimates of power have standard error <0.5%). In models 1-3, *one in twenty* foods (10 in data-set 1 and 4 in data-set 2) are selected at random in each replication from the foods on the FFQ. Family-wise error of Bonferroni correction at 5%.

Data-set and model[a]	Average Number of Cases	Sample Size	ESFA	PCA		
				Number of components		
				2	5	10
Data-set 1(FLAG)						
Model 1 (All foods negatively associated with disease)	61.7	300	11.5	53.1[b]	47.1[b]	42.1[b]
	123.2	600	52.3	75.6[b]	74.4[b]	72.4[b]
	246.3	1200	95.5	89.3	90.7	91.7
	492.4	2400	100.0	96.3	97.3	98.2
	984.8	4800	100.0	99.7	99.3	99.9
Model 2 (Half foods positively, half negatively associated with disease)	62.4	300	31.0	25.7	27.8	27.0
	124.7	600	85.7	41.7	49.9	53.7
	249.7	1200	99.8	60.2	70.6	80.3
	499.5	2400	100.0	73.8	87.5	96.5
	998.9	4800	100.0	85.5	96.5	99.6
Model 3 (All foods positively associated with disease)	66.8	300	67.8	67.1	67.4	67.4
	130.8	600	97.7	83.1	86.2	87.6
	258.7	1200	99.9	92.8	95.9	97.4
	510.5	2400	100.0	96.4	99.2	99.7
	1010.7	4800	100.0	99.2	99.5	100.0
Data-set 2 (ECRHS II)						
Model 1 (All foods negatively associated with disease)	51.6	300	8.8	19.3[b]	16.5[b]	17.3[b]
	103.3	600	38.8	32.4	38.6	36.7
	206.6	1200	89.6	50.9	64.0	68.4
	413.2	2400	100.0	68.3	83.3	92.5
	826.4	4800	100.0	82.3	94.5	98.9
Model 2 (Half foods positively, half negatively associated with disease)	52.3	300	43.1	23.9	30.1	32.4
	104.4	600	89.0	40.2	52.9	62.6
	209.0	1200	99.7	56.8	75.3	86.7
	418.6	2400	100.0	73.4	90.5	97.6
	837.1	4800	100.0	84.5	97.6	99.9
Model 3 (All foods positively associated with disease)	54.3	300	60.4	35.9	43.9	47.8
	106.7	600	96.6	54.3	69.7	78.1
	219.3	1200	100	71.6	86.9	95.2
	429.1	2400	100	84.3	96.1	99.4
	847.5	4800	100	91.8	99.0	99.9

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 1 all selected food intakes are negatively associated with disease; in Model 2 half the selected food intakes are negatively associated and half are positively associated; and in Model 3 all selected food intakes are positively associated with disease.[b] Power of PCA exceeds that of ESFA.

Table 5.2.3. Power (%) of exhaustive single food analysis (ESFA) and principal components analysis (PCA) to detect <u>any</u> association between diet and disease (all estimates of power have standard error <0.5%). In model 4 and 5, foods being included in a simplified *"Western"* dietary pattern (30 in data-set 1 and 10 in data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication. Family-wise error of Bonferroni correction at 5%.

Data-set and model[a]	Sample Size	ESFA	PCA		
			Number of Components		
			2	5	10
odds ratio=1.5					
Data-set 1(FLAG)					
Model 4	100	70.1 [b]	99.5 [b]	99.5 [b]	99.3 [b]
("Western" pattern	300	100	100	100	100
intake positively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Model 5	100	60.7 [b]	97.5 [b]	96.5 [b]	94.2 [b]
("Western" pattern	300	100	100	100	100
intake negatively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Data-set 2 (ECRHS II)					
Model 4	100	62.4 [b]	96.2 [b]	97.1 [b]	93.9 [b]
("Western" pattern	300	100	100	100	100
intake positively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Model 5	100	57.7 [b]	91.7 [b]	91.5 [b]	86.1 [b]
("Western" pattern	300	100	100	100	100
intake negatively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
odds ratio=1.3					
Data-set 1(FLAG)					
Model 4	100	44.0 [b]	97.1 [b]	97.8 [b]	94.2 [b]
("Western" pattern	300	100	100	100	100
intake positively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Model 5	100	23.8 [b]	92.8 [b]	92.9 [b]	84.8 [b]
("Western" pattern	300	99.9	100	100	100
intake negatively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100

Data-set 2 (ECRHS II)					
Model 4	100	55.5 [b]	77.6 [b]	74.3 [b]	66.6 [b]
("Western" pattern	300	100	99.9	100	99.9
intake positively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Model 5	100	34.1 [b]	57.5 [b]	55.3 [b]	55.5 [b]
("Western" pattern	300	100	100	100	100
intake negatively	600	100	100	100	100
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
		odds ratio=1.1			
Data-set 1(FLAG)					
Model 4	100	1.5	40.1 [b]	28.2 [b]	18.0 [b]
("Western" pattern	300	49.2	95.9 [b]	94.9 [b]	91.0
intake positively	600	95.6	99.9	100	100
associated with	1200	99.9	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Model 5	100	1.5 [b]	39.2 [b]	29.9 [b]	29.8 [b]
("Western" pattern	300	30.1	91.9	92.3	85.0
intake negatively	600	84.0	100	100	99.9
associated with	1200	100	100	100	100
disease)	2400	100	100	100	100
	4800	100	100	100	100
Data-set 2 (ECRHS II)					
Model 4	100	1.8 [b]	23.9 [b]	16.7 [b]	9.8 [b]
("Western" pattern	300	30.9 [b]	69.7 [b]	62.1 [b]	51.6 [b]
intake positively	600	74.9	94.0 [b]	92.2 [b]	88.0 [b]
associated with	1200	99.0	99.9	99.9	99.9
disease)	2400	100	100	100	100
	4800	100	100	100	100
Model 5	100	2.1 [b]	19.8 [b]	16.2 [b]	8.7 [b]
("Western" pattern	300	25.1	64.3	66.1	57.0
intake negatively	600	71.7	90.7	93.9	99.6
associated with	1200	99.3	99.8	100	99.9
disease)	2400	100	100	100	100
	4800	100	100	100	100

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 4, foods being included in a "Western" pattern (30 in Data-set 1 and 10 in Data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication, with pattern intake being positively associated with disease. In Model 5, foods being included in a "Western" pattern (30 in Data-set 1 and 10 in Data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication, with pattern intake being negatively associated.

5.3 Power and FDR with which ESFA and PCA could identify specific combinations of foods between diet and disease

Our multiple null hypotheses are evaluated in tables 5.3.1, 5.3.2, and 5.3.3 when 1 in 7 or 1 in 20 foods or a simplified **"Western"** dietary pattern were causally linked with disease.

In each scenario, when food intakes were causally associated with disease, ESFA had lower false discovery rate than PCA for identifying the specific combination of foods (Table 5.3.1 and 5.3.2). EFSA had greater power in all simulations except in some (17/242) of the simulations where the sample size was low (300 and 600).

When a simplified **"Western"** pattern intake was causally linked with disease, ESFA also had a higher power than PCA for identifying specific combination of foods (Table 5.3.3), but PCA had a lower false discovery rate (with also lower power).

Simulation results for average percentages of power and FDR estimates of ESFA didn't present a symmetrical pattern between model 1 (all selected food intakes are negatively associated with disease) and model 3 (all selected food intakes are positively associated with disease), and we get intermediate power estimates for Model 2 (half the selected food intakes are negatively associated and half are positively associated with disease). Simulation results for the ESFA method which were presented in Table 5.3.3 were more symmetrical compared to Tables 5.3.1 and 5.3.2.

Power and FDR of PCA for detecting specific effects of foods on disease is roughly the same or higher when all effects of foods are positive compared to when all effects are negative. In addition power and FDR is less when there is a mixture of positive and negative effects compared to when all effects of foods on disease are positive. Lower or roughly the same power is observed when all effect of foods are negative compared to where there is a mixture of positive and negative effects of foods on disease. In addition, power and FDR increases as the number of principal components and the size of the sample increases. Simulation results for PCA method which were presented in Table 5.3.3 presented an almost symmetrical pattern.

Table 5.3.1. Power[c] and false discovery rate (FDR)[d] estimates (%) of exhaustive single food analysis (ESFA) and principal components analysis (PCA) for detecting the foods that are causally linked to disease (all estimates of power and FDR have standard error <0.5%). In models 1-3, *one in seven* foods (30 in data-set 1 and 10 in data-set 2) are selected at random in each replication from the foods on the FFQ.

Data-set and model [a]	Sample Size	ESFA		PCA Number of Components					
				2		5		10	
		Power	FDR	Power	FDR	Power	FDR	Power	FDR
Data-set 1 (FLAG)									
Model 1 (All foods negatively associated with disease)	300	13.1	28.9	35.5	85.7	46.1	85.3	55.2 [b]	85.5
	600	42.9	50.1	34.3	86.3	47.1	86.9	56.8	86.1
	1200	72.6	64.3	36.2	86.1	49.5	86.1	58.6	85.8
	2400	87.6	73.2	37.2	86.2	49.6	86.3	59.3	86.1
	4800	93.9	78.9	38.2	86.5	51.5	86.1	60.9	86.0
Model 2 (Half foods positively, half negatively associated with disease)	300	49.1	70.3	14.6 [b]	42.3	20.7 [b]	63.9	24.6 [b]	76.1
	600	76.9	77.4	19.4	54.1	26.9	73.8	33.5	82.6
	1200	89.7	80.3	26.0	67.7	34.2	82.9	40.2	85.2
	2400	95.0	82.2	29.7	75.0	38.6	84.3	46.3	85.8
	4800	97.3	83.6	33.5	80.3	41.7	86.1	51.2	85.9
Model 3 (All foods positively associated with disease)	300	55.3	71.8	35.0	85.2	47.8	85.8	55.2	85.4
	600	77.5	77.8	36.0	86.4	48.0	86.2	57.8	85.7
	1200	88.6	80.5	36.8	86.2	49.6	86.2	59.0	85.9
	2400	93.8	82.6	37.4	86.5	50.0	86.3	60.2	85.9
	4800	98.7	83.5	39.0	86.4	50.1	86.4	60.2	86.1
Data-set 2 (ECRHS II)									
Model 1 (All foods negatively associated with disease)	300	21.7	37.8	21.0	46.5	29.3 [b]	64.9	34.4 [b]	75.4
	600	62.7	53.9	29.2	62.9	39.6	76.3	48.9	82.2
	1200	90.1	66.1	35.8	73.6	49.1	83.1	60.5	84.1
	2400	97.9	74.7	41.1	79.9	56.1	85.0	69.1	85.0
	4800	99.5	79.6	45.3	83.1	61.4	85.7	75.2	85.4
Model 2 (Half foods positively, half negatively associated with disease)	300	35.8	38.1	17.5	42.9	27.1	66.4	35.4	77.9
	600	70.6	54.9	24.6	57.0	39.6	77.4	48.5	83.3
	1200	90.1	67.2	31.7	68.1	45.6	83.3	60.0	84.5
	2400	96.2	75.5	38.3	75.9	52.6	85.3	68.3	85.3
	4800	98.2	80.0	42.9	80.6	58.7	85.9	74.0	85.6
Model 3 (All foods positively associated with disease)	300	53.3	48.0	27.2	58.6	39.5	75.3	47.9	80.3
	600	81.1	63.6	33.9	70.9	47.6	81.7	59.2	83.4
	1200	92.9	73.2	39.5	78.3	54.4	84.5	68.0	84.6
	2400	96.9	78.8	43.8	82.6	59.5	85.5	74.1	85.2
	4800	98.4	81.7	46.8	84.3	63.9	85.9	78.2	85.9

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 1 all selected food intakes are negatively associated with disease; in Model 2 half the selected food intakes are negatively associated and half are positively associated; and in Model 3 all selected food intakes are positively associated with disease. [b] Power of PCA exceeds that of ESFA, but FDR is also higher. [c] **Power** is defined as the number of true significant results identified by the method divided by the number of foods that are causally linked with disease. [d]**FDR** is defined as the number of false significant results identified by the method divided by the total number of significant results identified by the method. FDR of Simes procedure at 20%.

Table 5.3.2. Power[c] and false discovery rate (FDR)[d] estimates (%) of exhaustive single food analysis (ESFA) and principal components analysis (PCA) for detecting the foods that are causally linked to disease (all estimates of power and FDR have standard error <0.5%). In models 1-3, *one in twenty* foods (10 in data-set 1 and 4 in data-set 2) are selected at random in each replication from the foods on the FFQ.

Data-set and model [a]	Sample Size	ESFA		PCA Number of Components					
				2		5		10	
		Power	FDR	Power	FDR	Power	FDR	Power	FDR
Data-set 1 (FLAG)									
Model 1 (All foods negatively associated with disease)	300	6.5	22.7	21.5 [b]	60.6	26.5 [b]	70.8	29.9 [b]	80.5
	600	38.2	52.1	27.3	77.2	33.1	84.3	38.0	89.5
	1200	80.1	72.8	31.3	88.0	39.4	91.6	46.6	93.6
	2400	95.7	84.8	33.7	92.9	44.0	94.0	50.9	94.7
	4800	99.1	90.2	36.1	94.5	46.2	95.0	54.7	94.9
Model 2 (Half foods positively, half negatively associated with disease)	300	12.1	21.8	10.8	34.4	14.7 [b]	52.5	18.6 [b]	70.0
	600	40.4	41.5	15.7	48..9	21.1	69.9	26.7	84.9
	1200	73.6	63.1	21.5	64.6	28.0	82.9	35.5	91.8
	2400	91.1	79.9	26.1	75.4	34.7	90.5	42.6	94.5
	4800	96.6	88.3	30.6	85.0	39.8	93.6	49.0	94.8
Model 3 (All foods positively associated with disease)	300	41.2	50.3	26.6	71.6	33.7	81.0	39.5	87.2
	600	76.4	68.9	30.2	83.4	39.4	90.3	46.9.	92.7
	1200	94.3	82.2	33..0	95.1	43.6	93.7	52.1	94.2
	2400	98.5	89.1	35.3	93.3	45.6	94.9	55.9	94.8
	4800	99.5	92.7	38.0	94.4	48.0	95.2	58.0	95.0
Data-set 2 (ECRHS II)									
Model 1 (All foods negatively associated with disease)	300	1.2	8.5	8.8 [b]	24.1	12.2 [b]	37.5	12.7 [b]	51.0
	600	11.2	19.9	13.8 [b]	36.6	17.9 [b]	53.3	22.7 [b]	68.6
	1200	50.7	38.6	20.9	54.1	26.5	71.6	31.9	83.3
	2400	85.9	62.0	27.5	74.1	34.0	85.5	41.7	93.1
	4800	96.2	81.6	33.1	83.5	40.6	92.1	49.8	96.5
Model 2 (Half foods positively, half negatively associated with disease)	300	29.7	27.8	12.8	32.2	21.2	55.2	29.2	72.2
	600	64.7	46.3	19.2	46.3	30.7	71.6	43.5	84.9
	1200	90.3	65.1	26.3	60.3	40.7	84.1	56.8	90.5
	2400	97.9	80.4	34.0	74.2	49.4	90.7	67.2	92.4
	4800	99.5	87.9	39.4	83.2	56.9	93.1	74.2	93.3
Model 3 (All foods positively associated with disease)	300	51.5	35.5	20.5	43.3	31.3	65.2	41.7	79.1
	600	86.6	56.1	28.0	58.6	40.6	80.2	54.9	88.5
	1200	98.3	74.7	34.6	71.9	49.4	88.8	66.0	91.7
	2400	99.8	85.7	39.9	82.3	56.1	92.3	73.9	92.8
	4800	99.9	90.2	44.6	88.2	61.3	93.6	78.2	93.5

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 1 all selected food intakes are negatively associated with disease; in Model 2 half the selected food intakes are negatively associated and half are positively associated; and in Model 3 all selected food intakes are positively associated with disease. [b] Power of PCA exceeds that of ESFA, but FDR is also higher. [c] **Power** is defined as the number of true significant results identified by the method divided by the number of foods that are causally linked with disease. [d]**FDR** is defined as the number of false significant results identified by the method divided by the total number of significant results identified by the method. FDR of Simes procedure at 20%.

Table 5.3.3. Power[c] and false discovery rate (FDR)[d] estimates (%) of exhaustive single food analysis (ESFA) principal components analysis (PCA) for detecting the foods that deemed to constitute a simplified *"Western"* dietary pattern and are causally linked to disease (all estimates of power and fdr have standard error <0.5%). In model 4, foods being included in a simplified *"Western"* pattern are used in each replication.

Data-set and model [a]	Sample Size	ESFA		PCA Number of Components					
				2		5		10	
		Power	FDR	Power	FDR	Power	FDR	Power	FDR
Data-set 1 (FLAG)									
Model 4 ("Western" pattern intake positively associated with disease)	300	88.6	76.7	55.3	82.4	74.8	80.7	79.8	81.7
	600	97.6	80.5	67.5	77.9 [c]	81.6	79.2 [c]	83.3	81.3
	1200	99.7	82.8	78.2	74.3 [c]	86.9	77.4 [c]	86.4	80.7 [c]
	2400	99.1	84.1	84.5	72.2 [c]	90.9	76.1 [c]	87.7	80.4 [c]
	4800	100	84.9	87.3	71.5 [c]	93.1	75.2 [c]	87.7	80.4 [c]
Model 5 ("Western" pattern intake negatively associated with disease)	300	92.2	77.7	54.0	82.6	75.6	81.4	80.8	82.3
	600	98.7	81.1	67.9	77.8 [c]	82.2	79.4 [c]	84.2	81.6
	1200	99.9	83.3	78.4	74.2 [c]	87.1	77.5 [c]	86.5	80.9 [c]
	2400	100	84.3	84.4	72.3 [c]	90.7	76.0 [c]	87.8	80.5 [c]
	4800	100	84.9	87.4	71.5 [c]	93.2	75.2 [c]	87.9	80.5 [c]
Data-set 2 (ECRHS II)									
Model 4 ("Western" pattern intake positively associated with disease)	300	96.1	56.3	60.8	74.7	73.4	72.0	84.5	73.3
	600	99.9	67.1	63.6	74.3	76.6	72.9	89.5	75.3
	1200	100	73.0	66.5	74.2	79.9	73.1	92.6	76.4
	2400	100	76.0	68.6	74.0 [c]	84.5	72.6 [c]	93.6	77.0
	4800	100	77.3	70.3	73.7 [c]	89.8	71.6 [c]	94.2	77.3
Model 5 ("Western" pattern intake negatively associated with disease)	300	96.8	60.3	61.0	74.6	74.1	74.0	84.2	74.0
	600	99.9	68.2	63.9	74.4	77.1	74.8	89.2	75.8
	1200	100	73.3	66.4	74.2	80.5	74.5	92.2	76.7
	2400	100	76.0	68.7	73.9 [c]	85.7	73.3 [c]	94.1	77.1
	4800	100	77.4	70.3	73.7 [c]	91.2	71.9 [c]	94.8	77.3

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 4, foods being included in a "Western" pattern (30 in Data-set 1 and 10 in Data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication, with pattern intake being positively associated with disease. In Model 5, foods being included in a "Western" pattern (30 in Data-set 1 and 10 in Data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication, with pattern intake being negatively associated. [c] FDR of PCA is lower than that of ESFA, but power is also lower. [c] **Power** is defined as the number of true significant results identified by the method divided by the number of foods that are causally linked with disease. [d]**FDR** is defined as the number of false significant results identified by the method divided by the total number of significant results identified by the method. FDR of Simes procedure at 20%.

5.4 Power and FDR with which ESFA and PCA could identify specific combinations of foods between diet and disease adjusted in three different ways

Table 5.4.1, 5.4.2, and 5.4.3 presents average estimates of percentages of power and false discovery rates for different sample sizes when exhaustive single food analysis is employed for 3 different ways of adjustment (see paragraphs 4.3.2, 4.3.3 and 4.3.4).

Attempting to control the FDR of ESFA by adjusting for principal components of diet or propensity scores was not successful for either datasets, especially for large sample sizes. However, adjusting for other foods that were significant in an unadjusted analysis controlled the FDR at around the nominal 20% level, though with some loss of power, particularly with low sample sizes.

Simulation results for average percentages of power estimates of adjusted ESFA didn't present a symmetrical pattern between model 1 (all selected food intakes are negatively associated with disease) and model 3 (all selected food intakes are positively associated with disease), and we get intermediate power estimates for Model 2 (half the selected food intakes are negatively associated and half are positively associated with disease) for small sample size scenarios (300, 600 and 1200). However for larger sample size scenarios (2400, 4800) they are roughly the same. Average parentages of FDR for the adjusted ESFA, exceeded the nominal level of 20% for the F.L.A.G survey data-set for specific sample size scenarios (300, 600, 1200). However, as the sample size increases above 2400, FDR was controlled around 20%. Not any other specific patterns across Tables 5.4.1, 5.4.2 and 5.4.3 were observed.

Simulation results of power and FDR for ESFA adjusted for 5 principal components were qualitatively similar with the simulation results that were observer for the unadjusted ESFA. (Tables in paragraphs 5.3). No clear and consistent simulation patterns for the average percentages of power and FDR were observed when the ESFA method was adjusted for propensity scores.

Table 5.4.1. Power[c] and false discovery rate (FDR)[d] estimates (%) of exhaustive single food analysis (ESFA) with different methods of adjustment for other foods (all estimates of power and FDR have standard error <0.5%). In models 1-3, *one in seven* foods (30 in data-set 1 and 10 in data-set 2) are selected at random in each replication from the foods on the FFQ. FDR of Simes procedure at 20%.

Data-set and model [a]	Sample Size	Adjusted for 5 principal components		Adjusted for foods that are significant in unadjusted analysis		Adjusted for propensity scores	
		Power	FDR	Power	FDR	Power	FDR
Data-set 1 (FLAG)							
Model 1 (All foods negatively associated with disease)	300	1.7	27.5	4.8	29.4	0.9	8.5
	600	19.4	44.5	24.5	48.8	17.5	33.6
	1200	59.6	53.7	50.9	35.3	66.3	47.1
	2400	85.5	64.2	78.2	27.2	99.2	58.5
	4800	95.1	72.6	91.4	24.3	98.6	69.7
Model 2 (Half foods positively, half negatively associated with disease)	300	6.5	19.1	5.0	18.5	0.1	1.2
	600	35.0	35.8	30.5	33.4	2.3	3.3
	1200	71.5	51.5	67.5	31.8	35.5	8.0
	2400	89.9	65.0	88.7	27.8	79.2	14.5
	4800	96.4	73.9	96.1	25.7	93.9	24.8
Model 3 (All foods positively associated with disease)	300	7.3	26.8	5.2	17.2	0.2	0.8
	600	33.0	43.4	24.6	33.7	1.8	11.2
	1200	67.5	57.3	64.5	41.9	24.9	19.4
	2400	87.9	68.5	86.6	30.4	91.2	54.0
	4800	95.6	75.6	95.3	27.4	94.2	54.9
Data-set 2 (ECRHS II)							
Model 1 (All foods negatively associated with disease)	300	9.6	30.2	4.2	15.9	5.7	7.7
	600	40.1	44.1	22.7	19.8	28.4	10.8
	1200	77.5	55.4	67.5	19.7	73.6	16.0
	2400	94.3	66.8	93.3	19.0	96.4	22.8
	4800	98.6	74.8	99.2	18.1	99.7	35.5
Model 2 (Half foods positively, half negatively associated with disease)	300	25.4	29.9	15.4	15.2	4.1	6.4
	600	61.5	45.7	47.2	19.6	26.5	10.3
	1200	86.9	59.9	82.1	20.0	72.2	16.1
	2400	96.0	70.8	96.3	19.9	94.5	23.7
	4800	98.7	77.7	99.0	20.4	99.1	33.7
Model 3 (All foods negatively associated with disease)	300	32.3	35.1	23.0	16.8	3.0	5.5
	600	67.0	52.4	60.7	20.7	29.5	11.7
	1200	88.0	66.0	90.7	21.4	66.8	22.5
	2400	95.7	74.7	98.8	21.4	93.2	37.7
	4800	98.2	79.5	99.9	21.4	98.6	58.0

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 1 all selected food intakes are negatively associated with disease; in Model 2 half the selected food intakes are negatively associated and half are positively associated; and in Model 3 all selected food intakes are positively associated with disease. [c] **Power** is defined as the number of true significant results identified by the method divided by the number of foods that are causally linked with disease. [d]**FDR** is defined as the number of false significant results identified by the method divided by the total number of significant results identified by the method. FDR of Simes procedure at 20%.

Table 5.4.2. Power[c] and false discovery rate (FDR)[d] estimates (%) of exhaustive single food analysis (ESFA) with different methods of adjustment for other foods (all estimates of power and FDR have standard error <0.5%). In models 1-3, *one in twenty* foods (10 in data-set 1 and 4 in data-set 2) are selected at random in each replication from the foods on the FFQ. FDR of Simes procedure at 20%.

Data-set and model [a]	Sample Size	Adjusted for 5 principal components		Adjusted for foods that are significant in unadjusted analysis		Adjusted for propensity scores	
		Power	FDR	Power	FDR	Power	FDR
Data-set 1 (FLAG)							
Model 1 (All foods negatively associated with disease)	300	0.7	13.4	1.1	11.2	0.7	10.2
	600	8.7	26.7	5.9	21.5	4.6	13.8
	1200	46.8	37.7	31.0	27.4	38.1	24.5
	2400	82.9	54.9	70.3	22.8	83.3	36.4
	4800	95.4	73.0	89.1	21.3	96.9	53.0
Model 2 (Half foods positively, half negatively associated with disease)	300	7.7	13.6	6.0	10.5	0.1	5.8
	600	32.8	26.8	27.9	21.1	4.1	5.4
	1200	69.3	45.6	63.6	23.0	35.4	10.4
	2400	91.1	67.4	86.0	23.2	77.5	14.1
	4800	97.4	82.8	94.6	20.5	93.4	21.0
Model 3 (All foods positively associated with disease)	300	12.7	16.1	11.4	19.3	0.1	2.3
	600	48.7	31.5	44.3	29.5	3.7	4.3
	1200	83.6	52.6	79.3	28.1	41.2	9.5
	2400	96.5	74.5	93.4	24.4	87.1	14.7
	4800	99.2	86.4	97.5	21.5	98.1	25.6
Data-set 2 (ECRHS II)							
Model 1 (All foods negatively associated with disease)	300	9.6	30.2	4.2	15.9	5.7	7.7
	600	40.1	44.1	22.7	19.8	28.4	10.8
	1200	77.5	55.4	67.5	19.7	73.6	16.0
	2400	94.3	66.8	93.3	19.0	96.4	22.8
	4800	98.6	74.8	99.2	18.1	99.7	35.5
Model 2 (Half foods positively, half negatively associated with disease)	300	25.4	29.9	15.4	15.2	4.1	6.4
	600	61.5	45.7	47.2	19.6	26.5	10.3
	1200	86.9	59.9	82.1	20.0	72.2	16.1
	2400	96.0	70.8	96.3	19.9	94.5	23.7
	4800	98.7	77.7	99.0	20.4	99.1	33.7
Model 3 (All foods positively associated with disease)	300	32.3	35.1	23.0	16.8	3.0	5.5
	600	67.0	52.4	60.7	20.7	29.5	11.7
	1200	88.0	66.0	90.7	21.4	66.8	22.5
	2400	95.7	74.7	98.8	21.4	93.2	37.7
	4800	98.2	79.5	99.9	21.4	98.6	58.0

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 1 all selected food intakes are negatively associated with disease; in Model 2 half the selected food intakes are negatively associated and half are positively associated; and in Model 3 all selected food intakes are positively associated with disease. [c] **Power** is defined as the number of true significant results identified by the method divided by the number of foods that are causally linked with disease. [d]**FDR** is defined as the number of false significant results identified by the method divided by the total number of significant results identified by the method. FDR of Simes procedure at 20%.

Table 5.4.3. Power[c] and false discovery rate (FDR)[d] estimates (%) of exhaustive single food analysis (ESFA) with different methods of adjustment for other foods (all estimates of power and FDR have standard error <0.5%). In model 4, foods being included in a *"Western"* pattern are used in each replication. FDR of Simes procedure at 20%.

Data-set and model[a]	Sample Size	Adjusted for 5 principal components		Adjusted for foods that are significant in unadjusted analysis		Adjusted for propensity scores	
		Power	FDR	Power	FDR	Power	FDR
Data-set 1(FLAG)							
Model 4 ("Western" pattern intake positively associated with disease)	300	4.3	52.8	2.0	18.5	0.5	21.3
	600	13.5	70.1	32.5	33.4	2.1	79.5
	1200	32.7	75.6	57.5	38.8	19.7	77.7
	2400	56.1	78.0	72.7	35.1	54.2	74.3
	4800	72.5	80.6	94.4	22.9	76.4	76.8
Model 5 ("Western" pattern intake negatively associated with disease)	300	3.2	43.9	3.2	22.5	1.1	13.8
	600	12.1	65.8	24.1	26.9	27.8	34.8
	1200	33.8	71.4	44.9	47.4	70.1	40.0
	2400	61.2	75.2	73.7	29.2	70.3	39.1
	4800	79.2	79.0	94.5	20.8	76.6	76.8
Data-set 2 (ECRHS II)							
Model 4 ("Western" pattern intake positively associated with disease)	300	15.6	55.1	12.4	23.9	3.6	5.2
	600	33.8	66.9	34.1	21.9	26.9	18.1
	1200	50.6	70.3	65.6	18.8	67.5	33.5
	2400	63.1	73.3	87.1	17.2	83.1	50.0
	4800	71.4	75.7	96.7	17.5	89.6	61.3
Model 5 ("Western" pattern intake negatively associated with disease)	300	14.7	54.5	9.2	22.5	6.4	6.1
	600	33.4	67.6	27.1	22.9	38.2	11.1
	1200	50.6	70.1	61.7	19.2	77.4	21.0
	2400	63.0	73.4	90.2	18.1	95.5	34.4
	4800	71.8	75.5	98.7	17.8	99.7	50.4

[a] Data-set 1 is from the FLAG survey (Shaheen et al., 2001); Data-set 2 is from the UK ECRHS II survey (Hooper et al., 2010). In Model 4, foods being included in a "Western" pattern (30 in Data-set 1 and 10 in Data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication, with pattern intake being positively associated with disease. In Model 5, foods being included in a "Western" pattern (30 in Data-set 1 and 10 in Data-set 2 – see paragraph 4.9.2 and 4.9.3) are used in each replication, with pattern intake being negatively associated. [c] **Power** is defined as the number of true significant results identified by the method divided by the number of foods that are causally linked with disease. [d]**FDR** is defined as the number of false significant results identified by the method divided by the total number of significant results identified by the method. FDR of Simes procedure at 20%.

6 Analysis of G.A^2.L.E.N data: effect of diet on asthma

6.1 Introduction

There is accumulating evidence from observational studies that high intakes of fruit, vegetable and oily fish have a protective effect in children and adults with asthma (Oien, Storro & Johnsen, 2010, Hodge et al., 1996, Fitzsimon et al., 2007). Additionally, asthma has been associated with dietary antioxidants and particularly with low intakes of vitamin C, D and E and selenium. (Shaheen et al., 2001, Hodge et al., 1996, Devereux, 2010). However, trials of supplementation have been unsuccessful and have provided contradictory results (Pearson et al., 2004, Fogarty et al., 2003, Devereux & Seaton, 2005, Shaheen et al., 2007). This may be because apparent effects in the observational studies were confounded by lifestyle factors or by other dietary components. In addition, there are a number of methodological problems in analysing the effects of single food and nutrient analysis, as described in paragraph 2.1.2.

Dietary patterns empirically derived with the use of Principal Components Analysis are provided by the literature as an alternative way of investigating associations between diet and asthma. Studies have provided only weak (Bakolis et al., 2010, Butler et al., 2006) or no statistical significant evidence (Hooper et al., 2010, Shaheen et al., 2009, Varraso et al., 2009) that any of these patterns were associated with asthma. However, associations between dietary patterns and wheeze (Takaoka & Norback, 2008), allergic rhinitis (Bakolis et al., 2010), positive skin prick tests (Hooper et al., 2010) and COPD (Varraso et al., 2007a, Varraso et al., 2007b) have been observed. These associations are limited by conceptual and methodological disadvantages of PCA as described by the literature (Newby & Tucker, 2004, Slattery & Boucher, 1998) and by paragraph 2.6.

The purpose of this chapter is to provide a comprehensive analysis of the association between dietary intake in adults across Europe in relation to self-reported asthma and other respiratory and allergic symptoms. In addition, we aim to compare and interpret the results from dietary pattern analysis with the use of PCA with our two-step ESFA procedure.

6.2 Materials and methods

6.2.1 Study design and population

GA^{2}LEN was an EU funded network of excellence which coordinated the study of genetic and environmental risk factors for asthma in adult and adolescent population across 17 centres in 11 European countries. The GA^{2}LEN follow up survey is a cross-sectional study amongst those previously contacted in baseline postal surveys that were willing to be contacted again. Invitation to follow-up was dependent on three groups of cases (those with asthma, those with sinusitis, and those with both asthma and sinusitis) and one group of controls (those with neither asthma nor sinusitis) drawn from the postal survey sample. Body and height measurements and skin prick tests (SPTs) to grass pollen, grass mix, Dermatophagoides pteronyssinus, Dermatophagoides farinae, cockroach (Blatella), olive, Alternaria, dog, Artemisia, birch, cat and Parietaria were conducted. In all centres permission to conduct this study was obtained from appropriate local ethics committees, and all participants signed a written consent form after being fully informed about the study.

6.2.2 Definitions of respiratory symptoms

Asthma was defined as present in those who had answered yes to having a diagnosis of asthma and either wheezing, waking up with chest tightness, waking up with shortness of breath, or waking at night with an attack of coughing in the previous 12 months. Also Chronic sinusitis (CRS) was diagnosed in those who reported that nose had been blocked, pain/pressure around forehead nose/eyes, discoloured nasal discharge (snot)/discoloured mucus in the throat, sense of smell been reduced/absent in the previous 12 months. A symptom-based definition of CRS, according to the epidemiological part of the European Position Paper on Rhinosinusitis and Nasal Polyps (EPO3S) criteria is suitable for the assessment of geographic variation in prevalence of CRS (Tomassen et al., 2011). Other respiratory symptoms included allergic rhinitis, which was defined as a positive response to the question “Do you have any nasal allergies including hay fever?” and eczema was defined as a positive response to the question “Did you ever have eczema or any kind of skin allergy?” Atopy was defined as any positive response (1mm more than the diluent control) to any of the allergens tested (grass pollen, grass mix, Dermatophagoides pteronyssinus, Dermatophagoides farinae, cockroach (Blatella), olive, Alternaria, dog, Artemisia, birch, cat and Parietaria).

6.2.3 Dietary assessment

6.2.3.1 Food Frequency Questionnaire (FFQ)

The GA^{2}LEN survey objective is to assess dietary intake across European countries using a single common and standardised method. It is the first study that the same standardised FFQ was administered across the European countries that took part. We designed the FFQ taking into account what other large international epidemiological studies have done to assess dietary intake in various European countries. In particular, the researchers from EPIC (Epidemiological Prospective Study in the causes of Cancer) facilitated all the FFQs used in each country. We also had valuable input from various patient associations and from lay members of the public representing each country.

FFQs recorded a consumption of 239 food items over the last 12 months ranging from never to two portions a day or more. We estimated weekly intake (g) of foods and food groups by multiplying frequency of consumption by the weight of standard portion sizes.

Prior to the FFQ being used in the GA^{2}LEN Nutrition survey, we successfully piloted and validated it in five countries representing different regions of Europe (Scandinavia (Helsinki), South Mediterranean (Athens), Central Europe (Brandenburg), East Europe (Lodz), North Mediterranean (Porto) (Garcia-Larsen et al., 2011). The FFQ showed a high level of repeatability for most nutrients. We also validated the FFQ comparing dietary n-3 fatty acids against specific fatty acids in plasma. We found a good correlation between n-3 fatty acids in diet compared with total plasma phospholipid n-3 fatty acids and with docosahexaenoic acid. This was observed both in the entire sample (ICC 0.40) and per country. These results indicate that the GA^{2}LEN FFQ is an appropriate tool to estimate dietary intake for a range of nutrients across Europe regardless of cultural and linguistic differences (Garcia-Larsen et al., 2011)..

6.2.3.2 Exclusions of dietary data

Respondents sometimes left individual items blank on the FFQs. This was assumed to denote zero intakes of these foods unless more than 20% of items were blank, in which case the FFQ was considered incomplete, and the subject was excluded from analyses. We have included in our analysis only food items which were consumed at least 1-3 times per week from more than 2% of the individuals in our final sample (see appendix X for further details).

Based on this criterion, xilopites were excluded from our analysis. In this case we used a more a higher threshold for excluding food items than what is recommended (Willet, 1998).

6.2.4 Statistical Analysis

6.2.4.1 Dietary patterns analysis with the use of PCA

First, in order to derive our dietary patterns, we ran a principal component analysis for all data combined and estimated our principal component scores as we described at paragraph 2.4.

Since there is an expectation of between country heterogeneity our principal component scores tend to be correlated within countries. An alternative method which takes into account this dependence of our principal component scores is to identify our dietary patterns with the use of PCA from an overall pooled correlation matrix using a meta-analysis method proposed by Hedges and Olkin (Hedges, 1985). Specifically, in each country k for each food item *i* and food item *j*, we evaluated the correlation matrix using the Pearson product-moment correlation coefficient r_{kij}. Because the approximate distribution of r_{kij} depends strongly on the value of the population correlation ρ_{kij}, each correlation coefficient was transformed using a Fisher transformation

$$z_{kij} = 0.5 \log((1+ r_{kij})/(1- r_{kij}))) \text{ , } i=1,..239 \quad j=1,..239 \quad k=1,..10$$

to give it an approximately normal distribution with asymptotic variance $1/(n_j-3)$, where n_k is the sample size for the country *k*. A weighted average of these values was then calculated

$$\sum_{k=1}^{10} w_k z_{kij} = w_1 z_{1ij} + ... + w_{10} z_{10ij,} \; i=1,..239 \quad j=1,..239$$

where the weights are $w_k = \dfrac{(n_k - 3)}{\sum_{l=1}^{10}(n_l - 3)}$

An inverse Fisher transformation was then applied to give a pooled correlation coefficient matrix. PCA was applied to the matrix of pooled correlation coefficients, giving us dietary pattern scores which could be used in all the 10 countries. This meta-analytic approach to PCA has previously been applied in the field of psychiatry (Smith, Mar & Turoff, 1998, Grube, Bilder & Goldman, 1998) and asthma epidemiology (Hooper et al., 2010).

For each country, we used multivariable logistic regression to investigate associations between the dietary patterns (in quintile groups) and respiratory outcomes adjusted for age, sex, smoking status and body mass index. We selected our confounders based on information in previous literature (Tricon et al., 2006). The effects of the dietary patterns were also adjusted for each other, because although principal components are uncorrelated, varimax rotations can introduce correlations between the dietary patterns. Regression results were pooled across countries using random effects meta-analysis, with a test for heterogeneity of regression coefficients (DerSimonian & Laird, 1986). Heterogeneity was summarised using the I^2 statistic (Higgins et al., 2003).

All analyses were weighted to the population that took part in the postal survey. The sampling probability weights for each subject in each subset of the survey data were computed by dividing the frequency of the subject's centre and case status in the postal survey by the frequency of the same centre and case status in the subset.

6.2.4.2 Two step Exhaustive Single Food Analysis (ESFA) for multicentre data

In multi-centre studies, because responses tend to be correlated within centres, a method which takes into account this dependence is a multilevel model (Goldstein, 2011, Hox, 2010). Specifically, associations of self reported respiratory outcomes with each individual food in the FFQ were assessed using a random intercept logistic model. The random intercept could be thought as the combined effect of all omitted individual-specific covariates that cause some individuals to be more prone to the potential respiratory outcomes than others and is used to model the unobserved heterogeneity between countries. Random intercept models treat countries as a random representative sample from a "larger" population of countries (Rabe-Hesketh, Sophia 2006, Hox, 2010).

We employed our two-step exhaustive single food analysis framework. First we ran the analysis controlling the false discovery rate at 20%. Second, we re-ran the same analysis additionally adjusting for foods that were identified as statistically significant at the first step and controlling the false discovery rate at 20%. Similar analyses were conducted for associations between each individual food and asthma, sinusitis, nasal allergies, eczema and atopy. All analysis was weighted to the population that took part in the postal survey and adjusted for age, sex, smoking status and body mass index (we selected our confounders based on previous literature (Tricon et al., 2006)). Sampling probability weights were

rescaled according to a method proposed by Rabe-Hesketh in order not to induce bias in our standard estimators (Rabe-Hesketh, Sophia 2006).

Statistical analyses were conducted using STATA 12 (Stata Corporation, College Station, Texas USA).

6.3 Results

6.3.1 Descriptive statistics

Centres in Palermo (Italy), Krakow (Poland) and Skopje (Macedonia) were excluded because of the small number of cases and additionally small number of individuals who completed the FFQ questionnaire (n=32, n=0 and n=26 respectively). We didn't merge the German centres (Berlin, Duisberg) into one country because of socio-economic and demographic differences between their populations. Our final sample included 3057 individuals living in 15 centres in 9 countries with full data on food frequency questionnaire data and confounders. Food item entitled xilopites was excluded from our analysis due to infrequent consumption (which leads to no disease relevance) across countries; only 18 individuals (18/3057=0.4 %) consumed xilopites (see Appendix X for further details). Further descriptive data on health outcomes and confounders are presented in Table 6.3.1.1.

Table 6.3.1.1. Description of sample

	Number/Total with information	Proportion (%) unless otherwise stated
Respiratory and Allergic Outcomes		
Asthma (%)	1078/3057	35.2
Chronic Sinusitis (%)	595/3057	19.4
Allergic Rhinitis (%)	1437/3057	47.0
Eczema (%)	1760/3057	57.5
Atopy (%)	1434/3057	46.9
Confounders		
Age (med, IQR)	3057	(median: 48 , IQR : 36-60)
Female (%)	1746/3057	57.1
Smokers (%)		
Never	1519/3057	49.6
Ever	991/3057	32.6
Current	539/3057	17.6
Body Mass Index (med, IQR)	3057	(median: 25.2 , IQR : 22.7-28.6)
Country (%)		
Belgium(Ghent)	146/3057	4.78

Denmark(Odense)	354/3057	11.58
Finland(Helsinki)	155/3057	5.07
North Germany(Berlin)	177/3057	5.79
South Germany(Duisberg)	194/3057	6.35
Holland(Amsterdam)	215/3057	7.03
Portugal(Coimbra)	259/3057	8.47
Poland(Katowice, Lodz, Krakow)	210/3057	6.87
UK(Southampton, London)	171/3057	5.59
Sweden(Umea, Uppsala, Gothenburg)	1176/3057	38.47

6.3.2 Empirically derived dietary patterns with the use of PCA

6.3.2.1 For all countries combined

Principal component analysis was applied to the FFQ data for all countries combined. This was performed in controls only to avoid potential bias. A varimax rotation was applied to improve the interpretability of the patterns obtained. The scree plot (Figure 6.3.2.1.1) was examined to aid the choice of number of patterns, but this choice was mainly based on the principal component interpretability. We extracted the first two components (dietary patterns), which explained 14.8% of the variance in the original 238 items (food item entitled xilopites was removed due to limited observations; 0.5% of the sample (12 individuals in total) consumed xilopites in 4 different countries). A component score was created for each individual for each of the principal components identified. Individual foods that correlated >0.3 or <-0.3 with the varimax rotated dietary patterns (principal components) are shown in Table 6.3.2.1.1. Two patterns were derived; a **"fruit, fish and vegetables"** pattern (containing higher levels of fish, fruit, vegetables and boiled chicken) and a **"meat, potatoes and sweets"** one (containing higher levels of high fat foods, potatoes and meats).

Figure 6.3.2.1.2 presents ten different scree plots when PCA was performed for each country separately, showing different number of patterns (4 in Belgium, 5 in Denmark, 2 in Finland, 4 in North Germany, 3 in South Germany, 3 in Holland, 4 in Portugal, 2 in Poland, 4 in UK and 3 in Sweden) that could potentially be derived empirically by country. Furthermore, we calculated the mean of **"fruit, fish and vegetables"** and **"meat, potatoes and sweets"** pattern intake for all the individuals for each country separately. We observed that there is variation of the **"meat, potatoes and sweets"** and **"fruit, fish and vegetables"** mean pattern intake scores between countries (Figure 6.3.2.1.3). Therefore, using pattern scores without accounting for between country heterogeneity would lead to false estimations and conclusions.

Table 6.3.2.1.1. Foods items which correlated >0.3 or <-0.3 with the identified dietary patterns for all the data combined (Food items that didn't correlated > 0.3 or <-0.3 with any of the two patterns excluded from the table)

	"fruit, fish and vegetables"	"meat, potatoes and sweets"
Food item		
peach	0.55	-
garlic	0.52	-
fresh vegetable/cereal soup	0.52	-
leek	0.51	-
plum	0.51	-
melon/watermelon	0.50	-
chard	0.49	-
carrot	0.49	-
onion	0.49	-
fresh white fish	0.49	-
herbs	0.48	-
cauliflower	0.48	-
pear	0.48	-
legumes, any	0.46	-
pumpkin	0.46	-
cabbage	0.46	-
nectarine	0.46	-
orange	0.46	-
courgette	0.45	-
broccoli	0.45	-
fresh fruit	0.45	-
grape	0.45	-
mango	0.44	-
apricot	0.44	-
lettuce	0.43	-
tomato	0.43	-
turnip	0.43	-
chickpeas	0.42	-
apple	0.42	-
cherry	0.42	-
pineapple	0.42	-
spinach	0.41	-
small game	0.41	-
olive oil	0.40	-
white	0.39	-
kiwi	0.39	-
fresh fatty fish	0.39	-
mandarin/tangerine	0.38	-
fava beans	0.37	-
celery	0.37	0.31

brussels spouts	0.37	-
forest fruits	0.37	-
lemon	0.37	-
aubergine	0.36	-
turkey	0.36	-
kidney	0.35	-
french beans	0.35	-
chicken boiled	0.35	-
vegetables, any	0.34	-
olive	0.34	-
smoked white fish	0.34	-
veal	0.33	-
sweet peppers	0.32	-
bitter melon	0.32	-
stuffed vegetables	0.32	0.44
potato tortilla	0.32	0.31
any poultry	0.31	-
chicken	0.31	0.36
parsnip	0.30	-
beetroot	0.30	-
Danish pastries	-	0.46
sweet rolls	-	0.45
custard cream	-	0.44
potatoes(boiled/mashed)	-	0.44
cakes	-	0.42
potato dumpling	-	0.39
meat stew	-	0.39
beef burger	-	0.38
total sweets or bonbons	-	0.37
canned fruit	-	0.37
halva	-	0.36
boiled sweets/toffees	-	0.35
pickled vegetables	-	0.35
smoked game	-	0.35
puddings	-	0.34
chocolate, any	-	0.34
cottage cheese	-	0.34
radish	-	0.33
bacon	-	0.33
eggs, any	-	0.33
table sugar	-	0.32
potato salad	-	0.32
doughnuts/other pastries	-	0.31
chocolate (plain)	-	0.31
butter, any	-	0.31
frankfurter	-	0.31

egg-based desserts	-	0.31
ketchup	-	0.30

Figure 6.3.2.1.1. Scree plot for overall data combined

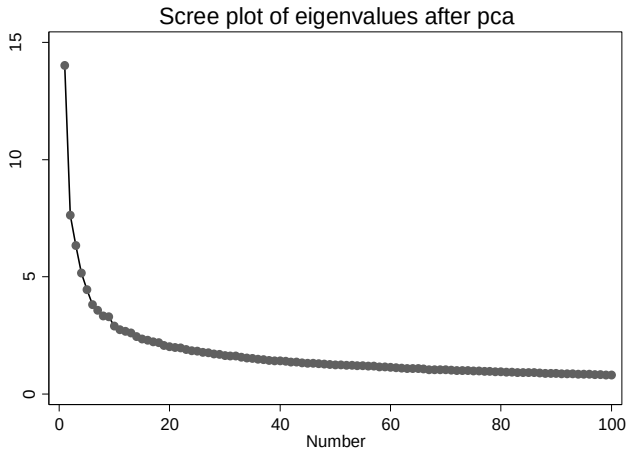


Figure 6.3.2.1.2. Scree plots for each country separately

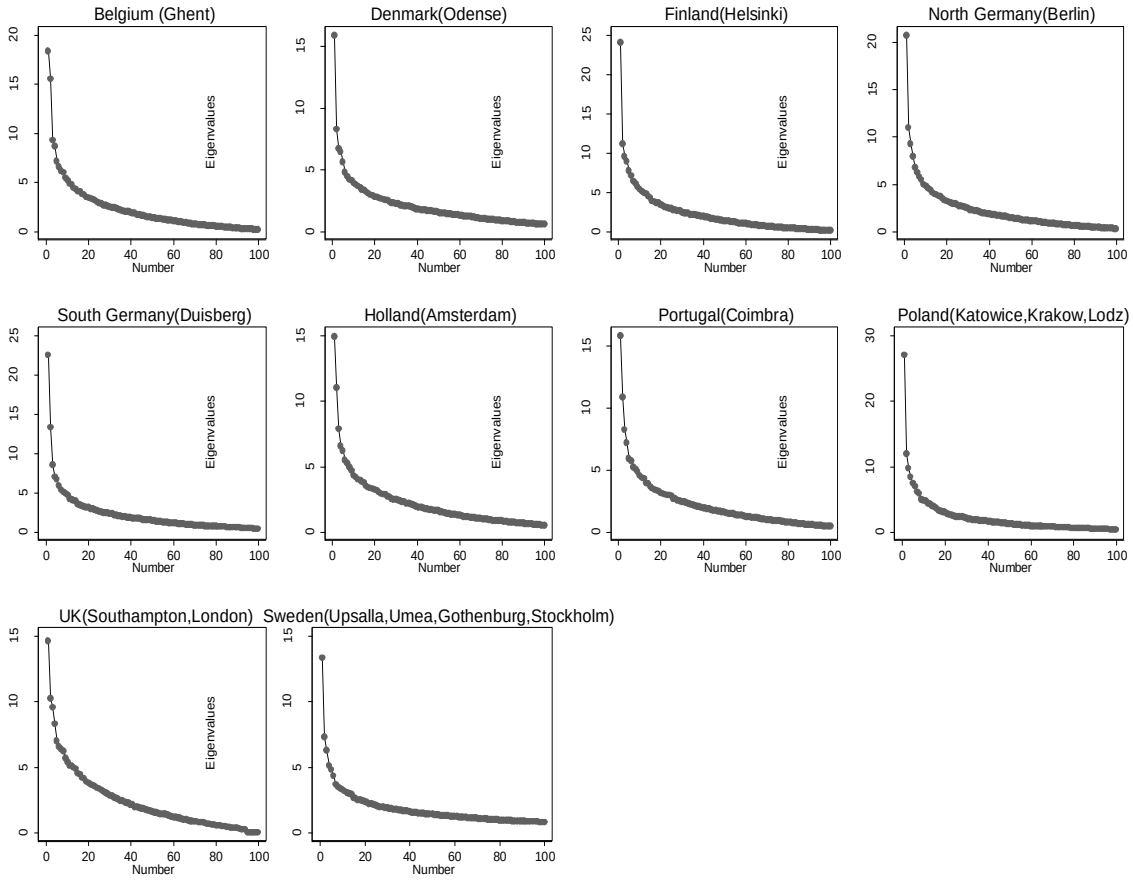
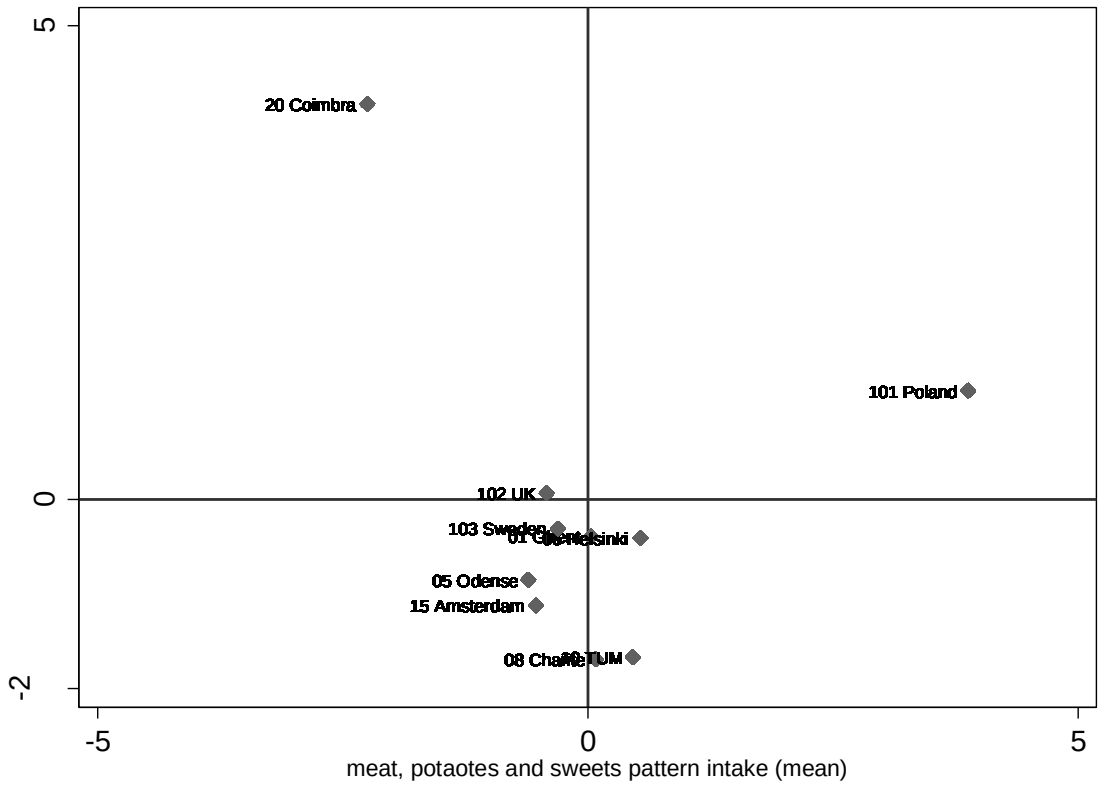


Figure 6.3.2.1.3. Mean individual "meat, potatoes and sweets" and "fruit, fish and vegetables" pattern intake for each country separately.



6.3.2.2 For multicentre data

The scree plot from the PCA (figure 6.3.2.2.1) showed a break in the curve after two or six or nine components. We derived two principal components since they provided a more meaningful interpretation. Table 6.3.2.2.1 shows how individual foods were correlated with each of these patterns at ten countries. This table shows little similarity between individual's patterns from country to country – that is, the patterns mean different things in different countries. This makes interpretation of the results more difficult, and is another way to represent the heterogeneity between countries because of the different local diets

Additionally, according to the table, at more than eight countries, the first pattern was characterized by high consumption of brown wholemeal bread, vegetables and fruits and fresh fatty fish and the second one was closely associated with intakes of white bread, cakes , muffins, any butter, chips/fries, beef and sausages , eggs, mayonnaise and crème fraiche. For simplicity purposes we labelled the first pattern as **"fruit, fish and vegetables"** and the second one as **"meat potatoes and sweets".** However, a closer investigation of the **"fruit,**

fish and vegetables" pattern reveals food items that are not only **"fish, fruit and vegetables"** or **"meat, potatoes and sweets"**. For example, in Table 6.3.2.2.1 a **"fish, fruit and vegetables"** pattern correlated highly with chicken in stews (Belgium, Finland), lamb (Belgium, South Germany, Sweden), tinned fatty fish (Belgium, Denmark, UK), turkey roast (Belgium, Finland, South Germany, Portugal, Poland), veal (Belgium, Finland), smoked game (Belgium, Finland, Portugal, Poland), sausages (Belgium), fresh fatty fish (Belgium, Denmark, Finland, North Germany, Portugal, UK and Sweden), full fat butter (Denmark), mayonnaise (Finland), pizza (North Germany), butter, any (Finland), crème fraiche (Finland, North Germany, South Germany, Poland, UK), meat pies (Finland, Poland), cured pork (Finland), frankfurter (Finland, Poland), bacon cubes (Finland) and smoked lamb (Finland, South Germany).

Figure 6.3.2.2.1. Scree plot for multicentre data

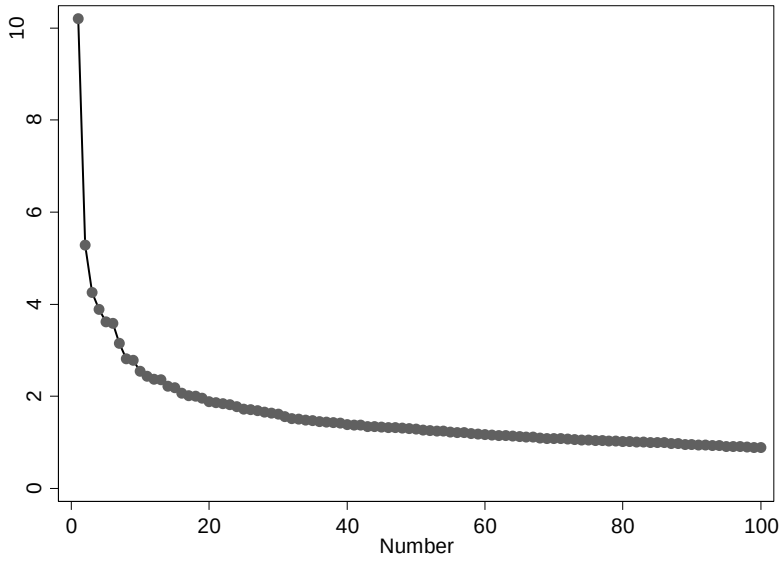


Table 6.3.2.2.1. How correlation coefficients of food items with identified dietary patterns vary between countries*

	Belgium (Ghent)		Denmark (Odense)		Finland (Helsinki)		North Germany Berlin(		South Germany Munich		Holland (Amsterdam)		Portugal (Coimbra)		Poland (Lodz) (Katowice)		UK (London) (Southampton)		Sweden (Karolinska) (Gothenburg) (Umea) (Uppsala)	
	I**	**II*****	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**
broccoli	**0.63**	-	**0.53**	-	**0.48**	-	**0.46**	-	**-**	-	**0.39**	-	**0.35**	-	**0.38**	-	**0.34**	-	**0.31**	-
tomato	**0.61**	-	**0.61**	-	**-**	-	**0.54**	-	**0.53**	-	**0.55**	-	**0.45**	-	**0.47**	-	**0.44**	-	**0.34**	-
pumpkin	**0.60**	-	**-**	-	**0.46**	-	**0.56**	-	**0.32**	-	**0.35**	-	**0.42**	-	**0.38**	-	**0.39**	-	**-**	-
radish	**0.59**	-	**0.40**	-	**-**	-	**-**	-	**-**	-	**0.35**	-	**0.44**	-	**0.42**	-	**-**	-	**0.33**	-
pineapple	**0.59**	-	**0.44**	-	**0.36**	-	**0.51**	-	**0.61**	0.35	**0.47**	-	**0.51**	-	**0.41**	-	**-**	-	**0.31**	-
cucumber	**0.58**	-	**0.44**	-	**-**	-	**0.35**	-	**0.58**	-	**0.39**	-	**0.32**	-	**0.31**	-	**0.35**	-	**-**	-
cauliflower	**0.58**	-	**0.43**	-	**0.63**	0.39	**0.38**	0.30	**-**	-	**0.33**	-	**0.35**	-	**0.49**	-	**-**	-	**0.50**	-
coleslaw	**0.57**	-	**0.31**	-	**0.59**	0.48	**0.34**	0.35	**-**	-	**0.48**	-	**-**	-	**0.32**	-	**-**	-	**-**	-
onion	**0.57**	-	**0.53**	-	**0.38**	-	**0.39**	-	**0.63**	0.32	**0.41**	-	**0.47**	-	**0.47**	-	**0.50**	-	**0.34**	-
caper	**0.54**	-	**-**	-	**0.50**	-	**0.37**	0.39	**-**	-	**0.32**	-	**-**	-	**-**	-	**-**	-	**-**	-
white bread	**0.54**	0.37	**-**	0.37	**0.52**	0.74	**-**	0.33	**-**	-	**-**	-	**-**	0.54	**0.37**	0.32	**-**	-	**-**	0.36
chickpeas	**0.52**	-	**-**	-	**0.34**	-	**-**	-	**-**	-	**-**	-	**-**	-	**-**	-	**0.42**	-	**0.50**	-
chicken in stews	**0.52**	-	**-**	-	**0.30**	0.36	**-**	0.33	**-**	-	**-**	-	**-**	-	**-**	-	**-**	-	**-**	-
pear	**0.51**	-	**0.43**	-	**0.49**	0.40	**-**	-	**0.49**	-	**-**	-	**0.32**	-	**0.41**	-	**-**	-	**0.34**	-
banana	**0.51**	0.37	**0.47**	-	**0.30**	-	**-**	-	**0.36**	-	**-**	-	**-**	-	**-**	-	**-**	-	**-**	-
melon watermelon	**0.51**	0.32	**-**	-	**0.42**	-	**-**	-	**0.61**	0.34	**0.59**	-	**0.46**	-	**-**	-	**-**	-	**0.34**	-
lamb	**0.51**	-	**-**	0.36	**-**	-	**-**	0.37	**0.35**	-	**-**	-	**-**	-	**-**	-	**-**	-	**0.33**	-
turkey roast	**0.51**	-	**-**	-	**0.48**	0.45	**-**	-	**0.53**	-	**-**	-	**0.50**	-	**0.34**	-	**-**	-	**-**	-
tinned fatty fish	**0.51**	-	**0.34**	-	**-**	0.34	**-**	-	**-**	-	**-**	-	**-**	0.33	**-**	-	**0.32**	-	**-**	-
Brussels sprouts	**0.50**	-	**0.42**	-	**0.38**	-	**0.42**	-	**0.38**	-	**0.37**	-	**-**	-	**0.50**	0.30	**-**	-	**0.36**	-
rhubarb	**0.50**	-	**-**	-	**0.48**	0.38	**0.37**	-	**-**	-	**-**	-	**-**	-	**-**	-	**-**	-	**-**	-
veal	**0.50**	-	**-**	0.34	**0.41**	0.66	**-**	0.44	**-**	-	**-**	0.39	**-**	0.52	**-**	0.48	**-**	-	**-**	-
carrot	**0.49**	-	**0.41**	-	**0.46**	0.32	**-**	-	**0.63**	-	**0.47**	-	**0.43**	-	**0.45**	-	**-**	-	**0.37**	-
smoked game	**0.49**	0.32	**-**	-	**0.55**	0.69	**-**	-	**-**	-	**-**	-	**0.30**	-	**0.39**	0.36	**-**	-	**-**	-

celery	**0.48**	-	**0.49**	-	**0.54**	-	**0.35**	-	**0.44**	-	**0.65**	-	-	-	**0.48**	-	**0.31**	-	**0.51**	-
nectarine	**0.48**	-	**0.35**	-	**0.50**	0.33	**0.50**	-	**0.44**	0.30	**0.33**	-	**0.38**	-	**0.51**	-	**0.31**	-	-	-
sausages	**0.48**	-	-	0.32	-	0.34	-	0.30	-	-	-	-	-	0.40	-	0.48	-	0.45	-	0.30
any legumes	**0.47**	-	-	-	**0.43**	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.52**	-
French beans	**0.47**	-	-	-	-	-	**0.52**	-	-	-	**0.44**	-	**0.33**	0.35	**0.40**	-	-	-	**0.36**	-
artichoke	**0.47**	-	-	-	-	-	**0.37**	-	-	-	-	-	-	-	-	-	-	-	**0.44**	-
apricot	**0.46**	0.33	**0.49**	-	-	-	**0.52**	-	**0.66**	0.40	**0.68**	-	**0.33**	-	**0.51**	-	-	-	**0.37**	-
parsnip	**0.44**	-	**0.45**	-	**0.42**	-	-	-	-	-	**0.33**	-	-	-	-	-	-	-	**0.43**	-
herbs	**0.43**	-	**0.54**	-	**0.56**	-	**0.50**	-	**0.70**	0.38	**0.33**	-	**0.53**	-	**0.62**	0.32	**0.59**	-	**0.62**	-
peach	**0.43**	-	**0.47**	-	**0.69**	0.67	**0.54**	-	-	-	-	-	**0.36**	-	**0.39**	-	**0.50**	-	**0.36**	-
fresh fatty fish	**0.43**	-	**0.30**	-	**0.46**	0.49	**0.44**	-	-	-	-	-	**0.41**	-	-	-	**0.30**	-	**0.41**	-
fresh white fish	**0.43**	-	**0.45**	-	-	-	**0.34**	-	-	-	-	-	-	-	**0.30**	-	-	-	**0.33**	-
fresh vegetable/cereal soup	**0.43**	-	-	-	**0.51**	0.48	-	-	**0.34**	0.38	-	-	**0.39**	-	**0.34**	0.36	**0.34**	-	**0.43**	-
leek	**0.42**	-	**0.62**	-	**0.66**	-	**0.62**	-	**0.35**	-	**0.61**	-	**0.62**	-	-	-	**0.38**	-	**0.45**	-
olive	**0.42**	-	**0.62**	-	**0.54**	0.43	**0.35**	-	**0.59**	-	**0.40**	-	-	0.31	**0.36**	-	-	-	**0.59**	-
ice cream	**0.42**	-	-	-	**0.43**	0.57	-	-	-	-	-	-	-	0.40	-	0.40	-	0.31	-	0.32
dressing sauces	**0.42**	-	-	-	-	-	-	-	-	-	-	0.44	-	-	-	-	**0.49**	-	-	0.31
marmalade	**0.41**	0.34	-	-	**0.31**	-	-	-	-	-	-	-	-	-	**0.42**	0.49	-	-	-	-
lettuce	**0.41**	-	**0.52**	-	**0.32**	-	**0.41**	-	**0.45**	-	**0.42**	-	-	-	-	-	**0.52**	-	**0.39**	-
mango	**0.41**	0.35	**0.45**	-	**0.54**	-	**0.56**	-	-	-	**0.57**	-	**0.55**	-	**0.42**	-	**0.46**	-	**0.40**	-
grapefruit	**0.41**	-	**0.36**	-	**0.38**	-	**0.31**	-	**0.65**	0.32	**0.47**	-	-	-	**0.44**	-	-	-	**0.32**	-
wild greens	**0.40**	0.45	**0.30**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
turnip	**0.40**	0.41	**0.38**	-	-	-	-	-	-	-	**0.35**	-	**0.39**	-	**0.45**	-	-	-	-	-
white sauce	**0.40**	0.33	**0.42**	-	-	0.38	**0.40**	0.38	-	0.34	**0.40**	-	-	-	**0.32**	-	-	-	**0.42**	-
hard cheese	**0.40**	-	-	-	-	-	-	-	**0.32**	-	**0.31**	-	-	-	-	0.50	-	-	-	-
sour milk	**0.39**	-	-	-	**0.40**	0.49	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Greek yoghurt	**0.39**	-	-	-	**0.38**	0.43	-	-	-	-	-	-	-	-	-	-	**0.31**	-	-	-
squeezed fresh fruit	**0.38**	-	**0.49**	-	**0.48**	-	**0.42**	-	**0.42**	-	**0.34**	-	**0.41**	-	**0.32**	-	**0.45**	-	**0.42**	-
canned fruit	**0.38**	-	-	-	-	-	-	-	**0.33**	-	-	-	-	-	**0.44**	0.53	-	-	-	-

fromage frais	**0.37**	-	-	-	-	-	-	-	-	-	-	-	**0.37**	-	-	-	-	-	-	-
mixed candies	**0.36**	0.38	-	0.34	-	-	-	-	-	-	-	-	-	-	-	-	-	0.41	-	0.31
sweet peppers	**0.34**	-	**0.55**	-	-	-	**0.34**	-	**0.55**	-	**0.48**	-	**0.53**	-	**0.48**	-	**0.43**	-	**0.35**	-
chocolate, any	**0.33**	0.39	-	0.48	-	-	-	-	-	-	-	-	-	0.50	-	0.49	-	-	-	-
fenugreek	**0.33**	0.39	-	-	-	-	**0.40**	-	-	-	**0.40**	-	-	-	-	-	-	-	**0.36**	-
taro	**0.33**	0.39	-	-	-	-	-	-	-	-	**0.40**	-	-	-	-	-	-	-	-	-
aubergine	**0.33**	-	**0.54**	-	**0.33**	-	**0.34**	-	**0.32**	-	**0.61**	-	**0.38**	-	**0.41**	-	**0.44**	-	**0.42**	-
bitter melon	**0.33**	0.39	**0.30**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
potato dumpling	**0.33**	0.39	-	-	-	-	-	-	-	0.43	-	-	-	0.30	-	0.37	-	-	-	-
grape	**0.33**	0.32	**0.41**	-	**0.50**	0.34	**0.66**	-	**0.41**	-	**0.49**	-	-	-	**0.45**	-	-	-	-	-
other game	**0.33**	-	-	-	**0.40**	0.73	-	-	-	-	-	-	-	-	-	-	-	-	-	-
chard	**0.32**	0.40	-	-	-	-	-	-	-	-	-	0.30	-	-	-	-	**0.35**	-	-	-
courgette	**0.32**	-	**0.44**	-	**0.61**	-	**0.42**	-	-	-	**0.62**	-	**0.57**	-	**0.45**	-	**0.55**	-	**0.56**	-
cherries	**0.32**	-	-	-	-	-	**0.47**	-	**0.30**	-	**0.37**	-	**0.45**	0.31	**0.52**	-	-	-	-	-
kiwi	**0.32**	-	**0.56**	-	-	-	**0.33**	-	**0.62**	-	-	-	**0.41**	-	-	-	-	-	-	-
mayonnaise	-	0.54	-	-	**0.56**	0.53	-	0.36	-	0.48	-	0.40	-	0.40	-	0.38	-	-	-	-
casserole	-	0.50	-	-	-	-	**0.34**	0.31	**0.31**	0.56	-	-	**0.31**	0.50	-	-	-	0.41	-	-
chips/ French fries	-	0.43	-	-	-	-	-	0.56	-	0.47	-	0.46	-	0.37	-	0.43	**-0.43**	0.37	-	0.33
hot/cold roast beef, boiled beef	-	0.43	-	-	**0.38**	0.65	-	0.40	-	0.40	-	-	-	-	-	0.57	-	-	-	0.35
total sweets or bonbons	-	0.40	-	0.45	-	-	-	-	-	-	-	-	-	0.32	-	-	-	0.50	-	0.43
low fat butter	-	0.40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
pizza	-	0.40	-	0.40	-	0.53	-	-	**0.33**	0.58	-	-	-	0.40	-	0.35	-	0.31	-	0.37
boiled sweets/toffees	-	0.37	-	0.30	-	-	-	-	-	0.31	-	-	-	-	-	-	**-0.33**	0.38	-	0.32
potatoes(boiled/mashed)	-	0.37	-	-	-	0.68	-	-	-	-	-	-	-	0.34	-	-	-	-	-	0.30
table sugar	-	0.35	-	-	-	0.34	-	-	-	0.31	-	-	-	-	-	-	**0.35**	0.38	-	0.36
chard	-	0.34	-	-	-	-	-	-	-	-	-	-	-	-	-	0.32	-	-	-	-
muffins	-	0.33	-	0.33	**0.39**	0.36	-	0.38	-	0.33	-	0.32	-	0.33	-	0.32	-	0.32	-	-
Greek cakes	-	0.33	-	-	-	-	-	-	**0.44**	-	-	-	-	-	-	-	-	-	-	-

any cakes or pastries	-	0.32	-	0.42	-	0.35	-	0.48	-	0.32	-	-	-	0.48	-	-	-	0.33	-	0.38
Italian biscuits	-	0.32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
margarine, any	-	0.32	-	-	-	-	-	-	-	-	-	0.31	-	-	-	-	-	-	-	-
half fat margarine	-	0.32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
any butter	-	0.32	-	0.32	**0.34**	0.45	-	0.45	-	-	-	-	-	0.44	-	0.34	-	0.30	-	-
condensed milk	-	0.32	-	-	-	-	-	-	-	-	-	-	-	0.52	-	-	-	0.33	-	-
custard cream	-	0.31	-	0.36	-	0.35	-	0.50	-	-	-	-	-	-	**0.36**	0.47	-	0.33	-	-
blended spreads	-	0.31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
egg based dishes	-	0.31	-	-	-	0.35	-	0.32	-	0.50	-	0.30	**0.32**	0.47	-	0.46	-	0.47	-	-
egg based desserts	-	0.31	-	0.32	-	0.44	-	0.38	-	0.37	-	0.33	-	0.40	-	0.45	-	0.56	-	0.32
single or sour cream	-	0.31	-	0.34	**0.33**	-	-	-	-	0.30	-	-	-	0.41	-	0.41	-	-	-	-
crème fraiche	-	0.31	-	-	**0.55**	0.64	**0.41**	-	**0.35**	0.35	-	0.38	-	-	**0.47**	0.47	**0.42**	-	-	-
wholemeal bread	-	-	-	-	-	-	-	-	-	-	-	-	**0.35**	-	-	-	-	-	-	-
white bread and rolls	-	-	-	0.31	-	0.30	-	0.43	-	0.43	-	0.31	-	0.36	-	0.32	-	-	-	0.38
rye bread	-	-	-	-	-	-	**0.52**	-	-	-	-	-	**0.31**	-	**0.38**	-	**0.36**	-	-	-
kneipp bread	-	-	-	-	-	-	-	-	**0.45**	0.38	-	-	-	-	-	-	-	-	-	-
nan, paratha	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.40	-	-
chapatti bread	-	-	-	-	-	0.31	-	-	**0.62**	0.45	-	-	-	-	-	-	-	-	-	-
breakfast cereal	-	-	-	-	-	-	-	-	-	-	**0.35**	-	-	-	-	-	-	-	-	-
couscous	-	-	**0.36**	-	**0.43**	0.40	-	0.30	-	-	-	-	-	-	-	-	**0.39**	-	-	-
pasta, any	-	-	-	-	-	-	-	0.37	-	-	**0.47**	-	-	0.39	-	-	-	0.30	-	0.33
refined pasta	-	-	-	-	-	-	-	-	-	-	**0.40**	0.43	-	0.34	-	-	-	0.31	-	0.32
wholemeal pasta	-	-	**0.31**	-	-	-	-	-	-	-	-	-	-	0.35	-	-	-	-	-	-
filled pasta	-	-	-	-	-	-	-	-	**0.37**	0.57	-	-	-	-	-	-	-	-	-	-
noodles	-	-	-	0.31	-	-	-	0.39	-	-	-	-	-	-	-	0.51	-	0.44	-	-
cakes	-	-	-	0.32	-	0.39	-	0.34	-	-	-	-	-	0.43	-	-	-	-	-	-
Danish pastries	-	-	-	0.42	-	-	-	-	-	-	-	-	-	0.44	-	0.36	-	-	-	0.32
sweet rolls	-	-	-	-	-	-	-	-	-	-	-	0.31	-	0.38	-	0.34	-	-	-	-
doughnuts	-	-	-	-	-	0.49	-	0.50	-	-	-	0.38	-	0.33	-	0.39	-	0.39	-	-

pudding and desserts	-	-	-	-	-	0.33	-	0.41	-	-	-	-	**0.34**	0.32	-	0.30	-	-	-	-
pancakes	-	-	-	0.36	-	0.39	-	0.49	-	-	-	0.42	-	-	-	-	-	-	-	-
sweet biscuits	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.31**	0.32	-	-	-	-
crisps, fries	-	-	-	-	-	-	-	-	-	-	**0.40**	-	-	0.32	-	0.35	-	-	-	-
thin biscuits	-	-	-	-	-	-	**0.33**	-	-	-	-	-	-	0.31	**0.34**	-	-	-	-	-
sweet biscuits	-	-	-	-	-	0.32	-	-	-	-	-	-	-	0.59	**0.35**	0.43	-	0.42	-	-
rice, any	-	-	**0.39**	-	**0.32**	0.32	-	-	-	-	-	-	-	0.46	-	-	-	-	-	-
white rice	-	-	-	-	-	-	-	0.35	-	-	-	-	-	0.43	-	-	-	-	-	-
brown rice	-	-	**0.49**	-	**0.54**	0.40	-	-	-	-	-	-	**0.35**	-	**0.44**	0.40	-	-	-	-
rice noodles	-	-	-	-	-	-	-	0.31	-	-	**0.31**	-	-	-	**0.53**	0.54	-	0.31	-	-
jam	-	-	-	-	-	0.32	-	-	-	-	-	-	-	-	-	0.33	-	-	-	-
honey	-	-	-	-	-	-	**0.34**	-	-	-	-	-	-	-	**0.33**	-	-	-	-	-
syrup spreads	-	-	-	-	-	-	-	-	-	-	-	0.33	-	0.40	**0.45**	0.48	-	-	-	-
apple spread	-	-	-	-	**0.39**	-	-	-	-	-	-	-	-	-	-	0.38	-	0.36	-	-
cereal bars	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.33	-	-
halva	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.32	-	-
water ice	-	-	-	-	-	-	-	0.45	-	-	-	-	-	-	-	-	-	-	-	-
chocolate snack bars	-	-	-	0.37	-	-	-	0.44	-	-	-	0.38	-	-	-	0.38	-	0.42	-	0.37
chocolate (plain)	-	-	-	-	-	-	-	-	-	-	-	-	-	0.39	-	-	-	-	-	-
vegetable oil	-	-	**0.43**	-	-	-	**0.31**	-	**0.33**	-	-	-	-	0.35	-	-	-	0.49	**0.30**	-
sunflower oil	-	-	-	-	-	-	**0.31**	-	**0.36**	0.41	-	-	-	-	-	-	-	0.47	-	-
olive oil	-	-	**0.45**	-	**0.34**	-	**0.44**	-	**0.43**	-	-	-	-	-	-	-	**0.42**	-	**0.43**	-
full fat butter	-	-	-	0.33	**0.33**	-	-	0.39	-	-	-	-	-	0.36	-	-	-	0.38	-	-
any nuts	-	-	**0.42**	-	-	-	**0.39**	-	-	-	**0.38**	-	**0.36**	-	-	-	-	-	**0.46**	-
peanuts	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.45**	-
cashew nuts	-	-	**0.44**	-	-	-	**0.31**	-	-	-	-	-	-	-	-	-	**0.30**	-	**0.42**	-
nut based spread nuts	-	-	-	-	-	-	-	-	-	0.30	-	-	-	-	-	-	-	-	-	-
kidney(red), black beans	-	-	-	-	**0.47**	-	-	-	-	-	-	-	-	-	**0.30**	-	-	-	**0.53**	-
lentils	-	-	-	-	**0.37**	-	-	-	-	-	**0.52**	-	-	-	-	-	**0.48**	-	**0.51**	-

cluster beans	-	-	-	-	-	-	-	-	-	-	**0.47**	-	-	-	-	-	-	-	-	-
fava beans	-	-	-	-	-	-	-	0.33	-	-	-	-	**0.36**	0.33	-	-	-	-	-	-
soya beans	-	-	-	-	**0.37**	-	-	-	-	-	**0.47**	-	-	-	**0.31**	-	-	-	-	-
vegetables, any	-	-	**0.51**	-	-	-	**0.36**	-	**0.52**	-	-	-	**0.41**	-	**0.40**	-	**0.39**	-	**0.32**	-
spinach	-	-	**0.42**	-	**0.41**	-	-	-	-	-	-	-	**0.32**	-	**0.38**	-	**0.38**	-	**0.48**	-
okra	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.33**	-	-	-	-	-
beetroot	-	-	-	-	-	-	**0.49**	-	-	-	**0.39**	-	**0.47**	-	**0.50**	0.30	-	-	**0.33**	-
sweet corn	-	-	-	-	-	-	-	-	**0.42**	0.33	**0.37**	-	-	-	**0.36**	0.32	-	-	-	-
asparagus	-	-	**0.34**	-	-	-	**0.53**	-	**0.32**	-	-	-	-	-	**0.46**	0.34	**0.45**	-	**0.49**	-
garlic	-	-	**0.58**	-	**0.41**	-	-	-	**0.62**	0.32	**0.30**	-	**0.50**	-	**0.45**	-	**0.60**	-	**0.35**	-
cabbage	-	-	**0.45**	-	**0.53**	0.31	-	-	-	0.33	**0.61**	-	-	-	**0.55**	0.32	-	-	**0.42**	-
stuffed vegetables	-	-	-	-	**0.49**	0.73	-	-	-	-	**0.35**	0.31	-	-	**0.38**	0.37	-	-	-	-
pickled vegetables	-	-	-	-	**0.37**	-	-	-	**0.35**	-	-	-	-	-	**0.33**	-	-	-	-	-
ginger	-	-	**0.41**	-	**0.48**	0.46	**0.39**	-	**0.56**	-	-	-	-	-	**0.54**	-	-	-	**0.31**	-
potatoes, all	-	-	-	-	-	-	-	-	-	0.33	-	-	-	0.39	-	0.44	-	0.37	-	-
potato salad	-	-	-	0.32	**0.45**	-	**0.40**	0.39	**0.59**	0.55	-	0.46	-	0.58	-	-	-	-	-	-
potato tortilla	-	-	-	-	-	-	-	-	-	-	-	-	-	0.32	-	-	-	-	-	-
sweet potato	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.32**	-	-	-
apple	-	-	**0.49**	-	**0.46**	-	**0.31**	-	**0.42**	-	-	-	**0.34**	-	**0.49**	-	-	-	**0.40**	-
avocado	-	-	**0.60**	-	**0.56**	-	**0.44**	-	**0.42**	-	**0.60**	-	-	-	-	-	**0.32**	-	**0.49**	-
forest fruits(berries/blueberries, strawberry)	-	-	-	-	**0.36**	-	**0.52**	-	**0.37**	-	-	-	**0.45**	-	**0.37**	-	**0.50**	-	**0.33**	-
plum	-	-	**0.42**	-	**0.45**	0.41	**0.56**	-	**0.68**	-	-	-	**0.50**	-	**0.46**	-	-	-	-	-
squeezed fresh fruit	-	-	-	-	-	-	-	-	**0.60**	-	-	-	-	-	**0.44**	0.31	-	-	-	-
lemon	-	-	**0.55**	-	**0.55**	0.34	**0.52**	-	**0.69**	0.34	**0.53**	-	-	-	**0.39**	-	**0.38**	-	**0.50**	-
orange	-	-	**0.56**	-	**0.38**	-	**0.58**	-	**0.61**	-	**0.35**	-	**0.35**	-	**0.31**	-	-	-	**0.33**	-
mandarin/tangerine	-	-	-	-	**0.40**	-	**0.57**	-	**0.35**	-	**0.41**	-	-	-	**0.40**	-	-	-	-	-
raisins	-	-	**0.54**	-	-	-	**0.39**	-	-	-	**0.35**	-	**0.35**	-	**0.33**	-	-	-	-	-
fig	-	-	**0.42**	-	-	-	**0.37**	-	**0.30**	-	**0.50**	-	**0.36**	-	**0.52**	0.31	**0.34**	-	-	-

prune	-	-	-	-	-	-	-	-	-	-	**0.43**	-	-	-	-	-	-	-	-	-
concentrated juice (sugar)	-	-	-	-	-	-	-	-	-	-	-	-	-	0.30	-	0.33	-	-	-	-
carbonated drinks	-	-	-	-	-	-	-	0.36	-	-	-	0.45	-	-	-	-	**-0.32**	-	-	-
tap water	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.44**	-	-	-
mineral water	-	-	-	-	**0.41**	0.38	-	-	-	-	-	-	-	-	-	-	-	-	-	-
soda (sugar)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.31	-	0.32	-	-
black tea	-	-	-	-	-	0.33	-	-	**0.35**	-	-	-	-	-	-	-	-	-	-	-
coffee	-	-	-	0.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
herbal tea	-	-	**0.42**	-	**0.42**	-	**0.32**	-	-	-	-	-	**0.33**	-	**0.30**	-	**0.43**	-	-	-
red wine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.34	-	-	-	-
rose wine	-	-	-	-	**0.35**	-	-	-	-	-	-	-	-	-	**0.36**	0.33	-	-	-	-
red meat, any	-	-	-	-	-	-	-	0.35	-	0.57	-	0.31	-	0.42	-	0.45	-	0.33	-	0.31
beef burger	-	-	-	-	**0.30**	0.67	-	0.45	-	0.43	-	0.39	-	0.36	-	0.51	-	-	-	0.43
minced meat in sauce	-	-	-	0.35	-	-	-	0.45	-	0.50	-	0.31	-	0.47	-	0.51	-	-	-	0.35
meat stew	-	-	-	-	**0.38**	0.62	-	-	-	-	-	-	-	0.43	**0.41**	0.64	-	0.42	-	-
pork cutlet, chop, steak, fillet	-	-	-	0.44	-	0.66	-	-	-	0.33	-	0.37	-	-	**0.34**	0.65	-	0.39	-	0.33
meat pies	-	-	-	-	**0.39**	0.76	-	-	-	-	-	0.34	-	-	**0.38**	0.32	**-0.30**	0.41	-	-
meat spreads e.g. rillets	-	-	-	-	-	0.42	-	0.48	-	-	-	-	-	-	-	0.57	-	-	-	-
small game	-	-	-	0.34	-	-	-	-	-	-	-	-	-	-	-	0.32	-	-	-	-
cured pork	-	-	-	-	**0.47**	0.48	-	0.34	-	-	-	-	-	-	-	0.32	-	-	-	0.33
salami/ gammon/ham	-	-	-	-	-	0.35	-	-	-	-	-	-	-	0.38	-	0.42	-	-	-	-
frankfurter	-	-	-	0.35	**0.31**	0.60	-	0.35	-	-	-	0.43	-	0.39	**0.33**	0.61	-	0.38	-	0.40
bacon, bacon cubes	-	-	-	0.49	**0.40**	0.70	-	0.38	-	-	-	0.40	-	0.56	-	0.59	-	0.39	-	0.41
smoked lamb	-	-	-	-	**0.56**	0.59	-	-	**0.49**	0.35	-	-	-	-	-	-	-	-	-	-
poultry, any	-	-	-	-	-	0.37	-	0.35	-	-	-	-	**0.34**	-	-	-	-	-	-	-
chicken in stews/breadcrumbs/pies	-	-	-	-	**0.45**	0.62	-	0.31	**0.41**	0.56	-	-	-	0.43	**0.37**	0.39	-	-	**0.31**	-
turkey in stews breadcrumbs/breadcru	-	-	-	-	**0.47**	0.62	**0.30**	-	**0.44**	0.41	-	-	-	0.44	**0.41**	0.36	-	0.30	-	-

mbs/pies																				
any smoked/cured poultry	-	-	-	-	**0.60**	0.62	-	-	-	-	-	0.32	-	-	-	-	-	-	-	-
liver/ pates	-	-	-	-	-	-	-	0.30	-	-	-	0.34	-	-	-	0.43	-	0.33	-	-
other offal (tongue/brain etc)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.54	-	-	-	-
fish, any	-	-	**0.39**	-	**0.35**	0.42	-	-	-	-	-	-	-	-	-	-	**0.38**	-	**0.43**	-
other fish	-	-	-	-	**0.49**	0.32	-	-	-	-	-	-	-	0.40	**0.35**	-	-	0.31	-	-
fresh crustaceans and molluscs	-	-	**0.41**	-	**0.33**	-	-	-	-	-	-	-	-	-	**0.39**	-	**0.38**	-	**0.34**	-
cured / smoked fatty fish	-	-	-	-	**0.50**	0.57	**0.35**	-	**0.32**	-	-	-	-	-	-	-	-	-	-	-
cured or smoked white fish	-	-	-	-	-	0.38	-	-	**0.32**	-	-	-	-	-	-	-	-	-	-	-
tinned crustaceans	-	-	-	0.33	-	-	-	-	-	-	-	-	-	-	**0.43**	-	-	-	-	-
eggs (fried/poached e.t.c)	-	-	-	0.34	-	0.35	-	-	-	0.51	**0.32**	-	**0.31**	-	-	0.42	-	0.30	-	-
milk, any	-	-	-	-	-	-	-	-	-	-	-	0.32	-	-	-	-	-	-	-	-
full fat milk	-	-	-	-	**0.42**	0.67	-	-	-	-	-	-	-	-	-	-	-	-	-	-
semi -skimmed milk	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.32	-	-	-	-
skimmed milk	-	-	-	-	-	-	-	-	-	-	-	-	**0.32**	-	-	-	-	-	-	-
yogurt, any	-	-	-	-	-	-	-	-	-	-	-	-	**0.37**	-	**0.37**	-	-	-	-	-
soya milk, any	-	-	-	-	-	-	-	-	-	-	**0.34**	-	-	-	**0.40**	-	-	-	-	-
viili yoghurt like fermented milk	-	-	-	-	**0.39**	0.40	-	0.35	-	-	-	-	-	-	**0.35**	-	-	-	-	-
tofu	-	-	-	-	-	-	-	-	-	-	**0.33**	-	-	-	**0.36**	-	-	-	-	-
cheese, any	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.40	-	-	-	-
soft cheese	-	-	-	-	-	-	-	-	-	-	**0.31**	-	-	-	**0.32**	0.53	**0.34**	-	-	-
semi hard cheese and hard cheeses(gouda emental e.t.c)	-	-	-	-	-	-	-	-	-	-	-	-	**0.36**	0.30	-	0.58	-	-	-	-
cottage cheese	-	-	**0.35**	-	-	-	**0.31**	-	**0.58**	0.40	-	-	-	-	**0.32**	-	-	-	-	-
semi hard and hard Greek cheese	-	-	**0.41**	-	-	-	-	-	**0.59**	0.47	-	-	-	-	-	0.37	-	-	-	-
fresh cheeses (e.g. mozzarella, feta e.t.c)	-	-	**0.40**	-	-	-	-	-	**0.40**	-	**0.32**	-	**0.34**	-	**0.41**	-	-	-	-	-

sour cream	-	-	-	-	**0.38**	0.38	**0.50**	-	-	-	**0.37**	-	-	-	-	-	-	-	-	-
double or clotted cream	-	-	-	-	**0.34**	0.59	-	-	-	-	-	-	-	-	-	0.41	-	-	-	-
ketchup	-	-	-	-	-	-	-	-	-	0.52	-	0.39	-	-	-	0.37	-	0.44	-	-
fresh meat or offal soup	-	-	-	-	**0.57**	0.76	-	-	**0.47**	0.41	-	-	-	-	**0.38**	0.54	-	-	-	-
moussaka	-	-	-	0.38	-	-	-	-	**0.42**	0.47	**0.31**	-	-	-	-	-	-	-	-	-

* Values are Pearson Correlation Coefficients between a food item and an identified dietary pattern. For clarity only food items that were correlated >0.30 or <-0.30 with a dietary pattern for each country were included in the table **I: fruit, fish and vegetables pattern, *** II: meat, potatoes and sweets pattern

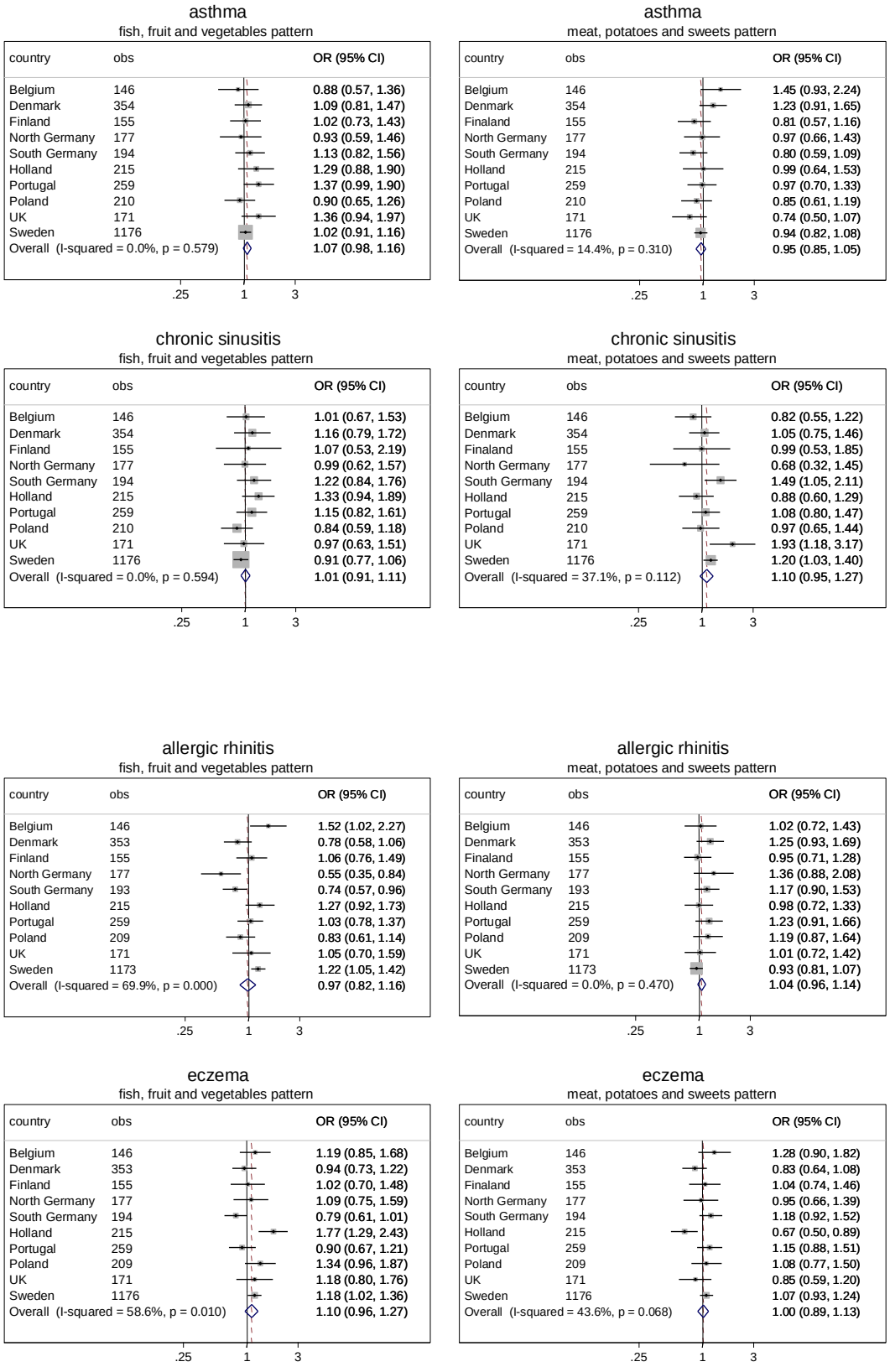
6.3.3 Dietary patterns and respiratory outcomes

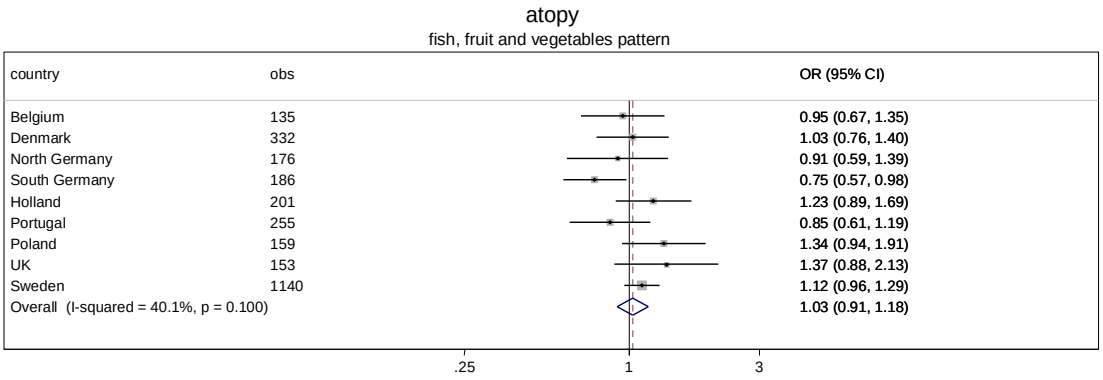
Forest plots from a random effect meta-analysis (Fig. 6.3.3.1) shows no overall statistical significant evidence that a **"fruit, fish and vegetables"** pattern was associated with asthma (P=0.13), atopy (P=0.62) chronic sinusitis (P=0.89), allergic rhinitis (P=0.77) or eczema (P=0.17), and similarly no evidence that the **"meat, potatoes and sweets"** pattern was associated with asthma (P=0.20), atopy (P=0.31), chronic sinusitis (P=0.19), allergic rhinitis (P=0.30) or eczema (P=0.98).

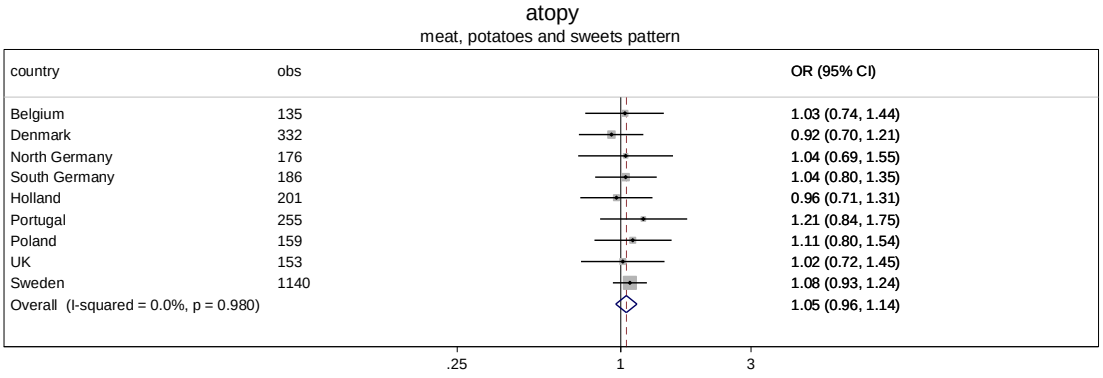
There was evidence of heterogeneity on the association of allergic rhinitis and eczema with the **"fruit, fish and vegetables"** pattern (I^2=69.9%; P <0.001 and I^2=58.6%%; P =0.01 respectively). In this case, it was evident from a visual inspection that the two German centres had qualitatively different results to the others, differences which were statistically significant; in fact, when countries were analysed separately, increased **"fruit, fish and vegetables"** pattern intake was associated with a decreased risk of allergic rhinitis in North Germany and South Germany (OR per quintile = 0.56; 95% CI 0.35, 0.84, OR per quintile =0.74; 95% CI 0.57, 0.96) and an increased risk in Sweden and Belgium (OR per quintile = 1.22; 95% CI 1.05, 1.42, OR per quintile =1.52; 95% CI 1.02, 2.27). A visual inspection of the forest plot of the association of eczema and **"fruit, fish and vegetables"** pattern showed that the Amsterdam centre had significantly different results than the other centres (OR per quintile = 1.77; 95%CI 1.29, 2.43).

There was evidence that a **"meat, potatoes and sweets"** pattern was statistically significant associated with chronic sinusitis in UK, South Germany and Sweden. Furthermore a "**fruit, fish and vegetables"** pattern was negatively associated in South Germany and North Germany and positively associated in Belgium with allergic rhinitis. A **"fruit, fish and vegetables"** pattern increased the risk and a **"meat, potatoes and sweets"** pattern decreased the risk of eczema in Holland (Figure 6.3.3.1).

Figure 6.3.3.1. Associations between the two dietary patterns and respiratory outcomes: results of meta-analyses. OR: odds ratio.




atopy
fish, fruit and vegetables pattern
country
obs
OR (95% CI)
Belgium 135 0.95 (0.67, 1.35)
Denmark 332 1.03 (0.76, 1.40)
North Germany 176 0.91 (0.59, 1.39)
South Germany 186 0.75 (0.57, 0.98)
Holland 201 1.23 (0.89, 1.69)
Portugal 255 0.85 (0.61, 1.19)
Poland 159 1.34 (0.94, 1.91)
UK 153 1.37 (0.88, 2.13)
Sweden 1140 1.12 (0.96, 1.29)
Overall (I-squared = 40.1%, p = 0.100) 1.03 (0.91, 1.18)
.25
1
3


atopy
meat, potatoes and sweets pattern
country
obs
OR (95% CI)
Belgium 135 1.03 (0.74, 1.44)
Denmark 332 0.92 (0.70, 1.21)
North Germany 176 1.04 (0.69, 1.55)
South Germany 186 1.04 (0.80, 1.35)
Holland 201 0.96 (0.71, 1.31)
Portugal 255 1.21 (0.84, 1.75)
Poland 159 1.11 (0.80, 1.54)
UK 153 1.02 (0.72, 1.45)
Sweden 1140 1.08 (0.93, 1.24)
Overall (I-squared = 0.0%, p = 0.980) 1.05 (0.96, 1.14)
.25
1
3

6.3.4 Exhaustive single food analysis and respiratory outcomes

The results from our two step exhaustive single food analysis where each of the 238 individual food items (xilopites was omitted due to limited observations) was separately entered into a random intercept logistic analysis are presented in smile plots in Fig. 6.3.4.1. To be considered significant with the false discovery rate controlled at 20%, an individual result here needs to be above the horizontal dashed line. Odds ratios of food items that were above the 20% threshold are shown in tables 6.3.4.1-6.3.4.5 for the analysis not adjusted for other foods at the first step (but adjusted for age, sex, body mass index and smoking status) and the adjusted analysis at the second step (adjusted for age, sex, body mass index and smoking status and for foods declared significant at the first step). Furthermore, our analysis was stratified by atopic status. An additional two-step ESFA was performed for the overall sample of adults controlling the FDR at 5%.

Overall, there was evidence of an association between individual food items and asthma, after taking proper account of the number of comparison- that is we set our false discovery rate to be at 20%. Specifically, for the adjusted analysis at the second step a number of vegetables (legumes, chard, lemon, radish), cherries, cured or smoked fatty fish, condensed milk, ice-cream, beans (kidney, red and black, wholemeal bread and pasta seem to decrease the risk of asthma, while couscous, turnip, ketchup, normal margarine, nut based spreads, other fish seafood, and thin biscuits (knackerbrod) seems to increase the risk (Table 6.3.4.1). Condensed milk, cured or smoked fatty fish, turnip and couscous remained statistically significantly associated with asthma, even when we stratified our analysis by non atopics and atopics or when we controlled the FDR for the overall sample at 5% (Table 6.3.4.1).

Furthermore, in our two-step ESFA (allowing the FDR at 20%) plant, vegetables and fruits (apple, chickpeas, fenugreek and cabbage), fresh white fish, raisins and sultan, sour cream, and meat stew were negatively associated with chronic sinusitis. Positive associations with chronic sinusitis were observed with table sugar, okra, pumpkin, turkey roasted or boiled, soya beans, and potatoes (mashed/boiled) (Table 6.3.4.2). Cabbage and Vili yogurt like fermented milk, okra and pumpkin remained significant in the majority of our ESFA models when we stratified for atopics, non-atopics and when we controlled the FDR of our ESFA at 5%.

Several food items such as fruits (peach, kiwi) butter, peanuts, beef burger, cured or smoked fatty fish, Greek Style yogurt, smoked game and tofu seems to be protective for allergic rhinitis (6.3.4.3). Condensed milk, rhubarb and lentils, bitter melon, crisp fried cakes, Greek style yoghurt were inversely associated with eczema (Table 6.3.4.4) for our two-step ESFA model (allowing the FDR at 20%). In the majority of the models (subgroup analysis for atopy and allowing the rate of our false discoveries to be at 5%), protective associations between peach, peanuts, and tofu and Allergic Rhinitis remained statistical significant. Similar results were observed between Bitter Mellon, Greek Style yogurt, smoked game and eczema.

Finally, atopy was negatively associated with smoked game, pancakes, parsnip, crisp fried cakes, Double clotted cream, moussaka, cured pork and positively associated with soft cheeses, legumes, white rice, beer, sweet potato, hot/cold/roast beef, broccoli, thin biscuits (knackerbrod), total sweets and couscous in our two-step ESFA (6.3.4.5). Crisp fried cakes, moussaka, beer, thin biscuits, hot/cold/roast beef remained statistical significant after controlling our rate of false discoveries at 5%.

Table 6.3.4.1. Associations between food items and asthma; odds ratio (OR) and P-values for statistical significant foods at first step and second step of Exhaustive Single Food Analysis (ESFA); ESFA (1st step) controlling the False Discovery Rate (FDR) at 20%; ESFA (2nd step) controlling the FDR at 20%; ESFA (2nd step) controlling the FDR at 20% for non-atopics; ESFA (2nd step) controlling the FDR at 20% for atopics; ESFA (2nd step) controlling the FDR at 5%.

ESFA (1st step)* FDR=20% all p-values< 0.024		ESFA (2nd step)** FDR=20% all p-values < 0.010		ESFA (2nd step)** non-atopics FDR=20% all p-values< 0.010		ESFA (2nd step)** atopics FDR=20% all p-values < 0.015		ESFA (2nd step)** FDR=5% all p-values < 0.001	
food	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**
cakes, any	0.79	refined pasta	0.81	cakes, any	0.70	**kidney and black beans**	**0.72**	**kidney (black beans)**	**0.85**
whole meal bread	0.84	**lemon**	**0.82**	small game	0.78	**lemon**	**0.75**	**lemon**	**0.85**
ice -cream	0.84	**kidney ,black beans**	**0.83**	apple spread	0.79	refined pasta	0.80	**cured or smoked fatty fish**	0.85
refined pasta	0.85	whole meal bread	0.83	freshly squeezed juice	0.81	radish	0.83	**cherries**	**0.86**
radish	0.86	**cured or smoked fatty fish**	0.85	grapefruit	0.81	**cherries**	**0.85**	legumes, any	0.89
condensed milk	**0.87**	radish	0.85	chicken boiled roasted	0.83	**cured or smoked fatty fish**	0.86	**condensed milk**	**0.89**
cured or smoked fatty fish	**0.88**	ice-cream	0.86	**condensed milk**	**0.85**	apricot	0.86	**thin biscuits (knackerbrod)**	**1.09**
chocolate(plain)	0.89	**cherries**	**0.87**	**cured or smoked fatty fish**	**0.88**	coffee	0.86	**turnip**	**1.13**
cherries	**0.90**	legumes, any	0.88	**couscous**	**1.12**	**condensed milk**	**0.87**	**couscous**	**1.17**
cottage cheese	0.90	**condensed milk**	**0.90**	single cream	1.18	chocolate plain	0.88		
nuts, any	0.91	chard	0.92	**turnip**	**1.27**	sour milk	0.88		
butter, any	0.92	other fish sea food	1.04	sunflower oil	1.27	normal margarine	1.11		
fenugreek	1.05	**thin biscuits (knackerbrod)**	**1.12**			minced beef meat	1.12		
fresh crustaceans	1.06	ketchup	1.12			fenugreek	1.12		

molluscs					
fresh white fish	1.06	nut based spreads	1.13	double clotted cream	1.13
cauliflower	1.06	normal margarine	1.13	**couscous**	**1.14**
bitter mellon	1.06	**turnip**	**1.14**	spirits	1.15
cured or smoked white fish	1.06	**couscous**	**1.21**	lentils	1.17
chard	1.07			mayonnaise	1.18
soya milk	1.08			crème fraiche	1.19
white rice	1.08			**turnip**	**1.20**
thin biscuits (knackerbrod)	**1.09**			half fat butter	1.23
herbs	1.09				
okra	1.10				
squeezed fresh fruit	1.10				
olive oil	1.10				
avocado	1.11				
fresh vegetable/cereal soup	1.12				
mango	1.13				
onion	1.13				
garlic	1.14				
courgette	1.14				
leek	1.16				
couscous	**1.19**				
turnip	**1.20**				

*Adjusted for age, sex, body mass index and smoking status. **Adjusted for age, sex, body mass index, smoking status and all the foods that were significant at the univariate ESFA (1st step) ***Bold type indicates foods which were statistically associated with asthma across 3 or more of the different procedures being used.

Table 6.3.4.2. Associations between food items and chronic sinusitis; odds ratio (OR) and P-Values for statistical significant foods at first step and second step of Exhaustive Single Food Analysis (ESFA) ;ESFA (1st step) controlling the False Discovery Rate (FDR) at 20%; ESFA (2nd step) controlling the FDR at 20%; ESFA (2nd step) controlling the FDR at 20% for non-atopics; ESFA (2nd step) controlling the FDR at 20% for atopics; ESFA (2nd step) controlling the FDR at 5%.

ESFA (1st step)* FDR=20% all p-values<0.015		ESFA (2nd step)** FDR=20% all p-values<0.011		ESFA (2nd step)** non-atopics FDR=20% all p-values<0.028		ESFA (2nd step)** atopics FDR=20% all p-values<0.019		ESFA (2nd step)** FDR=5% all p-values<0.001	
food	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**
sour cream	0.86	**cabbage**	**0.79**	rhubarb	0.57	raisins	0.66	**cabbage**	**0.87**
				custard		pickled		**viili yogurt like**	
mixed candies	0.88	raisin, sultan	0.80	cream	0.67	vegetables	0.72	**fermented milk**	**1.11**
apple	**0.88**	sour cream	0.81	**cabbage**	**0.71**	cucumber	0.73	**okra**	**1.13**
smoked lamb	1.05	**apple**	**0.81**	small game	0.77	peach	0.74	**pumpkin**	**1.14**
turnip	1.06	fresh white fish	0.84	celery	0.77	pear	0.74		
				bacon					
chapatti bread	1.06	chickpeas	0.85	cubes	0.78	fresh fatty fish	0.75		
				legumes					
cauliflower	1.10	fenugreek	0.88	,any	0.80	**cabbage**	**0.76**		
				fresh					
turkey roasted, boiled	1.10	meat stew	0.90	cheeses	0.80	fava beans	0.78		
viili yogurt like		turkey roasted,							
fermented milk	**1.13**	boiled	1.05	plum	0.81	fresh white fish	0.78		
				cottage		soda without			
table sugar	1.13	soya beans	1.10	cheese	0.81	sugar	0.79		
olive	1.14	**okra**	**1.11**	**apple**	**0.81**	**apple**	**0.79**		
aubergine	1.14	**pumpkin**	**1.11**	garlic	0.82	white wine	0.79		
okra	**1.16**	table sugar	1.11	sour cream	0.82	sweet corn	0.79		
		potatoes(boiled/							
pumpkin	**1.16**	mashed)	1.16	liver pates	0.87	chard	0.80		

carrot	1.16	salads	0.87	fresh fruit	0.80
courgette	1.16	cluster beans	1.10	garlic	0.80
olive oil	1.17	doughnuts			
white sauce	1.24	buns	1.14	small game	0.81
		pumpkin	**1.15**	fig	0.83
		okra	**1.20**	kidney black beans	0.84
		single cream	1.21	other offal	0.85
		any red meat	1.22	rice	0.85
		chocolates bars	1.24	soya beans	0.90
		white sauce	1.27	**viili yogurt like fermented milk**	1.07
				margarine, any	1.10
				okra	**1.15**
				low fat butter	1.18
				cakes	1.20
				mayonnaise	1.21
				apple spread	1.22
				juice with sugar	1.24
				full fat milk	1.25
				poultry, any	1.28
				courgette	1.33
				total sweets	1.35

*Adjusted for age, sex, body mass index and smoking status. **Adjusted for age, sex, body mass index, smoking status and all the foods that were significant at the univariate analysis (1st step) ***Bold type indicates foods which were statistically associated with chronic sinusitis across 3 or more of the different procedures being used.

Table 6.3.4.3. Associations between food items and allergic rhinitis; odds ratio (OR) and P-Values for statistical significant foods at first step and second step of Exhaustive Single Food Analysis (ESFA); ESFA (1st step) controlling the False Discovery Rate (FDR) at 20%, ESFA (2nd step) controlling the FDR at 20%, ESFA (2nd step) controlling the FDR at 20% for non-atopics, ESFA (2nd step) controlling the FDR at 20% for atopics, ESFA (2nd step) controlling the FDR at 5%.

ESFA (1st step)* FDR=20% all p-values<0.015		ESFA (2nd step)** FDR=20% all p-values<0.011		ESFA (2nd step)** non-atopics FDR=20% all p-values<0.015		ESFA (2nd step)** atopics FDR=20% all p-values<0.015		ESFA (2nd step)** FDR=5% all p-values<0.001	
food	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**
peach	**0.79**	**peach**	**0.72**	rhubarb	0.70	small game	0.84	**peach**	**0.77**
margarine, any	0.90	butter, any	0.83	asparagus	0.71	pickled vegetables	0.84	**peanuts**	**0.90**
low fat margarine	0.94	beef burger	0.84	**peach**	**0.75**	smoked fatty fish	0.85	tofu	0.92
peanuts	**0.93**	**peanuts**	**0.87**	butter, any	0.81	white wine	0.85	**mango**	**1.10**
mango	**1.05**	cured or smoked fatty fish	0.88	minced beef meat	0.82	legumes, any	1.12	**vegetable oil**	**1.17**
hot/cold roast beef, boiled beef	1.09	kiwi	0.89	kiwi	0.85	wholemeal bread	1.13	**any smoked/cured poultry**	**1.18**
any smoked/cured poultry	**1.09**	Greek style yogurt	0.90	celery	0.87	**mashed potato**	**1.14**	**single cream**	**1.25**
legumes, any	1.10	smoked game	0.91	radish	0.90	cherries	1.18		
cured pork	1.04	tofu	0.91	plum	0.91	tinned fatty fish	1.25		
spinach	1.06	**mango**	**1.05**	cured pork	1.12	double clotted cream	1.26		
pumpkin	1.09	skimmed milk	1.09	potato tortilla	1.14				
soya milk	1.12	soya milk	1.10	**any smoked/cured poultry**	**1.15**				
mashed potato	**1.13**	**vegetable oil**	**1.15**	potato dump	1.20				

artichoke	1.12	**any smoked/cured poultry**	**1.15**	wholemeal bread	1.20
wholemeal bread	1.16	**single cream**	**1.24**	soda with sugar	1.21
vegetable oil	**1.19**			prune	1.21
single cream	**1.21**			soya milk	1.21
				mashed potato	**1.23**

*Adjusted for age, sex, body mass index and smoking status. **Adjusted for age, sex, body mass index, smoking status and all the foods that were significant at the univariate analysis (1st step). ***Bold type indicates foods which were statistically significant associated with allergic rhinitis across 3 or more of the different procedures being used.

Table 6.3.4.4. Associations between food items and Eczema; odds ratio (OR) and P-Values for statistical significant foods at first step and second step of Exhaustive Single Food Analysis (ESFA); ESFA (1st step) controlling the False Discovery Rate (FDR) at 20%, ESFA (2nd step) controlling the FDR at 20%, ESFA (2nd step) controlling the FDR at 20% for non-atopics, ESFA (2nd step) controlling the FDR at 20% for atopics, ESFA (2nd step) controlling the FDR at 5%.

ESFA (1st step)* FDR=20% all p-values<0.009		ESFA (2nd step)** FDR=20% all p-values<0.015		ESFA (2nd step)** Non-atopics FDR=20% all p-values<0.014		ESFA (2nd step)** atopics FDR=20% all p-values<0.015		ESFA (2nd step)** FDR=5% all p-values<0.001	
food	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**	**food**	**OR**
crisp fried cakes	**0.86**	rhubarb	0.84	chips/ French fries	0.74	**crisp fried cakes**	**0.73**	rhubarb	0.83
bitter melon	**0.90**	**crisp fried cakes**	**0.86**	wholemeal bread	0.77	cottage cheese	0.80	crisp fried cakes	0.84
pizza	0.90	**bitter melon**	**0.89**	rhubarb	0.79	fresh white fish	0.83	**bitter melon**	**0.88**
Greek style yogurt	**0.91**	**Greek style yogurt**	**0.90**	cashew nuts	0.81	pickled vegetables	0.87	**Greek style yogurt**	**0.89**
kneipp bread	0.92	lentils	0.93	full fat milk	0.84	dressing sauces	0.87	**game, other**	**0.92**
game, other	**0.94**	**game, other**	**0.94**	courgette	0.88	red meat, any	0.88	lentils	0.93
tomato	1.10	condensed milk	0.94	wholemeal pasta	0.89	condensed milk	0.89	sour cream	1.10
wholemeal bread	1.12	turkey roasted/boiled	1.11	meat pies	0.89	viili yogurt like fermented milk	0.90		
turkey roasted/boiled	1.13	eggs all	1.12	chard	1.09	**game, other**	**0.90**		
sour cream	1.13	sour cream	1.12	brown wholemeal	1.09	chard	1.12		
honey	1.14			double clotted cream	1.14				
vegetable oil	1.17			casserole	1.15				
olive oil	1.19			Greek coffee	1.16				

potatoes	1.17
olive oil	1.19
prune	1.24
bread, any	1.25

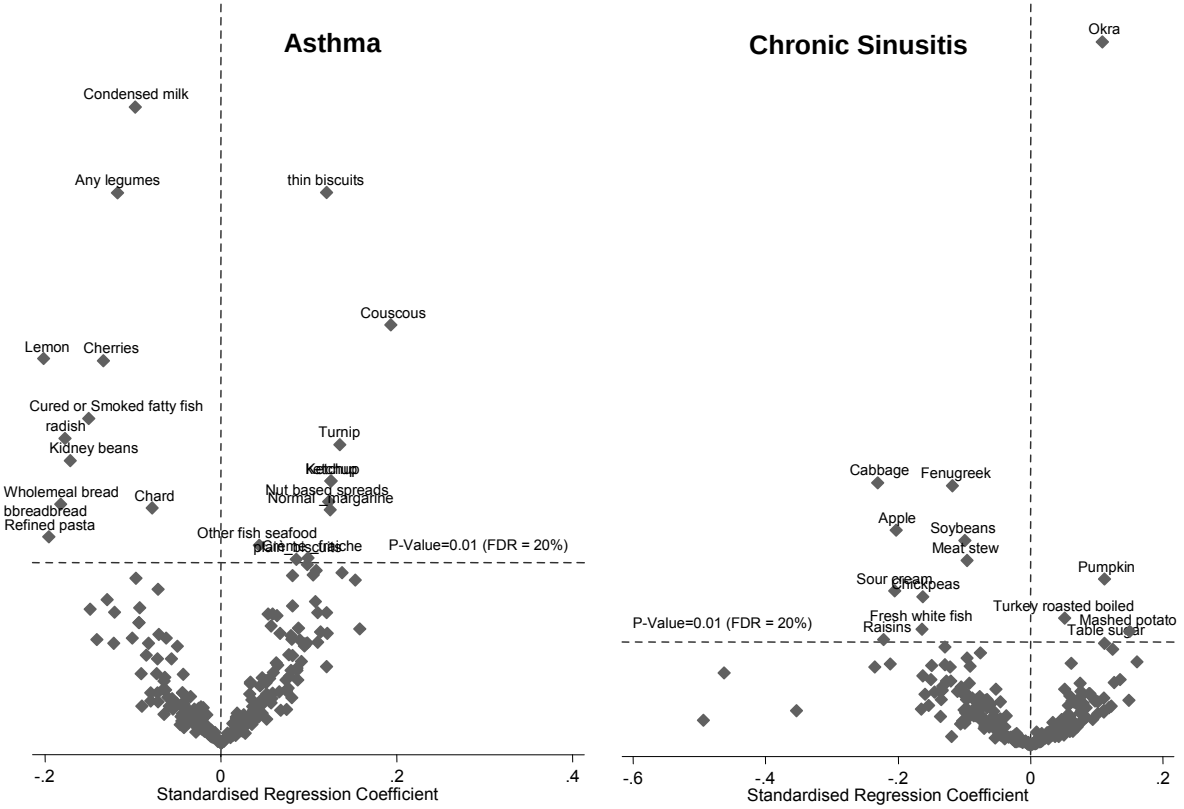
*Adjusted for age, sex, body mass index and smoking status. **Adjusted for age, sex, body mass index, smoking status and all the foods that were significant at the univariate analysis (1st step). ***Bold type indicates foods which were statistically significant associated with eczema across 3 or more of the different procedures being used

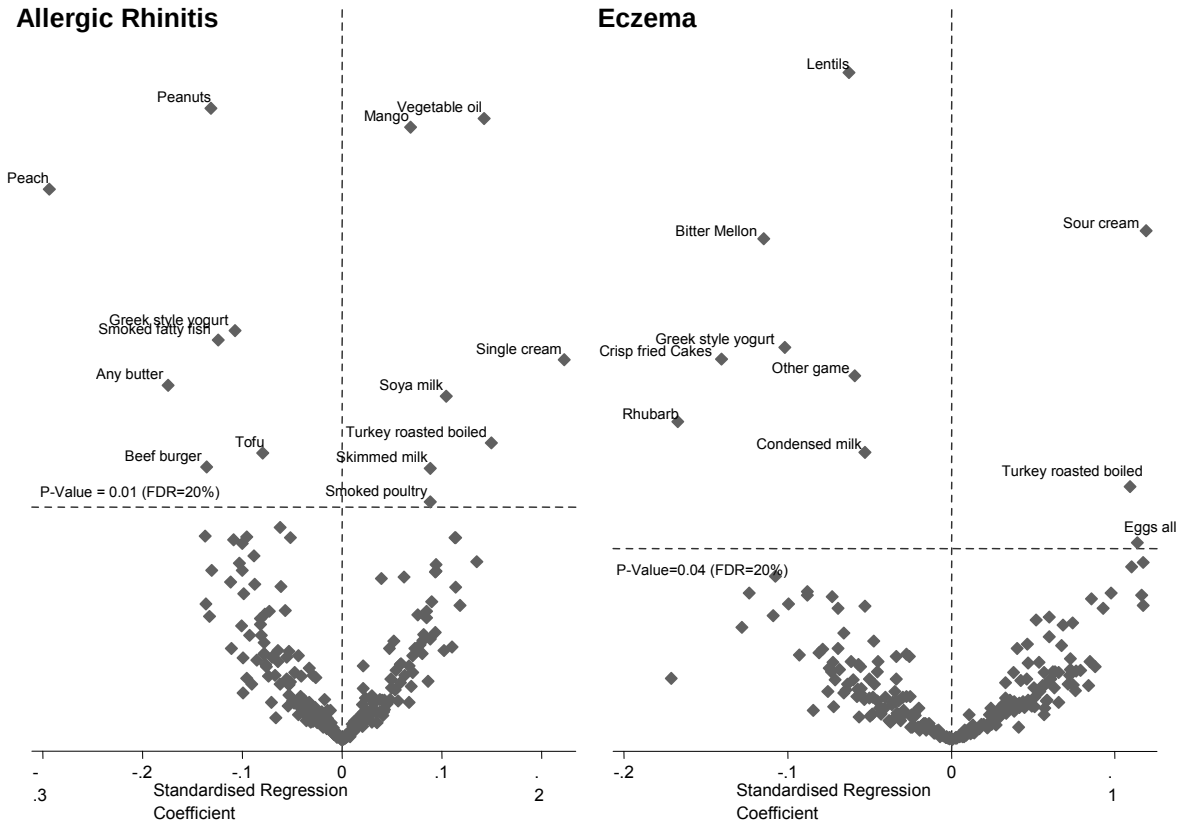
Table 6.3.4.5. Associations between food items and Atopy; odds ratio (OR) and P-Values for statistical significant foods at first step and second step of Exhaustive Single Food Analysis (ESFA); ESFA (1st step) controlling the False Discovery Rate (FDR) at 20%, ESFA (2nd step) controlling the FDR at 20%; ESFA (2nd step) controlling the FDR at 5%.

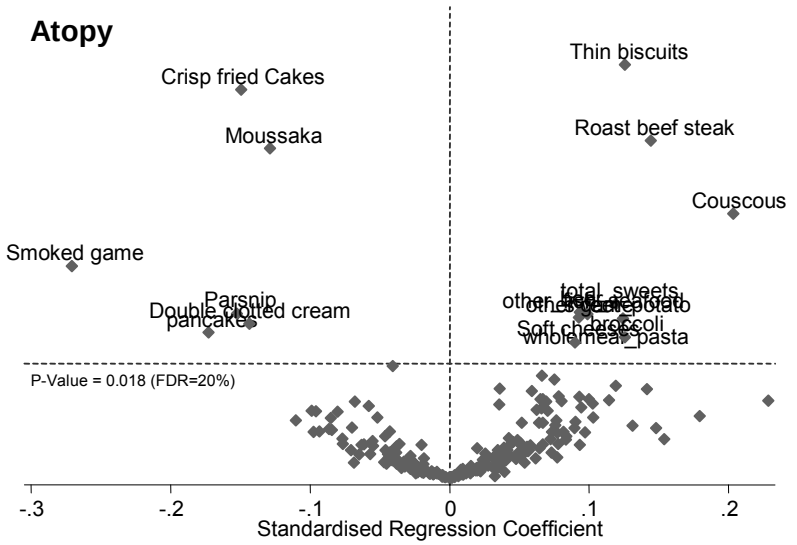
ESFA (1st step)* FDR=20% all p-values<0.015		ESFA (2nd step)** FDR=20% all p-values<0.011		ESFA (2nd step)** FDR=5% all p-values<0.001	
food	**OR**	**food**	**OR**	**food**	**OR**
parsnip	0.83	smoked game	0.81	**crisp fried cakes**	**0.86**
double or clotted cream	0.84	parsnip	0.85	**moussaka**	**0.88**
pancakes	0.85	pancakes	0.85	**beer**	**1.10**
carrot	0.86	**crisp fried cakes**	0.86	**thin biscuits (knackerbrod)**	**1.13**
crisp fried cakes	0.88	double or clotted cream	0.87	**hot cold roast beef/ boiled beef**	**1.14**
moussaka	0.88	**moussaka**	0.87		
water ice	0.90	cured pork	0.91		
crème fraiche	0.90	legumes, any	1.07		
turnip	0.92	white rice	1.08		
sweet potato	1.08	soft cheeses	1.09		
soft cheeses	1.08	**beer**	1.09		
thin biscuits (knackerbrod)	1.10	sweet potato	1.10		
beer	1.11	**hot cold roast beef/ boiled beef**	1.11		
total sweets	1.11	broccoli	1.11		
hot cold roast beef/ boiled beef	1.14	**thin biscuits (knackerbrod)**	1.12		
any red meat	1.15	total sweets	1.13		
white rice	1.22	couscous	1.20		

*Adjusted for age, sex, body mass index and smoking status. **Adjusted for age, sex, body mass index, smoking status and all the foods that were significant at the univariate analysis (1st step). ***Bold type indicates foods which were statistically significant associated with eczema across 3 or more of the different procedures being used.

Figure 6.3.4.1. Association of foods with asthma, chronic sinusitis, allergic rhinitis and eczema: smile plot showing P-value against standardized logistic regression coefficient for different foods. Points on the right represent positive associations; points on the left represent negative associations. To be considered significant, controlling the false discovery rate at 20 %, points must lie above the dotted line.




Allergic Rhinitis
Peanuts
Vegetable oil
Mango
Peach
Greek style yogurt
Smoked fatty fish
Single cream
Any butter
Soya milk
Turkey roasted boiled
Tofu
Beef burger
Skimmed milk
P-Value = 0.01 (FDR=20%)
Smoked poultry
-.3
-.2
-.1
0
.1
.2
Standardised Regression Coefficient
Eczema
Lentils
Bitter Mellon
Sour cream
Greek style yogurt
Crisp fried Cakes
Other game
Rhubarb
Condensed milk
Turkey roasted boiled
Eggs all
P-Value=0.04 (FDR=20%)
-.2
-.1
0
.1
Standardised Regression Coefficient


Atopy
Thin biscuits
Crisp fried Cakes
Roast beef steak
Moussaka
Couscous
Smoked game
Parsnip
Double clotted cream
pancakes
Soft cheeses
broccoli
whole meal pasta
P-Value = 0.018 (FDR=20%)
-.3
-.2
-.1
0
.1
.2
Standardised Regression Coefficient

7 Discussion and Conclusion

"I'm astounded by people who want to 'know' the universe when it's hard enough to find your way around Chinatown."

Woody Allen

7.1 Purpose

The purpose of this chapter is to:

- Interpret the results of our systematic review
- Interpret the results of our simulations
- Interpret the findings from our data analysis in the G.A^2.L.E.N study
- Identify and address potential limitations of our studies
- Present ideas for potential future research

7.2 Key points that are discussed in this chapter

- Methodological advantages of PCR that it can capture additive and interactive effects of diet on disease and it can solve the confounding issues between dietary exposures better than analysing any single food and nutrient are not critically evaluated.
- Application of Principal Components Analysis (PCA) of diet in nutritional epidemiology requires a lot of arbitrary and subjective decisions.
- Results of Principal Components Regression (PCR) studies are contradictory, inconsistent, and difficult to be interpreted and translated into an intervention.
- An Exhaustive Single Food Analysis (ESFA) is a better method at identifying combination of foods that are causally linked to disease.
- Even when a population Principal Component of diet was causally linked to disease, PCR could not outperform ESFA in identifying the combination of foods that were causally associated with disease.
- PCR is not better at detecting an additive effect of diet on disease than an ESFA.
- PCR doesn't resolve aspects of confounding and mulitcollinearity between food intakes better than an ESFA. An ESFA method adjusted for all food intakes which were significant in the unadjusted ESFA, controls rate of false discoveries to an acceptable level.
- Dietary patterns empirically derived with the use of PCA were not associated with respiratory outcomes in the G.A^2.L.E.N data-set.
- An ESFA method adjusted for all food intakes which were significant in an ESFA not adjusted for other foods (but adjusted for age, sex, bmi and smoking status) provide evidence of associations between a combination of foods and our respiratory outcomes in the G.A^2.L.E.N data-set.

- In the G.A^2.L.E.N analysis, diet- disease associations presented by ESFA adjusted for all food intakes which were significant in an ESFA not adjusted for other foods (but adjusted for age, sex, bmi and smoking status), could not be all causal and were difficult to be interpreted. Confounding still remains an important issue in our ESFA method and needs more careful consideration.

7.3 Interpretation of results from our Systematic Review

As was mentioned in Chapter 2 recent results have raised questions concerning the role of diet in the aetiology in certain chronic diseases and together with the challenges of analysing diets this has led to an explosion in the use of PCA. Ultimately, as was mentioned in the introduction, whether PCA can effectively detect associations between a combination of foods and disease across diverse populations (section 7.3.1), whether it is an objective tool of analysis (section 7.3.2) and whether this can be translated into an intervention (section 7.3.3) are three important indicators of the validity of the method and hence of their utility in epidemiological and clinical research.

7.3.1 Diet and disease associations with the use of posterior dietary patterns identified by PCA

Cardiovascular disease, obesity and hdl–cholesterol were consistently negatively associated with a pattern loosely defined by a high consumption of fruit, fish and vegetables. On the other hand, high consumption of a pattern consisting of red meat, chips and dairy products increased the risk.

However, there were major inconsistencies in the associations between dietary patterns and disease among the majority of the studies in our systematic review. Evidence regarding the association of disease outcomes (which appeared in more than one study) and dietary patterns were inconclusive about the beneficial or harmful effects of patterns in relation to adenoma, five different types of cancers, type II diabetes, respiratory symptoms, metabolic syndrome, overall mortality and myocardial infarction. For example, breast cancer risk was inversely associated (Hirose et al., 2007, Agurs-Collins et al., 2009), not associated (Fung et al., 2005, Kroenke et al., 2005, Robinson et al., 2004), and positively associated (Murtaugh et al., 2008) with a loosely named **"prudent"** pattern (Table C and D).

Inconsistencies may be partly explained by differences in sex and age since dietary patterns were highly associated with them (see paragraph 3.2.9) in diverse populations (see paragraph 3.2.2) or by the fact that PCA-derived components may be not explaining as much variation as possible in important nutrients that are potentially associated with disease risk. However,

these inconsistencies could be attributed to methodological issues such as that dietary patterns identified with the use of PCA are implicit measures of lifestyle and are not reproducible.

Dietary patterns confounded by lifestyle factors

First, as we can observe from paragraph 3.2.9, Table C and D, patterns associated with a high intake of **fruit, vegetables, and fresh fish** has been associated with elements of lifestyle such as vitamin and supplement use, income, social class, strenuous exercise, education, partnership status and non-smoking status which are considered to lead to a healthier life. On the other hand, individuals who consume patterns correlated highly with **meat, chips, fatty and sugary foods** are more likely to lead a less healthy lifestyle (e.g. non-exercisers, smokers and drinkers). So, failure of dietary pattern analysis with the use of PCA to consistently associate combination of foods with disease across diverse populations could be due to confounding and mulitcollinearity with cultural or lifestyle factors. This is also highlighted by the fact that in several studies the dietary pattern and health association was markedly attenuated by control for confounders (Kant, 2004, Ambrosini et al., 2008, Mannisto et al. 2005, Shaheen et al. 2009).

Reproducibility

Second, there is a vague language on the labelling of dietary patterns and the food items that constitute them which could produce spurious reproducibility when reporting.

Decisions on the label appear to be equally influenced by food items and by prevailing notions of diet and disease associations, which could be different from one place to another. For example , a **"prudent"**, **"healthy"**, **"health conscious"**, **"heart healthy"**, **"healthy" "healthy/Mediterranean"**, **"vegetables, fruit and fish"** or **"vegetarian"** pattern and generally patterns which were correlated highly with different types of fish, fruit and vegetables were reported in all PCA studies which used food items or food groups . These patterns were mainly characterised by high intakes on food items that are considered as beneficial for human health. A **"Western"**, **"Western-like"**, **"Western-type"**, **"processed"**, **"junk food"**, **"junk"**, **"fast food"** or **"unhealthy"** and in general patterns which were correlated highly in **high fat** and **processed foods**, **red meat** and **chips** were observed in the majority of the populations being studied and were characterised by the items which are considered to have a harmful effect on human health (Table C and D).

However, each of these so-named healthy or unhealthy patterns derived from PCA differed from each other to greater or lesser degree on the combination of foods that were singled out (Table D). For example, people who had higher intakes of a **"prudent"** pattern in one study had higher intakes of cruciferous and other vegetables, fruit, whole grains, low-fat dairy products, cereals, beans, fish, and poultry (Agurs-Collins et al., 2009) and in another study had a higher intake of tea, fruits, butter, vegetable oil, breakfast cereals, fish and seafood and reduced-fat products (Kesse-Guyot et al., 2009). In similar manner, a **"Western"** pattern was constituted of high fat dairy, eggs, processed meat, refined grains, meat, potato, fried foods (Imamura et al., 2009) but also of pasta with meat, beef taco, beef burrito, pizza, meatballs, hamburger, fried potatoes, baked potatoes, mashed potatoes, pancake, bagels in another one (Wu et al., 2009) (Table D).

One of the main reasons for these inconsistencies is that principal component loadings differ from one study to another, so different food items contribute differently to the nutrient composition of the pattern. For example, although cereals are a consistent element of a **"prudent"** pattern in the majority of studies in our review its principal component loading is 0.28 (Agurs-Collins et al., 2009) or 0.39 (Kesse-Guyot et al., 2009) or 0.44 (Murtaugh et al., 2007) or 0.13 (Crozier et al., 2006). Hence, contribution of cereals to the label of a **"prudent"** pattern depends solely on an arbitrary cut-off point of the principal component loading of a food on a pattern. More importantly, cereals contribute differently to the nutrient composition of each of these **"prudent"** patterns (Table D).

In conclusion, appearance of qualitatively "similar" dietary patterns in different studies could be the reason of inconsistent and potentially misleading information on associations between diet and disease across studies. Reasons for these inconsistencies may be more apparent if studies were reporting dietary patterns quantitatively and not qualitatively.

In addition, there are several methodological considerations which are raised with the use of PCA, which call for the subjective judgment of the researcher. In the next section, a detailed discussion will be provided.

7.3.2 Methodological considerations of dietary pattern analysis with the use of PCA

There are several methodological issues concerning principal component analysis as applied in nutritional epidemiology. Decisions made by the researchers for preparing the data before entering the PCA and when they employ PCA involve subjectivity and have an impact on the number, type and label of patterns that are derived, reported, and analyzed. Several studies in this field have provided useful commentary on the issue of subjectivity and other methodological considerations (Newby & Tucker, 2004, McCann et al., 2001b, Slattery & Boucher, 1998). In this paragraph, we will discuss the results from these studies along with the results from our systematic review of Chapter 3, and provide some suggestions of how to apply PCA in nutritional epidemiology.

Preparation of data before entering the PCA

Specifically, as presented in section 3.2.3 several studies decided to group the number of food items from the Food Frequency Questionnaire being used into predefined food groups before beginning the pattern analysis. Some of the food items were of such limited number (e.g. < 20) (Takaoka & Norback, 2008, Igbal et al., 2008) that further collapsing was not necessary. As in Newby (2004), in our systematic review, food grouping schemes depended on a priori knowledge and on the research hypothesis under investigation.

Mc Cann (2001) (McCann et al., 2001a) and Costacou (2003) (Costacou et al., 2003) stated that a smaller number of input variables included in the PCA procedure could affect the principal component solution and explain a greater percentage of the variance intake compared with a larger number of input variables in the same study. However, this conclusion is incorrect. This was highlighted from our systematic review and Newby (2004), since studies didn't provide different results on the number of patterns and even more importantly to the total variance being explained when they grouped food items before the application of PCA (Table B).

Specifically, if you aggregate, you inevitably explain a smaller percentage of the original variance with the same number of dietary patterns. McCann found that the percentage of variance in the aggregated variables was greater, but the problem is that by aggregating you throw away a percentage of the original variance. Furthermore, in this paper, the authors found that although the number and type of dietary patterns did not change the relationship

between the dietary patterns and cancer risk was substantially attenuated when using broad food groups, suggesting that greater detail in food groupings is important. Therefore retaining a large number of foods and reducing the number of subjective decisions required seems preferable especially since the multivariate statistical approaches are actually data reduction techniques. Few studies that collapsed food items to food groups identified patterns that explained a percentage of total variance much higher (>55%) above the median (median: 24%, IQR:19.9-31.3) (Ambrosini et al., 2008, Romaguera et al., 2008, Panagiotakos et al., 2007b). So, if there is not a strong clinical hypothesis to group the food items, doesn't seem to be the need to collapse food items to food groups before entering the PCA.

In Table B, the scale of the FFQ data had a range between 5 and 11 points (median value was 7). Willet (1998) suggested that most investigators used a range of an FFQ between 5 and 10 and that five choices are likely to be too few and will usually result in a serious loss of information. Furthermore, as a rule of thumb, a scale of ten (never, once a month or less, 2-3 times per month, once per week, 2-4 times per week, 5-7 times per week, over one per day, 2-3 times per day, 4-6 times per day, over 6 times per day) for the FFQ seems reasonable because it provides detailed description on the high frequency end. Nevertheless, a food consumed less than once per week makes relatively little contribution to nutrient intake (Willet, 1998). Moreover, as mentioned in Chapter 2, PCA is a linear procedure which provides more accurate estimates when variables are approximately continuous and normally distributed, so a detailed scale of 10 items for the FFQ is more appropriate than smaller scales.

In observational studies in nutritional epidemiology it is advisable to adjust for energy intake (Willet 1998, Jakes et al, 2004), as this may itself be a risk factor, may alter the effect of a food or a nutrient on the potential outcome, or may reflect the variation of nutrient intake between individuals. In our systematic review, a few studies adjust for energy intake using the residual method as advised by Willett (1998) (Table B). However, Northstone et al. 2008 (Northstone, Emmett & Rogers, 2008) demonstrated that it is not necessary to adjust for energy intake before entry into a PCA analysis to determine dietary patterns when using food frequency questionnaire data and that effects of total energy intake can be estimated at a later stage in the analytical process. In addition, in another study (DiBello et al., 2008) adjustment of the input variables for energy intake by the residual method seemed not to affect the factor

solution. This agrees with results from an older study (Balder et al., 2003), where the principal component solution was not substantially affected by energy adjustment.

Decisions on how to derive a realistic number of dietary patterns with the use of PCA

The majority of the studies in our systematic review and in the review by Newby (2004) claimed that they derived a number of dietary patterns by

i) Examining the percentage of total variance that the dietary patterns explain in the original dataset. As mentioned in paragraph 3.2.7, the majority of the studies in our systematic review derived from 2 to 10 (median: 3, IQR: 2-4) dietary patterns which altogether explained a percentage of total variance with a median value of 24%. The majority of these studies explain around 20% or 30% of the total variance (Table C), so the subset of principal components that is chosen contains little information about the variance of the original food intake variables. It is crucial to know how small percentage of total variance can be taken as serious information loss, and it is something that it needs further research in PCA studies of nutritional epidemiology.

ii) Employing the Kaiser criterion, that is retaining all the dietary patterns with corresponding eigenvalues above 1. However, as mentioned in paragraph 2.6 in nutritional studies, a number of studies that decided the number of patterns based on the cut-off point for eigenvalues used a different cut-off point (median value:1.6, IQR:1.25-2) in order to retain a small and interpretable number of dietary patterns.

iii) A screeplot.

Our systematic review shows that the number of dietary patterns derived with the use of PCA are based on arbitrary notions of how meaningful and interpretable is a dietary pattern rather than to any of the other objective criteria mentioned above. Interpretation of the dietary pattern could be aided with the use of component loadings, component correlation coefficients and the appropriate method of rotation. However, again these methods rely on the subjective judgment of the researcher.

For example, Lutsey (2008) examined associations between dietary intake and metabolic syndrome after evaluation of the eigenvalues and their interpretability. All values > 2.0 were retained, resulting in a 2-factor solution and principal component loadings with absolute

values ≥ 0.20 were shown in a table. If a lower cut-off point for eigenvalues was used, a larger number of dietary patterns would be derived. Furthermore labeling of the patterns would be different if a higher cut off point was selected for principal component loadings; **"prudent"** pattern would be characterized from fewer vegetables and not from yogurt.

Moreover, from our analysis in paragraph 6.3.2.1 if we strictly derived our dietary patterns according to the criteria mentioned above which are usually being applied in PCA studies then

- 129 dietary patterns should be derived if we wanted dietary patterns to explain a percentage of total variance of the original food items > 80%, (data are not shown)
- 69 dietary patterns should be derived if we wanted to retain all eigenvalues >1 (data not shown).
- 2, 6, 9 or 11 dietary patterns should be derived according to the screeplot

Label of our dietary patterns will also be different if

- We used a cut-off point of >0.5 or <-0.5 for our component correlations, our pattern could be easily consisted only of 3 fruits and 2 vegetables and labeled differently (see table 6.3.2.2.1)
- We used a different method of rotation than varimax; this had an effect on our results since slightly different patterns were derived. However, rotation of the components doesn't lead to different dietary patterns (Bountziouka et al. 2012)

Standardisation of PCA to improve comparability between studies

Overall, our systematic review highlights the fact that it is the investigator's decision of how many dietary patterns should be derived and what food items are considered to be highly correlated with them in order to be interpretable. Furthermore, there are differences between the studies of how they are reporting empirically derived dietary patterns. This inherent methodological subjectivity of PCA and the lack of detailed reporting in every step of the application of the method in the literature is an important reason for observing inconsistencies between the studies. Improved reporting of PCA and additional

methodological studies on the method, as applied in nutritional epidemiology, will help to decrease the impact of subjectivity on the findings. Investigators must report analytically all decisions that were made in each stage of the analysis; beginning with the grouping and managing of the dietary data, with how patterns are presented and analyzed and how this affects the interpretation of our results and study findings. Some suggestions for directions of how the PCA method should be applied in nutritional epidemiology are summarized below with the use of Table E at the end of the thesis.

Food frequency questionnaires were used in 93.2% of the papers identified from our systematic review. Our view is that this is due to the fact that FFQ are easy to administrate and inexpensive survey instruments. However, FFQs are prone to measurement error and this need to be taken into account before or after the principal component analysis when using the instrument (Kipnis et al., 2008). Grouping food items into food groups is reported in the 28% of the studies. Our view is that by grouping several food items into food group's it i) increases the number of subjective decisions during the application of PCA especially since the multivariate statistical approaches are actually data reduction techniques, and ii) more importantly leads to information loss or underestimation of the results (McCann et al., 2001b). Although median number of the scale of the FFQ was 7, using more sensitive scales in the FFQ instrument could give a better description of diet in a continuous scale and a better approximation of population diet, since PCA was originally developed for the normal multivariate distribution and samples from it. A scale of 10 as a rule of thumb in the FFQ instrument is suggested in order variation of diet consumption of individuals to be captured (Willet et al, 1998). Median value of the items included in the FFQ instrument was 92. Our view is that food frequency questionnaire should comprise as many food items as possible with adequate contribution to nutrient intake (McCann et al., 2001b) to examine adequately the food consumption of the population under study. Our systematic review presented a small number of studies that converted food frequency data into grams/day (13.4%) and fewer studies standardised their food intake data (4.2%). The way that the food frequency data are converted and which food composition tables were used is essential for the dissemination of the results. In addition, since FFQ dietary data are highly skewed and could express different measurements of food item intake, (e.g. bread could be measured in grams/per day and beer in pints /per day), standardisation of the food intake variables is required and is strongly advised (also see paragraph 2.4.4). Very few studies reported adjustment for energy intake

(3%) with residual method before the PCA. Our view is that because total energy intake could potential alter the results between the outcome and the exposure (Willet et al., 1998) should be included as a confounder in the analysis.

Scree plot provides an objective tool for deciding on the number of Principal components or dietary patterns to be derived. Other criteria's, such as the total percentage of variance being explained should be above 80% and eigenvalues should be above one, could result in a large number of dietary patterns which could be difficult to be interpreted. However, investigators should always report the total variance being explained by the dietary patterns, since as mentioned above a small percentage could be considered as serious information loss. Principal components were rotated in the majority of the studies (55%). Rotation method doesn't seem to affect dramatically the derived dietary patterns; however different methods should be carried out in the studies and not only varimax for validity purposes. In addition, studies should always report if their method of rotation results into uncorrelated components or not (i.e. varimax rotation introduces correlation between the components to enhance interpretability). Using a cut-off point of 0.3 for principal component loading is not an objective tool for labeling the dietary patterns. This arbitrary cut-off point is derived from the fact that moderate and high correlations between variables are defined with a value higher than 0.3. Investigators should always examine lower or higher cut off points for labeling their dietary patterns. In addition, inconsistencies in the findings may be reduced if studies report dietary patterns quantitatively and not qualitatively.

Validation studies that include a reproducibility component are warranted for both methods examining the identified dietary patterns in a sub-sample of the original sample as well as using confirmatory factor analysis if there is a strong a-priori hypothesis could support the original findings of PCA. However more validation methods should be conducted in the studies (only 24% of the studies validated the reproducibility of their dietary patterns). Finally, additional research is needed on how decisions at the different stages of PCA could affect the dietary pattern solution.

7.3.3 Translation of posterior dietary patterns identified by PCA into an intervention

The relative importance of dietary patterns contributing to the effect of the overall dietary score cannot be ascertained without examining components included in the score. For example, the US Department of Agriculture has proposed the Health Eating Index (HEI) as

an index for monitoring dietary quality in the United States (Kant, 2004). However, a score based on several components makes it impossible to determine which areas of the diet need attention. Individual components contributing to the pattern must be examined to determine which ones should be targeted for intervention. PCA provides a similar score of a pattern based on the individual's consumption on a variety of foods and this score is associated with several health outcomes. However, it is unclear whether foods that are singled out from the PCA method and contribute more to the dietary pattern could make altogether a good dietary intervention.

7.3.4 Conclusions

Dietary patterns empirically derived with the use of PCA are difficult to replicate across studies because of a varying number, range, and type of variables estimated from a variety of dietary measurement methods before entered into the analysis procedure. Similarly, differences in the statistical analytical decisions and in the labelling of patterns across studies lead to incomparability as we observed in the previous sections. Qualitative and quantitative results from our systematic review present all the major shortcomings that PCA has in nutritional epidemiology, and provide strong evidence of support to our fundamental PhD assumption that less complicated methods than PCA could be equally or more effective at detecting diet-disease associations and provide better guidance for designing clinical trials and improve comparability of the findings between studies.

7.4 Interpretation of results from our simulation study

A simulation study was designed to evaluate the usefulness of PCA in dietary pattern analysis and PCA was compared to an ESFA procedure.

In some scenarios ESFA had greater power than PCA to detect an association of diet with disease. Allowing for multiple testing using a Benjamini and Hochberg approach, ESFA also typically had higher power and lower FDR for identifying the combination of foods that were causally linked with disease than a PCA in which combination of foods were singled out if they correlated highly (>0.3 or <-0.3) with a significant dietary pattern. However, unadjusted ESFA and PCA had an uncontrolled false discovery rate, which increased with increasing power. This was the result of confounding between foods, many of which were highly correlated with each other. The false discovery rate of ESFA could be controlled at a fixed

level by adjusting for foods that were significant in an unadjusted ESFA, with surprisingly good power.

Even when a simplified “Western” dietary pattern was the real culprit, PCA could not outperform ESFA in reconstructing the foods that were linked with disease. Allowing for multiple testing, ESFA had greater power and similar false discovery rate with PCA for identifying combination of foods which constitute the simplified “Western” dietary pattern that was causally associated with disease. False discovery rates were extremely high for both methods. However, again ESFA could be controlled at a desirable fixed level of 20% by adjusting for foods that were significant in an unadjusted ESFA. These findings were replicated in two different FFQ data-sets.

Slattery et al. 1998 (Slattery & Boucher, 1998) concluded that the soundness of using PCA to identify eating patterns would be better understood when more epidemiologists had begun to use the method and when a thorough understanding of the individual data elements has been obtained. However, there are no clear conclusions to be drawn on the usefulness of the method from research conducted during the last 13 years as Newby (2004) and our systematic review presented. One of the main reasons is the vague language used to justify PCA.

As mentioned above, one of the main justifications that researchers have used is that dietary patterns analysis with the use of PCA allows for the examination of the additive effects of dietary exposures which are too small to be detected by their own, or are diluted by confounding (Jacques & Tucker, 2001, Randall et al., 1990). We would formalise these views in two ways - a "weak" sense and a "strong" sense.

The “weak” sense would be that PCA as a method is more likely than analysis of single foods to detect an additive effect of diet if there is one. As we observed in tables in paragraph 5.2 by estimating the power with which ESFA and PCA could detect *any* association between diet and disease, this was not the case. ESFA had better power in detecting *any* statistically significant effect between diet and disease, in 85% of our cases.

The “strong” sense would be that if disease is associated with a combination of intake of W and intake of X and intake of Y and intake of Z, then we are more likely to find effects of some or all of W, X, Y and Z by doing a PCA than by detecting this combinations of foods with the help of ESFA. To examine this, we estimated the power and false discovery rate

with which ESFA and PCA could detect specific combinations of foods which are causally linked to disease. As we can observe from the tables in paragraph 5.3, ESFA was more effective (higher power and lower or similar FDR) at detecting associations in this "strong" sense than PCA allowing for multiple testing and adjusting for energy intake in 93% of our cases.

It is common to try and control the FDR at a low level (Benjamini & Hochberg, 1995). We have used a nominal FDR of 20%; in genetic studies, where the use of FDR is well-established, FDRs between 5% and 20% are recommended depending on the circumstances. This corresponds to a search for candidate-gene effects requiring further replication, rather than for definitive evidence (Benjamini & Yekutieli, 2005). Note that 20% is still well below the FDR of individual hypothesis testing using $P<0.05$ as a cut-off (Ioannidis, 2005). However, it is concerning that the observed FDRs of ESFA (nominally controlled at 20%) and of PCA both increase in an uncontrolled fashion as the sample size and power increase (Tables in paragraph 5. 3). It is worth noting that when one in seven foods are causally linked with disease, as here, an FDR of around 86% would be achieved by selecting "significant" foods entirely at random. The uncontrolled FDR occurs because all food intakes are correlated to some extent with the causal foods, leading to false positive findings (more so as power increases). More formally, the general conditions of Benjamini & Hochberg procedure are not met, and therefore we cannot expect the FDR to be controlled at 20%. In fact, as the sample size and power increases, it becomes easier and easier to detect all these indirect effects of other foods that are associated with the foods that have a causal effect, so we get more and more significant results, and more of these are false discoveries, because they are only indirect effects. So the FDR gets higher with increasing sample size, up to the maximum value it can take. We tried a variety of approaches to control for other foods in ESFA, and found that the FDR could be successfully controlled at the nominal level by adjusting for foods which were significant in a univariate analysis. To achieve given power, this required around twice the sample size of an unadjusted ESFA (with its inflated FDR), and four times the sample size from our original calculation, *i.e.* for an unadjusted analysis with criterion $P<0.05$. The problem with this adjustment is that with small sample sizes, the method won't always be able to make the correct adjustment. As the sample size increases we get more power to detect the things we should be adjusting for, so the adjustment becomes more effective. However, there are two competing effects: firstly, as sample size increases the FDR

tends to increase, because we get more power to detect indirect effects of foods; secondly, as sample size increases the adjustment becomes more effective, so the FDR starts to be controlled at the desired level. This is why we see FDRs increasing above 20% for lower sample sizes of 300, 600 and 1200 and then decreasing again even below 20% for large sample sizes of 2400 and 4800.

Our models included 4 or 10 or 30 foods or 2 different simplified **"Western"** dietary patterns (consisted of 30 foods from the FLAG study and 10 in the ECRHS II study) which were causally linked with disease. In each scenario, this corresponds to around 1 in 7 foods (30 foods in FLAG study and 10 foods in the ECRHS II study and 30 foods consisting of the simplified **"Western"** dietary pattern in FLAG study) and 1 in 20 foods (10 foods in FLAG study and 4 foods in the ECRHS II study and 10 foods consisting of the simplified **"Western"** dietary pattern in the ECRHS II study). Similar qualitatively findings were obtained from all of our simulations. We did not consider interactive effects of foods in our models: this requires further investigation. As mentioned in Chapter 2 dietary constituents interact with each other in complex ways to impact health, for example, on breast cancer risk and hypertension (Messina et al., 2001, Sacks et al., 1995). PCA is often recommended as a way of dealing with interactions between foods (Hu, 2002, Varraso et al., 2009). However, as was mentioned in paragraph 2.5, it is questionable whether the linear combinations of food intakes produced by PCA adequately address the issues of modelling interactions.

Although we considered a model based on a **"Western"** dietary pattern, there is no reason why food intakes with truly causal effects should be only food intakes that are highly correlated with each other. Hence, in our simulation study, foods may or may not be correlated between each other when they are associated with disease risk. Where disease risk is explained by other factors that are confounded with diet, this confounding is likely to be also at the level of a dietary pattern. As we highlighted in paragraph 7.3, a **"prudent"** dietary pattern, for example, is associated with older age (Agurs-Collins et al., 2009), female sex (Robinson et al., 2009), non-smoking (Fung et al., 2001), higher income (Perrin et al., 2005), higher educational level (Raberg Kjollesdal, Holmboe-Ottesen & Wandel, 2010), exercise (Lopez-Garcia et al., 2004) and supplement use (Heidemann et al., 2008). These factors are likely to be associated with a number of food intakes contributing to a **"prudent"** pattern rather than with any one of these foods in particular. This is another reason to adjust each food effect for others found to be significant, as we suggest: this should help control for both

measured and unmeasured confounding. We did not include non-dietary confounders in our models because there are just too many different potential confounders and models for their effects that might be considered. We suspect, however, that as long as there are foods with truly causal effects, the findings presented in this paper will generalise to situations where other confounders have been explicitly adjusted for.

There is an asymmetry in the patterns of our Monte Carlo simulation results when different models (1, 2 and 3) have an effect on disease for ESFA method. Although the numbers of affected individuals are slightly lower for models where the power estimates are lower, this doesn't seem to be the only explanation for this asymmetry (see Tables 5.2.1, 5.2.2,). Models with all positive effects give different results to models with all negative effects; the situation that is being simulated is not symmetric, for example, the distribution of any given food intake cannot be expected to be symmetric but positively skewed as we observed in Figure 4.11.1. Thus, if the power to detect any effect of diet is much greater when all effects are positive than when all effects are negative, it is expected that when there is a mixture of positive and negative effects we could get intermediate power, because there are still a number of positive effects, which are easier to detect than negative ones. Similar patterns are observed when we are estimating the power and FDR of ESFA to detect the specific foods that are causally linked to disease (Tables 5.3.1 an 5.3.2).

However, average percentages of power of PCA in Tables 5.2.1, 5.2.2, 5.3.1 and 5.3.2 present a more symmetrical pattern when different models (1, 2 and 3) have an effect on disease. This is due to the fact that principal component scores are not expected to be as positively skewed as food intake data (Figure 4.11.1). However, we could not find an adequate explanation why this is not consistent for the ECHRS II survey data-set simulation scenario in tables 5.2.2 and 5.3.2.

Furthermore, in tables 5.2.1, 5.2.2 and 5.2.3 power increases with the number of principal components and sometimes it decreases. For example, power of PCA is lower when results for 10 principal components are compared to results for 2 and 5 principal components. This is only reported when power of PCA outperforms ESFA for 2 and 5 principal components. When power of ESFA outperforms PCA for 2 and 5 components then power of PCA for 10 principal components increases. This pattern in our simulation results is observed because as

the number of principal components increase the power of PCA shifts towards the power of ESFA.

There is a growing interest in designing dietary interventions around foods rather than nutrients (Jacobs et al., 2009) and around particular foods rather than dietary patterns (Jacobs et al., 2009, Mann, 2010). Specifically, Jacobs suggests that the evidence for beneficial effects of a **"prudent"** diet comes from interventions which only modified the intake of one or two foods (Jacobs et al., 2009). Mc Cann (2001) found that fruits and vegetables alone provided the highest discrimination among endometrial cancer cases and controls compared with the other methods of characterisation and highlighted the fact that sophisticated techniques may be unnecessary in studies of diet and disease. Mann and Aune, evaluating the evidence that fruit and vegetables can prevent diabetes, have called for more studies looking at effects of *specific* fruits and vegetables (Mann, 2010).

In our study, we decided to vary the particular parameters that affect the decision of a researcher when employing PCA in nutritional data. Our main sources of information were our systematic review for years 2004-2009 (chapter 3 of the thesis) and two other reviews published previously (Kant, 2004, Newby & Tucker, 2004). We decided that varying the number of our principal components from 2 to 10 components was reasonable since the upper limit for the derived principal component that were used by the literature was 10. We decided to use an odds ratio of 1.5, because this is a reasonable effect in epidemiological studies which are using dietary patterns and a false discovery rate of 20% in order not to penalize too heavily our findings.

Since ESFA outperforms PCA in our simulation study, dealing with high-dimensional multivariate dietary exposures could be treated as a problem of variable model selection that is, finding the nonzero regression coefficients in an unknown regression model. Our adjusted ESFA is similar to the iterative sure independent screening method (ISIS) for ultra-high dimensional data (Fan & Lv, 2008). Other forms of penalised likelihood estimation methods have been developed in the last decade to cope with high-dimensional data and have been lately reviewed by Fan & Lv (2010). This could be potentially useful in nutritional epidemiological studies, and further research is needed.

In conclusion, an FFQ-wide study of associations between food intakes and disease risk out-performs an analysis of dietary patterns derived from PCA. Analysing each food

adjusting for others allows truly causal effects to be identified with a low rate of false discoveries, and surprisingly good power. Although PCA has proved extremely popular in nutritional epidemiology to date, our simulation study questions its routine use in this context.

7.5 Analysis of GA^{2}LEN dataset: effect of diet on asthma

In this multicentre cross-sectional study, empirically derived dietary patterns (**"fish, fruit and vegetables"** and **"meat, potatoes and sweets"**) with the use of a meta-analytic approach to PCA were not associated with respiratory symptoms, after controlling for potential confounders. In our two-step exhaustive single food analysis approach, several food items were statistically associated with respiratory and allergic symptoms. However, our two-step ESFA results were not all of them in line with biological plausible hypotheses.

Our non-statistically significant findings of associations between our dietary patterns and asthma were consistent with other observational studies of 54,672 French women (Varraso et al., 2009), 52,325 male and female adult Chinese Singaporeans (Butler et al., 2006) and 1453 adults living in Greenwich (Bakolis et al., 2010). However, in two large US prospective cohort studies of 42,917 men and 72,043 US women, weak associations were reported between a **"prudent"** pattern and adult onset asthma for the female population only. There were no associations with the other dietary patterns and adult onset asthma (Varraso et al., 2007a, Varraso et al., 2007b). Furthermore our results were consistent with the results of 12008 pregnant women in the Avon Longitudinal Study of Parents and Children (ALSPAC) (Shaheen et al. 2009), where no associations were observed between a **"health conscious"** (similar to our **"fruit, fish and vegetables"**) and **"processed"** (similar to our **"meat, potatoes and sweets"**) pattern and asthma, atopy and eczema. Comparison with other studies of adults and respiratory and atopic outcomes is difficult as no other studies have analysed dietary patterns using PCA in this setting.

The heterogeneous effect on the association of the **"fruit, fish and vegetables"** pattern and eczema and allergic rhinitis is not easy to explain. Diet is strongly socially, environmentally and culturally patterned with specificities between countries and social groups (Galobardes, Morabia & Bernstein, 2001, Varasso, 2012). As Shaheen (2009) highlighted in the results from the Avon Longitudinal Study of Parents and Children (ALSPAC), a **"health conscious"** diet was univariately associated with eczema, total IgE, FEV_1 and negatively associated with persistent wheeze and asthma. However on controlling for numerous potential confounders these effects were attenuated and become non significant; went from an odds ratio of 0.90 (95% CI: 0.84-0.96; p-value< 0.001) to an odds ratio of 0.96 (95% CI: 0.88-1.05; p-value

=0.37) for the **"health conscious"** pattern; and from an odds ratio of 1.14 (95% CI: 1.07-1.22; p-value<0.001) to an odds ratio of 1.02 (95% CI: 0.94-1.10; p-value=0.69) for the **"processed"** one. Heterogeneity in multi-centre studies can also suggest alternative explanations for apparent effects of diet observed in single centers, such as uncontrolled confounding, and would make us cautious of progressing to a trial (Burney et al., 2008).

Because a dietary pattern is acting as a proxy to individual foods associated with respiratory outcomes then some heterogeneity in its effect might be due to heterogeneity in its associations with these foods; there was a great amount of variation in the correlations of individual foods with the **"fruit, fish and vegetables"** pattern as observed in table 6.3.2.2.1.

A meta-analytic approach to deriving dietary patterns across a number of centres has been investigated once before (Hooper et al., 2010). This method can be successful in identifying common dietary patterns, as well as evidence for heterogeneity in the effects of those patterns. Heterogeneity in observational studies of diet can sometimes argue against progressing to trials.

When ESFA was applied to our data, we found protective effects of fruits, whole grains, nuts, fish and vegetables on asthma (any legumes, chard, cherries, radish, lemon, wholemeal bread), atopy (parsnip), chronic sinusitis (apple, fenugreek, cabbage, fresh white fish, raisins) allergic rhinitis (peach, rhubarb, smoked fatty fish, asparagus, kiwi, peanuts) and eczema (bitter melon, rhubarb). Hard fruit intake has been negatively associated with asthma (Shaheen et al., 2001) impaired lung function (Butland, Fehily & Elwood, 2000) and chronic obstructive pulmonary disease (COPD) (Tabak et al., 2001a, Shaheen et al., 2010). However, longitudinal evidence for hard fruit intake relate only to COPD (McKeever & Britton, 2004). In addition, vegetable intake (carrots, tomatoes, leafy vegetables) has been related to reduced asthma (Romieu et al., 2006). There is evidence that intake of fish may protect against asthma symptoms (Laerum et al., 2007) and impaired lung function and COPD in adults (Tabak et al., 2001b). However, results on the beneficial effect of fish intake have been contradictory (Thien et al., 1996). Finally, fish, fruits and vegetables are essential components of a Mediterranean diet, which other recent work has found to be associated with improved asthma control in adults (Barros et al., 2008).

These protective associations are in line with current hypotheses and may be explained by the high antioxidant content of these food items (Halliwell, 1996). High levels of vitamin C

and flavonoids are present in citrus and hard fruits, phenolic acids, flavonoids, phytic acid, avenanthramides, vitamin E and selenium in whole grains (Devereux & Seaton, 2005, Slavin, 2004, Seaton, Godden & Brown, 1994), and n-3 fatty acids, n-6 fatty acids in fish (Schnappinger et al., 2009, Sausenthaler et al., 2009). However, these results are not confirmed by experimental studies where no effect has been demonstrated for Vitamin C, E and magnesium on asthma (Pearson et al., 2004, Fogarty et al., 2003).

Contrary to prevailing paradigms, we also found some evidence, for a positive association between fruits and vegetables intake with respiratory and allergic outcomes and a negative association between condensed milk with asthma and eczema. Recent studies suggested that a higher intake of antioxidants might promote the development of allergic disease (Murr et al., 2005) and whole milk intake might reduce asthma symptoms (Woods et al., 2003). However these results need further replication.

The explanation of other associations found between food items and respiratory and allergic outcomes are hard to be explained. Exhaustive single food analysis identified a number of foods which have not previously been considered to be associated with potential respiratory outcomes, and there are no obvious mechanisms that might link them. Although some of these may represent new and important findings, our analysis is limited by the presence of unmeasured, reverse and qualitative confounding.

Specifically, a lot of the highly significant foods have odds ratios close to 1, so it could be argued that the smallest bit of unmeasured or residual confounding or bias might explain their associations with respiratory and allergic outcomes. A simulation study by Fewel & Davey Smith & Sterne (2007), showed that bias in the exposure effect estimate increases as the amount of residual and unmeasured confounding increases, especially when confounders are uncorrelated with each other. Adjusting only for age, gender, body mass index and smoking status and food intakes that were statistically significant in an unadjusted ESFA didn't seem to control adequately for potential confounders.

In addition, we did not control our analysis for total energy intake. As we highlighted in paragraph 4.3.1, adjustment for total energy intake is important and should be considered because the level of energy intake might be a risk factor and distort the effect of a food on the potential respiratory and allergic outcome (Willett et al., 1997). Furthermore the effect of diet on allergic disease should be adjusted for other factors related to individual characteristics

such as social class, level of education, culture, ethnic background, language, as well as psychological factors (Tricon et al., 2006). In a systematic review by Nurmatov et al. (2012) on the effect of confounding in studies of diet and allergies in children, authors suggested that future studies should be adjusted for maternal characteristics, birth measurements, socioeconomic characteristics, other dietary factors and environmental exposures. Unfortunately in our analysis, information on cultural, socioeconomic and psychological factors was not available at the time of the analysis.

One other limitation is that there are differences in the correlation matrices between the centres. It is very difficult to present correlations between 238 food items for each of the ten centres. However, we presented in the thesis correlations of individual food items with our meta-analysed dietary patterns (Table 6.3.2.2.1) empirically derived with PCA. This allows us to investigate whether the correlation structure looks the same in different centres. Correlations of each food item with the dietary pattern are very different in each centre and this has implications in our meta-analysis results. In each centre, the higher the difference in the correlation values between food items, the higher the difference in the correlation matrices that are used in the PCA and the greater the heterogeneity of the dietary patterns that are meta-analysed; thus, it is doubtful the appropriateness of pulling dietary patterns from all the centres together. Appropriate statistical methods to derive and synthesize dietary patterns among different centres is a topic for further investigation.

Furthermore, because of the cross-sectional nature of our study there is the possibility of reverse confounding; people with respiratory symptoms may alter their diet in order to be healthier. For example, this is highlighted when we stratified our analysis for non atopics, associations between any legumes, chard, cherry, lemon, wholemeal bread and asthma didn't remain statistical significant. Finally, qualitative confounding is also apparent to our analysis; specifically in the univariate analysis chard is assumed to increase the risk of asthma, but when we adjust for the food intakes that were statically significant in an unadjusted ESFA (analysis not adjusted for other foods, but adjusted for age, sex, body mass index and smoking status) chard has a protective effect on asthma (Table 6.3.4.1).

FDR is controlled at 20% and 5%, so these are all foods for which there is some evidence of an association with health outcomes, not a clear definitive list. In our analysis, since we control the rate of our false discoveries at 20% and 5%, 20% and 5% of those discoveries are

expected to be false and consequently 80% and 95% are expected to be true. So, for example, in the results of the adjusted ESFA after controlling our FDR at 20%, 18 foods were expected to be associated with asthma (Table 6.3.4.1). Hence we can expect 14 of these to be genuinely associated with asthma and 4 of them to be false discoveries. Similarly, we can expect 11 (out of 14) foods to be genuinely associated with chronic sinusitis, 12 (out of 15) with allergic rhinitis, 8 (out of 10) with eczema and 14 (out of 17) with atopy. As expected, we had a smaller list of food items when we control our FDRs at 5%. Majority of these foods were indentified when FDR was at 20%. However, again a lot of these associations couldn't be easily clinically explained; positive association between couscous and turnip with asthma; positive associations of okra and pumpkin with chronic sinusitis; negative association of moussaka with atopy.

In order to examine associations between diet and disease, we employed a random intercept logistic model to take into account the dependence of individuals between countries. One of the major assumptions of the tests of significance used in the multilevel models is normality of the error distributions involve. Another assumption is the sufficient number of sample size for the higher-level variables. A potential limitation for using the multi-level model approach is that our countries are not a random sample of a population of countries, and that the number of our countries in the sample may be small. However, random intercept and random coefficient models are a flexible and powerful way to tackle the effect of clustering in country-level data (Localio et al., 2001) and it has been suggested that 10 or more groups in the second level (in our case country-level), provide accurate standard error estimates for the regression coefficients of the fixed part of the multilevel model (Maas and Hox, 2004).

An additional analysis was performed for the investigation of the consistency of effects across countries, reported in our adjusted ESFA controlling the FDR at 5%, of food items on respiratory and allergic outcomes. Hence, two types of approaches were employed: a random effects model and a meta-regression approach. From the random effects model all associations remained statistically significant with similar effect sizes (see Appendix IX). In the meta-regression model only the association between turnip and asthma appeared to be heterogeneous between centres (I^2=57.4%; P <0.05).

Comparing the results of the meta-regression approach with the results from the random intercept and the random effects modeling approach we observed that effect sizes were

similar. Associations that remained borderline significant was between; smoked fatty fish and asthma, couscous and asthma (see Appendix III); okra and chronic sinusitis (see Appendix IV); peach and allergic rhinitis, vegetable oil and allergic rhinitis, smoked poultry and allergic rhinitis, sour cream and allergic rhinitis (see Appendix V); rhubarb and eczema, crisp fried cakes and eczema, bitter melon and eczema, Greek style yogurt and eczema (see Appendix VI); crisp fried cakes and atopy and moussaka and atopy (see appendix VII). However, multilevel and meta-regression techniques are conceptually different. Multilevel analysis fits a hierarchical model and estimates the effect of a food item on disease by using individual-level observations which are clustered within the countries, while meta-regression estimates a separate effect of a food item on disease in each country , and estimates an overall pooled effect based on these per country effect sizes. In addition, for our meta-regression approach associations reported are pretty consistent in all counties. This is an additional argument of these associations to be real findings and even close to be causal according to Bradford-Hill (1965) criterion of consistency (although a confounder might be closely attached to the same food in all countries).

Although the use of adjusted ESFA obtained a set of foods predictive of the respiratory status of the individuals, the foods identified failed to represent a number of foods which are in line with current biological hypothesis or present some new streams in the nutritional research. It's possible that confounding between lifestyle factors and individual foods explains some of the findings, and that some of the statistical methods of analysis are just not appropriate (such as the assumption of random effect of site in our regression models or the correlation structure of food intakes which is very different in different sites). Although, less complicated methods than PCA such as variable selection methods seem promising in the field of nutritional epidemiology, our G.A.L.E.N analysis results suggest that challenges remain in this field.

In conclusion, we found no firm, consistent evidence for an association of dietary patterns empirically derived with the use of PCA with respiratory and allergic outcomes. When we employed ESFA in our data-set, a number of foods were associated with respiratory and allergic symptoms. However, our results may be affected by unmeasured confounders associated with dietary choices and statistical methodological issues so they must be interpreted with caution and need further replication.

7.6 Conclusions

Slattery in 1998 concluded that the soundness of using principal component analysis to identify eating patterns will be better understood when more epidemiologists have begun to use the method. In 2008, ten years after the explosion of the use of PCA in dietary studies, and with almost 100 papers employing the PCA method, Slattery claimed that data-driven dietary patterns identified empirically from PCA characterized the diet associated disease risk, provided more seemingly consistent associations and could guide diet and disease investigations better than anyone food or nutrient. All of these arguments are questionable.

Specifically, our systematic review provides evidence of highly inconsistent data-driven dietary patterns, empirically derived with the use of PCA, across different studies. This may be due to the fact that dietary patterns are strongly confounded by lifestyle factors. In addition, reviewing the literature from our systematic review and the review from Kant (2004) and Newby (2004) we concluded that there is a number of different subjective decisions taken by the researchers (because of the methodological subjectivity of PCA as employed in nutritional epidemiology) from study to study. Hence, data-driven dietary patterns cannot be easily translated into an intervention and be comparable across different studies.

Our simulation study provides quantitative evidence that questions the use of PCA for detecting combination of foods that are causally associated with disease and highlights the high risk of false positive findings of the method. Moreover, we suggest that the best way to analyse nutritional data assuming that a combination of foods are associated with disease in an additive way or as a principal component of the population is to 1) run an exhaustive search for associations between individual food intakes and disease (ESFA procedure) with an appropriate statistical model, allowing for multiple testing and adjusting for energy intake and then 2) to re-run an ESFA procedure by further adjusting for these foods that are statistically significant in the first round of univariate analysis.

In the GA^{2}LEN survey analysis study, PCR method was unsuccessful on detecting associations between diet and the allergic and respiratory outcomes. On the other hand, ESFA detected a number of associations, but not all of them were necessarily causal. Specifically, ESFA gives a list of variables with a 20% false discovery rate and not a definite list.

However, lack of biological plausible associations could be also due to unmeasured confounding or that assumptions of the statistical methods might not be met.

Our fundamental assumption was that in order for PCA and its extension PCR to be useful methods in nutritional epidemiology, they should identify combinations of foods in the population that are causally linked to disease, be reproducible and can be translated into an intervention. None of the three are possible with the use of the PCA and PCR methods as we proved from our systematic review and our simulation study.

We are not to claiming that different elements of diet are associated with disease *only* in the way that we specified or that there are not complex biological interactions between nutrients and foods or that all other data reduction techniques that associate a pattern with disease are not useful. The purpose of the thesis is to provide a critique and start a discussion on the inappropriateness of PCA as a prevailing method in nutritional epidemiology to detect associations between diet and disease, and emphasize on the need for future research (paragraph 7.7) in this field.

Paraphrasing Lehman on the theatre of Bertolt Brecht “We are highly interested in the questions that led to the explosion of dietary pattern analysis with the use of PCA, but we are not satisfied anymore by the answers provided by the method”.

7.7 Ideas for future research and implications of these findings

Analysis of nutritional data is exceedingly challenging, given that the aim is to tease apart complex data patterns, estimate the effect of overall diet, and tackle confounding of dietary exposures and estimate possible higher-order interactions between nutrients or foods consumed in combination. Possible ways to address all these issues in the future without the use of a PCA approach are

- To evaluate other existing multivariate methods applied in nutritional epidemiology. This may be done by comparing Cluster analysis, Treelet transform, Reduced Rank Regression, Gaussian mixture modelling with an exhaustive analysis of single foods (ESFA).

- To expand our simulation model in a way that different foods are causally linked with disease in an interactive way, and evaluate existing multivariate methods by comparing them with an exhaustive analysis of single foods (ESFA).

- To use other datasets for our Monte Carlo simulations with people living outside UK.

- To research into new methods for analyzing nutritional data with the use of statistical techniques applied for high dimensional data such as Sparse PCA (Jolliffe 2003), supervised Principal Components (Bair et al., 2004), Iterative Sure Independence Screening (Fan and Lv, 2008), Shrinkage and Selection via the Lasso (Tibshirani, 1996) and causal models (Galea et al.2010, Vineis., 2006).

- To continue the analysis of G.A^2.L.E.N data-sets for nutrients and other respiratory and allergic outcomes.

- To aggregate food items into food groups of the similar composition and exchangeability (e.g. biscuits and crackers) that are consumed each with an average frequency or less than once per week in G.A^2.L.E.N.

- To apply our exhaustive single food analysis method adjusted for foods that were significant in an unadjusted analysis (analysis not adjusted for other foods, but adjusted for age, sex, body mass index and smoking status) in other observational studies apart from G.A^2.L.E.N.

Tables of systematic review

Table A. Justifications for the use of Principal Component Analysis instead of a single nutrient/ food approach in dietary pattern analysis studies in nutritional epidemiology (Sorted by Year of Publication and Author). Blank cells describe not available information on that argument. Proportions of papers that provide this justification is given.

Author	Interactive, antagonistic and synergistic effects of foods consumed in combination. (proportion of papers : 25/163=15.3%)	Additive effects of foods consumed in combination which are too small to detect when they are examined separately (proportion of papers: 14/163=8.5%)	Confounding and mulitcollinearity from lifestyle factors and dietary exposures (proportion of papers: 52/163=31.9%)	Multiple testing problems (proportion of papers: 7/163=4.2%)	Public health recommendations (proportion of papers: 14/163=8.5%)	Complexity of diet (proportion of papers: 25/163=15.3%)	Better evaluation / representation of overall diet (proportion of papers: 30/163=18.4%)
(Agurs-Collins et al., 2009)			X				X
(Akbaraly et al., 2009)							
(Ambrosini et al., 2009)		X	X				
(Bakolis et al., 2010)		X	X	X			
(Bastos et al., 2010)			X				
(Bertuccio et al., 2009)			X	X			
(Brantsaeter et al., 2009)							
(Cottet et al., 2009)					X		
(Craig et al., 2010)			X				
(Cutler et al., 2009)			X		X	X	
(Deshmukh-Taskar et al., 2009)					X	X	
(Erber et al., 2010)							
(Hamer & Mishra, 2010)	X	X					
(He et al., 2009)			X				
(Hooper et al., 2010)	X		X				
(Hughes et al., 2009)							

(Imamura et al., 2009)			X				
(Jackson et al., 2009)							
(Kesse-Guyot et al., 2009)			X			X	
(Kontogianni et al., 2009)	X		X				X
(Qi et al., 2009)							
(Lutsey et al., 2009)							
(Muller et al., 2009)							X
(Nettleton et al., 2009)							
(Noel et al., 2009)							
(Oddy et al., 2009)					X		X
(Paradis et al., 2009)	X						X
(Reedy et al., 2010)							
(Rezazadeh, Rashidkhani & Omidvar, 2010)							X
(Robinson et al., 2009)	X					X	
(Shaheen et al., 2009)			X	X			
(Touvier et al., 2009)							
(Uusitalo et al., 2009)						X	
(Vujkovic et al., 2009a)							
(Vujkovic et al., 2009b)							
(Wiles et al., 2009)							
(Wu et al., 2009)						X	
(Ambrosini et al., 2008b)(2 papers)							X
(Ambrosini et al., 2008a)							
(Arkkola et al., 2008)						X	
(Borland et al., 2008)					X		
(Butler et al., 2008)			X				
(Campbell, Sloan & Kreiger, 2008)			X				
(Chang et al., 2008)	X		X				X
(Crozier et al., 2008)	X		X				X
(De Stefani et al., 2008a)							
(De Stefani et al., 2008b)							

(D'Souza et al., 2008)	X		X				
(Edefonti et al., 2008)			X	X			
(Engeset et al., 2009)	X					X	
(Esmaillzadeh & Azadbakht, 2008a)	X		X				X
(Esmaillzadeh & Azadbakht, 2008b)	X	X	X				X
(Feinstein et al., 2008)							
(Flood et al., 2008)	X		X				
(Heidemann et al., 2008)						X	
(Hooper et al., 2008)	X		X				
(Iqbal et al., 2008)	X				X		
(Keskitalo et al., 2008)							
(Kim et al., 2008)						X	
(Knudsen et al., 2008)			X				
(Kubo et al., 2008)	X						
(Lau et al., 2008)	X					X	
(Lutsey, Steffen & Stevens, 2008)							
(McNaughton et al., 2008)	X					X	
(McNaughton et al., 2008)	X					X	
(Murtaugh et al., 2008)							
(Nanri et al., 2008) (2 papers)	X		X			X	
(Nettleton et al., 2008b)	X						
(Nettleton et al., 2008a)							X
(Northstone & Emmett, 2008a)							
(Northstone, Emmett & Rogers, 2008a)							X
(Northstone, Emmett & Rogers, 2008b)			X				X
(Northstone & Emmett, 2008b)							
(Okubo et al., 2008)	X	X					
(Romaguera et al., 2008)							
(Sadakane et al., 2008)	X	X					
(Shi et al., 2008)	X						
(Takaoka & Norback, 2008)							

(Tseng et al., 2008)							
(Varraso et al., 2007b)	X						
(Yannakoulia et al., 2008)							X
(Akesson et al., 2007)							X
(Bamia et al., 2007)	X		X				
(Cai et al., 2007)	X		X			X	
(Cui et al., 2007)		X					
(Custodio das Dores et al., 2007)							
(Dalvi, Canchola & Horn-Ross, 2007)	X	X					
(De Stefani et al., 2007)			X				X
(Esmaillzadeh et al., 2007)							X
(Hirose et al., 2007)						X	
(Kim et al., 2007)							
(Maruapula & Chapman-Novakofski, 2007)							
(Masala et al., 2007)	.		X	X			X
(McNaughton et al., 2007)							
(Murtaugh et al., 2007)					X		
(Nettleton et al., 2007)	X						
(Okubo et al., 2007)	X	X					
(Panagiotakos et al., 2007a)	X		X		X		X
(Panagiotakos et al., 2007b)						X	X
(Robinson et al., 2007)							
(Sant et al., 2007)							
(Shimazu et al., 2007)	X		X				
(Shin, Oh & Park, 2007)							
(Takata et al., 2007)			X				
(Teucher et al., 2007)	X					X	X
(Varraso et al., 2007a)							
(Burt et al., 2006)							
(Butler et al., 2006)	X		X				
(Chen et al., 2006)							

(Crozier et al., 2006)	X		X		X		
(Cuco et al., 2006)			X				X
(Kesse, Clavel-Chapelon & Boutron-Ruault, 2006)			X				X
(Mizoue et al., 2006)	X		X				
(Naska et al., 2006)	X				X		
(Newby et al., 2006ba)		X	X				X
(Newby et al., 2006b)	X	X	X				
(Pala et al., 2006)	.			X		X	
(Paradis, Perusse & Vohl, 2006)	X					X	
(Ronco et al., 2006)							
(Schulze et al., 2006)							
(Waijers et al., 2006)	X	X	X		X	X	
(Weismayer, Anderson & Wolk, 2006)							
(Wu et al., 2006)	X						
(Zhang et al., 2006)							
(Balder, Goldbohm & van den Brandt, 2005)							X
(Cottet et al., 2005)	X		X				
(Engeset et al., 2005)						X	
(Fung et al., 2005)	X		X				
(Hoffmann et al., 2005)			X				
(Kroenke et al., 2005)							
(Mannisto et al., 2005)			X		X		
(Marchioni et al., 2005)	X		X			X	
(Michaud et al., 2005)	X						X
(Mikkila et al., 2005)	X	X					
(Mizoue et al., 2005)	X		X				
(Montonen et al., 2005)	X		X				X
(Nkondjock et al., 2005)	X						X
(Northstone & Emmett, 2005)							
(Park et al., 2005)	X		X				
(Perrin et al., 2005)	X					X	

(Rashidkhani et al., 2005)	X		X				
(Uusitalo et al., 2005)	X		X				
(Velie et al., 2005)							
(Yang, Kerver & Song, 2005)	X						
(Corrao et al., 2004)							
(De Stefani et al., 2004)							
(Dixon et al., 2004)	X						
(Fung et al., 2004a)							
(Fung et al., 2004b)	X						
(Kant et al. 2004)							
(Khani et al., 2004)	X		X	X			
(Kim et al., 2004)	X					X	
(Lopez-Garcia et al., 2004)							
(Newby et al., 2004) (Newby, Muller & Tucker, 2004) (2 papers)							
(Robinson et al., 2004)							
(Sieri et al., 2004)							X
(Togo et al., 2004)	X				X		
(Tseng et al., 2004)		X			X		
(Wu et al., 2004a)							

Table B. Application of PCA procedure to identify dietary patterns in nutritional epidemiology; details of each study; details on the dietary assessment instrument being used in each study; preparation of data before entering the PCA procedure in each study; applications of criteria of PCA for labelling and identifying the number of dietary patterns in each study; validation methods for retained dietary patterns in each study (Sorted alphabetically by study design and country). Summary statistics are provided for each paper.

Author	Details of each study	Details of the dietary assessment instrument being used in each study	Preparation of data before entering the PCA procedure in each study	Application of criteria of PCA for	Validation methods of retained dietary patterns in each study
	1. Name 2. Study design Case-control (proportion of papers:31/163=17.1%) Cohort (proportion of papers (31/163=19.0%) Cross sectional (proportion of papers 99/163=64%) Other (proportion of papers 4/163=2.4%) 3. Sample Size 4. Country being held	1. Definition of dietary assessment instrument a.FFQ (proportion of papers: 152/163=93.2%) b.24/48- hour recall ((proportion of papers: 4/163=2.4%) c. dietary records (proportion of papers:3/163=1.8%) d. diet history questionnaire (EPIC/DHQ) (proportion of papers:3/163=1.8%) 2. number of food items in each instrument (median value:92; IQR:21-204) 3. number of food groups in each instrument	1. Conversion of food frequency data to grams/d or grams /week (proportion of papers:22/163=13.4%) 2. Standardisation of food intake variables (proportion of papers:6/163=3.6%) 3. Food intake variables adjusted for energy intake by the residual method (proportion of papers:5/163=3.0%) * No details are given	Labeling the dietary patterns in each study 1. principal component loading and correlation coefficient cut off point s in each study (median value:0.3; 0.3-0.4) 2. Method of rotation in each study Orthogonal Varimax (proportion of papers: 85/163=52.1%) Oblique Promax (proportion of papers: 4/163=2.4%) Identifying thenumber of dietary patterns in	1. Randomly split sample in each study (proportion of papers: 8/163=4.9%) 2. Cronbach's alpha (proportion of papers: 5/163=3.0%) 3. Deriving dietary patterns separately for women and men (proportion of papers: 3/163=1.8%) 4. Barlet's test of sphericity (proportion of papers: 1/163=0.6%) 5. Kaiser - Meyen-Ollkin test (proportion of papers: 2/163=1.2%) 6. Identifying simplified dietary patterns (proportion of papers: 3/163=1.8%) 7. Use of Confirmatory/maximum likelihood factor analysis

		(median value (median value:38 ; IQR:19-69) **4. Scale of FFQ (median value:7; IQR: 5-10)**		**each study** **3. Scree Plot (proportion of papers: 81/163=49.1%)** **4. cut-off point s for eigen values (median value:1.6; IQR (1.2)** **5. dietary patterns interpretability (proportion of papers:70/163=42.9%)** **6. Van der Voet's test (proportion of papers: 1/163=0.6%)**	**(proportion of papers: 8/163=4.9%)** **8. φ coefficient for testing inter-correlation (proportion of papers: 2/163=1.2%)** **9. Stricter cut-off points for energy intake (proportion of papers: 1/163=0.6%)** **10. Different method of rotation (proportion of papers: 2/163=1.2%)** **11. Pearson Correlation coefficient in different time points (proportion of papers: 5/163=3.0%)**
(Iqbal et al., 2008)	1. INTERHEART study 2. Case-control study 3. 5761 Cases and 10646 controls 4. 52 Countries	1a. 2.19	1.	1. > |0.25| & 3. & 4. Eigenvalue > 1 & 5.	*
(Ambrosini et al., 2008b) (Ambrosini et al., 2008a)	**1. *** **2. Case-control study** **3. 546 Cases and 447 controls** **4. Australia**	**1a.** **2.101** **3.74** **4.Scale of 10:''never'' to ''3 or more times per day.'**	*	**1.>0.3 2a & 3. & 4. >1. & 5.**	*
(Marchioni et al., 2005)	1. * 2. Case-control study 3. 517 4. Brazil	1a. 2.27	*	1>0 2a. 3. & 4 & 5.	*
(Campbell, Sloan & Kreiger, 2008)	1. National Enhanced Cancer Surveillance System 2. Case-control study 3. 2813 4. Canada	1a. 2.69 4. Scale of 10: from ''never or less than once per month,'' to ''six or more times per day''	*	2a.. & 3. & 4 >3. & 5	*
(D'Souza et al., 2008)	1. * 2. Case-control study	1a. 2.151	*	1. >0.30 2a. & 3 & 4 & 5	*

	3. A total of 149 Cases and 251 controls were included for the study. 4. Canada	4. Scale of 8 "never or less than once per month" to "6 per day."			
(Nkondjock et al., 2005)	1. NECSS study 2. Case-control study 3. 585 4. Canada	1a. 2.69	*	1.>|0.30| & 2a 3. & 4 >1 & 5	*
(Cui et al., 2007)	1. Shanghai Breast Cancer Study 2. Case-control study 3. 1556 controls 4. China	1a. 2.76	*	1.>|0.20 |& 2a 3. & 4 >1 & 5 & 7	*
(Di bello et al. 2008)	1. * 2. Case-control study 3. 3574 Cases and controls 4. Costa Rica	1a. 2.135 3.43	3	5. & 6.	1.
(Hooper et al., 2010)	1. UK European Community Respiratory Health Survey I (ECRHS-I) 2. Case- control study 3. 1174 4. Europe	1a. 2.158 (German) 198 (UK) 204 (Norway174) 4.frequencies from never to five portions a day	1. & 2.	1. >|0.30| 2a. & 3.	*
(Bertuccio et al., 2009)	1. * 2. Case-control study 3. 230 patients with incident, histological confirmed gastric cancer and 547 frequency-matched controls, 4. Italy	1a. 2.78	1.	1. ≥|0.63| & 2a.	2. & 4. & 5.
(Edefonti et al., 2008)	1. * 2. Case-control study 3. Cases were 2,569 breast cancers and 1,031 ovarian. Controls were 3,413 women from the same hospital network 4. Italy	1 a. 2.78	*	1. >0.63 2a & 3. & 4. >1. & 5.	7.
(Corrao et al., 2004)	1. * 2. Case-control study 3. 481 selected controls, 152 were healthy subjects and 329 not 4. Italy	1a 2.93 4.Scale 0f 10:(3 times/d, to never/rarely)	3	4.>1	3.
(Jackson et al., 2009)	1. * 2. Case-control study 3. 204 histological confirmed newly diagnosed prostate cancer Cases	1a. 3.33	1.	1. >|0.40| & 2a. & 3. & 4. >1	*

	and 204 individually matched urology clinic controls 4. Jamaica				
(Hirose et al., 2007)	1. Aichi Cancer Center (HERPACC). 2. Case-control study 3. 1885 4. Japan	1a. 2.31	*	2a 3. & 4 >1 & 5	*
(Vujkovic et al., 2009a)	1. * 2. Case-control study 3. 161 4. Netherlands	1a. 2.195 3.22	3	2a.	*
(Vujkovic et al., 2009b)	1. * 2. Case–control study 3. * 4. Netherlands	1a. 2.200 3.16	*	2a.	*
(Bastos et al., 2010)	1. * 2. Case-control study 3. 591 incident Cases of gastric adenocarcinoma and 1463 community controls. 4. Portugal	1a. 2.82	*	2a. & 3. & 4. >1	*
(Bakolis et al., 2010)	1. Food and Lifestyle and Asthma in Greenwich (F.L.A.G) study 2. Case-control study 3. 1453 individuals (599 Cases and 854 controls) 4. UK	1a. 2.217 4.Scale of 10: from 'never', to 'two or more times per day')	1 & 2.	1. >|0.3| & 2a. & 3. & 5.	*
(De Stefani et al., 2008a) **(De Stefani et al., 2008b)** **(De Stefani et al., 2007)**	**1. *** **2. Case-control studies** **3. 861 (2008a)** **255 Cases and 501 hospitalized controls (2008b)** **Cases 290 Controls and 290 Cases** **(2007)** **4. Uruguay**	**1a.** **2.64** **3.27** **4. Scale of 6: never to more than one time per day.**	*	**1. >0.30 & 2b. & 3. & 4. >1.0**	**7.& 11**
(De Stefani et al., 2004)	1. All patients with newly diagnosed and microscopically confirmed gastric carcinomas, admitted for diagnosis and treatment in the four major hospitals 2. Case-control study 3. 240 Cases and 960 controls. 4. Uruguay	1a. 2.191	*2	2a. & 3. & 4 & 5	11.

(Ronco et al., 2006)	1. 2. Case –control study 3. 442 newly diagnosed and microscopically confirmed Cases with breast cancer and 442 hospitalized controls 4. Uruguay	1a. 2.64	*	1.>|0.29 | & 2a 3. & 4 >1 & 5	
(Wu et al., 2009)	1. * 2. Case-control study 3. 2172 4. USA	1a 2. 174	*	1. >|0.30| & 2a & 3. & 4. > 1 & 5.	*
(Kubo et al., 2008)	1. * 2. Case-control study 3. 296 Cases were matched to persons with gastro esophageal reflux disease (308 without Barrett's esophagus and to population controls (309 4. USA	1a. 2. 110	*	1. > 0.35 & 3. & 4 >1 & 5	*
(Murtaugh et al., 2008)	1. Four-Corners Breast Cancer (FCBC) study 2. Case-control study 3. (757 Cases, 867 controls) and non-Hispanic white women (1524 Cases, 1598 controls) 4. USA	1a. 3. 69	*	1. > 0.35& 2a.	*
(Dalvi, Canchola & Horn-Ross, 2007)	1. * 2. Case-control study 3. 647 Cases and 633 controls 4. USA	1a. 2. 103	*	1.>|0.35 |& 2a 3. & 4 >1 & 5	*
(Maruapula & Chapman-Novakofski, 2007)	1. Health and Nutrition of the Elderly in Botswana. 2. Cross-sectional study 3. 1086 4. Africa	1a. 2.21 4. Scale of 4:"eat less," "more," "same," and "never ate."	*	1.|>0.5| 2a. & 3. & 4. >1. & 5.	*
(Romaguera et al., 2008)	1. * 2. Cross-sectional study 3. 1236 4. Argentina	1a. 2.46 3.13	3	1. >0.2. & 4.>1	*
(McNaughton et al., 2008)	1. Australian National Nutrition Survey (NNS) 2. Cross-sectional study 3. 764 4. Australia	1a. 2.127 4.Scale of 9:''never or less than once a month'' to ''6 or more times per day''		1. >0.3 & 2. & 3. & 4 >1.25 & 5	2. & 6.

(Chen et al., 2006)	1. Health Effects of Arsenic Longitudinal Study (HEALS), 2. Cross-sectional study 3. 11116 4. Bangladesh	1a. 2.39	1.	1.>|0.15 | & 2a 3. & 4 >1.5 & 5	*
(Custodio das Dores et al., 2007)	1. * 2. Cross-sectional study 3. 115 4. Brazil	1a. 2.97	1.	1.>|0.40 | & 2a 3. & 4 >1 & 5	*
(Paradis et al., 2009)	1. * 2. Cross-sectional study. 3. 664 4. Canada	1. 2.91	*	1. >|0.30| 2a. & 3. & 4.>1	*
(Paradis, Perusse & Vohl, 2006)	1. * 2. Cross – sectional study 3. 197 women and 129 men 4. Canada	1a. 2.91	1.	1.>|0.30 | & 2a 3. & 4 >1 & 5	*
(He et al., 2009)	1. China National Nutrition and Health Survey 2. Cross-sectional study 3. 56442 4. China	1a.	*	2b.	*
(Togo et al., 2004)	1. MONICA study (Monitoring of Trends and Determinants in Cardiovascular Diseases). 2. Cross-sectional study 3. 2436 aged 30–60 y attended all three examinations 4. Denmark	1a. 2. 3.21	*	1.>|0.30 | & 2a & 3. & 4 >1 & 5	
(Naska et al., 2006)	1. Data Food Networking (DAFNE) project Standardized household budget surveys (HBS) 2. Cross-sectional study 3. 94564 households of the ten countries under study, 15251 households whose composition did not fit in any of the predefined 4. Europe	1e)	*	1.>|0.20 | & 2a 3. & 4 >1 & 5	3.
(Montonen et al., 2005)	1. The Finnish Mobile Clinic Health Examination Survey 2. Cross-sectional study 3. 4304 4. Finland	1a. 1.23	*	1.>|0.30 | & 2a 3. & 4 >1 & 5	1. & 7.
(Perrin et al., 2005)	1. MONICA study (Monitoring of Trends and Determinants in	1c. 2.15	*	1.>|0.25 | & 2a & 3. & 4 >1 & 5	6. & 7. & 10.

	Cardiovascular Diseases). 2. Cross-sectional study 3. 3508 4. France				
(Kontogianni et al., 2009)	1. * 2. Cross-sectional study 3. 220 4. Greece	1f.	*	1a. >|0.3| & 2a.	*
(Yannakoulia et al., 2008) (Panagiotakos et al., 2007b)	1. Attica study 2. Cross-sectional study 3. 453 men and 400 women (2008) 1514 men and 1528 women (2007) 4. Greece	1d. 2.156 3.22 (2007b)	*	1. >0.3 2008 >|0.40 |2007 2a 3. & 4 >1 & 5	9. & 10.
(Panagiotakos et al., 2007a)	1. MEDIS (Mediterranean Islands) 2. Cross-sectional study 3. 300 men and women from Cyprus, 142 from Mitilini, 100 from Samothraki, and 104 from Kefalonia islands. 4. Cyprus and Greece	1a. 2.15	*	1.>|0.3| & 2a. & 3. & 4. >1.	*
(Rezazadeh, Rashidkhani & Omidvar, 2010)	1. * 2. Cross-sectional study 3. 460 4. Iran	1a. 2.168 3.37	*	2a. 3. & 4. >1.0	*
(Esmaillzadeh & Azadbakht, 2008a) (Esmaillzadeh & Azadbakht, 2008b) (Esmaillzadeh et al., 2007)	1. * 2. Cross-sectional study 3. 486 4. Iran	1a. 2.168 3. 41	*	2a. & 3. & 4 >1. & 5	*
(Okubo et al., 2008) **(Okubo et al., 2007)**	**1. *** **2. Cross-sectional study** **3. 3760 (2008)** **3770(2007)** **4. Japan**	**1a** **2.148** **3.30**		**1.>|0.20 |& 2a 3. & 4 >1 & 5**	**1.**
(Sadakane et al., 2008)	1. * 2. Cross-sectional study 3. 6886 (in the analysis on blood pressure) and 7641 (in the analysis on serum lipids) 4. Japan	1a. 2.30 4. Scale of 5: 1: seldom to 5: almost every day.	*	2a. & 3. & 4 > 1.	1.
(Takaoka & Norback, 2008)	1. * 2. Cross-sectional study 3. 153	1a. 2. 11	*	2 a.	*

	4. Japan	4.Scale of five , 0 : never to 5 : almost daily			
(Mizoue et al., 2006) (Mizoue et al., 2005)	1. Self-Defense Forces Health Study 2. Cross-sectional study 3. 2141 4. Japan	**1a. 2.74 4.Scale of 7:from ‘‘never/,1 time/mo’’ to ‘‘2–3 times/ d.’’**	*****	**1.>|0.30 | & 2a 3. & 4 >1 & 5**	*****
(Kim et al., 2007)	1. Korean National Health and Nutrition Survey 2. Cross-sectional study 3. 1257 4. Korea	1b.	*	*	*
(Shin, Oh & Park, 2007)	1. * 2. Cross-sectional study 3. 1441 preschool children 4. Korea	1a. 2.100 3.31 4. Scale of 9:from “rarely use’ to ‘have three or more times per day ”	1.	2a & 3. & 4 >2 & 5	*
(Craig et al., 2010)	1. * 2. Cross –sectional study 3. 2352 4. Scotland	1a. 2.65	*	1. >|0.3| & 2a. & 3.	*
(Crozier et al., 2008)	1. * 2. Cross-sectional study 3. 617 women in early pregnancy 4. UK	1a. 2.100 3.49	*	*	*
(Teucher et al., 2007)	1. * 2. Cross-sectional study 3. 3262 4. UK	1a. 2.131 3.54	1. & 2. & 3	3. & 5	*
(Burt et al., 2006)	1. * 2. Cross-sectional study 3. 1021 4. USA	1a. 2.20	*	3 & 5	*
(Yang, Kerver & Song, 2005)	1. * 2. Cross-sectional study 3. 263 men, 234 women 4. USA	1a. 2.22	1.	1. >0.30 or <-0.20 & 2a & 3. & 4 >1.25 & 5	*
(Hughes et al., 2009)	1. * 2. Cohort study. 3. 1119 4. Australia	1a. 2.129 4. Scale of 9:‘‘never’’ to ‘‘>1 times per day.’’.	*	1. >|0.15| 2a. & 3. & 4.>1	*

(Muller et al., 2009)	1. The Melbourne Collaborative Cohort Study (MCCS) 2. Cohort study 3. 1018 incident prostate cancer Cases 4. Australia	1a. 2.121	*	4. >2.0	*
(Oddy et al., 2009)	1. The Western Australian Pregnancy Cohort Study 2. Cohort study 3. 1860 4. Australia	1a. 2.1 3.26	*	1. > |0.30| & 4. > 1	*
(Shi et al., 2008)	1. * 2. Cohort study 3. The total sample included 1308 men and 1541 women, f them 711 participants were from the urban area. 4. China	1a. 2.33 3.25	*	1. >0.2. & 2a. & 3. & 4.& 5.	*
(Cai et al., 2007)	1. The Shanghai Women's Health Study 2. Cohort study 3. 74942 4. China	1a. 2.71 4.Scale of 5:daily, weekly, monthly, yearly, or never)	*	1.>|0.30 | & 2a 3. & 4 >1 & 5	*
(Knudsen et al., 2008)	1. Danish National Birth Cohort (DNBC) 2. Cohort study 3. 44 612 women 4. Denmark	1a 3.36	*	2a & 3. & 4. & 5.	*
(Bamia et al., 2007) (Masala et al., 2007) (Pala et al., 2006) (Waijers et al., 2006)	**1. EPIC-Elderly project 2. Cohort study 3. 74607(2007a) 5611(2007b) 47749(2006) 4990(2005) 4. Europe**	**1a 1d.(2007b) 2.120(2007b) 188, 217 and 140(2006a) 178(2006b) 148(2005) 3. 22 (2007a) 17(2006b) 57(2006)**	*	**1.>|0.40 |2007a >|0.30| 2007b >|0.20| 2006b & 2a 3. & 4 >1 & 5**	*
(Mannisto et al., 2005) (Dixon et al., 2004)	**1. DIETSCAN project. Three of these Cohort studies (NLCS, ORDET and SMC) who had female participants were included in this specific breast cancer study.**	**1a. 2. 51 (2005) 64 (2004)**	*	**1.>|0.35 | & 2a & 3. & 4 >1 & 5**	*

	2. Cohort study **3. 3123 (2005)** **61463 (2004).** 4. **Europe**				
(Uusitalo et al., 2009) **(Arkkola et al., 2008)**	**1. Finnish Type 1 Diabetes Prediction and Prevention (DIPP) Nutrition Study** **2. Cohort study.** **3. 3360(2009 study)** **3730 (2008 study)** **4. Finland**	**1a** **2.181** **3. 52**	*****	**1. >0.2. & 2a. & 3.**	*****
(Mikkila et al., 2005)	1. * 2. Cohort study. 3. 1200 and 1037 4. Finland	1b.	1.	2a & 3. & 4 >1 & 5	*
(Keskitalo et al., 2008)	1. * 2. Cohort study 3. 2009 4. Finland	1a. 2.24 4.(5 categories (1=never–5=several times a day)	*	3. & 4. & 5.	2.
(Cottet et al., 2009) **(Touvier et al., 2009)** **(Varraso et al., 2009)** **(Kesse, Clavel-Chapelon & Boutron-Ruault, 2006)**	**1. E3N [Etude Epide´miologique aupre`s de Femmes dela Mutuelle Ge´ne´rale de l'Education Nationale] study** **2. Cohort study** **3.** **62372(Cottet 2009 study)** **64252(Touvier 2009 study)** **56,881 (Varraso 2009 study)** **516 adenoma Cases (175 high-risk adenomas) 4,804 polyp-free women** **4. France**	**1d.** **2.208** **3.57 (Cottet 2009 study)** **46 (Touvier 2009 study)** **56 (Varraso 2009 study)** **40 (Kesse 2006 study)**	*****	**1.** **> |0.25| Cottet 2009 study** **>0.2 Touvier 2009 study** **>0.4 Varraso 2009 study** **0.20 Kesse 2006 study** **2a.** **3.** **4. >1.25** **5.**	*****
(Kim et al., 2008)	1. Kohala Health Research Project 2. Cohort Study 3. 1257 participants 4. Hawaii	1a. 2.166 4. Scale of 6: ("never," to "2 times a day or more").	*	2a. & 3. & 4. >3	*
(Sant et al. 2007)	1. ORDET (Hormones and Diet in Etiology of Tumors) 2. Cohort study 3. 8,861 4. Italy	1a. 2.107 3.34	*	1.>|0.25 | & 2a 3. & 4 >1 & 5	1. & 7.
(Sieri et al., 2004)	1. * 2. Cohort study	1a. 2.100	*	1.>|0.25 | & 2a & 3. & 4 >1 & 5	

	3. 8984 4. Italy	3.49			
(Nanri et al., 2008)	1. * 2. Cohort study 3. 3243 men and 4,667 women 4. Japan	1a 2.49 4. Scale of 7:(1–3 times/mo, to 3 times/d)	*	2a. & 3. & 4. >1.68 & 5.	*
(Shimazu et al., 2007)	1. Ohsaki National Health Insurance (NHI) 2. Cohort study 3. 40547 4. Japan	1a. 2.40 4. Scale of 5:(almost never to almost every day	*	1.>|0.25|& 2a 3. & 4 >1 & 5	*
(Kim et al., 2004)	1. Japan Public Health Center (JPHC) study 2. Cohort study 3. 54,498 residents 27,063 men and 27,435. 4. Japan	1a 2.60	1.	2a. & 3. & 4>1.5 & 5	*
(Uusitalo et al., 2005)	1. NCD Prevention Programme 2. Cohort study 3. 561 men and 554 women 4. Mauritius	1a. 2.67 4 Scale of 8: from "never/seldom" to "4 times/d."	*	2a & 3. & 4 >1 & 5	*
(Balder, Goldbohm & van den Brandt, 2005)	1. The Netherlands Cohort Study on Diet and Cancer 2. Cohort study 3. 58279 4. Netherlands	1a 1c. 2.150 3.51	*	2a. & 3. & 4 >1 & 5	*
(Brantsaeter et al., 2009)	1. Norwegian Mother and Child Cohort Study (MoBa) 2. Cohort study 3. 23423 4. Norway	1a. 2.255 3.58	*	1. > |0.3| & 2a & 3. & 4. > 1.6	*
(Engeset et al., 2009) (Engeset et al., 2005)	**1. NEPIC part of Norwegian Women and Cancer (NOWAC) 2. Cohort study 3. 37 212(2009) 35553 (2005) 4. Norway**	**1a. 2.86 3.50**	*	**1. >|0.30| & 2a. & 3. & 4. > 1.5**	*
(Noel et al., 2009)	1. Boston Puerto Rican Health Study 2. Cohort study 3. 1167 4. Puerto Rico	1a. 2.126 3.34	1.	2a. & 3. & 4.	1.

(Butler et al., 2008)	1. Singapore Chinese Health Study 2. Cohort study 3. 63257 4. Singapore	1a. 2.165	*	1. >|0.3| & 2a. & 3. & 5.	*
(Butler et al., 2006)	1. Singapore Chinese Health Study 2. Cohort study 3. 623 4. Singapore	1a. 2.165	*	1.>|0.30 | & 2a & 3. & 4 >1 & 5	*
(Lopez et al. 2009)	1. Seguimiento Universidad de Navarra (SUN) study 2. Cohort study 3. 11195 4. Spain	1a. 2.136	*	1. >|0.30| & 2a & 3. & 4. > 1 & 5.	*
(Cuco et al., 2006)	1. * 2. Cohort study 3. 11 000 4. Spain	1c.	3	1.>|0.20 | & 2a 3. & 4 >1 & 5	*
(Akesson et al., 2007) **(Newby et al., 2006ba)** **(Newby et al., 2006b)** **(Weismayer, Anderson & Wolk, 2006)** **(Rashidkhani et al., 2005)** **(Khani et al., 2004)**	1. Swedish Mammography Cohort 2. Cohort study 3. 24444 (2007) 33840(2006a) 33840(2006b) 66651(2006c) 61431(2005) 66651(2004) 4. Sweden	**1a.** **2.67 in 1987 and 97 in 1997** **3.26 food groups(2005)** **4. Scale of 8: from "never/seldom" to "4 times/d."**	**1.** **3**	**1>|0.15| & >|0.20 |& 2a 3. & 4 >1 & 5**	**1. & 2. & 7. & 11.**
(Robinson et al., 2009)	1. Hertfordshire Cohort Study 2. Cohort study 3. 3217 4. UK	1d. 3.51	*	1.>0.15	*
(Akbaraly et al., 2009)	1. Whitehall II study 2. Cohort study 3. 3486 4. UK	1a. 2.127 3.37 4. Scale of 9: from 'never, or less than once per month' to 'six or more times per day'.)	1.	*	*
(Ambrosini et al., 2009)	1. Raine Study 2. Cohort study 3. 2900 4. Australia	1a. 2.212 3.38	*	1. >|0.30| & 2a & 3. & 4. > 1. & 5.	*

(Hamer & Mishra, 2010)	1. Low Income Diet and Nutrition Survey (LIDNS) 2. Cohort study 3. 3728 4. UK	1b. 2.51	*	1. >|0.30| & 2a.	2.
(Shaheen et al., 2009) **(Wiles et al., 2009)** **(Feinstein et al., 2008)** **(Northstone, Emmett & Rogers, 2008a)** **(Northstone, Emmett & Rogers, 2008b)** **(Northstone & Emmett, 2008b)** **(Northstone & Emmett, 2005)**	1. Avon Longitudinal Study of Parents and Children (ALSPAC) 2. Cohort study 3. 14062(shaheen 2009) 951(Wiles 2009) 13988(Feinstein 2008) 14 541 (Northstone 2008a) 6271 (Northstone 2008b) 12035 ((Northstone and Emmet 2008b) 8515 ((Northstone & Emmett, 2005) 4. UK	**1a.** **2. 90** **3. 43 (Shaheen 2009)** **34 to 41 (Wiles 2009)** **43 to 54 (Feinstein 2008)** **44 (Northstone 2008a)** **44(Northstone 2008b)** **44(Northstone and Emmet 2008b)** **(Northstone 2005)** **4. Scale of 5: from never or rarely to more than once a day**	*	**1.>|0.30| & 2a 3. & 4 >1 & 5**	*
(Borland et al., 2008) **(Robinson et al., 2007)** **(Crozier et al., 2006)** **(Robinson et al., 2004)**	1. Southampton Women's Survey (SWS) study 2. Cohort study 3. 6129 (2008 study) 434 (2007 study) 617(2006 study) 3779 (2004 study) 4. UK	**1a.** **2.100** **3.49**	**3**	**1.>|0.20 | & 2a & 3. & 4 >1 & 5**	*
(McNaughton et al., 2007)	1. The Medical Research Council (MRC) National Survey of Health and Development (NSHD, also known as the 1946 British Birth Cohort) 2. Cohort study 3. 1265 4. UK	1 c. 3.126	*	1.>|0.25 | & 2a & 3.& 4 >1 & 5	*
(Agurs-Collins et al., 2009)	1. Black Women's Health Study (BWHS) 2. Cohort study 3. 1144 incident Cases 4. USA	1a. 2.69 3.29 4.Scale of 9 : from "never or 1 per month" to "2 or more per day" for each food	*	1. >|0.40| & 2a. & 3. & 4 & 5.	*

(Cutler et al., 2009)	1. Project EAT study 2. Cohort study 3. At time 1, 4746. At Time 2, 2516 participants 4. USA	1a. 2.152	*	2a. & 3. & 4.>1	*
(Deshmukh-Taskar et al., 2009)	1. Bogalusa Heart Study (BHS) 2. Cohort study 3. 995 4. USA	1a. 2.31 4. Scale of 6: from 'never or less than once a month' to 'five or more times per day.'	1.	1. > |0.30 | & 3. & 4. >1.	*
(Erber et al., 2010) (Takata et al., 2007) (Park et al., 2005)	**1. Multiethnic Cohort (MEC) study 2. Cohort study 3. 36256 men and 39256 women (2010 study) 3512(2007 study) 195278 (2005 study) 4. USA**	**1a. 2.200**	*	**1.>|0.60 | & 2a. & 3. & 4 >1.25 & 5**	**7**
(Imamura et al., 2009)	1. Framingham Offspring Study 2. Cohort study 3. 2,879 4. USA	1a. 2.126 3.40 4.Scale of 9	*	3. & 4. > 1.0	*
(Qi et al., 2009) (Varraso et al., 2007b) (Wu et al., 2006) (Michaud et al., 2005) (Wu et al., 2004a)	**1. The Health Professionals Follow-Up Study (HPFS) 2. Cohort study 3. 1996 Cases with 1337 controls (2009 study) 42917 (2007 study) 47725 (2006 study) 51529(2005 study) 20888 (2004 study) 4. USA**	**1a. 2.131 3.40 4. Scale of 9: from "almost never" to "6 times/d."**	*	**1.>|0.30 |& 2a 3. & 4 >1 & 5**	*
(Lutsey et al., 2009)	1. Iowa's Women Health Study (IWHS) 2. Cohort study 3. 1950 4. USA	1a. 2.127 3.39 4. Scale of 9: from never or 1 serving per month to ≥6 servings per day.	*	2a. & 4. >2.0 & 5.	*
(Nettleton et al., 2009) (Nettleton et al.,	**1. Multi-Ethnic Study of Atherosclerosis(MESA) 2. Cohort study**	**1a. 2.120 3.47**	**1.**	**2a. & 3. & 4 >1 & 5**	**8.**

2008b) **(Nettleton et al., 2008a)** **(Nettleton et al., 2007)**	**3.** **5316 (2009 study)** **5011 (2008b study)** **5042 (200a study)** **5089 (2007 study)** **4. USA**	**4. Scale of 9: from "rare or never" to a maximum of "2 times per day**			
(Reedy et al., 2010) **(Flood et al., 2008)**	**1. National Institute of Health NIH–AARP Diet and Health Study,** **2. Cohort study** **3.** **492306 (2009 study)** **49282 (2008 study)** **4. USA**	**1a.** **2. 204** **4. Scale of 10 : "never" to " 6 times/d" for beverages and from "never" to " 2 times/d" for solid foods as**	**1. & 2. & 3.**	**2a. & 3. & 5.**	*****
(Chang et al., 2008)	1. California Teachers Study 2. Cohort study 3. 311 4. USA	1a. 2.112	*	2a. & 3. & 4. & 5.	*
(Heidemann et al., 2008) **(Varraso et al., 2007a)** **(Schulze et al., 2006)** **(Zhang et al., 2006)** **(Fung et al., 2005)** **(Kroenke et al., 2005)** **(Fung et al., 2004a)** **(Fung et al., 2004b)** **(Lopez-Garcia et al., 2004)**	**1. NHS I and II (Nurses' Health Study)** **2. Cohort study** **3. 72 113(2008)** **51670 (2006a)** **13110(2006b)** **71058(2005a)** **2619(2005b)** **69554(2004a)** **71768(2004b)** **121700(2004c)** **72043(2007)** **4. USA**	**1a.** **2.116 and 133** **3.37 to 39** **4. Scale of 9: from "never or less than once per month" to "6 or more times per day.**	*****	**2a. & 3. & 4. > 1, >2.75 & 5.**	**11**
(Lutsey, Steffen & Stevens, 2008)	1. Atherosclerosis Risk in Communities (ARIC) study 2. Cohort study 3. 3782 incident Cases 4. USA	1a. 2.66 3.32 4 Scale of 9: from never or 1 time a month to 6 times a day.	*	1. >0.2. & 2. & 3. & 4 >2.0 & 5	*
(Tseng et al., 2008)	1. Minnesota Breast Cancer Family Study 2. Cohort study 3. 3147 4. USA	1a. 2. 153 4. Scale of 8: from ''never or less than once per month'' to ''six or more times per day.''	1.	1. >0.20 & 2. & 3. & 4. >1.	7.

(Meyerhardt et al., 2007)	1. NCI sponsored Cancer and Leukemia Group B (CALGB) 2. Cohort study 3. 1009 4. USA	1a. 2.131 3.39 4. Scale of 9: from never to 6 or more times per day.	*	4.>1.5 & 5.	*
(Velie et al., 2005)	1. Breast Cancer Detection Demonstration Project (BCDDP) 2. Cohort study 3. 280000 4. USA	1a 2.61	1 & 2 & 3	1.>|0.20 | & 2a & 3. & 4 >1 & 5	*
(Newby et al., 2004) **(Newby, Muller & Tucker, 2004**	**1. Baltimore Longitudinal Study (BLS)** **2. Cohort study** **3. 459** **4. USA**	**1a.** **2.100** **3.49**	*	**1.>|0.25 | & 2a & 3 & 4 >1 & 5**	*
(Tseng et al., 2004)	1. NHANES 2. Cohort study 3. 136 4. USA	1a 2.26	*	1.>|0.20 | & 2a & 3. & 4 >1 & 5	
(Lau et al., 2008)	1. Inter99 study 2. population based randomised intervention study 3. Baseline data of 3372 women and 3191 men (30–60 years old) from the population-based survey Inter99 was used. 4. Denmark	1a. 2.198 3.34	*	1. >0.40 & 2b.	1. & 11.
(Cottet et al., 2005)	**1. European Cancer Prevention (ECP) Intervention Study** **2. a randomized trial of calcium and fiber supplementation** **3. 277 men and 165 women** **4. Europe**	1a. 3.50	*	1.>|0.25 | & 2a & 3. & 4 >1 & 5	6.
(Kesse-Guyot et al., 2009)	1. SU.VI.MAX (Supple´mentation en Vitamines et Mine´rau Antioxydants) study 2. a randomized,double-blind, placebo-controlled primary prevention trial 3. 2463 women and 2731 men. 4. France	1b.	*	1. >|0.30| & 2a & 3. & 4. > 1. & 5.	1.
(Murtaugh et al.,	1. * 2. Population-based control	1a. 3. 68	*	1.>|0.35 |& 2a &3. & 4 >2 & 5	3. & 8.

2007)	participants 3. 871 Hispanic and 1599 non-Hispanic 4. USA				

*Information are not provided

Table C. Number of dietary patterns, label of dietary patterns, foods that correlated highly with empirically derived dietary patterns in each study. Furthermore associations of dietary patterns with disease outcomes and socio-demographic characteristics are presented (Sorted alphabetically by health outcome).

Author	**Investigated health outcome in each study**	**Number of dietary patterns , label of dietary patterns and foods that correlated highly with the dietary patterns** **1. Number of dietary patterns (median: 3, IQR: 2-4)** **2. Label of each pattern/foods and food groups that correlated highly with this pattern** **3. Percentage of total variance of original food items being explained by the dietary patterns in each study. (median: 24%, IQR: 19.9-31.3).**	**Effect of dietary pattern on investigated health outcomes in each study** **Dietary pattern label: (Effect estimate , 95% CI, P for trend)**	**Direction of association of dietary patterns with socio-demographic variables in each study**
(Erber et al., 2010)	Type 2 Diabetes	1. 3 2. **Fat and Meat**: fat, meat, eggs, and cheese **Vegetables** : vegetables and also fruits **Fruit and Milk**: milk, yogurt, cheese, and fruits 3. 30%	**Fat and Meat**: significantly associated with diabetes risk in men (HR=1.40; 95% CI: 1.23- 1.60; P for trend<0.0001) and women (HR=1.22; 95% CI: 1.06- 1.40; P for trend=0.004) when comparing extreme quintiles) **Other patterns :** Not Statistically Significant	*
(Imamura et al., 2009)	Type 2 Diabetes	**1. 3** **2. Western:** High fat dairy, eggs , processed meat, refined grains, meat, potato, fried foods **Prudent:** reduced fat dairy , fruits , vegetables , sweet baked goods **Alcohol :** white wine , fruits , vegetables ,fish and vegetables	**Alcohol:** inversely associated with T2D risk (HR= 0.33, 95% CI: 0.17-0.64) **Other patterns :** Not Statistically Significant	*
(Qi et al., 2009)	Type 2 Diabetes	1. 2 2. **Prudent**: vegetables, fruit, legumes, whole grains, fish, and poultry. **Western**: processed meat, red meat, butter, high-fat dairy products, eggs, and refined grains	**Western:** A significant interaction (P < 0.02) was observed between the T2D. The multivariable odds ratios (ORs) of T2D across increasing quartiles for the Western dietary pattern were 1.00, 1.23 (95% CI: 0.88- 1.73), 1.49 (1.06-2.09), and 2.06 (1.48-2.88) among men with a high GRS (12 risk alleles; P for trend < 0.01). **Other patterns :** Not Statistically Significant	**Prudent:** positively associated with older age, physical activity, and inversely associated with smoking **Western** : positively associated with drinking and smoking
(Kim et al., 2008)	Type 2 Diabetes	1. 3 2. **Factor 1**: fruits, vegetables, and bean products. **Factor 2** : corned beef and cabbage, rice,	**Factor 2:** positively associated with T2DM (OR 1.30, 95% CI 1.03–1.68)	**Factor 1** : negatively correlated with BMI, smoking, and positively correlated with years of education and physical activity **Factor 2** : positively correlated with BMI, smoking,

		steamed shell fish, Filipino ethnic foods and local Hawaiian foods **Factor 3:** French fries, fast-foods hamburgers, pizza, chips, soda, pasta, and salad dressings.		and negatively correlated with income level, years of education, physical activity **Factor 3** : positively correlated with income and education and negatively correlated with BMI
(McNaughton et al., 2008)	Type 2 Diabetes	1. 3 2. **Fruit, salad, cereals, and fish** **High fat and sugar** **Vegetables** 3. 11.9, 5.9, and 3.9%	**Fruit, salad, cereals, and fish** : inversely associated with diastolic blood pressure (P<0.001) **High fat and sugar**: associated with increased risk of type 2 diabetes (HR top quartile 2.95 (95% CI: 2.19 –3.97); adjusted for age, sex, and energy misreporting). This relationship was attenuated after adjustment for ethnicity, employment grade, health behaviours (smoking, alcohol use, and physical activity) but remained significant after further adjustment for blood pressure and BMI (HR: 1.51 (95%CI: 1.10 –2.09)). **Other patterns**: Not Statistically Significant	**High fat and sugar** : positively associated with being male **Vegetables** : positively associated with rural region of residence **Fruit, salad, cereals, and fish** : inversely associated with age
(Nanri et al., 2008)	Type 2 Diabetes	1. 4 2. **Healthy:** vegetables, fruit, soy products, fish, and yogurt. **High-fat**: meat, processed meat, mayonnaise, and egg **Seafood:** seafood, shellfish, salted fish guts, fish roe, and fish-paste products **Westernized breakfast**: bread, margarine, and coffee and low intakes of rice and miso soup 3. 30.5%	**Westernized breakfast**: inversely related to A1C concentrations (P for trend 0.02 in both men and women); the multivariate-adjusted ORs for the highest versus lowest quintiles were 0.60 (95% CI: 0.43– 0.84) and 0.64 (95%CI: 0.46–0.90) for men and women, respectively. **Seafood** : positively associated with A1C concentrations in men only (P trend 0.01) **Other patterns:** Not Statistically Significant	**Healthy:** positively correlated with older age, physical activity, leisure time, and negatively associated with smoking status and alcohol drinking. **High-fat:** associated positively with younger age and smoking. **Seafood**: positively associated with alcohol drinking and higher BMI. **Westernized breakfast:** positively associated with younger age , lower BMI, no smoking status and physical activity
(Nettleton et al., 2008a)	Type 2 Diabetes	1. 5 **2. Fats and processed meats** **Vegetables and fish beans, Tomatoes** **Refined grains** **Whole grains and fruit**	**Vegetables and fish beans, Tomatoes** **Refined grains:** associated with an 18% greater risk (HR per 1-score SD 1.18 (95% CI 1.06 –1.32); P trend 0.004) **Whole grains and fruit** associated with a 15% lower diabetes risk (HR: 0.85 [0.76–0.95]; P- trend 0.005.	*
(Montonen et al., 2005)	Type 2 Diabetes	1. 2 2. **Prudent:** fruits and vegetables **Conservative:** butter, potatoes, and whole milk.	**Prudent :** associated with reduced risk of type 2 diabetes (RR 0.72 (95% CI: 0.53-0.97; p for trend = 0.03) **Conservative** : associated with increased risk of type 2 diabetes (RR 1.49 (95% confidence interval: 1.11, 2.00; p for trend = 0.01)	*
(Fung et al., 2004a)	Type 2 Diabetes	1. 2 2. **Prudent**: fruits, vegetables, whole grains, fish, poultry, and low-fat dairy **Western**: red and processed meats, refined grains, sweets and desserts, and high-fat	**Western:** associated with increase drisk of type 2 diabetes RR for type 2 diabetes of 1.49 (95% confidence interval [CI], 1.26-1.76, *P* for trend, <.001) when comparing the highest to lowest quintiles of the Western pattern	*

		dairy products.	**Other patterns:** Not Statistically Significant	
(Varraso et al., 2007a)	Respiratory and Allergic symptoms (COPD)	1. 2 2. **Prudent:** fruit (fresh apples or pears; oranges; peaches, apricots, or plums; strawberries; cantaloupes; blueberries; grapefruits), vegetables (broccoli, eggplant, cauliflower, coleslaw, carrots, raw spinach, celery, string beans, romaine leaf lettuce, yellow squash, cooked spinach, iceberg head lettuce, tomatoes, mushrooms, Brussels sprouts, mixed vegetables, garlic, beans lentils, beets), poultry (chicken or turkey without skin), and fish. **Western:** French fries, hamburger, cured meats (processed meats, hot dogs, bacon), sweets and desserts (home-baked cake, doughnuts, brownies, ready-made sweet rolls, home-baked pies, pancakes or waffles), and refined cereals (white bread, pasta)	**Prudent** : negatively associated with risk of newly diagnosed COPD (relative risk (RR) for highest compared with lowest quintile: 0.75; 95% CI: 0.58- 0.98; P for trend 0.02) **Western:** positively associated with risk of COPD (RR for highest compared with lowest quintile: 1.31; 95% CI: 0.94-1.82; P for trend 0.02.	**Prudent**: positively associated with physically activity lower body mass index, non-smoking, and being a female **Western**: positively associated with being men, white, and current smoker.
(Butler et al., 2006)	Respiratory and Allergic Symptoms (new cough in phlegm)	1. 2 2. **Vegetable–fruit–soy:** vegetable, fruit, and soy food intake; of the 32 foods included in the pattern, 23 were vegetables, five were soyfood items, and four were fruit items. **Meat–dim sum:** chicken, pork, fish, rice and noodle dishes, and preserved foods	**Meat–dim sum**: positively associated with new-onset cough with phlegm (OR, 1.43; 95% CI: 1.08-1.89, p for trend 0.02). **Other patterns:** Not Statistically Significant	**Vegetable–fruit–soy**: was positively correlated with people who were less likely to, to smoke, and to lack formal education. **Meat–dim sum** : : was positively correlated with people who were more likely to be younger, male, current smokers, and have had formal education
(Hooper et al., 2008)	Respiratory and allergic symptoms (asthma, atopy)	1. 2 2. **Urban/rural dimension:** Foods which were strongly negatively correlated with the (characteristic of rural diets) were pumpkin leaves, young pumpkin, cabbage and wild leaves and berries. Foods which were strongly positively associated with the urban/rural dimension of, diet (characteristic of urban diets) were fried potatoes, carrots, tinned fruit salad, chicken, sausages, yoghurt, packet custard and jelly. **Second convenience rotated dimension** 3. 25%	The **Urban** component of diet was strongly associated with positive skin tests even after adjusting for urban residence (OR: 2.1;95CI: 1.2–3.7, P-value= 0.009) **Other patterns**: Not Statistically Significant	*

(Bakolis et al., 2010)	Respiratory and Allergic Symptoms (Asthma)	1. 5 2. **Prudent:** wholemeal bread and rolls, yoghurt, cheese, fish, salad vegetables, pasta, couscous, vegetable dishes and different types of dressing. **Vegetable and Fruit**: vegetables and fruits **Western:** white bread and rolls, chips, roast potatoes, baked beans, processed meats, bacon, ham, crisps, meat dishes, fried snacks, chocolate bars, sponge puddings and cakes, ketchup, coke. **Traditional**: vegetables, pork, beef, liver and lamb, and a low intake of naan paratha and Bombay mix. **Vegetarian**: cream crackers, cre`me fraiche, macaroni cheese , chick peas, houmous, lentils, nut roast, vegetables, nuts and seeds 3. 25%	**Vegetarian:** positively associated with asthma [adjusted odds ratio comparing top vs. bottom quintile of pattern score (OR= 1.43, 95% CI: 0.93–2.20), P trend 0.075). T**raditional:** associated negatively with asthma [OR= 0.68, 95%CI: 0.45–1.03., P trend< 0.071). **Prudent**: positively associated with chronic bronchitis (OR= 2.61 (95%CI: 1.13–6.05, P trend 0.025). **Other patterns :** Not Statistically Significant	*
(Hooper et al., 2010)	Respiratory and allergic symptoms (asthma)	1. 2 **2. Meat and potato:** meat, beef, pork, bacon, sausage and fried egg/scrambled egg/omelette intake at all the centres and also with intake of potato or chips, bread, butter, biscuits and cakes. **Fish, fruits and vegetables:** several fruits and less consistently with intakes of a number of vegetables and fish. **3.** 11.2%	**All the patterns :** Not Statistically Significant	*
(Shaheen et al., 2009)	Respiratory and Allergic Symptoms (Asthma)	1. 5 2. **Health conscious:** salad, fruit, fruit juices, rice, pasta, oat/bran based breakfast cereals, fish, pulses, cheese, non-white bread **Traditional:** vegetables, red meat, poultry **Vegetarian:** meat substitutes, pulses, nuts, herbal tea **Processed**: meat pies, sausages, burgers, fried foods, pizza, chips, crisps, white bread, eggs, baked beans **Confectionery:** chocolate, sweets, biscuits, cakes, puddings 3. (31.6%)	*	**Health conscious:** positively associated with education, age and non-white women and negative associated with parity, being single, non-working women, smokers and overweight before pre-pregnancy. **Processed:** Opposite associations
(Varraso et al., 2009)	Respiratory and Allergic Symptoms (Asthma)	1. 3 **2. Prudent:** fruits and vegetables **Western:** pizza/salty pies, dessert, cured meats and pasta	**Western**: associated with an increased risk of reporting frequent asthma attacks (highest versus lowest tertile odds ratio (OR) 1.79, 95%confidence interval (CI) 1.11–3.73).	*

		Nuts and Wine : nuts and seeds, salty biscuits, olives, wine, and fortified wine	**Nuts and Wine**: was associated with a decreased risk of reporting frequent asthma attacks (highest versus lowest tertile OR 0.65, 95% CI 0.31–0.96). **All the patterns:** Not Statistically Significant	
(Varraso et al., 2007b)	Respiratory and allergic symptoms (COPD)	1. 2 2. **Prudent:** fruits, vegetables, fish, poultry and whole grain products **Western:** refined grains, cured and red meats, desserts and sweets, French fries, eggs and high-fat dairy products.	**Prudent** : inversely associated with the risk of newly diagnosed COPD (RR for highest vs. lowest quintile 0.50 (95% CI 0.25 - 0.98, p for trend = 0.02) **Western** : positively associated with the risk of newly diagnosed COPD (RR for highest vs. lowest quintile 4.56 (95% CI: 1.95- 10.69, p for trend < 0.001)	**Prudent** :positively associated with people who were more physically active, less likely to be current smokers and took more multivitamin supplements, **Western** : positively associated with higher BMI, less physical activity, less smoking and fewer multivitamin supplements consumption
(Takaoka & Norback, 2008)	Respiratory and Allergic Symptoms (wheeze and respiratory infections)	1. 5 2. **The first factor**: fruit, raw vegetable and cooked vegetable **The second factor** : fast food, soft drink and juice **The third factor** : meat, fish and seafood **The fourth factor** : milk and yoghurt consumption **The fifth factor** : butter and rapeseed oil	**Second factor (fast food, juice and soft drinks)** : positively related to wheeze and respiratory infections (OR:1.19 95%CI: (1.04–1.37), P trend < 0.01) **Fifth Factor :** positively related to wheeze and respiratory infections (OR:2.17 (1.31–3.59) P trend < 0.003)	
(Bamia et al., 2007)	Overall Mortality	1. 1 **2. plant-based** 3. 14.6%	**Plant-based**: associated with a lower overall mortality, a one standard deviation increment corresponding to a statistically significant reduction of 14% (95% confidence interval 5–23%).	*
(Masala et al., 2007)	Overall mortality	1. 4 2. **Prudent:** cooked vegetables, legumes, fish, and seed oil **Pasta & Meat**: pasta and other grains, tomato sauce, red and processed meats, added animal fat, white bread and wine; on the other hand, this pattern showed a low consumption of yoghurt. **Olive Oil & Salad:** olive oil as added fat, raw vegetables (tomatoes, leafy and root vegetables), soups and white meat (chicken and turkey) **Sweet & Dairy**: sugar, cakes, ice-cream, coffee, eggs, butter, milk and cheese. 3. 21%	**Olive Oil & Salad**: inversely associated with overall mortality. After adjustment for gender, age and caloric intake, overall mortality was reduced by approximately 50% in the highest quartile and a significant trend emerged (P<0·008). **Other patterns:** Not Statistically Significant	**Prudent**: positively associated with people who were more frequently females, more educated, more likely to be single, former smokers and obese **Pasta & Meat:** positively associated with people who were more likely among married males, current smokers, overweight or obese subjects. **Olive oil & Salad:** positively associated with people who were more frequently males, married, with a higher school education, leaner and more physically active. **Sweet & Dairy**: positively associated with people who were more likely to have a higher education, to be more physically active and with a normal weight.
(Waijers et al., 2006)	Overall mortality	1. 3 2. **Mediterranean-like:** pasta and rice, sauces, fish, and vegetables in combination with	**Healthy Traditional** was associated with a lower mortality rate (Women in the highest tertile of this pattern had ab30%	**Mediterranean-like:** positively associated with people who were younger, higher educated, and

		vegetable oils, wine, and other cereals. Foods such as potatoes, bread, and margarine, contributed negatively to this component. **Traditional Dutch:** meat, potatoes, vegetables, eggs and alcoholic beverages. It was low in intakes of dairy products, sweets, and pastries. **Healthy Traditional:** vegetables, fruit, dairy products, potatoes, and legumes, and also non-alcoholic beverages. It was low in intakes of butter and alcoholic beverages 3. 25%	lower mortality risk than those in lowest tertile (95% CI for the hazard ratio: 0.52- 0.95) **Other patterns :** Not Statistically Significant	more often former smokers **Traditional Dutch**: positively associated with people who had a lower level of education, were more current smokers, and were more overweight. **Healthy Traditional:** positively associated with people who were less educated, more likely nonsmokers, had higher BMIs, and were more physically active
(Vujkovic et al., 2009b)	Other(Spina Bifida)	1. 1 **2. Mediterranean**	**Mediterranean:** positively associated with increased risk of spina bifida of offspring in mothers (OR= 3.5 (95% CI: 1.5–7.9).	**Mediterranean:** positively associated with higher maternal age at birth of the index child, higher education and more alcohol consumption in the preconception period
(D'Souza et al., 2008)	Other(Chron's Disease)	1. 4 2. **Traditional Western (girls)**: meat, fried food items, fast foods, snacks, and desserts **Prudent (girls):** vegetables, fruits, dairy products, eggs, olive oil, dark breads, grains, fish, and nuts **Cheese-Snack (girls):** cheese, snacks, and desserts and this. **Beverage (girls):** beverages (tea, coffee, coke, and milk-shakes), some organ meats, and salsa **Partial Western (boys)**: beverages, fast foods, snacks, white bread, meat sandwiches, and dessert items. **Prudent(boys):** vegetables, fruits, yogurt, olive oil, fish white rice, tofu, grains, and nuts, a pattern that was similar to that in girls **Avoidance (boys)**: avoidance of fast foods and snacks. **Meat(boys):** beef and pork, mashed potatoes, and avoiding dark bread	**Traditional Western:** associated with Chorn's Disease (OR= 4.7, 95% CI 1.6 –14.2) **Prudent**, : inversely associated with CD in both genders (girls: OR =0.3, 95% CI 0.1– 0.9; boys: OR= 0.2, 95% CI 0.1– 0.5)	*
(Kubo et al., 2008)	Other(Barrett's Esophagus)	1. 2 2. **Western**: French fries, pizza, hamburgers,	**health-conscious:** inversely associated with Barrett's esophagus (OR : 0.35, 95% confidence interval: 0.20- 0.64)	*

		and tacos), soft drinks, beer/liquor, and coffee and was low in tofu, cooked cereals, fruits, and water. **Health-conscious**: fruits and vegetables, non-fried fish, and tofu and was low in meat, salty snacks, fried foods, and soft drinks. 3. 12.7%	**Other patterns**: Not Statistically Significant	
(Lutsey et al., 2009)	Other (venous thromboembolism)	1. 2 2. **Prudent** : vegetables, fruit, and poultry **Western** : processed meat, non-cereal whole grains, and added fats and oils	**All the patterns:** Not Statistically Significant	**Prudent** : positively associated with people who were physically active, less likely to be current smokers and took more multivitamin supplements, **Western** : positively associated with people who had higher BMI, were less physically active, were more likely to smoke and took fewer multivitamin supplements than men
(Fung et al., 2004b)	Other (Stroke)	Same as (Fung et al., 2004a)	**Western:** associated with an increased relative risk (RR) of 1.58 (95% CI, 1.15 -2.15; *P*<0.0002 for trend) for total strokes and 1.56 (95% CI, 1.05 - 2.33; *P*<0.02 for trend) **Prudent**: associated with decreased relative risk (RR) of 0.78 (95% CI, 0.61- 1.01) for total stroke and 0.74 (95% CI, 0.54 - 1.02) for ischemic stroke.	*
(Vujkovic et al., 2009a)	Other (semen quality)	1. 2 **2. Health Conscious:** fruits, vegetables, fish and whole grains **Traditional Dutch:** meat, potatoes and whole grains and low intakes of beverages and sweets	**Health Conscious:** inversely correlated with tHcy in blood (b = -0.07, P < 0.02) and seminal plasma (b = -1.34, P < 0.02) and positively with vitamin B6 in blood (b =0.217, P <0.01) **Traditional Dutch:** positively correlated with red blood cell folate (b = 0.06, P < 0.04) and sperm concentration (b = 13.25, P = 0.01).	*
(Feinstein et al., 2008)	Other (School attentaiment)	1. 2 2. **Junk food:** high-fat processed foods (sausages, burgers and poultry products), snack foods high in fat and/or sugar (such as crisps, sweets, chocolate, ice lollies and ice creams) fizzy drinks and the number of takeaway meals **Health Conscious**: vegetarian foods, nuts, salad, rice, pasta, fruit, cheese, fish, cereal	**Junk food :** negatively associated with the level of school attainment(P<0.05)	*
(Rashidkhani et al., 2005)	Other (Renal Cell Carcinoma)	1. 3 2. **Healthy**: vegetables, tomato, fish, fruits, poultry, and whole grains. **Western**: sweets, processed meat, refined grains, margarine/butter, high fat dairy products, fried potatoes, soft drinks, and	**Drinker**: associated with decreased risk of Renal Cell Carcinoma risk (RR comparing the 2nd and 3rd with the first tertile, 0.56; 95% CI, 0.34–0.95; and 0.72; 95% CI, 0.42–1.22, respectively, P 0.08 by Wald test). **Other patterns:** Not Statistically Significant	*

		meat **Drinker:** Alcoholic beverages (wine, liquor, beer) and snacks 3. 25%		
(Takata et al., 2007)	Other (mammographic density)	1. 2 **2. Fat and meat.** **Vegetables Fruit and milk**	**Fat and meat**: positively associated with mammographic densities than those with lower scores (P or trend=0.21) **Vegetables Fruit and milk**: weakly inversely associated with mammographic densities only among Japanese women (P for trend = 0.13 and 0.03, respectively).	*
(Kontogianni et al., 2009)	Other (lumbar spine bone mineral density**)**	**1. 10** **2. Component 1: dairy, cereals, red meat, and olive oil consumption** **Component 2: fish and olive oil and low intake of red meat and products** **Component 3: poultry and nuts and low intake of red meat and red meat products** **Component 4,5: alcohol** **Component 6:legumes** **Component 7: Sweets** **Component 8: .fruit drinks** **Component 9: Coffee** **Component 10: Soft Drinks** **3. 80%**	**Component 3:** positively associated with lumbar spine bone mineral density (beta coefficient= 0.185, P < 0.017 and total body bone mineral content (beta coefficient= 0.140 P < 0.048). **Other patterns :** Not Statistically Significant	*
(Chen et al., 2006)	Other (Hypertension)	1. 3 2. **Balanced:** steamed rice, red meat, small fish, fruit, and vegetables **Animal protein:** fish, eggs, milk, poultry, red meat (beef and mutton), bread, and fruit. **Gourd and root vegetable:** squashes, pumpkin, sweet potato, radish	**Balanced:** associated with decreased risk of hypertension (Adjusted prevalence odds ratios for general hypertension were 1.00 (reference), 0.81 (95% CI: 0.79, 0.97, 0.82 (0.68, 0.97, 0.79 (0.66, 0.94, and 0.71 (0.59, 0.85 (P for trend 0.01)) **Animal protein:** associated with increased risk of hypertension (Prevalence odds ratios for general hypertension were 1.00 (reference), 1.30 (1.01, 1.52., 1.20 (1.01, 1.47, 1.22 (1.00, 1.44, and 1.21 (1.03, 1.49 (P for trend = 0.23)) **Other patterns:** Not Statistically Significant	**Animal protein** : positively associated with the prevalence of cigarette smoking and markers of socioeconomic status, including educational attainment, television ownership, and land ownership **Gourd and root vegetable** : inversely associated with television ownership
(He et al., 2009)	Other (higher glucose abnormalities)	1. 4 2. **Green Water:** (like the rice area in the Southeast) **Yellow Earth**: their food is mainly produced on the dry and hilly land, like the mountain area in the Northwest **New Affluence**	**New Affluence and Yellow Earth:** positively associated with higher glucose abnormalities **(**prevalence ratio 1.22 (95% CI 1.04 –1.43)) and 2.05 (1.76 –2.37).	*

		Western Adopter : Western- oriented food style		
(Mizoue et al., 2006)	Other (Glucose tolerance abnormality)	1. 3 2. **DFSA (high-dairy, high-fruit and -vegetable, high-starch, low-alcohol):** fermented dairy products, milk, confectioneries, bread, fruits, and vegetables, a local alcoholic beverages **Animal food**: various kinds of animal foods, including red meat, poultry, seafood excluding fish, processed meat and fish products, and fried or broiled foods. **Japanese:** soybean products, seaweeds, pickles, and green tea, vegetables, and fish 3. 24%	**DFSA (high-dairy, high-fruit and -vegetable, high-starch, low-alcohol):** significantly and inversely associated with a glucose tolerance abnormality (impaired fasting glucose, impaired glucose tolerance, or type 2 diabetes) for the 2nd, 3rd, and 4th quartiles were 0.80 (95% CI : 0.62–1.04), 0.71 (95% CI : 0.54–0.92), and 0.51 (95% CI : 0.38–0.67), respectively, compared with the lowest quartile). **Other patterns :** Not Statistically Significant	**DFSA (high-dairy, high-fruit and -vegetable, high-starch, low-alcohol):**:positively associated with higher levels of leisure-time physical activity and smaller amounts of alcohol, nonsmoking status **Animal food:** positively associated with higher BMI and consumed larger amounts of alcohol **Japanese:** positively associated with higher levels of leisure-time physical activity, consumed greater amounts of alcohol, and had a higher proportion of nonsmokers.
(Tseng et al., 2008)	Other (Breast Density)	1. 3 **2. Fruit–vegetable– cereals** **Salad–sauce–pasta/grain:** pasta, rice, and such salad and sauce vegetables as mushrooms, garlic, peppers, lettuce, onions, and tomatoes. **Meat-starch:** French fries, fried chicken and fish, meat, white bread, cheese, eggs, and sweets.	**Fruit–vegetable–cereal**: inversely associated with breast density among premenopausal women (b = -0.13, p = 0.09; interaction p = 0.009) and current smokers, (b = -0.30, p = 0.02; interaction p = 0.05), while the salad–sauce–pasta/grain was inversely associated with breast density among current smokers (b = -0.27, p = 0.06; interaction p = 0.006).	**Fruit–vegetable–cereal and salad–sauce–pasta/grain :** inversely associated with age, college education, living in a large city or a suburb of a large city, with former rather than never smoking, and was positively associated with alcohol intake **Meat–starch:** positively associated with those who were younger, less well educated, more likely to live in a rural area and to smoke and less likely to use multivitamins or to exercise.
(Shin, Oh & Park, 2007)	Other (Better health status)	1. 3 2. **Korean healthy:** vegetables, kimchi (spicy raw vegetables), seaweeds, beans, fruits, milk and dairy products **Animal foods:** beef, pork, poultry and fish as well as fast food including hamburgers and pizza. **Sweets**: high intakes of ice cream, sweet drinks, chocolate, sweet baked goods and sugary foods. 3. 27.4%	**Korean healthy :** associated with better health status(as compared with the lowest quintile, the multivariate-adjusted OR of the highest quintile for health status inferior or similar to their peers was 0·59 (95% CI 0·42, 0·84) **Other patterns :** Not Statistically Significant	**Korean healthy**: positively related with people who were more likely to be from the households with higher income and food expenditure, and had mothers with better nutrition attitude. **Animal foods:** positively related with people who tended to be older and overweight and their households spent more money on buying food.
(Ambrosini et al., 2008a)	Other (benign prostatic hyperplasia)	Same as (Ambrosini et al., 2008b)	**Vegetable:** associated with lower risk of having prostatic hyperplasia (OR 0.78, 95% confidence interval 0.63–0.98).	*
(Brantsaeter et al., 2009)	Other (Pre-eclampsia)	1. 4. 2. **Vegetable:** vegetables, cooking oil, olive oil, fruits and berries, rice, and chicken **Processed:** processed meat products, white bread, French fries, salty snacks, and sugar-	**Vegetable:** associated with lower risk of Pre-eclampsia (relative risk (OR) for tertile 3 vs. tertile 1: 0.72; 95% CI: 0.62, 0.85). **Processed:** associated with increased risk (OR for tertile 3	**Vegetable**: positively associated with maternal age, education, and height, supplement use and inversely with BMI, and smoking. **Processed**: inversely associated with maternal age,

		sweetened drinks and high negative loadings on oily fish, high-fiber breakfast cereals, and lean fish. **Potato and Fish Pattern:** cooked potatoes, processed fish, lean fish, fish spread and shellfish, and margarine. **Cakes and Sweets Pattern:** cakes, waffles and pancakes, buns, ice cream, sweet biscuits, sweets, and chocolate. 3. 18%	vs. tertile 1: 1.21; 95% CI: 1.03, 1.42). **Other patterns :** Not Statistically Significant	supplement use education, and height and positively associated with BMI and smoking,
(Nettleton et al., 2008b)	Other (Lower spot urine collection)	1. 4 2. **Fats and Processed Meat:** added fats, processed meat, fried potatoes, and desserts **Vegetables and Fish** :several vegetable groups, fish, soup, Chinese foods, red meat, poultry, and soy), Beans, Tomatoes, and Refined Grains (beans, tomatoes) **Refined grains:** dairy foods, avocado/guacamole, and red meat) **Whole Grains Fruit** (whole grains, fruit, nuts and seeds, green leafy vegetables, and low-fat dairy foods).	**Whole grains, fruit, vegetables,** and **low-fat dairy foods**: associated with 20% lower spot urine collection across quintiles (P for trend 0.004) (renal vascular integrity).	*
(Okubo et al., 2007)	Other (Lower constipation)	1. 4 **2. Healthy** **Japanese traditional** **Western** **Coffee and dairy products**	**Japanese traditional:** associated with a significantly lower prevalence of functional constipation. In comparison with the lowest quintile, the multivariate adjusted odds ratio was 0.52 (95%CI: 0.41-0.66, p for trend < 0.001). **Other patterns :** Not Statistically Significant	*
(Okubo et al., 2008)	Obesity (BMI)	1. 4. 2. **Healthy** : Green Vegetables , White vegetables , Mushrooms , Seaweeds, Potatoes, Fish and selfish, Fruit , Salted Vegetables **Japanese traditional:** Soy products, Rice , Miso soup , Fruit and vegetable juices **Western** : Meats, fats and oils , seasonings , Processed meats, Eggs, butter **Coffee and dairy products**: Eggs , sugary foods , dairy products	**Healthy:** significantly associated with a lower risk of BMI>25 (OR of the highest quintile vs. lowest, 0.57; 95% CI: 0.37–0.87; P for trend 0.05). **Japanese traditional** and **Western:** significantly associated with an increased risk of BMI>25 (OR: 1.77; 95% CI: 1.17–2.67; P for trend o0.01 and OR: 1.56; 95% CI: 1.01–2.40; P for trend: 0.04, respectively). **Other patterns:** Not Statistically Significant	**Healthy:** positively associated with smaller number for current smokers and larger number for dietary supplement users and dieters
(McNaughton et al., 2007)	Obesity (BMI)	1. 3 2. **ethnic foods and alcohol (women) meat, potatoes, and sweet (women) foods; and fruit, vegetables, and**	**Fruit, vegetables, and dairy**: inversely associated with BMI (P < 0.004, waist circumference (P < 0.0007, blood pressure (P: 0.02., and was positively associated with red cell folate (P < 0.03.	*

		dairy(women) **ethnic foods(men)** **alcohol(men)**	**Ethnic foods and alcohol**: inversely associated with blood pressure (P: 0.008, whereas the meat, potatoes and sweet foods pattern was positively associated with glycated hemoglobin (P < 0.01. **Mixed:** inversely associated with waist circumference (P < 0.02. and blood pressure (P <0.01., whereas there were Not Statistically Significant associations with the ethnic foods and alcohol pattern. **Other patterns :** Not Statistically Significant	
(Newby et al., 2006ba)	Obesity (BMI)	1. 4 2. **Healthy:** Vegetables , fruits and fish **Western/Swedish:** meat , potatoes refined grains **Alcohol, Sweets** **High-fat dairy and coffee**	**Healthy:** associated with decreased risk in BMI (b: 20.18 kg/m2 for a 1 unit increase in SD score, CI: 20.26 to 20.10; P, 0.0001), whereas normal weight and overweight women who increased their Healthy pattern score had smaller increases in BMI (20.05 kg/m2 and 20.11 kg/m2, respectively; P, 0.05 for both). **Other patterns :** Not Statistically Significant	*
(Craig et al., 2010)	Obesity	1. 3 **2. Fruit and vegetables** **Snacks** **Fish and sauce** 3. 15%	**Snacks**: lowest factor score in obese children (P for linear trend 0·047). **Fish and Sauce:** highest factor score in obese children (P for linear trend 0·023).	*****
(Paradis et al., 2009)	Obesity	1. 2 2. **Western:** refined grains, French fries, red meats, condiments, processed meats and regular soft drinks **Prudent**: non-hydrogenated fat, vegetables, eggs and fish and seafood.	**Western** : positively associated with obesity (OR:1.82, 95% CI 1.16– 2.87) **Prudent**: inversely associated with obesity (OR: 0.62, 95% CI 0.40–0.96).	**Western:** positively associated with younger age **Prudent** : positively associated with older age
(Esmaillzadeh & Azadbakht, 2008b)	Obesity	Same as(Esmaillzadeh & Azadbakht, 2008a)	**Iranian:** was associated with increased risk of obesity (subjects in the highest quintile had greater odds of being centrally obese, either before (OR = 2.15; 95% CI ¼ 1.18–3.90) or after (OR =2.08; 95% CI: 1.09–3.65) control for confounders)	Same as(Esmaillzadeh & Azadbakht, 2008a)
(Shi et al., 2008)	Obesity	1. 4. 2. **Macho**: animal foods and alcohol **Traditional:** rice and fresh vegetable and inversely on wheat flour **sweet tooth** :cake, milk, yoghurt and drinks **Vegetable- rich:** whole grains, fruits, root vegetables, fresh and pickled vegetables, milk, eggs and fish.	**Vegetable-rich:** independently associated with obesity. Compared with the lowest quartile of vegetable-rich pattern, the highest quartile had higher risk of general obesity (men, prevalence ratio (PR): 1.82, 95% confidence interval (CI): 1.05–3.14; women, PR: 2.25, 95% CI: 1.45–3.49).	**Vegetable- rich :** associated positively with education and negatively with income
(Murtaugh et	Obesity	1. 5 2. **Western :** high-fat dairy foods, refined grains and refined grain snacks, gravy and	**Western** : was associated with higher prevalence of overweight and obesity	*

al., 2007)		sauces, fast foods (fries, beef sandwiches, and chicken), bacon and sausage, potatoes, margarine, polyunsaturated fats, high-fat/high-sugar desserts, and red meats **Native Hispanic:** Mexican cheeses, Mexican soups, Mexican meats, legumes, Mexican tomato-based sauces, and tomato based sauces **Prudent:** low-fat dairy, whole grain cereals, fruits (canned, dried, and fresh), fruit juices, legumes, vegetables, broth soups, and nuts **Mediterranean:** liquor, poultry, fish and shellfish, vegetables, salad, Greens and high-fat salad dressings. **Dieter:** avoiding high-fat dairy products, high-fat salad dressing, cola beverages, and butter, and using low-fat dairy, low-fat margarine, low-fat and fat-free salad dressings, low-fat/high-sugar desserts, diet cola, other diet beverages, and sugar substitutes	**Prudent** : was associated with a 29% lower prevalence of overweight **Other patterns :** Not Statistically Significant	
(Pala et al., 2006)	Obesity	1. 3 2. **Prudent:** cooked vegetables, pulses, cabbage, seed oil and fish **Pasta & meat**: pasta, tomato sauce, red meat, processed meat, bread and wine **Olive oil & salad** raw vegetables, olive oil, soup and chicken **Sweet & dairy** sugar, cakes, ice cream, coffee and dairy 3. 21%	**Pasta & meat** :associated with increased BMI: ($P < 0.001$).	**Prudent:** positively associated with education **Pasta and meat**: inversely associated with education
(Lopez-Garcia et al., 2004)	Obesity	1. 2. 2. **Prudent:** vegetables, fruit, legumes, whole grains, fish, and poultry. 3. **Western:** red meat, processed meat, refined grains, sweets, desserts, French fries, and high-fat dairy products.	**Prudent**: inversely associated with plasma concentrations of CRP ($P<0.02$) and E-selectin (< 0.001) after adjustment for age, body mass index (BMI), physical activity, smoking status, and alcohol consumption. **Western** : showed a positive relation with CRP ($P < 0.001$), interleukin 6 ($P<0.006$), E-selectin ($P<0.001$), sICAM-1 ($P <0.001$), andsVCAM-1($P<0.008$) after adjustment for all confounders except BMI; Atheroscheloris	**Prudent** : subjects were more physically active and smoked less
(Newby, Muller & Tucker, 2004	Obesity	1. 6 2. **Healthy, fiber-rich food pattern (factor 1):** reduced-fat dairy products, fruit, and fibre and loaded moderately on fruit juice, non-white bread, nuts and seeds, whole	**Factor 1**: inversely associated with annual change in BMI(OR=0.51; 95% CI: ~(0.82, 0.20); *P* =0.05; *P* for trend =0.01)in women and inversely associated with annual change in waist circumference (OR:1.06; 95% ; *P* =0.05; *P* for trend =0.04) in both sexes	*

		grains, and beans and legumes **Protein and alcohol** **Sweets** **Vegetable fats and vegetable** **Fatty meats** **Eggs, bread and soup**		
(Togo et al., 2004)	Obesity	**1.** 2 for men and 2 for women **2. Green** **Sweet** **Traditional** **Green** **Sweet-Traditional**	**Sweet and Sweet-Traditional**: inversely associated with baseline BMI. **Traditional**: inversely associated with subsequent 11- and 5-y BMI change, respectively. **Other patterns:** Not Statistically Significant	*****
(Di bello et al. 2008)	Myocardial infraction	1. 5 2. **Vegetable**: fruit, dark yellow vegetables, green leafy vegetables, other vegetables, and polyunsaturated oil and a low intake of palm oil **Factors 2 and 3** vegetables and high-fat dairy products including whole milk, ice cream, and cheese **Factor 4:** palm oil and coffee pattern, was characterized by high intakes of coffee, sugar, and palm oil **Factor 5:** alcohol, legumes, and polyunsaturated oil pattern	**Vegetable:** associated with a significantly decreased adjusted risk of 28% of myocardial infarction. **Factor 4:** associated with a 38% increased risk of myocardial infarction **Other patterns :** Not Statistically Significant	*
(Iqbal et al., 2008)	Myocardial infraction	1. 3 2. **Oriental:** tofu and soy and other sauces 3. **Western:** fried foods , salty snacks , eggs and meat **Prudent:** fruit and vegetables	**Prudent**: associated inversely with Acute Myocardial Infarction, (Compared with the first quartile, the adjusted ORs were 0.78 (95% CI 0.69 to 0.88 for the second quartile, 0.66 (95% CI 0.59 to 0.75 for the third, and 0.70 (95% CI 0.61 to 0.80) for the fourth (P for trend 0.001). **Western** :U-shaped associated with AMI (compared with the first quartile, the adjusted OR for the second quartile was 0.87 [95% CI 0.78 to 0.98], whereas it was 1.12 [95% CI 1.00 to 1.25] for the third quartile and 1.35 [95% CI 1.21 to 1.51] for the fourth quartile; P for trend 0.001) **Other patterns**: Not Statistically Significant	*
(Akesson et al., 2007)	Myocardial Infraction	1. 4 2. **Healthy:** vegetables, fruit, and legumes **Western/Swedish:** red meat, processed meat, poultry, rice, pasta, eggs, fried potatoes, and fish **Alcohol:** wine, liquor, beer, and some snacks	**Healthy** and the **alcohol**: statistically significantly associated with the risk of primary MI **Healthy**: associated with 71% increased risk compared with the highest quintile (P for trend=.004. In the lowest quintile of the **alcohol** dietary pattern, the relative risk of MI was 1.64 (95% confidence interval, 1.09-2.47 compared with the	*

		Sweets: sweet baked goods, candy, chocolate, jam, and ice cream	highest quintile (P for trend=.002). **Other patterns:** Not Statistically Significant	
(Ambrosini et al., 2009)	Metabolic syndrome	1. 2 **2. Western:** meat ,fast food, potato**,** soft drinks, cakes, high fat products **Healthy**: whole grain vegetables and fish 3. 88%	**Western:** associated with greater odds for the 'high risk metabolic cluster (highest vs. lowest quartile (OR= 2.50 ,95% CI: 1.05- 5.98;p for trend < 0.02) **Other patterns:** Not Statistically Significant	*
(Noel et al., 2009)	Metabolic Syndrome	**1.** 3. 2. **Meat and French fries** : meat, processed meat, French fries, pizza and Mexican foods, eggs, alcohol, and other grains and pasta, and low loadings from reduced-fat dairy, fruit and fruit juice, hot and cold cereal, citrus fruit and juice, poultry and vegetables. **Traditional**: high in beans and legumes, rice, and oil, and low in high-fat dairy, condiments, and nuts and seeds. **Sweets**: candy, sugar and chocolate candy, soft drinks, sugary beverages, sweet baked goods, dairy desserts, and salty snacks, and low loadings from fish, poultry, vegetables, oils, and soups.	**Sweets:** After excluding individuals with diabetes, associated with metabolic syndrome (OR: 1.8, 95% CI: 1.03, 3.3). **Other patterns:** Not Statistically Significant	**Sweets**: positively associated with younger age, less physical activity, smoking, men, lower vitamin and medication use.
(Lutsey, Steffen & Stevens, 2008)	Metabolic Syndrome	1. 2 2. **Western**: refined grains, processed meat, fried foods, and red meat. **Prudent:** Cruciferous and carotinoid vegetables, fruit, fish, and poultry. 3. 19.9%	**Western** : adversely associated with incident Metabolic Syndrome highest vs. lowest quintile, OR: 1.18 (1.03–1.37) P for trend 0.03) **Other patterns**: Not Statistically Significant	*****
(Esmaillzadeh et al., 2007)	Metabolic Syndrome	Same as (Esmaillzadeh & Azadbakht, 2008a)	**Healthy:** subjects in the highest quintile of pattern scores had a lower odds ratio for the metabolic syndrome (odds ratio: 0.61; 95% CI: 0.30, 0.79; P for trend 0.01. and insulin resistance (0.51; 0.24, 0.88; P for trend < 0.01than did those in the lowest quintile). **Western**: highest quantile of pattern had greater odds for the metabolic syndrome (1.68; 1.10, 1.95; P for trend < 0.01) and insulin resistance (1.26; 1.00, 1.78; P for trend < 0.01) **Iranian:** significantly associated only with abnormal glucose homeostasis (1.19; 1.04, 1.59; P 0.05)	Same as (Esmaillzadeh & Azadbakht, 2008a)

(Oddy et al., 2009)	Mental Health and Behavioural problems (depression)	1. 2 2. **Healthy:** fruits and vegetables and low intakes of crisps and confectionary. **Western:** meat ,crisps and fast food items and soft drinks 3. 50% and 34%	**Western:** associated with higher total (b=2.20, 95% CI=1.06, 3.35), internalizing (withdrawn/depressed) (b=1.25, 95% 31 CI=0.15, 2.35) and externalizing (delinquent/aggressive) (b=2.60, 95% CI=1.51, 3.68) CBCL scores (mental health) **Prudent :** Not Statistically Significant for the adjusted case	*
(Yannakoulia et al., 2008)	Mental and Health problems (anxiety score)	1. 6 2. **Healthful dietary:** vegetables, cereals, fish and dairy. **Vegetarian:** vegetables, fruits, nuts and potatoes **Sweets and soft drinks** **Low-fat/Low-sugar:** products such as low-fat dairy products, poultry and light soft drinks **Western-type:** potatoes, red meat and coffee. **Cereals and legumes:** cereals and legumes and low alcohol intake 3. 69.3% .’	**Sweets intake**, **Meat and products intake**, positively associated with anxiety score in females (P<0.05) **Cereals and Legumes :** negatively associated with anxiety score (P<0.05)	*
(Akbaraly et al., 2009)	Mental and Health Problems (depression)	1. 2 2. **Whole food:** vegetables, fruits and fish. **Processed food**: sweetened desserts, chocolates, fried food, processed meat, pies, refined grains, high-fat dairy products and condiments.	**Whole food :** associated with lower odds of CES–D depression (OR = 0.74, 95% CI 0.56–0.99) **Processed food:** associated with an increased odds of CES–D depression (OR = 1.58, 95% CI 1.11–2.23).	*
(Wiles et al., 2009)	Mental and Behavioural Health	1. 3 2. **Junk** : high-fat processed foods (burgers, coated poultry) and snack foods high in fat and/or sugar (such as crisps, and chocolate), **Health Conscious** :rice, pasta, salad and fruit **Traditional:** diet of meat, potatoes and vegetables	**All the patterns:** Not Statistically Significant	*
(Deshmukh-Taskar et al., 2009)	HDL-Cholesterol	1. 2 2. **Western Dietary Pattern**: refined grains, French fries, high-fat dairy products, dishes with cheese, red meats, processed meats, eggs, snacks, sweets and desserts, sweetened beverages and condiments **Prudent Dietary Pattern**: grains, legumes, vegetables (i.e. cruciferous, other leafy and dark-yellow vegetables), tomatoes, fruits,	**Prudent:**: positively associated with Insulin sensitivity (P < 0.005) **Western**: inversely associated with Serum HDL cholesterol (P<0.005)	*

		100% fruit juices, low-fat dairy products, poultry, clear soups and low-fat salad dressings) 3. 31%.		
(Hamer & Mishra, 2010)	HDL-cholesterol	1. 4. 2. **Fast food:** pasta , chicken , burger , kebab, crisps , chocolate , tea , coffee **Health aware:** wholemeal bread ,oily fish ,vegetables, yogurt **Traditional:** white bread ,eggs , bacon , sausages, beer **Sweet:** breakfast cereals, biscuits , buns, tea, coffee , beer 3. 16.5%	**Health aware:** inversely associated with concentrations of homocysteine (mean = -2.43, 95%CI: -3.41,-1.45, P trend 0.01) and, and positively with HDL-cholesterol (highest vs. lowest tertile mean=0.08 (95%CI ,0.03, 0.12; P trend=0.09) **Other patterns :** Not Statistically Significant	**Fast food** : positively associated with younger age , smoking, employment and single life
(Panagiotakos et al., 2007b)	HDL- Cholesterol	1. 6 2. **Component 1/healthfu**l : low-fat products such as fish, vegetables, legumes, cereals, and fruits; **Component 2/ high glycemic index and** **Component 3/high-fat :** red or white meat and meat products, and potatoes **Component 4/pasta:** pattern that included consumption of pasta **Component 5/dairy products and eggs** **Component 6/sweets** **Component 7/alcohol**	**Component 1** : inversely associated with waist circumference, systolic blood pressure, triglycerides, positively associated with high-density lipoprotein cholesterol levels, and inversely with the likelihood of the metabolic syndrome (odds ratio [OR] 0.87, 95% confidence interval [CI] 0.79 to 0.97, **Components 2 and 6** : positively correlated with the previous indexes, and the likelihood of having the metabolic syndrome (OR 1.13, 95% CI 1.05 to 1.21 and OR 1.26, 95% CI 1.21 to 1.33) **Other patterns :** Not Statistically Significant	**Component 1** was inversely associated with age and positively associated with male sex and physical activity status. **Component 2** was positively associated with male sex. **Component 3** was positively associated with age income and physical activity level. **Component 4** was positively associated with age and smoking habits and **Component 5** was positively associated with male sex, age.
(Sadakane et al., 2008)	HDL Cholesterol	1. 3 2. **Vegetable:** vegetables, potatoes, soybeans products tofu and fermented soybeans, fruits, sea weeds, citrus, beans, and dried fish **Meat:** processed meats, beef, pork, poultry, steamed fish paste, high-fat products, and butter **Western:** breads, butter, and yoghurt, and lower intakes of rice, salty products, and miso soup 3. 28.5%	**Vegetable**: associated with higher HDL cholesterol. **Meat:** associated with higher total and HDL cholesterol. **Western** : associated with higher total, HDL, and LDL cholesterol	**Vegetable**: positively associated with people who were older, married, and less likely to smoke. **Meat:** positively associated with people who were younger, married, highly educated, and likely to smoke and drink alcohol. **Western:** positively associated with people who were younger, highly educated, less likely to smoke and drink alcohol, but were physically inactive.
(Zhang et al., 2006)	Diabetes	1. 2 2. **Prudent:** fruits, vegetables, whole grains, fish, and poultry **Western:** meat products (red meat and	**Western:** increased risk of gestational diabetes mellitus (RR: 1.63 (95% CI 1.20–2.21,p for trend=0.001) **Prudent** : RR: 1.39 (95% CI 1.08–1.80, p for trend=0.018.	**Western** : positively associated with people who tended to smoke more, consume less fibre, and engage in less physical activity

		processed meat), refined grains and high-fat dairy 3. 15.2%		**Prudent** : positively associated with people who tended to smoke less, consume more fibre, and to be more physically active
(Cai et al., 2007)	CVD, T2D, Stroke, Diabetes	1. 3 2. **Vegetable-rich:** green beans and yard long beans 3. **Fruit-rich:** fruits **Meat-Rich:** meat, poultry, and animal organs	**Fruit-rich**:decreased risk of all causes of death(HR: 0.94 (95% CI: 0.89–0.98) **Meat-rich** : associated with increased risk of diabetes (HR 1.18, 95% CI: 0.98 –1.42)	*
(Nettleton et al., 2009)	CVD	1. 4 2. **Fats and Processed Meat:** added fats, processed meat, fried potatoes, and desserts **Vegetables and Fish** :several vegetable groups, fish, soup, Chinese foods, red meat, poultry, and soy, Beans, Tomatoes, and Refined Grains (beans, tomatoes) **Refined grains:** high-fat dairy foods, avocado/guacamole, red meat, and whole grains **Fruit:** whole grains, fruit, nuts and seeds, green leafy vegetables, and low-fat dairy foods.	**Fats and Processed Meat** : associated with a greater risk of incident CVD (hazard ratio quintile 5 compared with quintile 1: 1.82; 95% CI: 0.99, 3.3) **Fruit**: associated with a lower risk of CVD (0.54; 95%CI : 0.33, 0.91) **Other patterns :** Not Statistically Significant	*
(Esmaillzadeh & Azadbakht, 2008a)	CVD	1. 3 2. **Healthy:** fruits, vegetables, tomatoes, poultry, legumes, cruciferous and green leafy vegetables, tea, fruit juices, and whole grains **Western:** refined grains, red meat, butter, processed meat, high-fat dairy products, sweets and desserts, pizza, potatoes, eggs, hydrogenated fats, and soft drinks. Low in other vegetables and low fat dairy products **Iranian:** refined grains, potato, tea, whole grains, hydrogenated fats, legumes, and broth 3. 24%	**Healthy:** decreased risk of dyslipidemia (odds ratio (OR), 0.36; 95% CI, 0.19–0.53), hypertension (OR, 0.33; 95% CI, 0.17–0.60), at least 1 (OR, 0.30; 95% CI, 0.18–0.58, and at least 2 risk factors (OR, 0.39; 95% CI, 0.20–0.77) compared with the lowest quintile. **Western**: increased cardiovascular risk factors (OR, 2.59–3.11; $P < 0.05$). **Iranian**: significantly associated with dyslipidemia (OR, 1.73; 95% CI, 1.02–2.99 and at least 1 risk factor (OR, 1.89; 95% CI, 1.05–3.20).	**Healthy**: positively associated with more physical activity and greater fiber intake but lower energy and cholesterol intakes. **Western** positively associated with less physical activity and lower fiber intake but greater energy and cholesterol intakes. **Iranian**: positively associated with age, physical activity, lower energy intake
(Heidemann et al., 2008)	CVD	1. 2 2. **Prudent:** vegetables, fruit, legumes, fish, poultry, and whole grains. **Western:** red meat, processed meat, refined grains,french fries, and sweets/desserts	**Prudent**: associated with a 28% lower risk of cardiovascular mortality (95% CI: 13 - 40) and a 17% lower risk of all-cause mortality (95% CI: 10 -24) **Western** : associated with a higher risk of mortality from cardiovascular disease 22%(95% CI: 1 – 48), cancer (16%; 95% CI, 3 - 30), and all causes 21%; 95% CI: 12 - 32)	**Prudent:** associated with people who were slightly older, exercised more, were less likely to be smokers, were more likely to use hormone replacement therapy and multivitamin supplements, and had a more advantageous nutrient profile than those with lower sores for this pattern.

				Western: associated with people who were younger, less physically active, were more likely to smoke, were less likely to use hormone replacement therapy and multivitamin supplements, and had a more unfavorable nutrient profile than those who scored low on this pattern.
(Panagiotakos et al., 2007a)	CVD	1. 5 2. **Component 1:** low-fat products, like fish, vegetables, legumes, greens, and salads **Component 2:** red or white meat and meat products, pasta, and potatoes **Component 3:** cereals and sweets **Component 4:** dairy products and fruits **Component 5:** alcoholic beverages 3. 56.3%	**Component 1, Component 3, and component 5** were associated with lower likelihood of having increased burden of CVD (P <0.01), irrespective of various potential confounders. **Other patterns :** Not Statistically Significant	**Component 1**: inversely associated with age and number of cigarettes smoked and positively associated with physical activity status. **Component 2**: positively associated with age, male sex and inversely associated with income and physical activity level.
(Shimazu et al., 2007)	CVD	1. 3 2. **Japanese:** soybean products, fish, seaweeds, vegetables, fruits and green tea **Animal food**: animal-derived foods (beef, pork, ham, sausage, chicken, liver and butter), coffee and alcoholic beverages. **high-dairy, high-fruit-and-vegetable, and low-alcohol:** dairy products milk and yoghurt), margarine, fruits and vegetables (carrot, pumpkin and tomato 3. 26.2%	**Japanese:** associated with a lower risk of CVD mortality (hazard ratio of the highest quartile vs. the lowest, 0.73; 95% confidence interval: 0.59–0.90; P for trend: 0.003... **Animal food:** associated with an increased risk of CVD. **Other patterns :** Not Statistically Significant	*
(Tseng et al., 2004)	Cancer(prostate)	1. 3 2. **Vegetable fruit:** vegetables, fruits, fish, and shellfish **Red meat-starch**: red meats, potatoes cheese, salty snacks, and desserts **Southern pattern**: cornbread, grits, sweet potatoes, okra, beans, and rice.	**Southern** : decreased risk of prostate cancer (3rd versus 1st tertile relative risk, 0.6; 95% confidence interval, 0.4 – 1.1; trend P < 0.08) **Other patterns:** Not Statistically Significant	*
(Chang et al., 2008)	Cancer(Ovarian)	1. 5 2. **Plant-based :** Vegetables and Fruits **High- Protein / High-Fat:** processed meat , butter, ice-cream, cheese and potatoes **High-Carbohydrate:** pizza, spaghetti , cheese **Ethnic :** Beans , lentils , rice , potatoes **Salad and wine** 3. 35.0%	**plant-based:** increased relative risk of ovarian cancer was 1.65 (95% confidence interval: 1.07–2.54; P for trend=0.03) **Other patterns:** Not Statistically Significant	*

(De Stefani et al., 2008b)	Cancer(Lung)	1. 3 2. **Antioxidants** **Carbohydrates** **High Meat**	**Antioxidants**: inversely associated with lung cancer risk (OR 0.69, 0.51-0.96 **High-meat**: associated with a strong increase in risk (OR 2.90, 95% CI 1.91-4.40). **Other patterns :** Not Statistically Significant	**High meat**: positively associated with younger agerural living, inversely correlated with education and associated with smoking intensity and duration and positively correlated with mate drinking **Carbohydrates:** inversely associated with age. **Antioxidants:** associated with socio demographic variables, smoking and alcohol drinking.
(Campbell, Sloan & Kreiger, 2008)	Cancer(gastric)	1. 2 2. **Prudent**: vegetables, fruits, and fish **Western**: soft drinks, French fries, white bread, hamburger, eggs, bacon, doughnuts, and hot dogs 3. 23%	**Prudent:** associated with decreased risk of gastric cancer in women (odds ratio (OR): 0.58, 95% confidence interval (CI): 0.37, 0.92) **Western**: associated with increased risk of gastric cancer in women (OR: 1.86, 95% CI: 1.20, 2.89 and men (OR: 1.44, 95% CI: 1.03, 2.02.)	*
(De Stefani et al., 2004)	Cancer(gastric)	1. 3 2. **Starchy**: total grains and tubers. **Healthy:** white meat, dairy foods, desserts, raw vegetables, and fruits. **Mixed pattern**: red meat, processed meat, eggs, and pulses.	**Starchy :** increased risk of gastric cancer (odds ratio (OR): 4.1, 95% confidence interval CI: 2.6–6.6) .	*
(Reedy et al., 2010)	Cancer(Colorectal)	1. 3 2. **fruits and vegetables** **fat reduced**: **meat and potatoes**	**Meat and potatoes:** associated with increased risk of colorectal cancer risk. **Other patterns :** Not Statistically Significant	*
(Butler et al., 2008)	Cancer(Colorectal)	1. 2 2. **Vegetable–fruit–soy:** vegetable, fruit, and soyfood intake; of the 32 foods included in the pattern, 23 were vegetables, five were soyfood items, and four were fruit items. **Meat–dim sum:** chicken, pork, fish, rice and noodle dishes, and preserved foods.	**All the patterns:** Not Statistically Significant	**meat–dim sum** : positively associated with being male, have higher education, report any weekly physical activity, be a heavy smoker, drink alcohol, and consume more saturated fat, compared to individuals in the first quartile **Vegetable-fruit–soy** positively associated with physical activity and education level and inversely with heavy smoking.
(Edefonti et al., 2008)	Cancer(Breast)	1. 4 2. **Animal products**: animal protein and animal fat, calcium, cholesterol, saturated fatty acids, riboflavin, zinc and phosphorus **Vitamins and Fiber:** vitamin C and total fiber, total folate, potassium, b-carotene equivalent, soluble carbohydrates and vitamin **Unsaturated fats**: vegetable fat and vitamin E, monounsaturated and	**Animal products** pattern and the **Unsaturated fats** pattern :negatively associated with breast cancer (OR= 0.74, 95% CI: 0.61–0.91) and (OR= 0.83, 95% CI: 0.68–1.00), **Starch-rich** associated with breast cancer (OR = 1.34, 95% CI: 1.10–1.65) **Vitamins and fiber** : inversely associated with ovarian cancer (OR = 0.77, 95% CI: 0.61–0.9) **Starch-rich pattern**: positively associated with ovarian cancer (OR = 1.85, 95% CI: 1.37–2.48)	*

		polyunsaturated fatty acids **Starch-rich:** starch, vegetable protein and sodium.		
(Mannisto et al., 2005)	Cancer(Breast)	1. 3. 2. **Vegetables:** vegetables, legumes, fruit, pasta, fish and oil **Pork, Processed meat, Potatoes – PPP:** pork, beef, processed meats, potatoes, rice, poultry, liver, butter/low-fat margarine, pasta and coffee 3. 23.2%, 29.0% and 21.8%	**All patterns** : Not Statistically Significant	*
(De Stefani et al., 2008a)	Cancer(Bladder)	1. 3 2. **Sweet beverages:** coffee, tea, and added sugar. **Prudent:** fresh vegetables, cooked vegetables, and fruits. **Western:** red meat, fried eggs, potatoes, and red wine 3. 25.1%	**Sweet beverages:** associated with risk of bladder cancer (OR 3.27, 95% CI: 1.96–5.45) **Prudent pattern**: associated with risk of bladder cancer (OR 2.35, 95% CI: 1.42–3.89. **Other patterns :** Not Statistically Significant	*
(Cottet et al., 2005)	Cancer (reduced adenoma)	1. 3 2. **Mediterranean:** olive oil, fresh fruit, vegetables, legumes, lean meat and fresh fish. **Western:** potatoes, fats, delicatessen products, high-fat meat, beer, rice and pasta, refined bread and cereals, nuts, sodas. **Snacks:** high fat delicatessen, high-fat cheese, desserts and sweets, beer, soda and mineral water 3. 21.9%	**Mediterranean:** associated with reduced adenoma recurrence (second tertile: adjusted odds ratio (OR) = 0.50, 95% confidence interval (CI) = 0.18–1.42; third tertile: adjusted OR=0.30, 95% CI= 0.09–0.98; P for linear trend= 0.04). **Other patterns:** Not Statistically Significant	*
(Jackson et al., 2009)	Cancer (prostate)	1. 4. 2. **Healthy** : vegetables, fruits and peas and beans **Carbohydrate** white bread and refined cereals **Sugary foods and Sweet Baked Products** **Organ meat and fast food**: high fat dessert, organ meat, fast food and salty snacks 3. 24.5%	**All the patterns:** Not Statistically Significant	**Carbohydrate** : positively associated with lower use of vitamins , less obesity **Sugary foods and sweet baked goods** positively associated with smoking and were less likely to have tertiary education. Men with high intakes of the **Organ meat and fast food** : positively associated with obesity
(Muller et al., 2009)	Cancer (Prostate)	1. 4 2. **Mediterranean:** meats, vegetables, and fruits, and avoidance of cakes and sweet biscuits	**All the patterns:** Not Statistically Significant	*

		Vegetable: vegetables **Meat & Potatoes**: meats and potato cooked in fat **Fruit & Salad** : high intake of salad greens and fruit.		
(Ambrosini et al., 2008b)	Cancer (prostate)	1. 3 2. **Vegetable:** fresh and tinned tomatoes plus jam, honey, and apples **Western**: full cream milk, white bread, cakes, potato crisps, French fries (chips), eggs, red and processed meats, hamburgers, fried or takeaway fish, and full alcohol beer **Health-conscious**: steamed and grilled fish, tinned fish, chicken, rice, pasta, legumes, and tofu. It also had moderate positive loadings for bean sprouts, nuts, yoghurt, ricotta cheese, red wine, and white wine)	**Western:** associated with increased risk for prostate cancer (Men in the highest quartile for **Western** pattern score had an OR of 1.82 (95% confidence interval: 1.15–2.87, trend p < 0.02) Results were similar for aggressive cases and attenuated for non-aggressive cancers. **Other patterns :** Not Statistically Significant	*
(Wu et al., 2006)	Cancer (Prostate)	1. 2 2. **Prudent:** fruits, vegetables, whole grains, fish, and poultry **Western:** meat products (red meat and processed meat), refined grains	**All the patterns :** Not Statistically Significant	**Prudent**: positively related with subjects who were younger and more likely to engage in regular physical exercise and to be of Southern European origin. They were also less likely to be smokers. **Western**: positively related with men who were more likely to be older and to smoke and less likely to be of Southern European origin, or to exercise regularly.
(Michaud et al., 2005)	Cancer (Pancreatic)	1. 3 2. **Prudent:** vegetables, legumes, fruit, whole grains, fish, and poultry **Western:** red meat, processed meat, refined grains, French fries, high-fat dairy products, sweets and desserts, and high-sugar drinks	**All patterns :** Not Statistically Significant	*
(Nkondjock et al., 2005)	Cancer (Pancreatic)	1. 3 2. **Western:** processed meats, sweets and desserts, refined grains and potatoes **Fruits and vegetables:** fresh fruits and cruciferous vegetables **Drinker** liquor, wine and beer 3. 16.7% (8.2%, 4.4% and 4.1%, respectively)	**fruits and vegetables** : was associated with a 49% reduction in pancreatic risk among men (OR : 0.51, 95% CI: 0.29–0.90, $p < 0.004$) **Other patterns :** Not Statistically Significant	*****

(Balder, Goldbohm & van den Brandt, 2005)	Cancer (Lung)	1. 5 2. **Salad vegetables:** vegetable items, several fruit items, pasta, rice, poultry, fish, and oil. **Cooked vegetables:** cooked leaf vegetables, cabbages, legumes, and carrots. **Pork, processed meat, and potatoes:** also coffee and low-fat margarine, **Sweet foods**: cakes and cookies, sweet sandwich spread, sweets and candies, and (straw)berries **White/brown bread :** brown/ wholemeal bread types and apples and pears 3. 23%	**Salad vegetables:** was associated with decreased risk of lung cancer (rate ratios (RR) Q5, 0.75; 95% confidence interval (CI): 0.55-1.01], after multivariate adjustment. **Sweet foods:** was inversely associated with lung cancer risk (RR Q5, 0.62; 95% CI, 0.43-0.89. **Other patterns:** Not Statistically Significant	*
(Sant et al. 2007)	Cancer (HER-2-positive cancers)	1. 4 2. **Salad Vegetables:** raw vegetables and olive oil **Western:** potatoes, ravioli, red and processed meat, eggs, butter, seed oil (as added fat) and cakes. **Canteen:** pasta, tomato sauce, olive oil and wine. **Prudent:** cooked vegetables, rice, poultry, fish and low consumption of alcohol. 3. 21%	**Salad vegetables**: decreased risk wit HER-2- positive cancers (RR 5 0.25, 95% CI: 0.10–0.64, for the highest tertile; p trend < 0.001, much stronger than for HER-2-negative cancers (heterogeneity < 0.039). **Other patterns :** Not Statistically Significant	*
(Bastos et al., 2010)	Cancer (gastric)	1. 3 2. **Dietary pattern I:** fruits and dairy products, and low consumption of alcoholic beverages. **Dietary pattern II:** fruit, salads, vegetables, dairy products, fish and meat. **Dietary pattern III:** most food groups and low vegetable soup intake.	**Dietary pattern II (low consumption of fruit, salads, vegetables, dairy products, fish and meat)**: associated with higher gastric cancer compared to Pattern I (OR=1.68, 95%CI: 1.31-2.14). Similar associations for cardiac and non-cardiac gastric cancer. **Other patterns :** Not Statistically Significant	*
(Bertuccio et al., 2009)	Cancer (gastric)	1. 4. 2. **Animal products:** animal protein, riboflavin, cholesterol, phosphorus, calcium, and zinc. **vitamins and fiber:** vitamin C, total fiber, potassium, total folate, β-carotene equivalents, and soluble carbohydrates **vegetable:** polyunsaturated fatty acids, vitamin E, monounsaturated fatty acids, linoleic acid, and linolenic acid **Starch-rich:** starch, vegetable protein, and	**Animal products:** associated positively with gastric cancer (OR= 2.13; 95% CI, 1.34-3.40) **Starch-rich:** associated positively with gastric cancer (OR= 1.67; 95% CI, 1.01-2.77) **Vitamins and fiber:** associated inversely with gastric cancer (OR=0.60; 95% CI, 0.37-0.99)	*

		sodium.		
(Kim et al., 2004)	Cancer (gastric)	1. 3 2. **Prudent:** vegetables, fruits, soy products, seaweeds, mushroom, milk, beans and yogurt **Traditional:** pickled vegetables, salted fish and roe, fish, rice and miso soup, alcoholic beverages (sake, shochu and beer) for men and thus was called the traditional dietary pattern. **Western:** meat, poultry, cheese, bread and butter	**Prudent:** inverse associated between the healthy pattern and gastric cancer risk in women (rate ratio for highest quartile (RR) : 0.56; 95% CI :0.32–0.96; *p* for trend <0.03) **Traditional:** significantly associated with the increased risk of gastric cancer in both genders (for men, RR: 2.88, 95% CI: 1.76–4.72; for women, RR: 2.40, 95% CI: 1.32–4.35). **Other patterns:** Not Statistically Significant	**Prudent** :positively related with subjects who were more physically active and smoked less
(Flood et al., 2008)	Cancer (Colorectal)	1. 3 2. **First factor (men/women): fruit and vegetables** **Second factor(men): fat-reduced foods, diet foods, and lean meats** **Second factor (women): high-fat foods, red meats, and potatoes.** **Third factor(men): processed foods (sausages and French fries) and sweets** **Third factor(women): chicken, milk , mayo** 3. 35.1% and 34.2%	**Fruit and vegetables**: associated with decreased risk of colorectal cancer (relative risk (RR) for quintile Q 5 versus Q1: 0.81; 95% CI: 0.70, 0.93; P for trend 0.004). **Red meat and potatoes**: associated with increased risk: men (RR: 1.17; 95% CI: 1.02, 1.35; P for trend 0.14 and women (RR: 1.48; 95% CI: 1.20, 1.83; P for trend 0.0002)	**Fruit and vegetable**: associated with lower BMI, physically activity, education, non-smoking, and less alcohol. **Red meat and potatoes**: associated with higher BMI, increased energy intake, decreased physical activity, a lower likelihood of being a college graduate, and increased smoking for both men and women.
(Kesse, Clavel-Chapelon & Boutron-Ruault, 2006)	Cancer (Colorectal)	1. 4 2. **Healthy:** raw and cooked vegetables, legumes, fruit, yogurt, fresh cheese, breakfast cereals, sea products, eggs, and vegetable oils (olive oil and others) and by a low consumption of sweets **Western:** potatoes, pizza and pie, sandwiches, legumes, sweets, cakes, cheese, bread, rice, pasta, processed meat, eggs, and butter **Drinker:** sandwiches, snacks, coffee, processed meat, sea products, wine, and other alcoholic beverages, as well as by a low consumption of soup and fruit, strongly associated with ethanol intake. **Meat eaters:** potatoes, legumes, coffee, meat, poultry, vegetable oils (except olive oil), and margarine and negatively	**Western**: positively associated with adenoma (For quartile 4 versus quartile 1, an increased risk of adenoma was observed with high scores of the Western pattern (multivariate relative risk (RR): 1.39, 95% confidence interval: 1.00, 1.94; ptrend <0.03) and the drinker pattern (RR: 1.42, 95%confidence interval: 1.10, 1.83; ptrend < 0.01). **Other patterns :** Not Statistically Significant	*

		associated with tea, olive oil, and breakfast cereals.		
(Mizoue et al., 2005)	Cancer (Colorectal)	1. 3 **2. DFSA: A high-dairy, high-fruit and -vegetable, high starch, low-alcohol Animal food Japanese**	**DFSA: Dairy products and fruits and vegetables with low alcohol consumption:** associated with decreased risk of colorectal adenomas. **Other patterns:** Not Statistically Significant	*
(Meyerhardt et al., 2007)	Cancer (Colon)	1. 2 2. **Prudent**: fruits, vegetables, whole grains, legumes, poultry, and fish **Western:** refined grains, processed and red meats, desserts, high-fat dairy products, and French fries	**Western:** associated with a significantly worse disease-free survival (colon cancer recurrences or death). Compared with patients in the lowest quintile of Western dietary pattern, those in the highest quintile experienced an adjusted hazard ratio (AHR) for disease-free survival of 3.25 (95% CI: 2.04-5.19; *P* for trend< .001). Also associated with a similar detriment in recurrence free survival (AHR, 2.85; 95% CI, 1.75-4.63) and overall survival (AHR, 2.32; 95% CI, 1.36-3.96)), comparing highest to lowest quintiles (both with *P* for trend <.001). **Other patterns:** Not Statistically Significant	**Prudent**: positively associated with physical activity lower body mass no smoking cigarettes **Western**: positively correlated with white men, and past or current smokers.
(Wu et al., 2004a)	Cancer (Colon)	1. 2 **2. Prudent :** fruits , vegetables , poultry **Western** : meat,eggs , high-sugar foods	**Prudent** : weakly and Not Statistically Significantly associated with decreased risk of colon cancer or distal colon adenoma (highest versus lowest quintile: colon cancer: multivariate adjusted relative risk (RR) = 0.84 (95% confidence interval (CI) (0.64–1.10); p trend 0.37) distal adenoma: multivariate odds ratio (OR):0.88 (95% CI: 0.73–1.08; p trend=0.12 **Western**: positively associated with colon cancer and distal adenoma (colon cancer: RR:1.27 (95% CI:0.96–1.69, p trend=0.05; distal adenoma: OR=1.28 (95% CI: 1.05–1.56, p trend 0.01)	*
(Cottet et al., 2009)	Cancer (Breast)	1. 2. 2. **Western/alcohol:** processed meat and meat products (ham, offal), French fries, appetizers, sandwiches, rice/pasta, potatoes, pulses, pizza/pies, canned fish, eggs, crustaceans, alcoholic beverages, cakes, mayonnaise, and butter/cream. **Healthy/Mediterranean:** vegetables and fruits, fish and crustaceans, olives, and sunflower oil. traits (fish, fruits, vegetables, olive oil, essentially vegetables, fruits, seafood, olive oil, and sunflower oil)	**Western/alcohol:** positively associated with breast cancer risk (hazard ratio = 1.20, 95% CI: 1.03, 1.38; P < 0.007 for linear trend). **Healthy/Mediterranean:** negatively associated with breast cancer risk (hazard ratio = 0.85, 95% CI: 0.75, 0.95; P < 0.003 for linear trend),	**Western/alcohol:** positively associated with younger age, decreasing prevalence of null parity, decreasing duration of breastfeeding, increasing prevalence of overweight, greater height, a higher proportion of relatives with a history of breast cancer, a higher proportion of oral contraceptive use, biennial Pap smears and a higher prevalence of current smoking. **Healthy/Mediterranean :** positively associated with older age, higher education, a higher prevalence of overweight, a higher proportion of personal history of benign breast disease, increasing use of

				menopausal hormone therapy, increasing duration of breastfeeding, an increasing proportion of annual Pap smears, higher levels of physical activity, and an increasing proportion of former smokers.
(Wu et al., 2009)	Cancer (breast)	1. 3 2. **Western meat/starch:** pasta with meat, beef taco, beef burrito, pizza, meatballs, hamburger, fried potatoes, baked potatoes, mashed potatoes, pancake, bagels **Ethnic meat/starch:** pork/fish soups, liver, pork spareribs, salted/dried fish, fried shellfish, chicken wings, rice, fried/Spanish rice, fried noodles, fried dim sum, and other foods **Vegetables/soy:** green beans/peas, carrots, cabbage, bean sprouts, green peppers, bok choy, fresh tofu, fresh soybeans, soybean milk, and other items	**Western:** increased risk of breast cancer (OR: 2.19; 95% CI: 1.40, 3.42; P for trend < 0.0005). **Other patterns:** Not Statistically Significant	*
(Murtaugh et al., 2008)	Cancer (Breast)	1. 5 2. **Western:** high-fat dairy foods, refined grains, gravy and sauces, fast foods, red and processed meats, potatoes, margarine, polyunsaturated fats, and high-fat and high-sugar desserts. **Native Mexican:** Mexican cheeses, soups, meat dishes, legumes, and tomato-based sauces **Prudent:** low-fat dairy, whole grains, fruit and fruit juice, legumes, vegetables, and soups **Mediterranean**: liquor, poultry, seafood, vegetables, salad greens, and high-fat salad dressings **Diet**: high-fat dairy products and salad dressing, cola beverages, and butter and with using low-fat dairy, margarine, and salad dressings, as well as low-fat high-sugar desserts, diet beverages, and sugar substitutes	**Western:** increased risk of breast Cancer (odds ratio for highest versus lowest quartile: 1.32; 95% CI: 1.04, 168; P for trend 0.01) **Prudent:** increased risk of breast cancer (OR: 1.42; 1.14, 1.77; P for trend< 0.01) **Native Mexican** : increased risk of breast cancer (OR 0.68; 0.55, 0.85; P for trend < 0.01) **Mediterranean** : increased risk of breast Ccancer (0.76; 0.63, 0.92; P for trend < 0.01) **Other patterns:** Not Statistically Significant	*
(Cui et al., 2007)	Cancer (Breast)	1. 2 2. **Vegetable-soy:** variety of different vegetables, soy-based products, and freshwater fish.	**Meat-sweet:** associated with increased breast cancer risk in postmenopausal Chinese women. (4^{th} versus 1st quartile: odds ratio, 1.3; 95% confidence interval, 1.0- 1.7; Ptrend =	**Vegetable-soy**: positively associated with people who were more likely to be physically active. **Meat-sweet**: positively associated with people who

		Meat-sweet: various meats, primarily pork but also poultry, organ meats, beef and lamb, and shrimp, saltwater fish, and shellfish, as well as candy, dessert, bread, and milk	0.03., but only in postmenopausal women, specifically among those with estrogen receptor–positive tumors (4th versus 1st quartile: odds ratio, 1.9; 95% confidence interval, 1.1-3.3; Ptrend = 0.03). **Other patterns:** Not Statistically Significant	were younger, better educated, had later menopausal age, and were less likely to have had their first live birth before age 25 years.
(Hirose et al., 2007)	Cancer (Breast)	1. 4 2. **Prudent:** vegetables and fruit, soybean curd, fish and milk **Fatty**: meat and fatty foods **Japanese:** cooked rice for breakfast and miso soup **Salty:** pickles, dried or salted fishes and salty foods 3. 21%	**Prudent**: inverse association with breast cancer risk (Women in the highest quartile of the prudent dietary pattern scores, had a 27% decreased risk of breast cancer compared with those in the lowest (95% CI: 0.63–0.84, p for trend < 0.0001).	*
(Kroenke et al., 2005)	Cancer (Breast)	1. 2 2. **Prudent:** fruits, vegetables, whole grains, legumes, poultry, and fish **Western:** refined grains, processed and red meats, desserts, high-fat dairy products, and French fries.	**All patterns:** Not Statistically Significant	*
(Velie et al., 2005)	Cancer (breast)	1. 3 2. **vegetable-fish/ poultry-fruit :** vegetables and broiled or baked fish and chicken and low intakes of sweets and white bread **beef/pork-starch**: pork, beef, processed meat, French fries, and eggs and low intakes of bran cereal, skim milk, broiled or baked fish and chicken, and dark bread **traditional southern** : traditional rural southern US foods 3. 12.5%	**All patterns:** Not Statistically Significant	**vegetable-fish/poultry-fruit** : was associated with higher education
(Sieri et al., 2004)	Cancer (Breast)	1. 4 2. **Salad vegetables**: raw vegetables and olive oil **Western:** potatoes, red meat, eggs, butter, seed oil (as added fat) and cakes **Canteen**: pasta, tomato sauce, and wine. **Prudent:** cooked vegetables, rice, poultry, fish, and low consumption of alcohol	**salad vegetables** : was associated with significantly lower breast cancer incidence (RR = 0.66, CI 95% = 0.47 , 0.95)comparing highest with lowest tertile) with a significant linear trend (P = 0.016) **Other patterns:** Not Statistically Significant	*****
(Ronco et al., 2006)	Cancer (Breast)	1. 6 2. **Traditional**: boiled meat, grains, cooked vegetables and tubers **Healthy**: white meat, raw vegetables,	**Western:** positively associated with breast risk cancer ((OR 1.31, 95% CI 1.13–1.51) **Traditional** (OR 0.77, 95% CI 0.64–0.93	*****

		cooked vegetables and total fruits **Western:** fried meat, barbecue and processed meat **Stew:** boiled meat and legumes. **High-Fat:** dairy foods and eggs and was labelled **Drinker:** alcohol 3. 58.3%	H**ealthy** (OR 0.84, 95% CI 0.73–0.98 and **stew** (OR 0.83, 95% CI 0.71–0.98 diets were significantly protective. **Other patterns :** Not Statistically Significant	
(De Stefani et al., 2007)	Cancer (Laryngeal carcinoma)	1. 6 **2. (Pattern 1)Traditional** **(Pattern 2)Healthy:** fish, fresh vegetables, total fruits and tea **Third pattern:** cheese, butter, mayonnaise and custard **(Pattern 4)High – Fat:** processed meat, soft drinks and negative loading with whole milk. **(Pattern 5)Drinker pattern :** beer, wine and hard liquor and can be labelled as the drinker pattern **(Pattern 6)Western** : fried meat, barbecued meat, processed meat and fried eggs 3. 36.4%	**(Pattern 5) Drinker** : directly associated with risk of laryngeal carcinoma (OR 3.8, 95% CI 1.9–7.5, whereas the Pattern 2 **(Pattern 2) Healthy** was protective to cancer (OR 0.6, 95% CI 0.3–1.2. **(Pattern 6) Western**: significant increase in risk of laryngeal cancer 3.2 (95% CI 1.6–6.2.). **Other patterns:** Not Statistically Significant	**Pattern 1:** positively associated with those who were frequently lived in a rural environment, and were less educated and smoked significantly more than those in the other patterns. **Pattern 2** was positively associated with education and high body mass and inversely associated with tobacco smoking. **Pattern 3** was directly associated with total energy and total fat intake. **Pattern 4** was inversely associated with tobacco smoking and b-carotene, **Pattern 5** was strongly associated with pack-years of tobacco smoking, black tobacco use and hand-rolled cigarettes preference. This pattern was also highly correlated with mate drinking.
(Agurs-Collins et al., 2009)	Cancer (Breast)	1. 2 2. **Western:** refined grains, high-fat dairy products, meat and processed meat, eggs, margarine, butter and mayonnaise, potato, French fries, sweets, soda, and snacks. **Prudent:** cruciferous and other vegetables, fruit, whole grains, cereals, beans, low-fat dairy products, fish, and poultry. 3. 22%	**Prudent:** associated with lower breast cancer risk overall (P for trend = 0.06) highest vs. lowest quintile (OR=0.86, 95% CI: 0.68, 1.08). **Western:** Not Statistically Significant	**Western:** positively associated with younger age, weight , lower education, smoking and drinking , less exercise **Prudent** : positively associated with older age , higher educational status, less likelihood to smoke, higher levels of strenuous physical activity
(Dixon et al., 2004)	Cancer (Colon)	1. 2. 2. **Vegetarian:** vegetables and legumes, citrus fruit and berries, pasta and rice, poultry and fish, and oil and salad dressings **PPP**: pork, processed meats, potatoes, and coffee. 3. 5.6% in the men to 9.7% in the women	**PPP:** associated with an increased risk of colon cancer in women(quintile multivariate relative risk: 1.62;95%CI: 1.12, 2.34;*P*for trend<0.01) **PPP**: associated with an increased risk of colorectal cancer in men (quintile 4multivariate relative risk: 2.21; 95% CI: 1.07, 4.57; *P* for trend<0.05). **Other patterns:** Not Statistically Significant	*
(Hughes et	Actinic Keratoses	1. 3 2. **the vegetable-fruit potato-sweets-meat**	**All the patterns:** Not Statistically Significant	*

al., 2009)		**the alcohol-snacks**		
(Cutler et al., 2009)	*	1. 3 2. **Vegetable pattern:** zucchini, squash, and eggplant, kale and greens, spinach, and peas andlima beans. **Fruit pattern:** oranges and grapefruit, apples and apple sauce, pears, grapes, bananas, strawberries, cantaloupe and melons, peaches, and plums and apricots **Sweet/Salty Snack Food pattern:** chocolate bars, other candy bars, candy with chocolate, brownies, cake, potato chips, and nachos	*	*
(Kesse-Guyot et al., 2009)	*	1. 4 2. **Alcohol and Meat products:** meat and poultry, processed meat, beer and cider, alcoholic beverages and wine **Prudent:** fruit, tea, vegetable oil, breakfast cereals, vegetables, fish and seafood, and reduced-fat products **Convenience**: candy and cookies, dessert, sweetened beverages, croissants and ready-to-eat products **Starch, sauces and vegetables:** pasta and rice, vegetable oil, fat-free or low-fat sauces and to a lesser extent with high-fat sauces and vegetables. Low consumption of breakfast cereals	*	**Alcohol and meat products:** positively associated with low education, smoking and overweight in both genders, as well as with abdominal obesity in women **Prudent:** positively correlated with high education and being older than 55 years and negatively correlated with current smoking overweight. **Convenience:** inversely related to age and positively related to higher education in both genders. In men, higher scores were related to living alone and an urban residence. **Starch, sauces, and vegetables**: associated with high education and an urban residence in men only.
(Lopez et al. 2009)	*	1. 2 **2. Western:** meat ,eggs , proceed foods , fast food , sweets, potatoes **Mediterranean:** poultry , fruits and vegetables	*	*
(Rezazadeh, Rashidkhani & Omidvar, 2010)	*	1. 2 2. **Healthy** : vegetables, fruits, yellow vegetables, cruciferous vegetables, tomato, low fat dairy products, yogurt drink, poultry, olive, nuts, fruit juice, potato, coffee, and garlic **Unhealthy:** processed meat, mayonnaise, soft drinks, sweets, refined grains, snacks, industrial juice, red meat, nuts,hydrogenated	*	**Healthy:** associated positively with university degree , housing size, family income, physical activity, women, non-manual classes , non-smoking, non –residence in Tehran **Unhealthy:** negatively associated with ethnicity

		fats, butter, French fries, high-fat dairy products, egg, organ meats, and sugar 3. 14.5%		
(Robinson et al., 2009)	*	1. 2 **2. Prudent :**fruit, vegetables, oily fish and wholemeal cereals **Traditional:** vegetables, processed and red meat, fish and puddings	*	**Prudent :** positively associated with women, non-manual classes and non-smoking **Traditional** : positively associated with men, men and women who had partners and alcohol consumption
(Touvier et al., 2009)	*	1. 3 2. **Processed/starchy meat**: biscuits, bread, butter/cream, cakes, French fries, marmalade/honey, pulses, pizza/pies, potatoes, processed meat, rice, pasta, semolina, sandwiches, dairy-based sweet puddings, sugar/confectionary, low consumption of vegetables **Fruit/vegetables**: breakfast cereals, eggs, fruits, ham, pulses, offal, olive oil, seafood, vegetable oil, vegetables, yoghurt/ cottage cheese **Alcohol/meat products**: alcohol beverages, wine, meat, offal, poultry/rabbit, processed meat, appetizers, artificial sweeteners, coffee and French fries, as well as with low consumption of fruits, soup and marmalade/honey.	*	**Fruit/vegetables** : positively associated with supplement use , smoking, **Processed meat/Starchy** foods and **Alcohol/meat** inversely associated with supplement use , smoking,
(Uusitalo et al., 2009)	*	1. 7 2. **Healthy:** vegetables , fish , rice , eggs , meat dishes , creamy **Fast food:** processed meat , chocolate , sweets , sausage , eggs **Traditional bread:** meat, cereals , bread **Traditional meat :** meat , sausage potatoes **Low-fat:** low fat cheese , meat and soft drinks **Coffee** **Alcohol and butter:** beer wine and butter	*	*
(Arkkola et al., 2008)	*	1. 6 2. **Healthy:** Vegetables, pasta , eggs , fruits **Fast food:** processed and fried foods and snacks **Traditional bread/Traditional meat**	*	*

		Low-fat **Coffee** 3. 29.5%		
(Borland et al., 2008)	*	1. 2 2. **Prudent:** vegetables, fruit, wholemeal bread, rice/pasta, yoghurt and breakfast **High – energy:** Puddings, cakes/biscuits, potatoes/chips, vegetables, fruit, red/processed meat, fish, eggs, oils and full-fat spreads	*	**Prudent:** associated with strenuous exercise taken and partnership status. **High-energy**: associated with lower social class.
(Crozier et al., 2008)	*	1. 2 2. **Prudent:** fruit and vegetables, wholemeal bread, rice and pasta, yoghurt, cheese, fish and reduced-fat milk, low in white bread, added sugar, tinned vegetables, full-fat milkand crisps. **Western:** processed meat, cakes and biscuits, puddings, Yorkshire puddings and savoury pancakes, chips, roast and boiled potatoes, sugar, sweets and chocolate.	*	*
(Engeset et al., 2009)	*	1. 6 **2. Traditional fish eaters Healthy Average** **Less fish** **Less healthy** **Western** **Traditional bread eaters** **Alcohol users** 3. 23.7%	*	*
(Keskitalo et al., 2008)	*	1. 4 **2. Healthy:** Vegetables , rice , chicken , yogurt, fish **High-fat:** Processed food , pizza , creamy foods **Sweets :** Salty snacks , chocolate **Meats:** Sausage , meat	*	
(Knudsen et al., 2008)	*	1. 2 **2. Western :** red and processed meat, high-fat dairy **Health Conscious:** vegetables, fruits, poultry and fish.	*	**Western:** positively associated with smoking **Health Conscious:** negatively associated with smoking
(Lau et al., 2008)	*	1. 2 2. **Traditional:** high loadings on pate´ or high-fat meat for sandwiches, mayonnaise salads,	*	*

		red meat, potatoes, butter and lard low-fat fish, low-fat meat for sandwiches, and sauces. **Modern:** high loadings on vegetables, fruit, mixed vegetable dishes, vegetable oil and vinegar dressing, poultry, and pasta, rice and wheat kernels.		
(Northstone & Emmett, 2008a)	*	1. 2 patterns consistently over time **2. Processed** **Traditional** **Health conscious**	*	*
(Northstone, Emmett & Rogers, 2008a)	*	1. 5 2. **Health Conscious:** salad, fresh fruit, rice, pasta, fish, pulses, and non-white bread **Traditional:** vegetables and to some extent red meat and poultry **Processed**: meat pies, sausages and burgers, fried foods, pizza and chips **Confectionery**: chocolate, sweets, biscuits, cakes and other pudding **Vegetarian:** meat substitutes, pulses, nuts and herbal tea 3. 31·3%	*	**Health conscious:** positively associated with increasing education and age and non-white women and negatively associated with increased parity, single, non-working women, those who smoked and who were overweight pre-pregnancy. **Processed:** Opposite associations
(Northstone, Emmett & Rogers, 2008b)	*	Same as (Northstone, Emmett & Rogers, 2008a)	Same as (Northstone, Emmett & Rogers, 2008a)	Same as (Northstone, Emmett & Rogers, 2008a)
(Northstone & Emmett, 2008b)	*	Same as (Northstone, Emmett & Rogers, 2008a)	Same as (Northstone, Emmett & Rogers, 2008a)	Same as (Northstone, Emmett & Rogers, 2008a)

(Romaguera et al., 2008)	*	1. 2 2. **Western-like:** beef, common bread, chicken and sugary drinks **Andean-like:** vegetable products, cereals, tubers, legumes and fruit together with some animal products such as eggs or cheese. herbal teas 3. 61%	*	*
(Custodio das Dores et al., 2007)	*	1. 4 2. **Factor 1:** kidney beans, soybean oil, **Factor 2:** serum triglycerides, PT, INR, and plasma phylloquinone. **Factor 3:** Triglycerides, PT, INR, plasma phylloquinone **Factor 4:** phylloquinone intake 56.4%	*	*
(Kim et al., 2007)	*	**1.** 3 **2. Korean traditional:** rice and kimchi, fish and seaweed, and legumes **Western :** flour and bread, pizza and hamburgers, snacks and cereals, sugars and sweets, meats, and beverages **Modified:** Rice and dumplings	*	
(Maruapula & Chapman-Novakofski, 2007)	*	1. 5 2. **Beer:** types of beers **Meat and Fruit n**: red and white meat. **Vegetable and Bread**: green leafy vegetables and other vegetables and the bread group. **Seasonal Produce:** named because of the seasonality of the food items contained in this dietary pattern (pumpkin, melon, and watermelon) **Milk, Tea, and Candy:** only these food items.	*	**Beer** : positively related with elderly women, those attending church, and those living with grandchildren **Vegetable and Bread**: more common among grandparents living with children and those living in towns (urban). .
(Nettleton et	*	**1. 4**	*	*

al., 2007)		**2. Fats and Processed Meat:** added fats, processed meat, fried potatoes, and desserts **Vegetables and Fish** :several vegetable groups, fish, soup, Chinese foods, red meat, poultry, and soy), Beans, Tomatoes, and Refined Grains (beans, tomatoes) **Refined grains,** high-fat dairy foods, avocado/guacamole, and red meat),and the whole Grains **Fruit** **3. 29.1%**		
(Robinson et al., 2007)	*	1. 3 2. **First component** : vegetables, fruit, meat and fish, other home-prepared foods and greater breast milk, and a low frequency of consumption of commercial baby foods in jars and lower consumption of formula milk.. **Second component**: bread, savoury snacks, biscuits, squash, breakfast cereals and chips, but by low frequency of consumption of breast milk, baby rice and cooked and tinned fruit. **Third component**: 'wet' commercial baby foods, most commonly available in jars, but by a low frequency of consumption of dried commercial baby foods. **21%**	*	*
(Dalvi, Canchola & Horn-Ross, 2007)	*	1. 4 2. **Plant-based:** of fruits and vegetables **Western:** convenience and processed foods and sweets. **Ethnic:** tortillas, cornbread, beans, cheese dishes such as macaroni and cheese, and traditional Latino dishes, e.g., tacos and burritos **phytoestrogen-rich**: tofu, sprouts, miso soup, and selected vegetables. 3. 15.4%	*	*
(Teucher et al., 2007)	*	1. 5 2. **Fruit and vegetable:** fruit, allium and cruciferous vegetables; low intakes of fried potatoes **High alcohol:** beer, wine and allium vegetables; low intakes of high fiber breakfast cereals and fruit. **Traditional English:** fried fish and	*	*

		potatoes, meats, savoury pies and cruciferous vegetables **Dieting**: low-fat dairy products, low-sugar soda; low intake of butter and sweet baked products. **Low meat:** baked beans, pizza and soy foods; low intakes of meat, other fish and seafood, and poultry 3. 22%		
(Burt et al., 2006)	*	1. 2 2. **Liquid** **Solid food**	*	*
(Crozier et al., 2006)	*	1. 2 2. **Prudent:** fruit and vegetables, wholemeal bread, rice and pasta, yoghurt and breakfast cereals, and low intakes of chips and roast potatoes, sugar, white bread, red and processed meat, full-fat dairy products, crisps, Yorkshire puddings and savoury pancakes, confectionery, tea and coffee, tinned vegetables, cakes and biscuits, and soft drinks **Western/Overall diet:** fruit and vegetables, puddings, meat and fish, eggs and egg dishes, cakes and biscuits, full-fat spread, cooking fats and salad oils, and potatoes.	*	*
(Cuco et al., 2006)	*	1. 2 **2. Sweetened beverages and sugars** **Vegetables and meat** 3. 11.62% and 15.06% ‘	*	**Vegetables and meat**: negatively associated with the BMI during the preconception period and positively associated with age in weeks 10 and 38 of pregnancy
(Naska et al., 2006)	*	1. 2 2. **First component**: fruits, vegetables and cereals to meat, fish and dairy products **Second component**: beverages, alcoholic and non-alcoholic 3. 15-20%	*	*
(Newby et al., 2006b)	*	Same as (Newby et al., 2006a)	Same as (Newby et al., 2006a)	Same as (Newby et al., 2006a)
(Paradis, Perusse &	*	1. 2 **2. Western:** red meats, poultry, processed meats, refined grains, and dessert **Prudent:** vegetables, fruits, and fish and	*	**Western** : negatively associated with age and positively associated with physical activity, smoking and personal income

Vohl, 2006)		other seafood 3. 19.9% and 17.2%		**Prudent** : positively associated with physical activity and negatively associated with BMI and
(Schulze et al., 2006)	*	1. 2 2. **Prudent:** fruits, vegetables **Western:** red and processed meats, refined grains, sweets and desserts, and potatoes whole grains, fish, poultry, and salad dressing	*	*
(Weismayer, Anderson & Wolk, 2006)	*	1. 3 2. **Healthily:** fruits, tomatoes, vegetables, cereal, and fish **Western :** meat, processed meat, fried potatoes, soft drinks, and sweets **Alcohol:** beer, wine, and liquor consumption as well as snacks	*	*
(Engeset et al., 2005)	*	1. 6 **2. Traditional fish eaters Healthy Average** **Less fish** **Less healthy** **Western** **Traditional bread eaters** **Alcohol users** **3.** 23.7%	*	**Traditional fish** eaters and the **Traditional bread eaters**: positively associated with lower income and lower education. **Healthy and the alcohol users** : positively associated with higher income and south and east location **Alcohol users**: positively associated with current smokers.
(Fung et al., 2005)	*	1. 2 2. **Prudent:** fruits, vegetables, whole grains, fish and poultry **Western**: processed and red meats, refined grains, sweets and desserts.	*	*
(Hoffmann et al., 2005)	*	1. 2 2. **First PCA:** potatoes, vegetables, legumes, bread, all types of meat, eggs, sauces and soups **Second PCA:** vegetables, fruits, dairy products, other cereals, vegetable oils and non-alcoholic beverages, low consumption of alcoholic beverages other than wine. 3. 11·3 and 8·4%	*	*
(Marchioni et al., 2005)	*	1. 2 **2. Prudent:** fruits and vegetables **Traditional:** cereals and snacks **Snacks:** dairy products and processed meat**.**	*	*

(Mikkila et al., 2005)	*	1. 3 2. **Traditional Finnish foods**: rye, potatoes, milk, butter, sausages and coffee, and low consumption with fruit, berries and dairy products other than milk. **Health-conscious foods:** vegetables, legumes and nuts, tea, rye, cheese and other dairy products, and also alcoholic beverages.	*	*
(Northstone & Emmett, 2005)	*	1. 3. 2. **Junk** : high-fat processed foods (sausages, burgers, coated poultry) and snack foods high in fat and/or sugar (such as crisps, sweets, chocolate, ice lollies and ice creams) **Traditional**: meat and vegetables and **Vegetarian:** meat substitutes, pulses and nuts.	*	**Junk**: was significantly more likely in white children, where maternal education level was low and where the child had more siblings. **Traditional:** was more likely in girls, where the mother had a partner and in no vegetarians (both mother and child). **Vegetarian:** was more likely with increasing levels of education and increasing maternal, age.
(Park et al., 2005)	*	1. 3 2. **Fat and Meat**: discretionary fat, meat, eggs, and cheese. **Vegetables**, **Fruit and Milk:** milk and yogurt and fruit groups. 3. 63%	*	**Fat and Meat** : positively associated with current smokers **Vegetables and Fruit and Milk** : positively associated with physical activity
(Perrin et al., 2005)	*	1. 2 2. **Western:** sugar and sweets, grains, butter, added fats, eggs, dairy products, potatoes, cheese and fruit. **Prudent:** fruit, vegetables, olive oil, dairy products and fish 3. 26·7%	*	**Prudent:** positively associated with region, educational and income tax levels, leisure-time physical activity and smoking status.
(Uusitalo et al., 2005)	*	1. 3 **2. Western:** **Bread/butter Traditional** **High protein and Margarine/milk**	*	**Western :** positively related with people who were younger, educated and wealthier subjects
(Yang, Kerver & Song, 2005)	*	1. 3 **2. Vegetable/fruit** **Traditional Korean** **Acculturated American**	*	**Traditional Korean:** was negatively associated with length of residence in the U.S. for both men and women ($p < 0.01$).
(Corrao et al., 2004)	*	1. 2 2. **PC1** **PC2** 3. 75%	*	*

(Khani et al., 2004)	*	1. 2 2. **Healthy:** vegetables, fruits, fish, poultry, tomato, whole grains, cereal and low-fat dairy products. **Western**: processed meat, meat, refined grains, sweets, margarine, high-fat dairy, potatoes, and soda. **Drinker:** wine, liquor, and beer and snacks.	*	*
(Newby et al., 2004)	*	1. 3 **2. Healthy** **Western** **Third component, canteen:** pasta, tomato sauce, and wine.	*	*
(Robinson et al., 2004)	*	1. 1 2. **Pattern 1:** chips and roast potatoes, sugar, white bread, red, and processed meat and full-fat dairy products	*	**Pattern 1:** positively associated with educational attainment ,smoking, watching television, lack of strenuous exercise, and living with children

*no information provided

Table D. Foods items which correlated >0.3 or <-0.3 with a "western" or a "Prudent" dietary pattern in different studies (Food items that didn't correlated > 0.3 or <-0.3 with any of the two patterns excluded from the table; Sorted alphabetically with year of publication and author; Studies that didn't identify a "Prudent" or "western" pattern were excluded from the table)

	Western (foods that deemed to constitute the pattern with corresponding principal component loadings)	**Prudent (foods that deemed to constitute the dietary pattern with corresponding principal component loadings)**
(Agurs-Collins et al., 2009)	eggs 0.41 french fries 0.55 high-fat dairy products 0.46 margarine, butter, and mayonnaise 0.40 potato 0.36 processed meat 0.62 refined grains 0.47 snacks 0.45 soda 0.42 sweets 0.47 total meat 0.65	beans 0.49 cruciferous vegetables 0.65 fish 0.48 fruits 0.61 juice 0.34 low-fat dairy products 0.39 other vegetables 0.75 pasta 0.35 poultry 0.36 soup 0.41 tomatoes 0.48 whole grains 0.54
(Ambrosini et al., 2009)	cakes, biscuits 0.34 confectionery 0.46 crisps 0.39 full fat dairy products 0.30 potato, fried e.g. french fries 0.39 potato, not fried 0.34 processed meats 0.41 red meat 0.46 refined grains 0.42 sauces and dressings 0.34 soft drinks 0.37 takeaway foods 0.53	
(Bakolis et al., 2010)	butter beans/broad beans roast potatoes 0.30 ham 0.31 ice cream 0.31	avocado 0.37 bean sprouts 0.30 broccoli 0.38 carrots 0.33

	pork – roast, chops 0.32 pork stew, casserole 0.32 omelette/scrambled egg 0.32 fruit pies, tarts, crumbles 0.32 beef stew, casserole, mince, curry 0.34 sponge cakes 0.34 fried fish in batter/breadcrumb 0.35 baked beans 0.36 chocolate biscuits 0.36 sandwich/cream biscuits 0.36 corned beef, spam, luncheon meat 0.37 white bread and rolls 0.39 bacon 0.40 fried egg 0.40 milk chocolate 0.40 bread crumbed e.g. chicken nuggets 0.41 crisps 0.41 chocolate snack bars 0.44 meat pizza 0.46 other fried snacks 0.46 chips 0.47 sausages – beef, pork 0.48 beef burger, hamburger 0.53 pies/pasties/sausage rolls/meat 0.53	celery 0.30 cheese – cheddar, brie, edam 0.31 courgettes, marrow, squash 0.47 couscous 0.36 currants, raisins, sultanas 0.40 french type dressing 0.54 fresh oily fish 0.40 garlic 0.49 green beans, runner beans 0.37 leeks 0.42 lettuce 0.46 mayonnaise 0.35 mixed bean casserole/ratatouille 0.45 mushrooms 0.44 onions 0.49 parsley – flat leaf 0.37 parsnips and turnips 0.34 peppers – red, green, yellow 0.52 spinach 0.38 sweetcorn 0.39 tinned oily fish 0.33 tofu – bean curd 0.30 tomatoes – raw, canned, sauce 0.39 vegeburgers 0.31 vegetable – lasagne/moussaka 0.43 vegetable pies/samosas 0.39 vegetable pizza 0.46 watercress, mustard and cress white fish not fried 0.38 white pasta 0.42 wholemeal bread and rolls 0.33 wholemeal pasta 0.33
(Deshmukh-Taskar et al., 2009)	condiments 0.40 dishes with cheese 0.58 eggs 0.39 french fries 0.53	100% fruit juices 0.43 clear soups 0.36 cruciferous vegetables 0.70 dark-yellow vegetables 0.70

	high-fat dairy products 0.53 processed meats 0.59 red meats 0.50 refined grains 0.43 snacks 0.53 sweetened beverages 0.44 sweets and desserts 0.54	fruits 0.64 green leafy vegetables 0.69 legumes 0.61 low-fat dairy products 0.36 low-fat salad dressings 0.49 other vegetables –0.74 poultry 0.40 tomatoes 0.58 whole grains –0.46
(Imamura et al., 2009)	high-fat dairy 0.32 high-fat dairy desserts 0.35 nuts and seeds 0.30 eggs 0.39 processed meat 0.61 meat 0.55 refined grains 0.34 chocolate 0.38 sweet baked goods 0.46 0.37 chowder/cream soup 0.32 soda 0.38 pizza, sandwich, casserole 0.36 potato or corn chips 0.48 fried foods 0.48	reduced-fat dairy 0.39 fruits 0.50 whole grains 0.35 refined-grain cereal 0.33 sweet baked goods 0.37 miscellaneous sweets 0.42
(Kesse-Guyot et al., 2009)		fruits 0.50 tea 0.34 butter 0.47 vegetable oil 0.30 breakfast cereals 0.39 vegetables 0.49 fish and seafood 0.30 reduced-fat products 0.44
(Lutsey et al., 2009)	processed meat, noncereal whole grains added fats oils	vegetables fruit poultry

(Oddy et al., 2009)	takeaway foods 0.53 confectionery 0.46 red meat 0.46 refined grains 0.42 processed meats 0.41 potato, fried e.g. french fries 0.39 crisps 0.39 soft drinks 0.37 cakes, biscuits 0.34 potato, not fried 0.34 sauces and dressings 0.34 full fat dairy products0.30	yellow or red vegetables 0.56 leafy green vegetables 0.49 tomato 0.49 cruciferous vegetables 0.48 other vegetables 0.66 fresh fruit 0.48 legumes 0.43 wholegrain 0.39 fish, steamed, grilled or tinned 0.33
(Paradis et al., 2009)	refined grains 0.68 french fries 0.61 red meats 0.57 condiments 0.50 processed meats 0.50 regular soft drinks 0.48 pizza 0.44 snacks 0.37 potatoes other than french fried 0.35 legumes -0.31 fruits -0.44	non-hydrogenated fat 0.56 vegetables 0.52 eggs 0.46 fish and other seafood 0.45 wine 0.44 coffee 0.42 regular dairy products 0.37 desserts whole grains 0.32
(Robinson et al., 2009)	fruit vegetables oily fish wholemeal cereals	
e3n study		
(Varraso et al., 2009) (Touvier et al., 2009)	condiments and sauces 0.32 onions, garlic 0.73 dough and pastry 0.70 cream desserts 0.62 ice cream 0.60 processed meats 0.55 0.33 cakes, pies and pastries 0.45 pasta, rice and grain 0.40	fruity vegetables 0.89 root vegetables 0.85 cabbages 0.79 mushrooms 0.73 grain and peas 0.72 leafy vegetables (except cabbages) 0.71 stalk vegetables 0.70

	potatoes and other tubers 0.31 egg 0.30	fruits with beta carotene 0.61 fruits with citric 0.60 condiments and sauces 0.42 red meat 0.34 poultry 0.33 blue fish 0.32
(Kesse, Clavel-Chapelon & Boutron-Ruault, 2006)	potatoes 0.45 pizza and pies 0.48 sandwiches 0.32 legumes 0.32 sweets 0.42 cakes 0.41 pasta 0.63 rice 0.55 bread 0.37 processed meat 0.39	
(Cottet et al., 2009)	potatoes 0.33 pulses 0.29 rice, pasta, semolina 0.39 french fries 0.48 appetizers 0.45 pizza, pies 0.39 sandwiches 0.32 cakes 0.36 processed meat 0.59 ham 0.31 offal 0.29 eggs 0.36 canned fish 0.37 crustaceans 0.32 mayonnaise 0.39 butter, cream 0.31 high-alcohol beverages 0.37 wine 0.26	
(Ambrosini et al., 2008)		peas 0.30 cakes 0.33

		white bread 0.45 eggs 0.38 potato crisps 0.36 potato chips (french fries) 0.31 meat pie 0.46 hamburger 0.40 beef 0.42 lamb 0.32 pork 0.37 bacon 0.37 sausages 0.48 full cream milk 0.31 beer (full alcohol) 0.30 fish, fried or takeaway 0.41
southampton women's survey study (sws)		
(Borland et al., 2008)		vegetables fruit wholemeal bread rice/pasta yoghurt and breakfast cereals lower intakes of white bread roast potatoes/ chips red/processed meat full-fat milk full-fat spread, crisps confectionery sugar tea/coffee yorkshire puddings/ pancake tinned vegetables cakes and biscuits soft drinks high energy
(Crozier et al., 2006)		rice and pasta 0.21 white bread -0.22

		wholemeal bread 0.23 full-fat milk -0.20 processed meat -0.21 salad vegetables 0.25 other vegetables 0.23 vegetable dishes 0.21 chips and roast potatoes -0.27 other fruit 0.23 sugar -0.23
(Campbell, Sloan & Kreiger, 2008) Different Gender Men Women	potatoes: baked, broiled, mashed 33* french fries or fried potatoes 34* 34* white bread 35* 36* beef, pork, or lamb as a main dish 25* 36* hamburger 40* 46* hot dogs 37* 40* hamburger 40* 46* hot dogs 37* 40* luncheon meats (salami, bologna) 30* 30* smoked meat or corned beef 26* 25* bacon 34* 45* sausage 31*45* eggs 33* 41* cheese 29* cake 42* doughnuts, pastry 46* 36* pies 39* 35* ice cream 35* 28* chocolate 28* 29* potato chips 32* 28* butter on bread or vegetables 26* 29* mayonnaise or salad dressing on bread or in salads 27* 29*	tomato or vegetable juice 26* apples or pears 43* 38* oranges 34* 38* bananas 33* 40* cantaloupe 29* 35* other fruit, fresh or canned 40*39* tomatoes 41* 41* carrots 40* 54* broccoli 52* 58* cabbage, cauliflower, brussels sprouts 47* 51* spinach or other greens 44* 51* yellow squash 26*44* other vegetables 35* 47* soups with vegetables 32* 30* sweet potatoes 27* baked beans or lentils 30* rice 27* chicken or turkey 32* fish: fresh, frozen canned 33* 31*
(De Stefani et al., 2008a)	red meat 0.49 poultry -0.55 fish -0.52 wine 0.41 cheese -0.34	desserts 0.45 french bread -0.40 raw vegetables 0.41 cooked vegetables 0.55 citrus fruits 0.45

	fried eggs 0.34 wine 0.41	other fruits 0.51
nurses health study i and ii (nhs)		
(Heidemann et al., 2008) Different time points 1984 1986 1990 1994 1998	refined grains 0.58 0.52 0.46 0.50 processed meat 0.57 0.58 0.60 0.58 0.58 red meat 0.550.57 0.60 0.610.62 french fries0.470.48 0.47 0.470.48 condiments 0.45 0.32 0.350.32 0.36 sweets and desserts0.43 0.49 0.43 0.41 0.36 potatoes 0.39 0.36 0.30 0.29 0.33 high-fat dairy 0.37 0.42 0.48 0.47 0.45 pizza 0.35 0.35 0.36 0.34 0.39 mayonnaise 0.34 0.35 0.35 0.33 0.33 high-sugar beverages 0.32 0.32 0.32 0.29 0.30 eggs 0.30 0.33 0.42 0.44 0.40 margarine 0.29 0.27 0.28 0.31 0.25 snacks 0.28 0.32 0.28 0.32 0.32 butter 0.27 0.29 0.33 0.33 0.31 soups 0.22 0.29 0.31 0.34 0.32	other vegetables 0.680.72 0.69 0.71 0.71 green, leafy vegetables 0.65 0.64 0.64 0.64 0.63 cruciferous vegetables0.61 0.60 0.61 0.630.62 legumes 0.59 0.56 0.58 0.59 0.57 dark-yellow vegetables 0.58 0.63 0.65 0.64 0.62 fruit 0.58 0.59 0.57 0.58 0.59 fish 0.51 0.53 0.51 0.47 0.50 tomatoes 0.460.550.49 0.520.53 poultry 0.44 0.41 0.42 0.34 0.400.17 whole grains 0.39 0.39 0.41 0.40 0.40 · salad dressing0.36 0.38 0.34 0.31 0.33 low-fat dairy 0.32 0.32 0.32 0.33 olive oil na na 0.31 0.32 0.39
(Varraso et al., 2007a)	refined grains 0.74 desserts and sweets 0.60 cured meats 0.52 red meats 0.52 french fries 0.44 condiments 0.40 potatoes 0.39 pizza 0.36 full-fat dairy products 0.35 sweetened beverages 0.32 mayonnaise 0.31 margarine 0.30	other vegetables 0.68 leafy vegetables 0.63 cruciferous vegetables 0.61 fruit 0.60 yellow vegetables 0.60 legumes 0.55 fish 0.50 tomatoes 0.45 poultry 0.43 whole-grain products 0.41 low-fat dairy products 0.35 garlic 0.35 salad dressing 0.33
(Schulze et al., 2006) Different time points 1991 1995 1999	red meats 0.61 0.55 0.62 processed meats 0.58 0.54 0.49 french fries 0.50 0.51 0.54 refined grains 0.47 0.43 0.51	other vegetables 0.68 0.70 0.69 green, leafy vegetables0.67 0.65 0.61 dark-yellow vegetables 0.63 0.59 0.55 fruit 0.62 0.61 0.59

	sweets and desserts 0.42 0.43 0.39 potatoes 0.41 0.37 0.43 eggs 0.39 0.39 0.20 snacks 0.39 0.33 0.33 high-fat dairy products 0.36 0.34 0.31 margarine 0.36 0.38 0.32 pizza 0.340.39 0.40 mayonnaise0.34 0.36 0.31 sugar-sweetened soft drinks 0.30 0.28 0.30	cruciferous vegetables 0.58 0.58 0.57 tomatoes 0.54 0.44 0.46 legumes 0.50 0.53 0.50 fish and other seafood0.44 0.39 0.43 oil and vinegar salad dressing0.43 0.50 0.55 whole grains 0.43 0.43 0.41 poultry 0.40 0.24 0.31 garlic 0.38 0.41 0.43 water 0.38 0.39 0.39 condiments 0.31 0.30 -
(Fung et al., 2005)	refined grains 0.74 desserts and sweets 0.60 processed meats 0.52 red meats 0.52 french fries 0.44 condiments 0.40 potatoes 0.39 pizza 0.36 full-fat dairy products 0.35 sweetened beverages 0.32 mayonnaise 0.31 margarine 0.30	other vegetables 0.68 leafy vegetables 0.63 cruciferous vegetables 0.61 fruit 0.60 yellow vegetables 0.60 legumes 0.55 fish 0.50 tomatoes 0.45 poultry 0.43 whole grain products 0.41 low-fat dairy products 0.35 salad dressings 0.33 garlic 0.35
(Kroenke et al., 2005)	refined grains processed and red meats desserts high-fat dairy products french fries	fruits vegetables whole grains legumes poultry fish
(Fung et al., 2004a) Different time points 1984 1986 1990 1994	refined grains 0.58 0.57 0.52 0.44 processed meats 0.56 0.57 0.58 0.57 red meats 0.56 0.56 0.60 0.61 french fries 0.47 0.47 0.46 0.47 condiments 0.44 0.33 0.36 0.29 desserts and sweets 0.43 0.49 0.46 0.46 potatoes 0.41 0.39 0.34 0.34	other vegetables 0.69 0.75 0.68 0.68 leafy vegetables 0.68 0.68 0.67 0.60 cruciferous vegetables 0.61 0.59 0.61 0.63 yellow vegetables 0.56 0.60 0.64 0.66 fruits 0.55 0.55 0.55 0.62 fish 0.52 0.54 0.52 0.43 legumes 0.51 0.49 0.55 0.60

	full fat dairy products 0.36 0.40 0.45 0.43 pizza 0.35 0.34 0.35 0.33 sweetened beverages _0.33 0.34 0.33 0.33 margarine 0.33 0.31 0.33 0.34 mayonnaise 0.32 0.33 0.34 0.27 eggs 0.29 0.32 0.41 0.41 snacks 0.28 0.31 0.29 0.33 butter 0.24 0.25 0.29 0.27 cream soups— 0.30 0.32 0.35 water n/a n/a 0.30 0.35 olive oil n/a n/a 0.35 0.21	tomatoes 0.48 0.57 0.51 0.46 poultry 0.45 0.42 0.42 0.32 garlic 0.41 n/a n/a 0.26 salad dressings 0.39 0.41 0.38 0.24 whole grains 0.35 0.34 0.35 0.42 low-fat dairy products 0.28 0.26 0.28 0.37 fruit juice 0.21 0.22 0.19 0.26 nuts 0.16 0.20 0.15
(Lopez-Garcia et al., 2004) Different time points 1986 1990 1986/1990	processed meats 0.57 0.58 0.59 refined grains 0.57 0.52 0.58 red meats 0.56 0.60 0.61 sweets and desserts 0.49 0.45 0.49 french fries 0.48 0.46 0.49 high-fat dairy products 0.40 0.46 0.45 potatoes 0.38 0.33 0.39 pizza 0.34 0.36 0.37 sugar-containing beverages 0.34 0.33 0.36 mayonnaise and other creamy salad dressings 0.33 0.34 0.36 condiments 0.33 0.37 0.38 eggs 0.32 0.41 0.38 snacks 0.31 0.29 0.31 margarine 0.30 0.32 0.34 cream soup 0.30 0.31 0.32	other vegetables 0.73 0.68 0.73 green, leafy vegetables 0.66 0.65 0.68 dark-yellow vegetables 0.62 0.65 0.66 cruciferous vegetables 0.59 0.61 0.62 fruit 0.58 0.58 0.60 tomatoes 0.56 0.49 0.55 fish and other seafood 0.53 0.51 0.55 legumes 0.51 0.57 0.55 poultry 0.42 0.42 0.44 salad dressings 0.40 0.34 0.39 whole grains 0.38 0.41 0.41 low-fat dairy products 0.30 0.32 0.32
(Iqbal et al., 2008)	dairy 0.56 nuts 0.29 glv 0.32 raw vegetables other than glv 0.63 fruits 0.68 desserts 0.40	eggs 0.44 meats 0.39 fried foods 0.63 salty foods 0.61 sugar 0.32
(Kubo et al., 2008)	beef 59 french fries 53 hamburger 52	

	pizza 49 refried beans 49 mustard 47 tacos 45 white potatoes 44 pork chop 44 pork spareribs 44 chili 43 hot dog 43 salty snacks 43 spaghetti 40 fried chicken 38 bacon 36 sausage 36 jelly 35	
(Lutsey, Steffen & Stevens, 2008)	refined-grain bread, cereal, rice, and pasta 0.63 processed meat 0.63 fried foods 0.61 red meat 0.57 eggs 0.48 refined-grain desserts 0.43 soda and sweetened beverages 0.41 cheese and whole milk 0.38 legumes 0.35 sweets/candy 0.30 fat 0.30	cruciferous vegetables 0.62 carotenoid vegetables 0.60 fruit (no juice) 0.58 other vegetables 0.52 fish and seafood 0.46 poultry 0.43 dark leafy vegetables 0.43 whole grains 0.40 tomatoes 0.39 legumes 0.34 low-fat dairy 0.31
(Murtaugh et al., 2008)	low-fat dairy whole grains fruit and fruit juice legumes vegetables soups	high-fat dairy foods, refined grains gravy and sauces fast foods red and processed meats potatoes margarine polyunsaturated fat

		high-fat and high-sugar desserts
54 universities in japan		
(Okubo et al., 2008)	breads-0.28 confectionary -0.35 meats 0.60 fats and oils 0.58 seasonings 0.50 processed meats 0.45 eggs 0.30	
(Okubo et al., 2007)	confectionaries -0.33 fats and oils 0.60 meats 0.58 seasonings 0.51 processed meats 0.46 eggs 0.33	
(Romaguera et al., 2008)	beef (0.51) lamb (-0·61) common bread (0·42) bollo and tortilla (-0·52) chicken (0·42) animal fat (-0·46) fruit (0·38) creole potatoes (-0·43) sugary drinks (0·29) mote (-0·30) common potatoes (0·25) herbal teas (-0·32) yoghurt (0·25) llama (-0·27) green beans (0·23) vegetables (-0·20) sweet and milky desserts (0·21)	
(Sadakane et al., 2008)		high –fat products -0.32 bread 0.59 butter 0.53 rice -0.51

		salty products -0.48 miso-soup -0.43 yoghurt 0.42 fruits 0.30 milk 0.38
(De Stefani et al., 2007)	fried meat 0.62 barbeque meat 0.56 poultry -0.31 processed meat 0.32 boiled eggs 0.32 fried eggs 0.39	
(Hirose et al., 2007)		carrot 66 green leafy vegetables 65 potato 53 pumpkin 52 cabbage 48 soy bean curd (tofu) 46 fruit 45 raw vegetables 45 cooked/raw fish 30 milk 30 lettuce 36
(Kim et al., 2007)	flour and bread pizza and hamburgers snacks and cereals sugars and sweets meats beverages	
(Meyerhardt et al., 2007)	high-fat dairy 0.67 low-fat dairy 0.64 refined grains 0.60 condiments 0.51 red meat 0.53 sweets and desserts 0.53 margarine 0.50 processed meat 0.45	vegetables 0.72 leafy vegetables 0.71 yellow vegetables 0.67 cruciferous vegetables 0.65 legumes 0.56 fruit 0.55 light salad dressing 0.48 tomatoes 0.46 0.36

	potatoes 0.17 0.45 regular mayonnaise 0.35 butter 0.33 french fries −0.16 0.37 eggs 0.30 snacks 0.36 nuts 0.30	garlic 0.39 fish 0.46 poultry 0.37 fruit juice 0.35 whole grains 0.32 low-fat mayonnaise 0.31
(Murtaugh et al., 2007)	high-cholesterol eggs 0.32 high-fat dairy 0.41 refined grains (regular) 0.60 refined-grain snacks (regular) 0.43 refined-grain cereals (regular) 0.33 gravy and sauces 0.54 sauces, tomato-based 0.39 fast-food vegetables (french fries) 0.58 fast-food beef sandwiches, hamburgers 0.59 fast-food chicken 0.50 bacon, sausage, cold cuts 0.54 potatoes 0.55 margarine 0.40 polyunsaturated oils 0.39 sugar 0.43 high-fat, high-sugar desserts 0.41 no-fat, high-sugar desserts 0.30 meats 0.54 mexican meats 0.30	low-fat dairy 0.35 whole grains (regular) 0.34 whole-grain cereals 0.44 orange, grapefruit, citrus juices 0.40 fruit juices other than citrus 0.50 canned fruit 0.46 dried fruit 0.51 soups (broth or cream based) 0.37 soy beans, tofu 0.32 salad greens, lettuce 0.30 legumes, beans 0.36 nuts 0.37 tea, herbal 0.32 fresh fruit 0.60 vegetables 0.47
(Sant et al. 2007)	potatoes ravioli red and processed meat eggs butter seed oil cakes.	
health professionals study (hps)		
(Qi et al., 2009)	processed meat, red meat, butte	vegetables fruit

	high-fat dairy products eggs refined grains	legumes whole grains fish poultry
(Wu et al., 2004a) (Varraso et al., 2007b)	red meat 0.64 0.67 0.67 processed meat 0.60 0.62 0.61 refined grains 0.46 0.39 0.34 french fries 0.46 0.50 0.50 high-fat dairy 0.44 0.50 0.49 sweets and desserts 0.41 0.42 0.41 eggs 0.41 0.48 0.46 condiments 0.36 – 0.37 high-sugar drinks 0.33 0.32 0.29 snacks 0.33 0.31 0.30 mayonnaise 0.34 0.34 0.35 butter 0.32 0.35 0.36	other vegetables 0.74 0.70 0.69 dark-yellow vegetables 0.61 0.63 0.63 cruciferous vegetables 0.59 0.62 0.62 green, leafy vegetables 0.64 0.60 0.61 legumes 0.58 0.58 0.59 fruit 0.56 0.56 0.55 tomatoes 0.55 0.50 0.51 fish 0.47 0.45 0.42 whole grains 0.36 0.36 0.36 poultry 0.34 0.34 0.28 other salad dressing 0.35 0.36 0.24
(Michaud et al., 2005) Different Studies HPFS NHS	red meat .64 .53 processed meat .60 .54 french fries .45 .45 refined grains .45 .72 high-fat dairy products.44 .36 condiments .42 .44 eggs .42 .26 sweets and desserts .38 .58 mayonnaise .34 .33 snacks .33 .30 sugar drinks .32 .31 butter .32 .25 margarine .28 .28 potatoes .28 .38	vegetables .73 .69 leafy vegetables .63 .66 yellow vegetables .61 .58 cruciferous vegetables .59 .61 legumes .59 .53 fruit .58 .58 tomatoes .54 .46 fish .47 .51 whole grains .36 .39 poultry.34 .44 salad dressing .33 .37 low-fat dairy products — .32
(Pala et al., 2006)		other vegetables 0.69 legumes (pulses) 0.61 leaf vegetables cooked 0.54 onions, garlic 0.52 cabbage 0.43 fish 0.42

		crustaceans mullocs 0.42 mushrooms 0.37 seed oils 0.33 tomatoes cooked 0.32 fresh fruit (non-citrus) 0.32 nuts and seeds 0.30
(Paradis, Perusse & Vohl, 2006) Different Gender Men Women	red meats 0.78 0.80 butter 0.62 poultry 0.59 0.63 high-fat dairy products 0.56 processed meats 0.550.68 potatoes other than french fried0.47 refined grains0.46 0.38 condiments 0.44 french fries 0.44 mayonnaise 0.36 desserts0.33 0.30	vegetables 0.76 0.65 fruits 0.68 0.73 non-hydrogenated fat 0.62 fish and other seafood 0.33 wine 0.30 nuts0.48 legumes 0.47-0.41 whole grains 0.41 organ meats 0.37 fruit juices -0.33
(Ronco et al., 2006)		fried meat 0.81 barbecue 0.66 processed meat 0.47
(Cottet et al., 2005) (women only)	poultry 0.47 fish and crustaceans 0.32 high-fat delicatessen 0.37 vegetable fat 0.49 nuts 0.48 legumes 0.30 potatoes 0.67 refined bread and cereals 0.36 milk 0.30 sodas 0.35 beer 0.39 condiments 0.43	
(Marchioni et al., 2005)		meat 0.617 vegetables 0.817 fruits 0.651

(Montonen et al., 2005)		whole milk -0.30 red meat 0.32 yellow and red vegetables 0.64 green vegetables 0.63 fruit 0.62 vegetables, other 0.57 poultry 0.35 eggs 0.34
(Nkondjock et al., 2005)	processed meats 0.52 sweets and desserts 0.49 refined grains 0.48 potatoes 0.43 processed fish 0.41 organ meats 0.39 soft drinks 0.36 legumes and legume products 0.36 snacks 0.34 margarine 0.33 nuts 0.32	
(Perrin et al., 2005)	sugar and sweets 0·70 grains 0·54 butter 0·50 added fats 0·44 eggs 0·43 dairy products 0·37	fruit 0·56 vegetables 0·54 olive oil 0·48 fish 0·32 alcohol 0·44 high-fat meat 0·50 potatoes 0·55 dairy products 0·34
(Rashidkhani et al., 2005)	sweets 0.56 processed 0.55 refined grains 0.54 added fat 0.51 high-fat dairy 0.49 fried potatoes 0.41 soft drinks 0.4 meat beef 0.4 cooked 0.33	

(Khani et al., 2004)	sweets 0.56 processed meat 0.55 refined grains 0.54 margarine 0.51 high-fat dairy 0.49 fried potatoes 0.41 soda 0.40 meat 0.40 cooked potato 0.33	
(Uusitalo et al., 2005) Different time points 1998 1992 1998	rice -0.43 -0.34 -0.11 bread 0.27 0.20 0.12 pulses -0.30 -0.10 0.00 poultry 0.24 0.24 0.40 processed meat 0.15 0.14 0.40 fresh/frozen fish 0.14 0.25 0.40 butter -0.18 -0.10 -0.02 margarine 0.24 0.08 0.11 whole milk -0.16 -0.39 -0.09 skimmed/low-fat milk 0.33 ,0.41 0.17	
(Kim et al., 2004)	butter 0.40 0.37 mayonnaise 0.37 0.36 cheese 0.48 0.38 beef 0.54 0.45 pork 0.39 0.48 poultry 0.40 0.45 bacon 0.49 0.55 liver 0.46 0.38 soda beverages 0.35 0.42 fruit juice 0.39 0.40 vegetable juice 0.38 0.32 instant noodles 0.34 0.31 coffee 0.21 0.26 black tea 0.25 0.24	

Table E . Descriptive Statistics of our systematic review

Justifications for the use of PCA vs. Single Food or Nutrient Analysis	Number / total number of papers	Percentage (%) unless otherwise stated
- Interactive, antagonistic and synergistic effects of foods consumed in combination.	25/163	15.3
- Additive effects of foods consumed in combination which are too small to detect when they are examined separately	14/163	8.5
- Confounding and mulitcollinearity from lifestyle factors and dietary exposures	52/163	31.9
- Multiple testing problems	7/163	4.2
- Public health recommendations	14/163	8.5
- Complexity of diet	25/163	15.3
- Better evaluation / representation of overall diet	30/163	18.4
Study Design		
- Case-control	31/163	17.1
- Cohort	31/163	19.0
- Cross-sectional	99/163	64.0
- Other	4/163	2.4
Dietary assessment instrument		
- Food Frequency Questionnaire	152/163	93.2
- 24/48 hour recall	4/163	2.4
- Dietary records	3/163	1.8
- Diet history questionnaire	3/163	1.8
- number of food items in each instrument	163	median value:92 (IQR:21-204)
- number of food groups in each instrument	163	median value:38 (IQR:19-69)
- Scale of Food Frequency Questionnaire	163	median value:7 (IQR: 5-10)
Preparation of data before entering the PCA		
- Conversion of food frequency data to grams/d or grams/week	22/163	13.4
- Food items collapsed to food groups	42/163	28.2
- Standardisation of food intake variables	6/163	3.6

- Food intake variables adjusted for energy intake by the residual method	5/163	3.0
Labeling the dietary patterns in each study		
- principal component loading and correlation coefficient cut off points in each study	163	median value:0.3 (IQR: 0.3-0.4)
Method of rotation in each study		
- Orthogonal/ Varimax	85/163	52.1
- Oblique / Promax	5/163	3.0
Identifying the number of dietary patterns in each study		
- Scree Plot	81/163	49.1
- cut-off point s for eigenvalues	163	median value:1.6; (IQR: 1-2)
- dietary patterns interpretability	70/163	42.9
- Van der Voet's test	3/163	1.8
Validation methods of retained dietary patterns	40/163	24%
- Randomly split sample in each study	8/163	4.9
- Cronbach's alpha	5/163	3.0
- Deriving dietary patterns separately for women and men	3/163	1.8
- Barlet's test of sphericity	1/163	0.6
- Kaiser - Meyen-Ollkin test	2/163	1.2
- Identifying simplified dietary patterns	3/163	1.8
- Use of Confirmatory/maximum likelihood factor analysis	8/163	4.9
- φ coefficient for testing inter-correlation	2/163	1.2
- Stricter cut-off points for energy intake	1/163	0.6
- Different method of rotation	2/163	1.2
- Pearson Correlation coefficient in different time points	5/163	3.0
Number of dietary patterns	163	median: 3 (IQR: 2-4)
Percentage of total variance of original food items being explained by the dietary patterns in each study.	163	median: 24(IQR: 19.9-31.3)
Number of studies that examine associations with health outcomes	120/163	73.6
Number of studies that examine associations with socio-economic characteristics.	60/163	36.8
Number of studies which didn't associated Principal Components with any health outcome or socio-economic factor.	29/163	17.7

References

Agurs-Collins, T., Rosenberg, L., Makambi, K., Palmer, J. R. & Adams-Campbell, L. (2009) Dietary patterns and breast cancer risk in women participating in the Black Women's Health Study. *The American Journal of Clinical Nutrition.* 90 (3), 621-628.

Akbaraly, T. N., Brunner, E. J., Ferrie, J. E., Marmot, M. G., Kivimaki, M. & Singh-Manoux, A. (2009) Dietary pattern and depressive symptoms in middle age. *The British Journal of Psychiatry : The Journal of Mental Science.* 195 (5), 408-413.

Akesson, A., Weismayer, C., Newby, P. K. & Wolk, A. (2007) Combined effect of low-risk dietary and lifestyle behaviors in primary prevention of myocardial infarction in women. *Archives of Internal Medicine.* 167 (19), 2122-2127.

Ambrosini, G. L., de Klerk, N. H., Mackerras, D., Leavy, J. & Fritschi, L. (2008a) Dietary patterns and surgically treated benign prostatic hyperplasia: a case control study in Western Australia. *BJU International.* 101 (7), 853-860.

Ambrosini, G. L., Fritschi, L., de Klerk, N. H., Mackerras, D. & Leavy, J. (2008b) Dietary patterns identified using factor analysis and prostate cancer risk: a case control study in Western Australia. *Annals of Epidemiology.* 18 (5), 364-370.

Ambrosini GL, Huang RC, Mori TA, Hands BP, O'Sullivan TA, de Klerk NH, et al.(2010) Dietary patterns and markers for the metabolic syndrome in Australian adolescents. *Nutr Metab Cardiovasc Dis*.;20:274–283

Armstrong, B., and R. Doll (1975). Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int J Cancer 15*, 617–631.

Anderson, T. W. (2003) *An introduction to multivariate statistical analysis.* Wiley series in probability and statistics. 3rd edition. Hoboken, N.J. ; Great Britain, Wiley-Interscience.

Anderson, A. L., Harris, T. B., Tylavsky, F. A., Perry, S. E., Houston, D. K., Hue, T. F., Strotmeyer, E. S., Sahyoun, N. R. & Health ABC Study. (2011) Dietary patterns and survival of older adults. *Journal of the American Dietetic Association.* 111 (1), 84-91.

Appel, L. J., Moore, T. J., Obarzanek, E., Vollmer, W. M., Svetkey, L. P., Sacks, F. M., Bray, G. A., Vogt, T. M., Cutler, J. A., Windhauser, M. M., Lin, P. H. & Karanja, N. (1997) A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *The New England Journal of Medicine.* 336 (16), 1117-1124.

Arkkola, T., Uusitalo, U., Kronberg-Kippila, C., Mannisto, S., Virtanen, M., Kenward, M. G., Veijola, R., Knip, M., Ovaskainen, M. L. & Virtanen, S. M. (2008) Seven distinct dietary patterns identified among pregnant Finnish women--associations with nutrient intake and sociodemographic factors. *Public Health Nutrition.* 11 (2), 176-182.

Bair, Hastie,, Paul & Tibshirani (2004). Prediction by supervised principal components. Stanford University Department of Statistics Tech Report

Bakolis, I., Hooper, R., Thompson, R. L. & Shaheen, S. O. (2010) Dietary patterns and adult asthma: population-based case-control study. *Allergy.* 65 (5), 606-615.

Balder, H. F., Goldbohm, R. A. & van den Brandt, P. A. (2005) Dietary patterns associated with male lung cancer risk in the Netherlands Cohort Study. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 14 (2), 483-490.

Balder, H. F., Virtanen, M., Brants, H. A., Krogh, V., Dixon, L. B., Tan, F., Mannisto, S., Bellocco, R., Pietinen, P., Wolk, A., Berrino, F., Van den Brandt, P. A., Hartman, A. M. & Goldbohm, R. A. (2003) Common and country-specific dietary patterns in four European cohort studies. *The Journal of Nutrition.* 133 (12), 4246-4251.

Bamia, C., Trichopoulos, D., Ferrari, P., Overvad, K., Bjerregaard, L., Tjonneland, A., Halkjaer, J., Clavel-Chapelon, F., Kesse, E., Boutron-Ruault, M. C., Boffetta, P., Nagel, G., Linseisen, J., Boeing, H., Hoffmann, K., Kasapa, C., Orfanou, A., Travezea, C., Slimani, N., Norat, T., Palli, D., Pala, V., Panico, S., Tumino, R., Sacerdote, C., Bueno-de-Mesquita, H. B., Waijers, P. M., Peeters, P. H., van der Schouw, Y. T., Berenguer, A., Martinez-Garcia, C., Navarro, C., Barricarte, A., Dorronsoro, M., Berglund, G., Wirfalt, E., Johansson, I., Johansson, G., Bingham, S., Khaw, K. T., Spencer, E. A., Key, T., Riboli, E. & Trichopoulou, A. (2007) Dietary patterns and survival of older Europeans: the EPIC-Elderly Study (European Prospective Investigation into Cancer and Nutrition). *Public Health Nutrition.* 10 (6), 590-598.

Barros, R., Moreira, A., Fonseca, J., de Oliveira, J. F., Delgado, L., Castel-Branco, M. G., Haahtela, T., Lopes, C. & Moreira, P. (2008) Adherence to the Mediterranean diet and fresh fruit intake are associated with improved asthma control. *Allergy.* 63 (7), 917-923.

Bartholomew, D. J. & Steele. (2002) *The analysis and interpretation of multivariate data for social scientists.* Texts in statistical science series. Boca Raton, FL, Chapman & Hall/CRC Press.

Bastos, J., Lunet, N., Peleteiro, B., Lopes, C. & Barros, H. (2010) Dietary patterns and gastric cancer in a Portuguese urban population. *International Journal of Cancer.Journal International Du Cancer.* 127 (2), 433-441.

Benjamini, Y. & Yekutieli, D. (2005) Quantitative trait Loci analysis using the false discovery rate. *Genetics.* 171 (2), 783-790.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society.Series B (Methodological).* 57 (1), 289-300.

Bertuccio, P., Edefonti, V., Bravi, F., Ferraroni, M., Pelucchi, C., Negri, E., Decarli, A. & La Vecchia, C. (2009) Nutrient dietary patterns and gastric cancer risk in Italy. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 18 (11), 2882-2886.

Bohlscheid-Thomas, S., Hoting, I., Boeing, H. & Wahrendorf, J. (1997) Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. *International Journal of Epidemiology.* 26 Suppl 1, S71-81.

Boos D. & Hughes-Oliver J. (May, 2000 How Large Does n Have to be for Z and t Intervals? *The American Statistician.* Vol. 54, No. 2, pp. 121-128

Borland, S. E., Robinson, S. M., Crozier, S. R., Inskip, H. M. & SWS Study Group. (2008) Stability of dietary patterns in young women over a 2-year period. *European Journal of Clinical Nutrition.* 62 (1), 119-126.

Bountziouka V. and Panagiotakos D. (2012). The Role of Rotation Type used to Extract Dietary Patterns through Principal Component Analysis, on their Short-Term Repeatability. Journal of Data Science 10(2012), 19-36

Brantsaeter, A. L., Haugen, M., Samuelsen, S. O., Torjusen, H., Trogstad, L., Alexander, J., Magnus, P. & Meltzer, H. M. (2009) A dietary pattern characterized by high intake of vegetables, fruits, and vegetable oils is associated with reduced risk of preeclampsia in nulliparous pregnant Norwegian women. *The Journal of Nutrition.* 139 (6), 1162-1168.

Bradford-Hill, Austin (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, **58**: 295–300

Burney P, Potts J, Makowska J, et al. (2008) A case–control study of the relation between plasma selenium and asthma in European populations: a GA2LEN project. *Allergy* 63, 865–871.

Burt, B. A., Kolker, J. L., Sandretto, A. M., Yuan, Y., Sohn, W. & Ismail, A. I. (2006) Dietary patterns related to caries in a low-income adult population. *Caries Research.* 40 (6), 473-480.

Butland, B. K., Fehily, A. M. & Elwood, P. C. (2000) Diet, lung function, and lung function decline in a cohort of 2512 middle aged men. *Thorax.* 55 (2), 102-108.

Butler, L. M., Koh, W. P., Lee, H. P., Tseng, M., Yu, M. C., London, S. J. & Singapore Chinese Health Study. (2006) Prospective study of dietary patterns and persistent cough with phlegm among Chinese Singaporeans. *American Journal of Respiratory and Critical Care Medicine.* 173 (3), 264-270.

Butler, L. M., Wang, R., Koh, W. P. & Yu, M. C. (2008) Prospective study of dietary patterns and colorectal cancer among Singapore Chinese. *British Journal of Cancer.* 99 (9), 1511-1516.

Cadima, J., and Jolliffe, I. T. (1995) Loadings and Correlations in the Interpretation of Principal Components. *Journal of Applied Statistic.* 22, 203-214.

Cai, H., Shu, X. O., Gao, Y. T., Li, H., Yang, G. & Zheng, W. (2007) A prospective study of dietary patterns and mortality in Chinese women. *Epidemiology (Cambridge, Mass.).* 18 (3), 393-401.

Calvert, C., Cade, J., Barrett, J. H. & Woodhouse, A. (1997) Using cross-check questions to address the problem of mis-reporting of specific food groups on Food Frequency Questionnaires. UKWCS Steering Group. United Kingdom Women's Cohort Study Steering Group. *European Journal of Clinical Nutrition.* 51 (10), 708-712.

Campbell, P. T., Sloan, M. & Kreiger, N. (2008) Dietary patterns and risk of incident gastric adenocarcinoma. *American Journal of Epidemiology.* 167 (3), 295-304.

Carpenter, K. J. (1986) *The history of scurvy and vitamin C.* Cambridge, Cambridge University Press.

Cattell, R.B. (1966). The scree test for the number of factors. *Multiv. Behav. Res.*, **1**, 245–276.

Cattell, R.B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum Press.

Chang, E. T., Lee, V. S., Canchola, A. J., Dalvi, T. B., Clarke, C. A., Reynolds, P., Purdie, D. M., Stram, D. O., West, D. W., Ziogas, A., Bernstein, L. & Horn-Ross, P. L. (2008) Dietary patterns and risk of ovarian cancer in the California Teachers Study cohort. *Nutrition and Cancer.* 60 (3), 285-291.

Chatfield, C. & Collins, A. J. (1980) *Introduction to multivariate analysis.* Science paperbacks. London, Chapman and Hall.

Chatzi, L., Apostolaki, G., Bibakis, I., Skypala, I., Bibaki-Liakou, V., Tzanakis, N., Kogevinas, M. & Cullinan, P. (2007) Protective effect of fruits, vegetables and the Mediterranean diet on asthma and allergies among children in Crete. *Thorax.* 62 (8), 677-683.

Chen, Y., Factor-Litvak, P., Howe, G. R., Parvez, F. & Ahsan, H. (2006) Nutritional influence on risk of high blood pressure in Bangladesh: a population-based cross-sectional study. *The American Journal of Clinical Nutrition.* 84 (5), 1224-1232.

Comrey, A. L. (1988) Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology.* 56 (5), 754-761.

Corrao, G., Zambon, A., Bagnardi, V., Arico, S., Loguercio, C., D'Amicis, A. & Collaborative SIDECIR Group. (2004) Nutrient intakes, nutritional patterns and the risk of liver cirrhosis: an explorative case-control study. *European Journal of Epidemiology.* 19 (9), 861-869.

Costacou, T., Bamia, C., Ferrari, P., Riboli, E., Trichopoulos, D. & Trichopoulou, A. (2003) Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population. *European Journal of Clinical Nutrition.* 57 (11), 1378-1385.

Cottet, V., Bonithon-Kopp, C., Kronborg, O., Santos, L., Andreatta, R., Boutron-Ruault, M. C., Faivre, J. & European Cancer Prevention Organisation Study Group. (2005) Dietary patterns and the risk of colorectal adenoma recurrence in a European intervention trial. *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP).* 14 (1), 21-29.

Cottet, V., Touvier, M., Fournier, A., Touillaud, M. S., Lafay, L., Clavel-Chapelon, F. & Boutron-Ruault, M. C. (2009) Postmenopausal breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study. *American Journal of Epidemiology.* 170 (10), 1257-1267.

Craig, L. C., McNeill, G., Macdiarmid, J. I., Masson, L. F. & Holmes, B. A. (2010) Dietary patterns of school-age children in Scotland: association with socio-economic indicators, physical activity and obesity. *The British Journal of Nutrition.* 103 (3), 319-334.

Crozier, S. R., Inskip, H. M., Godfrey, K. M. & Robinson, S. M. (2008) Dietary patterns in pregnant women: a comparison of food-frequency questionnaires and 4 d prospective diaries. *The British Journal of Nutrition.* 99 (4), 869-875.

Crozier, S. R., Robinson, S. M., Borland, S. E., Inskip, H. M. & SWS Study Group. (2006) Dietary patterns in the Southampton Women's Survey. *European Journal of Clinical Nutrition.* 60 (12), 1391-1399.

Cuco, G., Fernandez-Ballart, J., Sala, J., Viladrich, C., Iranzo, R., Vila, J. & Arija, V. (2006) Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. *European Journal of Clinical Nutrition.* 60 (3), 364-371.

Cui, X., Dai, Q., Tseng, M., Shu, X. O., Gao, Y. T. & Zheng, W. (2007) Dietary patterns and breast cancer risk in the shanghai breast cancer study. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 16 (7), 1443-1448.

Custodio das Dores, S. M., Booth, S. L., Martini, L. A., de Carvalho Gouvea, V. H., Padovani, C. R., de Abreu Maffei, F. H., Campana, A. O. & Rupp de Paiva, S. A. (2007) Relationship between diet and anticoagulant response to warfarin: a factor analysis. *European Journal of Nutrition.* 46 (3), 147-154.

Cutler, G. J., Flood, A., Hannan, P. & Neumark-Sztainer, D. (2009) Major patterns of dietary intake in adolescents and their stability over time. *The Journal of Nutrition.* 139 (2), 323-328.

Dalvi, T. B., Canchola, A. J. & Horn-Ross, P. L. (2007) Dietary patterns, Mediterranean diet, and endometrial cancer risk. *Cancer Causes & Control : CCC.* 18 (9), 957-966.

de Lorgeril, M., Salen, P., Martin, J. L., Monjaud, I., Delaye, J. & Mamelle, N. (1999) Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. *Circulation.* 99 (6), 779-785.

De Stefani, E., Boffetta, P., Ronco, A. L., Deneo-Pellegrini, H., Acosta, G., Gutierrez, L. P. & Mendilaharsu, M. (2008a) Nutrient patterns and risk of lung cancer: a factor analysis in Uruguayan men. *Lung Cancer (Amsterdam, Netherlands).* 61 (3), 283-291.

De Stefani, E., Boffetta, P., Ronco, A. L., Deneo-Pellegrini, H., Acosta, G. & Mendilaharsu, M. (2008b) Dietary patterns and risk of bladder cancer: a factor analysis in Uruguay. *Cancer Causes & Control : CCC.* 19 (10), 1243-1249.

De Stefani, E., Boffetta, P., Ronco, A. L., Deneo-Pellegrini, H., Acosta, G. & Mendilaharsu, M. (2007) Dietary patterns and risk of laryngeal cancer: an exploratory factor analysis in

Uruguayan men. *International Journal of Cancer.Journal International Du Cancer.* 121 (5), 1086-1091.

De Stefani, E., Correa, P., Boffetta, P., Deneo-Pellegrini, H., Ronco, A. L. & Mendilaharsu, M. (2004) Dietary patterns and risk of gastric cancer: a case-control study in Uruguay. *Gastric Cancer : Official Journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association.* 7 (4), 211-220.

De Stefani, E., Deneo-Pellegrini, H., Boffetta, P., Ronco, A. L., Aune, D., Acosta, G., Mendilaharsu, M., Brennan, P. & Ferro, G. (2009) Dietary patterns and risk of cancer: a factor analysis in Uruguay. *International Journal of Cancer.Journal International Du Cancer.* 124 (6), 1391-1397.

DerSimonian, R. & Laird, N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials.* 7 (3), 177-188.

Deshmukh-Taskar, P. R., O'Neil, C. E., Nicklas, T. A., Yang, S. J., Liu, Y., Gustat, J. & Berenson, G. S. (2009) Dietary patterns associated with metabolic syndrome, sociodemographic and lifestyle factors in young adults: the Bogalusa Heart Study. *Public Health Nutrition.* 12 (12), 2493-2503.

Devereux, G. (2010) Session 1: Allergic disease: Nutrition as a potential determinant of asthma. *The Proceedings of the Nutrition Society.* 69 (1), 1-10.

Devereux, G. & Seaton, A. (2005) Diet as a risk factor for atopy and asthma. *The Journal of Allergy and Clinical Immunology.* 115 (6), 1109-17; quiz 1118.

DiBello, J. R., Kraft, P., McGarvey, S. T., Goldberg, R., Campos, H. & Baylin, A. (2008) Comparison of 3 methods for identifying dietary patterns associated with risk of disease. *American Journal of Epidemiology.* 168 (12), 1433-1443.

Dixon, L. B., Balder, H. F., Virtanen, M. J., Rashidkhani, B., Mannisto, S., Krogh, V., van Den Brandt, P. A., Hartman, A. M., Pietinen, P., Tan, F., Virtamo, J., Wolk, A. & Goldbohm, R. A. (2004) Dietary patterns associated with colon and rectal cancer: results from the Dietary Patterns and Cancer (DIETSCAN) Project. *The American Journal of Clinical Nutrition.* 80 (4), 1003-1011.

Drewnowski A, Henderson SA, Driscoll A, Rolls B. The dietary variety score: Assessing diet quality in healthy young and older adults. *J Am Diet Assoc*. 1997; 97:266-271.

Drion, E. F. (1961) Two Papers on Nutrition: The Intercorrelations between the Nutrients Consumed by a Group of Families in the Netherlands. *Journal of the Royal Statistical Society.Series A (General).* 124 (3), pp. 314-335.

D'Souza, S., Levy, E., Mack, D., Israel, D., Lambrette, P., Ghadirian, P., Deslandres, C., Morgan, K., Seidman, E. G. & Amre, D. K. (2008) Dietary patterns and risk for Crohn's disease in children. *Inflammatory Bowel Diseases.* 14 (3), 367-373.

Edefonti, V., Decarli, A., La Vecchia, C., Bosetti, C., Randi, G., Franceschi, S., Dal Maso, L. & Ferraroni, M. (2008) Nutrient dietary patterns and the risk of breast and ovarian cancers. *International Journal of Cancer.Journal International Du Cancer.* 122 (3), 609-613.

Edefonti, V., Randi, G., La Vecchia, C., Ferraroni, M. & Decarli, A. (2009) Dietary patterns and breast cancer: a review with focus on methodological issues. *Nutrition Reviews.* 67 (6), 297-314.

Efron, B. & Tibshirani, R. (1993; 1993) An introduction to the bootstrap. Monographs on statistics and applied probability. New York, Chapman & Hall.

Ender, P. B. (2002) *powerlog: command to perform logistic regression power analysis.* [Online] Available from: http://www.ats.ucla.edu/stat/stata/ado/analysis/ .

Engeset, D., Alsaker, E., Ciampi, A. & Lund, E. (2005) Dietary patterns and lifestyle factors in the Norwegian EPIC cohort: the Norwegian Women and Cancer (NOWAC) study. *European Journal of Clinical Nutrition.* 59 (5), 675-684.

Engeset, D., Dyachenko, A., Ciampi, A. & Lund, E. (2009) Dietary patterns and risk of cancer of various sites in the Norwegian European Prospective Investigation into Cancer and Nutrition cohort: the Norwegian Women and Cancer study. *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP).* 18 (1), 69-75.

Erber, E., Hopping, B. N., Grandinetti, A., Park, S. Y., Kolonel, L. N. & Maskarinec, G. (2010) Dietary patterns and risk for diabetes: the multiethnic cohort. *Diabetes Care.* 33 (3), 532-538.

Esmaillzadeh, A. & Azadbakht, L. (2008a) Food intake patterns may explain the high prevalence of cardiovascular risk factors among Iranian women. *The Journal of Nutrition.* 138 (8), 1469-1475.

Esmaillzadeh, A. & Azadbakht, L. (2008b) Major dietary patterns in relation to general obesity and central adiposity among Iranian women. *The Journal of Nutrition.* 138 (2), 358-363.

Esmaillzadeh, A., Kimiagar, M., Mehrabi, Y., Azadbakht, L., Hu, F. B. & Willett, W. C. (2007) Dietary patterns, insulin resistance, and prevalence of the metabolic syndrome in women. *The American Journal of Clinical Nutrition.* 85 (3), 910-918.

European Community Respiratory Health Survey II Steering Committee. (2002) The European Community Respiratory Health Survey II. *The European Respiratory Journal : Official Journal of the European Society for Clinical Respiratory Physiology.* 20 (5), 1071-1079.

Fahey, M. T., Thane, C. W., Bramwell, G. D. & Coward, W. A. (2007) Conditional Gaussian mixture modelling for dietary pattern analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 170 (1), 149-166.

Fan, J & Lv, J. (2008). Sure independence screening for ultra-high dimensional feature space. *Journal of the Royal Statistical Society: Series B.* 70(5),849-911

Fan & Lv (2010) A selective overview of variable selection in high dimensional feature space. *Statistica Sinica.* 20:101-148

Feinstein, L., Sabates, R., Sorhaindo, A., Rogers, I., Herrick, D., Northstone, K. & Emmett, P. (2008) Dietary patterns related to attainment in school: the importance of early eating patterns. *Journal of Epidemiology and Community Health.* 62 (8), 734-739.

Fewell, Z., Davey Smith, G. & Sterne, J. A. (2007) The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *American Journal of Epidemiology*. 166 (6), 646-655.

Fitzsimon, N., Fallon, U., O'Mahony, D., Loftus, B. G., Bury, G., Murphy, A. W., Kelleher, C. C. & Lifeways Cross Generation Cohort Study Steering Group. (2007) Mothers' dietary patterns during pregnancy and risk of asthma symptoms in children at 3 years. *Irish Medical Journal.* 100 (8), suppl 27-32.

Flood, A., Rastogi, T., Wirfalt, E., Mitrou, P. N., Reedy, J., Subar, A. F., Kipnis, V., Mouw, T., Hollenbeck, A. R., Leitzmann, M. & Schatzkin, A. (2008) Dietary patterns as identified by factor analysis and colorectal cancer among middle-aged Americans. *The American Journal of Clinical Nutrition.* 88 (1), 176-184.

Fogarty, A., Lewis, S. A., Scrivener, S. L., Antoniak, M., Pacey, S., Pringle, M. & Britton, J. (2003) Oral magnesium and vitamin C supplements in asthma: a parallel group randomized placebo-controlled trial. *Clinical and Experimental Allergy : Journal of the British Society for Allergy and Clinical Immunology.* 33 (10), 1355-1359.

Fung, T., Hu, F. B., Fuchs, C., Giovannucci, E., Hunter, D. J., Stampfer, M. J., Colditz, G. A. & Willett, W. C. (2003) Major dietary patterns and the risk of colorectal cancer in women. *Archives of Internal Medicine.* 163 (3), 309-314.

Fung, T. T., Hu, F. B., Holmes, M. D., Rosner, B. A., Hunter, D. J., Colditz, G. A. & Willett, W. C. (2005) Dietary patterns and the risk of postmenopausal breast cancer. *International Journal of Cancer.Journal International Du Cancer.* 116 (1), 116-121.

Fung, T. T., Rimm, E. B., Spiegelman, D., Rifai, N., Tofler, G. H., Willett, W. C. & Hu, F. B. (2001) Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *The American Journal of Clinical Nutrition.* 73 (1), 61-67.

Fung, T. T., Schulze, M., Manson, J. E., Willett, W. C. & Hu, F. B. (2004a) Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Archives of Internal Medicine.* 164 (20), 2235-2240.

Fung, T. T., Stampfer, M. J., Manson, J. E., Rexrode, K. M., Willett, W. C. & Hu, F. B. (2004b) Prospective study of major dietary patterns and stroke risk in women. *Stroke; a Journal of Cerebral Circulation.* 35 (9), 2014-2019.

Fung, T. T., Willett, W. C., Stampfer, M. J., Manson, J. E. & Hu, F. B. (2001) Dietary patterns and the risk of coronary heart disease in women. *Archives of Internal Medicine.* 161 (15), 1857-1862.

Galea S, Riddle M, Kaplan GA.(2010) Causal thinking and complex system approaches in epidemiology. *Int J Epidemiol*.;39(1):97–106

Galobardes, B., Morabia, A. & Bernstein, M. S. (2001) Diet and socioeconomic position: does the use of different indicators matter? *International Journal of Epidemiology.* 30 (2), 334-340.

Garcia-Larsen, V., Luczynska, M., Kowalski, M. L., Voutilainen, H., Ahlstrom, M., Haahtela, T., Toskala, E., Bockelbrink, A., Lee, H. H., Vassilopoulou, E., Papadopoulos, N. G., Ramalho, R., Moreira, A., Delgado, L., Castel-Branco, M. G., Calder, P. C., Childs, C. E., Bakolis, I., Hooper, R., Burney, P. G. & GA2LEN-WP 1.2 'Epidemiological and Clinical Studies'. (2011) Use of a common food frequency questionnaire (FFQ) to assess dietary patterns and their relation to allergy and asthma in Europe: pilot study of the GA2LEN FFQ. *European Journal of Clinical Nutrition.* 65 (6), 750-756.

Gex-Fabry, M., Raymond, L. & Jeanneret, O. (1988) Multivariate Analysis of Dietary Patterns in 939 Swiss Adults: Sociodemographic Parameters and Alcohol Consumption Profiles. *International Journal of Epidemiology.* 17 (3), 548-555.

Giovannucci, E., Liu, Y., Rimm, E. B., Hollis, B. W., Fuchs, C. S., Stampfer, M. J. & Willett, W. C. (2006) Prospective study of predictors of vitamin D status and cancer incidence and mortality in men. *Journal of the National Cancer Institute.* 98 (7), 451-459.

Goldstein, H. (2011) *Multilevel statistical models.* Wiley series in probability and statistics. 4th edition. Chichester, Wiley.

Gorst-Rasmussen, A., Dahm, C. C., Dethlefsen, C., Scheike, T. & Overvad, K. (2011) Exploring dietary patterns by using the treelet transform. *American Journal of Epidemiology.* 173 (10), 1097-1104.

Greenberg, E. R., Baron, J. A., Tosteson, T. D., Freeman, D. H.,Jr, Beck, G. J., Bond, J. H., Colacchio, T. A., Coller, J. A., Frankl, H. D. & Haile, R. W. (1994) A clinical trial of antioxidant vitamins to prevent colorectal adenoma. Polyp Prevention Study Group. *The New England Journal of Medicine.* 331 (3), 141-147.

Grube, B. S., Bilder, R. M. & Goldman, R. S. (1998) Meta-analysis of symptom factors in schizophrenia. *Schizophrenia Research.* 31 (2-3), 113-120.

Guthrie, H. A., Owen, G. M. & Guthrie, G. M. (1973) Factor analysis of measures of nutritional status of preschool children. *The American Journal of Clinical Nutrition.* 26 (5), 497-502.

Hadi, A.S. and Ling, R.F.(1998) Some Cautionary Notes on the Use of Principal Components Regression. *The American Statistician.* Vol. 52, No. 1.

Halliwell, B. (1996) Antioxidants in human health and disease. *Annual Review of Nutrition.* 16, 33-50.

Hamer, M. & Mishra, G. D. (2010) Dietary patterns and cardiovascular risk markers in the UK Low Income Diet and Nutrition Survey. *Nutrition, Metabolism, and Cardiovascular Diseases : NMCD.* 20 (7), 491-497.

Haines, P. S., Siega-Riz, A. M. & Popkin, B. M. (1999) The Diet Quality Index revised: a measurement instrument for populations. *Journal of the American Dietetic Association*. 99 (6), 697-704.

Hann, C. S., Rock, C. L., King, I. & Drewnowski, A. (2001) Validation of the Healthy Eating Index with use of plasma biomarkers in a clinical sample of women. *The American Journal of Clinical Nutrition*. 74 (4), 479-486.

He, Y., Ma, G., Zhai, F., Li, Y., Hu, Y., Feskens, E. J. & Yang, X. (2009) Dietary patterns and glucose tolerance abnormalities in Chinese adults. *Diabetes Care.* 32 (11), 1972-1976.

Hedges, L. V. (1985) *Statistical methods for meta-analysis.* Orlando, Academic Press.

Heidemann, C., Schulze, M. B., Franco, O. H., van Dam, R. M., Mantzoros, C. S. & Hu, F. B. (2008) Dietary patterns and risk of mortality from cardiovascular disease, cancer, and all causes in a prospective cohort of women. *Circulation.* 118 (3), 230-237.

Hendrickson, A. E., and P. O. White.(1964) Promax: a quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*. 17, 65-70.

Hennekens, C. H., Buring, J. E., Manson, J. E., Stampfer, M., Rosner, B., Cook, N. R., Belanger, C., LaMotte, F., Gaziano, J. M., Ridker, P. M., Willett, W. & Peto, R. (1996) Lack of Effect of Long-Term Supplementation with Beta Carotene on the Incidence of Malignant Neoplasms and Cardiovascular Disease. *The New England Journal of Medicine.* 334 (18), 1145-1149.

Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.).* 327 (7414), 557-560.

Hirose, K., Matsuo, K., Iwata, H. & Tajima, K. (2007) Dietary patterns and the risk of breast cancer in Japanese women. *Cancer Science.* 98 (9), 1431-1438.

Hodge, L., Salome, C. M., Peat, J. K., Haby, M. M., Xuan, W. & Woolcock, A. J. (1996) Consumption of oily fish and childhood asthma risk. *The Medical Journal of Australia.* 164 (3), 137-140.

Hoffmann, K., Boeing, H., Boffetta, P., Nagel, G., Orfanos, P., Ferrari, P. & Bamia, C. (2005) Comparison of two statistical approaches to predict all-cause mortality by dietary patterns in German elderly subjects. *The British Journal of Nutrition.* 93 (5), 709-716.

Hoffmann, K., Schulze, M. B., Schienkiewitz, A., Nothlings, U. & Boeing, H. (2004) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *American Journal of Epidemiology.* 159 (10), 935-944.

Hooper, R., Calvert, J., Thompson, R. L., Deetlefs, M. E. & Burney, P. (2008) Urban/rural differences in diet and atopy in South Africa. *Allergy.* 63 (4), 425-431.

Hooper, R., Heinrich, J., Omenaas, E., Sausenthaler, S., Garcia-Larsen, V., Bakolis, I. & Burney, P. (2010) Dietary patterns and risk of asthma: results from three countries in European Community Respiratory Health Survey-II. *The British Journal of Nutrition.* 103 (9), 1354-1365.

Hox, J. J. (2010) *Multilevel analysis: techniques and applications.* Quantitative methodology series. 2nd edition. London, Routledge.

Hu, F. B. (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology.* 13 (1), 3-9.

Hu, F. B., Rimm, E. B., Stampfer, M. J., Ascherio, A., Spiegelman, D. & Willett, W. C. (2000) Prospective study of major dietary patterns and risk of coronary heart disease in men. *The American Journal of Clinical Nutrition.* 72 (4), 912-921.

Hughes, M. C., Williams, G. M., Fourtanier, A. & Green, A. C. (2009) Food intake, dietary patterns, and actinic keratoses of the skin: a longitudinal study. *The American Journal of Clinical Nutrition.* 89 (4), 1246-1255.

Imamura, F., Lichtenstein, A. H., Dallal, G. E., Meigs, J. B. & Jacques, P. F. (2009) Confounding by dietary patterns of the inverse association between alcohol consumption and type 2 diabetes risk. *American Journal of Epidemiology.* 170 (1), 37-45.

Ioannidis, J. P. (2005) Why most published research findings are false. *PLoS Medicine.* 2 (8), e124.

Iqbal, R., Anand, S., Ounpuu, S., Islam, S., Zhang, X., Rangarajan, S., Chifamba, J., Al-Hinai, A., Keltai, M., Yusuf, S. & INTERHEART Study Investigators. (2008) Dietary patterns and the risk of acute myocardial infarction in 52 countries: results of the INTERHEART study. *Circulation.* 118 (19), 1929-1937.

Jackson, M., Walker, S., Simpson, C., McFarlane-Anderson, N. & Bennett, F. (2009) Are food patterns associated with prostate cancer in Jamaican men: a preliminary report. *Infectious Agents and Cancer.* 4 Suppl 1, S5.

Jacobs, D. R.,Jr, Gross, M. D., Steffen, L., Steffes, M. W., Yu, X., Svetkey, L. P., Appel, L. J., Vollmer, W. M., Bray, G. A., Moore, T., Conlin, P. R. & Sacks, F. (2009) The effects of dietary patterns on urinary albumin excretion: results of the Dietary Approaches to Stop Hypertension (DASH) Trial. *American Journal of Kidney Diseases : The Official Journal of the National Kidney Foundation*. 53 (4), 638-646.

Jacobs, D. R.,Jr & Steffen, L. M. (2003) Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *The American Journal of Clinical Nutrition.* 78 (3 Suppl), 508S-513S.

Jacques, P. F. & Tucker, K. L. (2001) Are dietary patterns useful for understanding the role of diet in chronic disease? *The American Journal of Clinical Nutrition.* 73 (1), 1-2.

Jakes, R. W., Day, N. E., Luben, R., Welch, A., Bingham, S., Mitchell, J., Hennings, S., Rennie, K. & Wareham, N. J. (2004) Adjusting for energy intake—what measure to use in nutritional epidemiological studies? *International Journal of Epidemiology.* 33 (6), 1382-1386.

Jeffers, J. N. R. (1967) Two case studies in the application of principal component analysis, *Applied Staristics.* 16, pp. 225-236.

Johnson, R. A. & Wichern, D. W. (1982) *Applied multivariate statistical analysis*. Prentice Hall series in statistics. Englewood Cliffs ; London, Prentice-Hall.

Jolliffe I.T (1995) Rotation of principal components: Choice of normalization constraints,*Journal of Applied Statistics.* 22(1),29-36

Jolliffe, I.T.,. Trendafilov, N. T and Uddin, M (2003). A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531–547

Jolliffe, I. T. (2010) *Principal component analysis.* Springer series in statistics. 2nd edition. New York ; London, Springer.

Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics.* 11 (1), 94.

Kaiser, H. (1958) *The varimax criterion for analytic rotation in factor analysis.* Springer New York.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.*, **20**, 141–151.

Kant, A. K. (2004) Dietary patterns and health outcomes. *Journal of the American Dietetic Association.* 104 (4), 615-635.

Kant, A. K., Schatzkin, A., Block, G., Ziegler, R. G. & Nestle, M. (1991) Food group intake patterns and associated nutrient profiles of the US population. *Journal of the American Dietetic Association.* 91 (12), 1532-1537.

Kant AK, Graubard BI & Schatzkin A (2004) Dietary patterns predict mortality in a national cohort: the National Health Interview Surveys, 1987 and 1992. *J Nutr* 134, 1793–1799.

Keskitalo, K., Silventoinen, K., Tuorila, H., Perola, M., Pietilainen, K. H., Rissanen, A. & Kaprio, J. (2008) Genetic and environmental contributions to food use patterns of young adult twins. *Physiology & Behavior.* 93 (1-2), 235-242.

Kesse, E., Clavel-Chapelon, F. & Boutron-Ruault, M. C. (2006) Dietary patterns and risk of colorectal tumors: a cohort of French women of the National Education System (E3N). *American Journal of Epidemiology.* 164 (11), 1085-1093.

Kesse-Guyot, E., Bertrais, S., Peneau, S., Estaquio, C., Dauchet, L., Vergnaud, A. C., Czernichow, S., Galan, P., Hercberg, S. & Bellisle, F. (2009) Dietary patterns and their sociodemographic and behavioural correlates in French middle-aged adults from the SU.VI.MAX cohort. *European Journal of Clinical Nutrition.* 63 (4), 521-528.

Khani, B. R., Ye, W., Terry, P. & Wolk, A. (2004) Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *The Journal of Nutrition.* 134 (6), 1541-1545.

Kim, H. S., Park, S. Y., Grandinetti, A., Holck, P. S. & Waslien, C. (2008) Major dietary patterns, ethnicity, and prevalence of type 2 diabetes in rural Hawaii. *Nutrition (Burbank, Los Angeles County, Calif.).* 24 (11-12), 1065-1072.

Kim, J. A., Kim, S. M., Lee, J. S., Oh, H. J., Han, J. H., Song, Y., Joung, H. & Park, H. S. (2007) Dietary patterns and the metabolic syndrome in Korean adolescents: 2001 Korean National Health and Nutrition Survey. *Diabetes Care.* 30 (7), 1904-1905.

Kim, M. K., Sasaki, S., Sasazuki, S., Tsugane, S. & Japan Public Health Center-based Prospective Study Group. (2004) Prospective study of three major dietary patterns and risk of gastric cancer in Japan. *International Journal of Cancer.Journal International Du Cancer.* 110 (3), 435-442.

Kim, S. & Popkin, B. M. (February 2006) Commentary: Understanding the epidemiology of overweight and obesity—a real global public health concern. *International Journal of Epidemiology.* 35 (1), 60-67.

Kipnis, V. & Freedman, L. S. (2008) Impact of Exposure Measurement Error in Nutritional Epidemiology. Journal of the National Cancer Institute. 100 (23), 1658-1659.

Knekt P, Ritz J, Pereira MA, O'Reilly EJ, Augustsson K, Fraser GE, et al.(2004) Antioxidant vitamins and coronary heart disease risk: a pooled analysis of 9 cohorts. *Am J Clin Nutr*.80:1508-20.

Knudsen, V. K., Orozova-Bekkevold, I. M., Mikkelsen, T. B., Wolff, S. & Olsen, S. F. (2008) Major dietary patterns in pregnancy and fetal growth. *European Journal of Clinical Nutrition.* 62 (4), 463-470.

Kontogianni, M. D., Melistas, L., Yannakoulia, M., Malagaris, I., Panagiotakos, D. B. & Yiannakouris, N. (2009) Association between dietary patterns and indices of bone mass in a sample of Mediterranean women. *Nutrition (Burbank, Los Angeles County, Calif.).* 25 (2), 165-171.

Kosuke Imai & David A van Dyk. (2004) Causal Inference With General Treatment Regime: Generalizing the propensity score. *Journal of the American Statistical Association* .99 (467), 854-854-866.

Kroenke, C. H., Fung, T. T., Hu, F. B. & Holmes, M. D. (2005) Dietary patterns and survival after breast cancer diagnosis. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* 23 (36), 9295-9303.

Kroke, A. (2004) Re: "Application of a new statistical method to derive dietary patterns in nutritional epidemiology". *American Journal of Epidemiology.* 160 (11), 1132; author reply 1132-3.

Kubo, A., Levin, T. R., Block, G., Rumore, G. J., Quesenberry, C. P.,Jr, Buffler, P. & Corley, D. A. (2008) Dietary patterns and the risk of Barrett's esophagus. *American Journal of Epidemiology.* 167 (7), 839-846.

Laerum, B. N., Wentzel-Larsen, T., Gulsvik, A., Omenaas, E., Gislason, T., Janson, C. & Svanes, C. (2007) Relationship of fish and cod oil intake with adult asthma. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology.* 37 (11), 1616-1623.

Laird, N. M. & SpringerLink. (2011) *The Fundamentals of Modern Statistical Genetics.* Statistics for Biology and HealthSpringer Science+Business Media, LLC, New York, NY.

Lau, C., Glumer, C., Toft, U., Tetens, I., Carstensen, B., Jorgensen, T. & Borch-Johnsen, K. (2008) Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study. *The British Journal of Nutrition.* 99 (5), 1089-1098.

Lee, C. N., Reed, D. M., MacLean, C. J., Yano, K. & Chiu, D. (1988) Dietary potassium and stroke. *The New England Journal of Medicine.* 318 (15), 995-996.

Lind, J. (1753). *A Treatise on the Scurvy*. Reprinted Edinburgh: Edinburgh University Press, 1953.

Liu, L., Nettleton, J. A., Bertoni, A. G., Bluemke, D. A., Lima, J. A. & Szklo, M. (2009) Dietary pattern, the metabolic syndrome, and left ventricular mass and systolic function: the Multi-Ethnic Study of Atherosclerosis. *The American Journal of Clinical Nutrition.* 90 (2), 362-368.

Localio, A. R., Berlin, J. A., Ten Have, T. R. & Kimmel, S. E. (2001) Adjustments for center in multicenter studies: an overview. *Annals of Internal Medicine*. 135 (2), 112-123.

Lopez, C. N., Martinez-Gonzalez, M. A., Sanchez-Villegas, A., Alonso, A., Pimenta, A. M. & Bes-Rastrollo, M. (2009) Costs of Mediterranean and western dietary patterns in a Spanish cohort and their relationship with prospective weight change. *Journal of Epidemiology and Community Health.* 63 (11), 920-927.

Lopez-Garcia, E., Schulze, M. B., Fung, T. T., Meigs, J. B., Rifai, N., Manson, J. E. & Hu, F. B. (2004) Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. *The American Journal of Clinical Nutrition.* 80 (4), 1029-1035.

Lutsey, P. L., Steffen, L. M. & Stevens, J. (2008) Dietary intake and the development of the metabolic syndrome: the Atherosclerosis Risk in Communities study. *Circulation.* 117 (6), 754-761.

Lutsey, P. L., Steffen, L. M., Virnig, B. A. & Folsom, A. R. (2009) Diet and incident venous thromboembolism: the Iowa Women's Health Study. *American Heart Journal.* 157 (6), 1081-1087.

Maldonado, G. & Greenland, S. (1997) The importance of critically interpreting simulation studies. Epidemiology (Cambridge, Mass.). 8 (4), 453-456.

Mann J, Aune D.(2010). Can specific fruits and vegetables prevent diabetes? *BMJ*;341:c4395

Mannisto, S., Dixon, L. B., Balder, H. F., Virtanen, M. J., Krogh, V., Khani, B. R., Berrino, F., van den Brandt, P. A., Hartman, A. M., Pietinen, P., Tan, F., Wolk, A. & Goldbohm, R. A. (2005) Dietary patterns and breast cancer risk: results from three cohort studies in the DIETSCAN project. *Cancer Causes & Control : CCC.* 16 (6), 725-733.

Marchioni, D. M., Latorre Mdo, R., Eluf-Neto, J., Wunsch-Filho, V. & Fisberg, R. M. (2005) Identification of dietary patterns using factor analysis in an epidemiological study in Sao Paulo. *Sao Paulo Medical Journal = Revista Paulista De Medicina.* 123 (3), 124-127.

Marks, G. B., Burney, P. G., Premaratne, U. N., Simpson, J. & Webb, J. (1997) Asthma in Greenwich, UK: impact of the disease and current management practices. *The European*

Respiratory Journal : Official Journal of the European Society for Clinical Respiratory Physiology. 10 (6), 1224-1229.

Maruapula, S. & Chapman-Novakofski, K. (2007) Health and dietary patterns of the elderly in Botswana. *Journal of Nutrition Education and Behavior.* 39 (6), 311-319.

Masala, G., Ceroti, M., Pala, V., Krogh, V., Vineis, P., Sacerdote, C., Saieva, C., Salvini, S., Sieri, S., Berrino, F., Panico, S., Mattiello, A., Tumino, R., Giurdanella, M. C., Bamia, C., Trichopoulou, A., Riboli, E. & Palli, D. (2007) A dietary pattern rich in olive oil and raw vegetables is associated with lower mortality in Italian elderly subjects. *The British Journal of Nutrition.* 98 (2), 406-415.

Maskarinec, G., Novotny, R. & Tasaki, K. (2000) Dietary Patterns Are Associated with Body Mass Index in Multiethnic Women. *The Journal of Nutrition.* 130 (12), 3068-3072.

Maas, C.J., Hox, J.J., 2004. The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Comp. Stat. Data Anal*. 46, 427–440

McCann, S. E., Marshall, J. R., Brasure, J. R., Graham, S. & Freudenheim, J. L. (2001a) Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. *Public Health Nutrition.* 4 (5), 989-997.

McCann, S. E., Weiner, J., Graham, S. & Freudenheim, J. L. (2001b) Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? *Public Health Nutrition.* 4 (4), 903-908.

McCance and Widdowson's the composition of foods (1991) 5th edn (plus supplements) ed. London: *The Royal Society of Chemistry and Ministry of Agriculture*, Fisheries and Food.

McKeever, T. M. & Britton, J. (2004) Diet and asthma. *American Journal of Respiratory and Critical Care Medicine.* 170 (7), 725-729.

McNaughton, S. A., Ball, K., Mishra, G. D. & Crawford, D. A. (2008) Dietary patterns of adolescents and risk of obesity and hypertension. *The Journal of Nutrition.* 138 (2), 364-370.

McNaughton, S. A., Mishra, G. D. & Brunner, E. J. (2008) Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. *Diabetes Care.* 31 (7), 1343-1348.

McNaughton, S. A., Mishra, G. D., Stephen, A. M. & Wadsworth, M. E. (2007) Dietary patterns throughout adult life are associated with body mass index, waist circumference, blood pressure, and red cell folate. *The Journal of Nutrition.* 137 (1), 99-105.

Mertz, W. (1984) Foods and nutrients. *Journal of the American Dietetic Association.* 84 (7), 769-770.

Messina, M., Lampe, J. W., Birt, D. F., Appel, L. J., Pivonka, E., Berry, B. & Jacobs, D. R.,Jr. (2001) Reductionism and the narrowing nutrition perspective: time for reevaluation and emphasis on food synergy. *Journal of the American Dietetic Association.* 101 (12), 1416-1419.

Meyerhardt, J. A., Niedzwiecki, D., Hollis, D., Saltz, L. B., Hu, F. B., Mayer, R. J., Nelson, H., Whittom, R., Hantel, A., Thomas, J. & Fuchs, C. S. (2007) Association of dietary patterns with cancer recurrence and survival in patients with stage III colon cancer. *JAMA : The Journal of the American Medical Association.* 298 (7), 754-764.

Michaud, D. S., Skinner, H. G., Wu, K., Hu, F., Giovannucci, E., Willett, W. C., Colditz, G. A. & Fuchs, C. S. (2005) Dietary patterns and pancreatic cancer risk in men and women. *Journal of the National Cancer Institute.* 97 (7), 518-524.

Michels, K. B. (2003) Nutritional epidemiology--past, present, future. *International Journal of Epidemiology.* 32 (4), 486-488.

Michels, K. B. & Schulze, M. B. (2005) Can dietary patterns help us detect diet-disease associations? *Nutrition Research Reviews.* 18 (2), 241-248.

Mikkila, V., Rasanen, L., Raitakari, O. T., Pietinen, P. & Viikari, J. (2005) Consistent dietary patterns identified from childhood to adulthood: the cardiovascular risk in Young Finns Study. *The British Journal of Nutrition.* 93 (6), 923-931.

Miller, R. G. (1981) *Simultaneous statistical inference.* Springer series in statistics. 2nd edition. New York, Springer-Verlag.

Mizoue, T., Yamaji, T., Tabata, S., Yamaguchi, K., Ogawa, S., Mineshita, M. & Kono, S. (2006) Dietary patterns and glucose tolerance abnormalities in Japanese men. *The Journal of Nutrition.* 136 (5), 1352-1358.

Mizoue, T., Yamaji, T., Tabata, S., Yamaguchi, K., Shimizu, E., Mineshita, M., Ogawa, S. & Kono, S. (2005) Dietary patterns and colorectal adenomas in Japanese men: the Self-Defense Forces Health Study. *American Journal of Epidemiology.* 161 (4), 338-345.

Montonen, J., Knekt, P., Harkanen, T., Jarvinen, R., Heliovaara, M., Aromaa, A. & Reunanen, A. (2005) Dietary patterns and the incidence of type 2 diabetes. *American Journal of Epidemiology.* 161 (3), 219-227.

Muller, D. C., Severi, G., Baglietto, L., Krishnan, K., English, D. R., Hopper, J. L. & Giles, G. G. (2009) Dietary patterns and prostate cancer risk. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 18 (11), 3126-3129.

Murr, C., Schroecksnadel, K., Winkler, C., Ledochowski, M. & Fuchs, D. (2005) Antioxidants may increase the probability of developing allergic diseases and asthma. *Medical Hypotheses.* 64 (5), 973-977.

Murtaugh, M. A., Herrick, J. S., Sweeney, C., Baumgartner, K. B., Guiliano, A. R., Byers, T. & Slattery, M. L. (2007) Diet composition and risk of overweight and obesity in women living in the southwestern United States. *Journal of the American Dietetic Association.* 107 (8), 1311-1321.

Murtaugh, M. A., Sweeney, C., Giuliano, A. R., Herrick, J. S., Hines, L., Byers, T., Baumgartner, K. B. & Slattery, M. L. (2008) Diet patterns and breast cancer risk in Hispanic and non-Hispanic white women: the Four-Corners Breast Cancer Study. *The American Journal of Clinical Nutrition.* 87 (4), 978-984.

Nanri, A., Mizoue, T., Yoshida, D., Takahashi, R. & Takayanagi, R. (2008) Dietary patterns and A1C in Japanese men and women. *Diabetes Care.* 31 (8), 1568-1573.

Nanri, A., Yoshida, D., Yamaji, T., Mizoue, T., Takayanagi, R. & Kono, S. (2008b) Dietary patterns and C-reactive protein in Japanese men and women. *The American Journal of Clinical Nutrition.* 87 (5), 1488-1496.

Naska, A., Fouskakis, D., Oikonomou, E., Almeida, M. D., Berg, M. A., Gedrich, K., Moreiras, O., Nelson, M., Trygg, K., Turrini, A., Remaut, A. M., Volatier, J. L., Trichopoulou, A. & DAFNE participants. (2006) Dietary patterns and their socio-demographic determinants in 10 European countries: data from the DAFNE databank. *European Journal of Clinical Nutrition.* 60 (2), 181-190.

Nettleton, J. A., Polak, J. F., Tracy, R., Burke, G. L. & Jacobs, D. R.,Jr. (2009) Dietary patterns and incident cardiovascular disease in the Multi-Ethnic Study of Atherosclerosis. *The American Journal of Clinical Nutrition.* 90 (3), 647-654.

Nettleton, J. A., Steffen, L. M., Ni, H., Liu, K. & Jacobs, D. R.,Jr. (2008a) Dietary patterns and risk of incident type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care.* 31 (9), 1777-1782.

Nettleton, J. A., Steffen, L. M., Palmas, W., Burke, G. L. & Jacobs, D. R.,Jr. (2008b) Associations between microalbuminuria and animal foods, plant foods, and dietary patterns in the Multiethnic Study of Atherosclerosis. *The American Journal of Clinical Nutrition.* 87 (6), 1825-1836.

Nettleton, J. A., Steffen, L. M., Schulze, M. B., Jenny, N. S., Barr, R. G., Bertoni, A. G. & Jacobs, D. R.,Jr. (2007) Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *The American Journal of Clinical Nutrition.* 85 (6), 1615-1625.

Newby, P. K., Muller, D., Hallfrisch, J., Andres, R. & Tucker, K. L. (2004) Food patterns measured by factor analysis and anthropometric changes in adults. *The American Journal of Clinical Nutrition.* 80 (2), 504-513.

Newby, P. K., Muller, D. & Tucker, K. L. (2004) Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *The American Journal of Clinical Nutrition.* 80 (3), 759-767.

Newby, P. K. & Tucker, K. L. (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutrition Reviews.* 62 (5), 177-203.

Newby, P. K., Weismayer, C., Akesson, A., Tucker, K. L. & Wolk, A. (2006a) Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. *The Journal of Nutrition.* 136 (10), 2580-2587.

Newby, P. K., Weismayer, C., Akesson, A., Tucker, K. L. & Wolk, A. (2006b) Long-term stability of food patterns identified by use of factor analysis among Swedish women. *The Journal of Nutrition.* 136 (3), 626-633.

Newson, R. & ALSPAC Study Team (2003). Multiple test procedures and smile plots. *Stata J*, (3), 109–132.

Nkondjock, A., Krewski, D., Johnson, K. C., Ghadirian, P. & Canadian Cancer Registries Epidemiology Research Group. (2005) Dietary patterns and risk of pancreatic cancer. *International Journal of Cancer.Journal International Du Cancer.* 114 (5), 817-823.

Noel, S. E., Newby, P. K., Ordovas, J. M. & Tucker, K. L. (2009) A traditional rice and beans pattern is associated with metabolic syndrome in Puerto Rican older adults. *The Journal of Nutrition.* 139 (7), 1360-1367.

Northstone, K. & Emmett, P. (2005) Multivariate analysis of diet in children at four and seven years of age and associations with socio-demographic characteristics. *European Journal of Clinical Nutrition.* 59 (6), 751-760.

Northstone, K., Emmett, P. & Rogers, I. (2008a) Dietary patterns in pregnancy and associations with socio-demographic and lifestyle factors. *European Journal of Clinical Nutrition.* 62 (4), 471-479.

Northstone, K. & Emmett, P. M. (2008a) Are dietary patterns stable throughout early and mid-childhood? A birth cohort study. *The British Journal of Nutrition.* 100 (5), 1069-1076.

Northstone, K. & Emmett, P. M. (2008b) A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis. *The British Journal of Nutrition.* 99 (5), 1099-1106.

Northstone, K., Emmett, P. M. & Rogers, I. (2008b) Dietary patterns in pregnancy and associations with nutrient intakes. *The British Journal of Nutrition.* 99 (2), 406-415.

Nothlings, U., Murphy, S. P., Wilkens, L. R., Boeing, H., Schulze, M. B., Bueno-de-Mesquita, H. B., Michaud, D. S., Roddam, A., Rohrmann, S., Tjonneland, A., Clavel-Chapelon, F., Trichopoulou, A., Sieri, S., Rodriguez, L., Ye, W., Jenab, M. & Kolonel, L. N. (2008) A food pattern that is predictive of flavonol intake and risk of pancreatic cancer. *The American Journal of Clinical Nutrition.* 88 (6), 1653-1662.

Nurmatov, U., Nwaru, B. I., Devereux, G. & Sheikh, A. (2012) Confounding and effect modification in studies of diet and childhood asthma and allergies. *Allergy.* 67 (8), 1041-1059.

Oddy, W. H., Robinson, M., Ambrosini, G. L., O'Sullivan, T. A., de Klerk, N. H., Beilin, L. J., Silburn, S. R., Zubrick, S. R. & Stanley, F. J. (2009) The association between dietary patterns and mental health in early adolescence. *Preventive Medicine.* 49 (1), 39-44.

Oien, T., Storro, O. & Johnsen, R. (2010) Do early intake of fish and fish oil protect against eczema and doctor-diagnosed asthma at 2 years of age? A cohort study. *Journal of Epidemiology and Community Health.* 64 (2), 124-129.

Okubo, H., Sasaki, S., Murakami, K., Kim, M. K., Takahashi, Y., Hosoi, Y., Itabashi, M. & Freshmen in Dietetic Courses Study II group. (2008) Three major dietary patterns are all independently related to the risk of obesity among 3760 Japanese women aged 18-20 years. *International Journal of Obesity (2005).* 32 (3), 541-549.

Okubo, H., Sasaki, S., Murakami, K., Kim, M. K., Takahashi, Y., Hosoi, Y., Itabashi, M. & Freshmen in Dietetic Courses Study II Group. (2007) Dietary patterns associated with functional constipation among Japanese women aged 18 to 20 years: a cross-sectional study. *Journal of Nutritional Science and Vitaminology.* 53 (3), 232-238.

Pala, V., Sieri, S., Masala, G., Palli, D., Panico, S., Vineis, P., Sacerdote, C., Mattiello, A., Galasso, R., Salvini, S., Ceroti, M., Berrino, F., Fusconi, E., Tumino, R., Frasca, G., Riboli, E., Trichopoulou, A., Baibas, N. & Krogh, V. (2006) Associations between dietary pattern and lifestyle, anthropometry and other health indicators in the elderly participants of the EPIC-Italy cohort. *Nutrition, Metabolism, and Cardiovascular Diseases : NMCD.* 16 (3), 186-201.

Panagiotakos, D., Bountziouka, V., Zeimbekis, A., Vlachou, I. & Polychronopoulos, E. (2007a) Food pattern analysis and prevalence of cardiovascular disease risk factors among elderly people from Mediterranean islands. *Journal of Medicinal Food.* 10 (4), 615-621.

Panagiotakos, D. B., Pitsavos, C., Skoumas, Y. & Stefanadis, C. (2007b) The association between food patterns and the metabolic syndrome using principal components analysis: The ATTICA Study. *Journal of the American Dietetic Association.* 107 (6), 979-87; quiz 997.

Paradis, A. M., Godin, G., Perusse, L. & Vohl, M. C. (2009) Associations between dietary patterns and obesity phenotypes. *International Journal of Obesity (2005).* 33 (12), 1419-1426.

Paradis, A. M., Perusse, L. & Vohl, M. C. (2006) Dietary patterns and associated lifestyles in individuals with and without familial history of obesity: a cross-sectional study. *The International Journal of Behavioral Nutrition and Physical Activity.* 3, 38.

Park, S. Y., Murphy, S. P., Wilkens, L. R., Yamamoto, J. F., Sharma, S., Hankin, J. H., Henderson, B. E. & Kolonel, L. N. (2005) Dietary patterns using the Food Guide Pyramid groups are associated with sociodemographic and lifestyle factors: the multiethnic cohort study. *The Journal of Nutrition.* 135 (4), 843-849.

Pastor, R. & Guallar, E. (2001) Re: "Use of two-segmented logistic regression to estimate change-points in epidemiologic studies". American Journal of Epidemiology. 153 (6), 615.

Paul, A. A., Southgate, D. A. & Buss, D. H. (1986) McCance and Widdowson's 'The composition of foods': supplementary information and review of new compositional data. *Human Nutrition.Applied Nutrition.* 40 (4), 287-299.

Pearson, P. J., Lewis, S. A., Britton, J. & Fogarty, A. (2004) Vitamin E supplements in asthma: a parallel group randomised placebo controlled trial. *Thorax.* 59 (8), 652-656.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*. 49 (12), 1373-1379.

Perrin, A. E., Dallongeville, J., Ducimetiere, P., Ruidavets, J. B., Schlienger, J. L., Arveiler, D. & Simon, C. (2005) Interactions between traditional regional determinants and socio-economic status on dietary patterns in a sample of French men. *The British Journal of Nutrition.* 93 (1), 109-114.

Premaratne, U. N., Sterne, J. A., Marks, G. B., Webb, J. R., Azima, H. & Burney, P. G. (1999) Clustered randomised trial of an intervention to improve the management of asthma: Greenwich asthma study. *BMJ (Clinical Research Ed.).* 318 (7193), 1251-1255.

Qi, L., Cornelis, M. C., Zhang, C., van Dam, R. M. & Hu, F. B. (2009) Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men. *The American Journal of Clinical Nutrition.* 89 (5), 1453-1458.

Rabe-Hesketh, S. & Skrondal, A. (2006) Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 169 (4), 805-827.

Raberg Kjollesdal, M. K., Holmboe-Ottesen, G. & Wandel, M. (2010) Associations between food patterns, socioeconomic position and working situation among adult, working women and men in Oslo. *European Journal of Clinical Nutrition.* 64 (10), 1150-1157.

Randall, E., Marshall, J. R., Graham, S. & Brasure, J. (1990) Patterns in food use and their associations with nutrient intakes. *The American Journal of Clinical Nutrition.* 52 (4), 739-745.

Rashidkhani, B., Akesson, A., Lindblad, P. & Wolk, A. (2005) Major dietary patterns and risk of renal cell carcinoma in a prospective cohort of Swedish women. *The Journal of Nutrition.* 135 (7), 1757-1762.

Reedy, J., Wirfalt, E., Flood, A., Mitrou, P. N., Krebs-Smith, S. M., Kipnis, V., Midthune, D., Leitzmann, M., Hollenbeck, A., Schatzkin, A. & Subar, A. F. (2010) Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study. *American Journal of Epidemiology.* 171 (4), 479-487.

Reineke, D. M., Baggett, J., & Elfessi, A. (2003). A note on the effect of skewness, kurtosis, and shifting on one-samplet and sign tests.*Journal of Statistics Education*,11, (3). Retrieved March 28, 2004, from www.amstat.org/publications/jse/v11n3/reineke.html

Rezazadeh, A., Rashidkhani, B. & Omidvar, N. (2010) Association of major dietary patterns with socioeconomic and lifestyle factors of adult women living in Tehran, Iran. *Nutrition (Burbank, Los Angeles County, Calif.).* 26 (3), 337-341.

Richman, M.B. (1986). Rotation of principal components. *J. Climatol.*, **6**, 293–335.

Robinson, S., Marriott, L., Poole, J., Crozier, S., Borland, S., Lawrence, W., Law, C., Godfrey, K., Cooper, C., Inskip, H. & Southampton Women's Survey Study Group. (2007) Dietary patterns in infancy: the importance of maternal and family influences on feeding practice. *The British Journal of Nutrition.* 98 (5), 1029-1037.

Robinson, S., Syddall, H., Jameson, K., Batelaan, S., Martin, H., Dennison, E. M., Cooper, C., Sayer, A. A. & Hertfordshire Study Group. (2009) Current patterns of diet in community-dwelling older men and women: results from the Hertfordshire Cohort Study. *Age and Ageing.* 38 (5), 594-599.

Robinson, S. M., Crozier, S. R., Borland, S. E., Hammond, J., Barker, D. J. & Inskip, H. M. (2004) Impact of educational attainment on the quality of young women's diets. *European Journal of Clinical Nutrition.* 58 (8), 1174-1180.

Romaguera, D., Samman, N., Rossi, A., Miranda, C., Pons, A. & Tur, J. A. (2008) Dietary patterns of the Andean population of Puna and Quebrada of Humahuaca, Jujuy, Argentina. *The British Journal of Nutrition.* 99 (2), 390-397.

Romieu, I., Varraso, R., Avenel, V., Leynaert, B., Kauffmann, F. & Clavel-Chapelon, F. (2006) Fruit and vegetable intakes and asthma in the E3N study. *Thorax.* 61 (3), 209-215.

Ronco, A. L., De Stefani, E., Boffetta, P., Deneo-Pellegrini, H., Acosta, G. & Mendilaharsu, M. (2006) Food patterns and risk of breast cancer: A factor analysis study in Uruguay. *International Journal of Cancer.Journal International Du Cancer.* 119 (7), 1672-1678.

Rosenbaum, P. R. & Rubin D. P. (1983) Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. 45 (2), 212-218.

Rothman KJ. (1990) No adjustments are needed for multiple comparisons. *Epidemiology*. 1, 43-46.

Rothman, K. J. & Greenland, S. (1998) *Modern epidemiology.* 2nd edition. Philadelphia, Lippincott-Raven.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* **127**, 757-763.

Sacks, F. M., Obarzanek, E., Windhauser, M. M., Svetkey, L. P., Vollmer, W. M., McCullough, M., Karanja, N., Lin, P. H., Steele, P. & Proschan, M. A. (1995) Rationale and design of the Dietary Approaches to Stop Hypertension trial (DASH). A multicenter controlled-feeding study of dietary patterns to lower blood pressure. *Annals of Epidemiology.* 5 (2), 108-118.

Sadakane, A., Tsutsumi, A., Gotoh, T., Ishikawa, S., Ojima, T., Kario, K., Nakamura, Y. & Kayaba, K. (2008) Dietary patterns and levels of blood pressure and serum lipids in a Japanese population. *Journal of Epidemiology / Japan Epidemiological Association.* 18 (2), 58-67.

Sant, M., Allemani, C., Sieri, S., Krogh, V., Menard, S., Tagliabue, E., Nardini, E., Micheli, A., Crosignani, P., Muti, P. & Berrino, F. (2007) Salad vegetables dietary pattern protects against HER-2-positive breast cancer: a prospective Italian study. *International Journal of Cancer.Journal International Du Cancer.* 121 (4), 911-914.

Sausenthaler, S., Loebel, T., Linseisen, J., Nagel, G., Magnussen, H. & Heinrich, J. (2009) Vitamin E intake in relation to allergic sensitization and IgE serum concentration. *Central European Journal of Public Health.* 17 (2), 79-85.

Schatzkin, A., Lanza, E., Corle, D., Lance, P., Iber, F., Caan, B., Shike, M., Weissfeld, J., Burt, R., Cooper, M. R., Kikendall, J. W. & Cahill, J. (2000) Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. *The New England Journal of Medicine.* 342 (16), 1149-1155.

Schnappinger, M., Sausenthaler, S., Linseisen, J., Hauner, H. & Heinrich, J. (2009) Fish consumption, allergic sensitisation and allergic diseases in adults. *Annals of Nutrition & Metabolism.* 54 (1), 67-74.

Schulz, M., Nothlings, U., Hoffmann, K., Bergmann, M. M. & Boeing, H. (2005) Identification of a food pattern characterized by high-fiber and low-fat food choices associated with low prospective weight change in the EPIC-Potsdam cohort. *The Journal of Nutrition.* 135 (5), 1183-1189.

Schulze, M. B., Fung, T. T., Manson, J. E., Willett, W. C. & Hu, F. B. (2006) Dietary patterns and changes in body weight in women. *Obesity (Silver Spring, Md.).* 14 (8), 1444-1453.

Schulze, M. B., Hoffmann, K., Kroke, A. & Boeing, H. (2003) An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *The British Journal of Nutrition.* 89 (3), 409-419.

Schwerin, H. S., Stanton, J. L., Riley, A. M.,Jr, Schaefer, A. E., Leveille, G. A., Elliott, J. G., Warwick, K. M. & Brett, B. E. (1981) Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. *The American Journal of Clinical Nutrition.* 34 (4), 568-580.

Seaton, A., Godden, D. J. & Brown, K. (1994) Increase in asthma: a more toxic environment or a more susceptible population? *Thorax.* 49 (2), 171-174.

Shaheen, S. O., Jameson, K. A., Syddall, H. E., Aihie Sayer, A., Dennison, E. M., Cooper, C., Robinson, S. M. & Hertfordshire Cohort Study Group. (2010) The relationship of dietary patterns with adult lung function and COPD. *The European Respiratory Journal : Official Journal of the European Society for Clinical Respiratory Physiology.* 36 (2), 277-284.

Shaheen, S. O., Newson, R. B., Rayman, M. P., Wong, A. P., Tumilty, M. K., Phillips, J. M., Potts, J. F., Kelly, F. J., White, P. T. & Burney, P. G. (2007) Randomised, double blind, placebo-controlled trial of selenium supplementation in adult asthma. *Thorax.* 62 (6), 483-490.

Shaheen, S. O., Northstone, K., Newson, R. B., Emmett, P. M., Sherriff, A. & Henderson, A. J. (2009) Dietary patterns in pregnancy and respiratory and atopic outcomes in childhood. *Thorax.* 64 (5), 411-417.

Shaheen, S. O., Sterne, J. A., Thompson, R. L., Songhurst, C. E., Margetts, B. M. & Burney, P. G. (2001) Dietary antioxidants and asthma in adults: population-based case-control study. *American Journal of Respiratory and Critical Care Medicine.* 164 (10 Pt 1), 1823-1828.

Shekelle RB, Lepper M, Liu S, Maliza C, Raynor WJ Jr, Rossof AH, et al. (1981) Dietary vitamin A and risk of cancer in the Western Electric study. *Lancet*.2:1185-90.

Shi, Z., Hu, X., Yuan, B., Hu, G., Pan, X., Dai, Y., Byles, J. E. & Holmboe-Ottesen, G. (2008) Vegetable-rich food pattern is related to obesity in China. *International Journal of Obesity (2005).* 32 (6), 975-984.

Shimazu, T., Kuriyama, S., Hozawa, A., Ohmori, K., Sato, Y., Nakaya, N., Nishino, Y., Tsubono, Y. & Tsuji, I. (2007) Dietary patterns and cardiovascular disease mortality in Japan: a prospective cohort study. *International Journal of Epidemiology.* 36 (3), 600-609.

Shin, K. O., Oh, S. Y. & Park, H. S. (2007) Empirically derived major dietary patterns and their associations with overweight in Korean preschool children. *The British Journal of Nutrition.* 98 (2), 416-421.

Sieri, S., Krogh, V., Pala, V., Muti, P., Micheli, A., Evangelista, A., Tagliabue, G. & Berrino, F. (2004) Dietary patterns and risk of breast cancer in the ORDET cohort. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 13 (4), 567-572.

Slattery, M. L., Berry, T. D., Potter, J. & Caan, B. (1997) Diet diversity, diet composition, and risk of colon cancer (United States). *Cancer Causes & Control* : CCC. 8 (6), 872-882.

Slattery, M. L., Boucher, K. M., Caan, B. J., Potter, J. D. & Ma, K. (1998) Eating Patterns and Risk of Colon Cancer. *American Journal of Epidemiology.* 148 (1), 4-16.

Slattery, M. L. & Boucher, K. M. (1998) The senior authors' response: Factor analysis as a tool for evaluating eating patterns. *American Journal of Epidemiology.* 148 (1), 20-21.

Slattery, M. L., Edwards, S. L., Boucher, K. M., Anderson, K. & Caan, B. J. (1999) Lifestyle and colon cancer: an assessment of factors associated with risk. *American Journal of Epidemiology.* 150 (8), 869-877.

Slattery ML.(2008) Defining dietary consumption: Is the sum greater than its parts? *Am J Clin Nutr*.88(1):14-15.

Slavin, J. (2004) Whole grains and human health. *Nutrition Research Reviews.* 17 (1), 99-110.

Smith, D. A., Mar, C. M. & Turoff, B. K. (1998) The structure of schizophrenic symptoms: a meta-analytic confirmatory factor analysis. *Schizophrenia Research.* 31 (1), 57-70.

Stegmann, M. B., Sjöstrand, K. and Larsen, R. (2006). Sparse modeling of landmark and texture variability using the orthomax criterion. International Symposium on Medical Imaging. San Diego, CA.

Sturmer, T., Joshi, M., Glynn, R. J., Avorn, J., Rothman, K. J. & Schneeweiss, S. (2006) A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of Clinical Epidemiology.* 59 (5), 437-447.

Tabak, C., Arts, I. C., Smit, H. A., Heederik, D. & Kromhout, D. (2001a) Chronic obstructive pulmonary disease and intake of catechins, flavonols, and flavones: the MORGEN Study. *American Journal of Respiratory and Critical Care Medicine.* 164 (1), 61-64.

Tabak, C., Smit, H. A., Heederik, D., Ocke, M. C. & Kromhout, D. (2001b) Diet and chronic obstructive pulmonary disease: independent beneficial effects of fruits, whole grains, and alcohol (the MORGEN study). *Clinical and Experimental Allergy : Journal of the British Society for Allergy and Clinical Immunology.* 31 (5), 747-755.

Takaoka, M. & Norback, D. (2008) Diet among Japanese female university students and asthmatic symptoms, infections, pollen and furry pet allergy. *Respiratory Medicine.* 102 (7), 1045-1054.

Takata, Y., Maskarinec, G., Park, S. Y., Murphy, S. P., Wilkens, L. R. & Kolonel, L. N. (2007) Mammographic density and dietary patterns: the multiethnic cohort. *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP).* 16 (5), 409-414.

Teo, Y. Y. & Chong, F. F. (2006) On the usage of principal components analysis and multiple testing. *American Journal of Respiratory and Critical Care Medicine.* 173 (5), 574; author reply 574-5.

Teucher, B., Skinner, J., Skidmore, P. M., Cassidy, A., Fairweather-Tait, S. J., Hooper, L., Roe, M. A., Foxall, R., Oyston, S. L., Cherkas, L. F., Perks, U. C., Spector, T. D. & MacGregor, A. J. (2007) Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies.* 10 (5), 734-748.

Thien FC,Woods RK,Walters EH.(1996) Oily fish and asthma-a fishy story? Further studies are required before claims can be made of a beneficial effect of oily fish consumption on asthma. *Med JAust*;164:135–6.

Tibshirani R.(1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(2):267–288

Togo, P., Osler, M., Sorensen, T. I. & Heitmann, B. L. (2004) A longitudinal study of food intake patterns and obesity in adult Danish men and women. *International Journal of Obesity and Related Metabolic Disorders : Journal of the International Association for the Study of Obesity.* 28 (4), 583-593.

Tomassen, P., Newson, R. B., Hoffmans, R., Lotvall, J., Cardell, L. O., Gunnbjornsdottir, M., Thilsing, T., Matricardi, P., Kramer, U., Makowska, J. S., Brozek, G., Gjomarkaj, M., Howarth, P., Loureiro, C., Toskala, E., Fokkens, W., Bachert, C., Burney, P. & Jarvis, D. (2011) Reliability of EP3OS symptom criteria and nasal endoscopy in the assessment of chronic rhinosinusitis--a GA(2) LEN study. *Allergy.* 66 (4), 556-561.

Touvier, M., Niravong, M., Volatier, J. L., Lafay, L., Lioret, S., Clavel-Chapelon, F. & Boutron-Ruault, M. C. (2009) Dietary patterns associated with vitamin/mineral supplement use and smoking among women of the E3N-EPIC cohort. *European Journal of Clinical Nutrition.* 63 (1), 39-47.

Trichopoulou A, Costacou T, Bamia C, Trichopoulos D(2004). Adherence to a Mediterranean diet and survival in a Greek population.*NEnglJMed*.348:2599-608.

Tricon, S., Willers, S., Smit, H. A., Burney, P. G., Devereux, G., Frew, A. J., Halken, S., Høst, A., Nelson, M., Shaheen, S., Warner, J. O. & Calder, P. C. (2006) Nutrition and allergic disease. *Clinical & Experimental Allergy Reviews.* 6 (5), 117-188.

Tseng, M., Breslow, R. A., DeVellis, R. F. & Ziegler, R. G. (2004) Dietary patterns and prostate cancer risk in the National Health and Nutrition Examination Survey Epidemiological Follow-up Study cohort. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 13 (1), 71-77.

Tseng, M., Vierkant, R. A., Kushi, L. H., Sellers, T. A. & Vachon, C. M. (2008) Dietary patterns and breast density in the Minnesota Breast Cancer Family Study. *Cancer Causes & Control : CCC.* 19 (5), 481-489.

Uusitalo, U., Arkkola, T., Ovaskainen, M. L., Kronberg-Kippila, C., Kenward, M. G., Veijola, R., Simell, O., Knip, M. & Virtanen, S. M. (2009) Unhealthy dietary patterns are associated with weight gain during pregnancy among Finnish women. *Public Health Nutrition.* 12 (12), 2392-2399.

Uusitalo, U., Sobal, J., Moothoosamy, L., Chitson, P., Shaw, J., Zimmet, P. & Tuomilehto, J. (2005) Dietary Westernisation: conceptualisation and measurement in Mauritius. *Public Health Nutrition.* 8 (6), 608-619.

Varraso, R., Fung, T. T., Barr, R. G., Hu, F. B., Willett, W. & Camargo, C. A.,Jr. (2007a) Prospective study of dietary patterns and chronic obstructive pulmonary disease among US women. *The American Journal of Clinical Nutrition.* 86 (2), 488-495.

Varraso, R., Fung, T. T., Hu, F. B., Willett, W. & Camargo, C. A. (2007b) Prospective study of dietary patterns and chronic obstructive pulmonary disease among US men. *Thorax.* 62 (9), 786-791.

Varraso, R., Kauffmann, F., Leynaert, B., Le Moual, N., Boutron-Ruault, M. C., Clavel-Chapelon, F. & Romieu, I. (2009) Dietary patterns and asthma in the E3N study. *The European Respiratory Journal : Official Journal of the European Society for Clinical Respiratory Physiology.* 33 (1), 33-41.

Velie, E. M., Schairer, C., Flood, A., He, J. P., Khattree, R. & Schatzkin, A. (2005) Empirically derived dietary patterns and risk of postmenopausal breast cancer in a large prospective cohort study. *The American Journal of Clinical Nutrition.* 82 (6), 1308-1319.

Vineis P, Kriebel D(2006). Causal models in epidemiology: past inheritance and genetic future. *Environ Health*;5:21.

Vujkovic, M., de Vries, J. H., Dohle, G. R., Bonsel, G. J., Lindemans, J., Macklon, N. S., van der Spek, P. J., Steegers, E. A. & Steegers-Theunissen, R. P. (2009a) Associations between dietary patterns and semen quality in men undergoing IVF/ICSI treatment. *Human Reproduction (Oxford, England).* 24 (6), 1304-1312.

Vujkovic, M., Steegers, E. A., Looman, C. W., Ocke, M. C., van der Spek, P. J. & Steegers-Theunissen, R. P. (2009b) The maternal Mediterranean dietary pattern is associated with a reduced risk of spina bifida in the offspring. *BJOG : An International Journal of Obstetrics and Gynaecology.* 116 (3), 408-415.

Waijers, P. M., Ocke, M. C., van Rossum, C. T., Peeters, P. H., Bamia, C., Chloptsios, Y., van der Schouw, Y. T., Slimani, N. & Bueno-de-Mesquita, H. B. (2006) Dietary patterns and

survival in older Dutch women. *The American Journal of Clinical Nutrition.* 83 (5), 1170-1176.

Ward, J. H.,Jr. (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association.* 58 (301), pp. 236-244.

Weismayer, C., Anderson, J. G. & Wolk, A. (2006) Changes in the stability of dietary patterns in a study of middle-aged Swedish women. *The Journal of Nutrition.* 136 (6), 1582-1587.

Wiles, N. J., Northstone, K., Emmett, P. & Lewis, G. (2009) 'Junk food' diet and childhood behavioural problems: results from the ALSPAC cohort. *European Journal of Clinical Nutrition.* 63 (4), 491-498.

Willett, W. C., Howe, G. R. & Kushi, L. H. (1997) Adjustment for total energy intake in epidemiologic studies. *The American Journal of Clinical Nutrition.* 65 (4 Suppl), 1220S-1228S; discussion 1229S-1231S.

Willett, W. (1998) *Nutritional epidemiology.* Monographs in epidemiology and biostatistics. 2nd edition. Oxford, Oxford University Press.

Woods, R. K., Walters, E. H., Raven, J. M., Wolfe, R., Ireland, P. D., Thien, F. C. & Abramson, M. J. (2003) Food and nutrient intakes and asthma risk in young adults. *The American Journal of Clinical Nutrition.* 78 (3), 414-421.

Wu, A. H., Yu, M. C., Tseng, C. C., Stanczyk, F. Z. & Pike, M. C. (2009) Dietary patterns and breast cancer risk in Asian American women. *The American Journal of Clinical Nutrition.* 89 (4), 1145-1154.

Wu, K., Hu, F. B., Fuchs, C., Rimm, E. B., Willett, W. C. & Giovannucci, E. (2004) Dietary patterns and risk of colon cancer and adenoma in a cohort of men (United States). *Cancer Causes & Control : CCC.* 15 (9), 853-862.

Wu, K., Hu, F. B., Willett, W. C. & Giovannucci, E. (2006) Dietary patterns and risk of prostate cancer in U.S. men. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 15 (1), 167-171.

Yang, E. J., Kerver, J. M. & Song, W. O. (2005) Dietary patterns of Korean Americans described by factor analysis. *Journal of the American College of Nutrition.* 24 (2), 115-121.

Yannakoulia, M., Panagiotakos, D. B., Pitsavos, C., Tsetsekou, E., Fappa, E., Papageorgiou, C. & Stefanadis, C. (2008) Eating habits in relations to anxiety symptoms among apparently healthy adults. A pattern analysis from the ATTICA Study. *Appetite.* 51 (3), 519-525.

Zhang, C., Schulze, M. B., Solomon, C. G. & Hu, F. B. (2006) A prospective study of dietary patterns, meat intake and the risk of gestational diabetes mellitus. *Diabetologia.* 49 (11), 2604-2613.

Appendices

I. Data Abstraction questions for the systematic review

The relevant information was abstracted from eligible studies and inserted into an Excel data collection form with prepared fields. Data collected included:

1. Title
2. Author
3. Justifications for the use of PCA instead of a single food or nutrient analysis
   (*1. Interactive and synergistic effects of foods consumed in combination. 2. Additive Effects of foods consumed in combination* 3. *Mulitcollinearity and confounding between dietary and lifestyle exposures 4. Multiple testing problems* 5. *Public health recommendation 6.* Complexity of diet, 7.Better evaluation / representation of overall diet)
4. Details of the Study
   (1.Population being used in each study 2. sample size being used in each study)
5. Country/Countries
6. Health outcome being investigated
7. Details of dietary assessment instrument being used
   (1.Instrument Definition a.FFQ b.24/48- hour recall c. dietary records d. diet history questionnaire (EPIC/DHQ)
   2. Number of food items of the instrument 3. Number of food groups of the instrument 4. Scale of FFQ)
8. Preparing the data before entering the PCA procedure
   (*1.Convertion to grams/d or grams /week 2. Standardisation of food intake variabless3. Food intake variables adjusted for energy intake by the residual method 4. Box-Cox transformation method)*
9. Criteria for labelling and identifying the number of dietary patterns
   (1. Cut off points 2. Method of rotation a. Varimax b. Oblique (premix) 3. Scree Plot 4. Eigenvalues 5. Principal component interpretability 6. Van der Voet's)
10. Application of PCA
    (1. Number of dietary patterns 2.Percentage of total variance being explained by the dietary patterns)
11. Foods included in dietary pattern 1 and label of the dietary pattern
12. Foods included in dietary pattern 2 and label of the dietary pattern
13. Foods included in dietary pattern 3 and label of the dietary pattern
14. Foods included in dietary pattern 4 and label of the dietary pattern
15. Foods included in dietary pattern 5 and label of the dietary pattern
16. Foods included in dietary pattern 6-10 and label of the dietary pattern
17. Associations of dietary patterns with health outcome
18. Associations of dietary patterns with socio-demographic variables

19. Validation methods *(1.Two random split sample 2. Cronbach's alpha 3. Deriving patterns separately for women and men4. Barlet's test of sphericity5. Kaiser - Meyen-Ollkin test 6. Simplified Dietary patterns 7. Confirmatory/maximum likelihood factor analysis 8. φ coefficient for testing inter-correlation 9. Stricter cut-off points for energy intake 10. Different method of rotation 11. Pearson Correlation different time points)*
21. Conclusions
22. Personal Criticism

II. Code used to develop and analyze simulation experiment

```
*** Code used to develop and analyze simulation data ***

*** Estimating power of ESFA and PCA  ***

program drop mybootun
program mybootun , rclass

display " Monte Carlo "
display " ------------------------------ "
display " Varying Effect Size and Principal Components  "
display " ------------------------------ "
display "  effect size is `4' , constant=`3'"
display " ------------------------------ "
display "  "" odds ratio is " exp(`4') ", cut off value=`7'"
display " ------------------------------ "
display " `6' random selected foods , `5' principal components , sample size= `2' , number of replications=`1', food items = `9’
"
display " ------------------------------ "

clear
quietly set memory 64m
set matsize 500
quietly postutil clear
local sum = 0
local sum1 =0

forvalues i=1(1)`1' {
```

```
quietly use "C:\Phd\Phd\Phd\Phd\number1.dta" ,replace
quietly sample `6', count

forvalues j=1(1)`6' {
                    local k`j' = number[`j']
                    }

quietly use "C:\Phd\echrsbootstm.dta" ,replace
quietly bsample `2'
local conf "  port`k1' "
local v=(`6'/2)
forvalues w=2(1)`v' {

                    local conf  "`conf' +  port`k`w''"

                    }

local z=`v'+1

forvalues w=`z'(1)`6' {

                       local conf  "`conf' - port`k`w''"

                       }

quietly gen p=1 /(1+exp(-(`3'+`4'*(`conf'))))
gen uniform=uniform()
quietly gen disease = 0
quietly replace disease = 1  if uniform <= p
quietly pca port6-port79, comp(`5')
quietly rotate, varimax
quietly predict pc1-pc`5'

postfile multipc1 coef pvalue using "C:\Phd\Phd\Phd\Phd\multipc2.dta" ,replace
forvalues i=1(1)`5' {

                     quietly xi: logit disease pc`i' alkcal
                     local b`i'=_b[pc`i']
```

```
                        local z`i'=_b[pc`i']/_se[pc`i']
                        local pval`i'= 2*normal(-abs(`z`i''))
                        post multipc1 (`b`i'') (`pval`i'')

                        }

        postclose multipc1
        postfile multisf1  coef pvalue using "C:\Phd\Phd\Phd\Phd\multisf2.dta" ,replace
        foreach var of varlist port6-port79 {

                                        quietly xi:logit disease `var' alkcal

                                        if e(df_m)> 1 {
                                                local bs=_b[`var']
                                                local zs=_b[`var']/_se[`var']
                                                local pvals=2*normal(-abs(`zs'))
                                                post multisf1 (`bs') (`pvals')
                                              }
                                        }
        postclose multisf1

        quietly use "C:\Phd\Phd\Phd\Phd\multipc2.dta" ,replace
        quietly multproc, pvalue(pvalue)

if  r(nreject) > 0 {

 local sum = `sum' + 1

                  }

  quietly use "C:\Phd\Phd\Phd\Phd\multisf2.dta" ,replace
  quietly multproc, pvalue(pvalue)

 if r(nreject)>0  {

  local sum1 = `sum1'+ 1
```

```
                                }

    }

local  prop  = `sum'/ `1'
local  prop1  = `sum1'/ `1'

display " proportion of any statistical effect is   `prop'  "
display " proportion of any statistical effect is   `prop1'  "

end

local i=log(1.5)
mybootun 10000 1200 -1.74 `i' 2 10 0.3 79
local i=log(1.5)
mybootun 10000 300 -1.74 `i' 2 10 0.3 79
local i=log(1.5)
mybootun 10000 4800 -1.74 `i' 2 10 0.3 79
local i=log(1.5)
mybootun 10000 1200 -1.74 `i' 5 10 0.3 79
local i=log(1.5)
mybootun 10000 300 -1.74 `i' 5 10 0.3 79
local i=log(1.5)
mybootun 10000 4800 -1.74 `i' 5 10 0.3 7
local i=log(1.5)
mybootun 10000 1200 -1.74 `i' 10 10 0.3 79
local i=log(1.5)
mybootun 10000 300 -1.74 `i' 10 10 0.3 79
local i=log(1.5)
mybootun 10000 4800 -1.74 `i' 10 10 0.3 79

*** Estimating Power and FDR of ESFA ***

program drop mybootunmix
program mybootunmix , rclass

display " Monte Carlo"
display " ------------------------------ "
```

```
display " Varying Effect Size "
display " ------------------------------ "
display " effect size is `4' , constant=`3' , Simes procedure"
display " ------------------------------ "
display " FDR at  " `8'*100 "%"  " level , " " odds ratio is " exp(`4') ", cut off value=`7'"
display " ------------------------------ "
display " `6' random selected foods , `5' principal components , sample size= `2' , number of replications=`1' , food items=`9'
"
display " ------------------------------ "

clear
quietly set memory 64m
set matsize 500
quietly postutil clear
postfile power1 power1 using "C:\Phd\power1.dta" ,every (1)replace
postfile fdr1 fdr1 using "C:\Phd\fdr1.dta" ,every(1) replace
forvalues i=1(1)`1' {

                    local S1=0
                    local R1=0

                    quietly use "C:\Phd\number1.dta" ,replace
                    quietly sample `6', count


                    forvalues j=1(1)`6' {

                                        local k`j' = number[`j']

                                        }

                    quietly use "C:\Phd\echrsbootstm.dta" ,replace
                    quietly bsample `2'

                                         local conf " port`k1' "

                    local v=(`6'/2)
```

```
 forvalues w=2(1)`v' {

                local conf  "`conf' + port`k`w'"

                }
 local z=`v'+1

 forvalues w=`z'(1)`6' {

                       local conf  "`conf' + port`k`w'"

                       }

quietly gen p=1 /(1+exp(-(`3'+`4'*(`conf'))))

 gen uniform=uniform()

 quietly gen disease = 0
 quietly replace disease = 1  if uniform <= p

 postfile multisf1  coef pvalue using "C:\Phd\multisf1.dta" ,replace

 forvalues k=6(1)79 {

                                     quietly xi:logit disease port`k' alkcal

                                     if e(df_m) > 1 {

                                                 local bs=_b[port`k']
                                                 local zs=_b[port`k']/_se[port`k']
                                                 local pvals=2*normal(-abs(`zs'))
                                                 post multisf1 (`bs') (`pvals')


                                                      }
                                            else  {

                                                 local bs=0
```

```
                                                    local pvals=1

                                                    post multisf1 (`bs') (`pvals')

                                                  }
                                  }
postclose multisf1

quietly use "C:\Phd\multisf1.dta" ,replace
quietly multproc, pvalue(pvalue) method(simes) reject(dummy1) puncor(`8')
quietly tab dummy

if r(nreject) > 0 {

                    forvalues i=1(1)`6' {

                                        local v=(`k`i''-5)

                                        if dummy1[`v']==1 {

                                                            local  S1 = `S1'+1

                                                  }

                                         }

                     local power=(`S1'/`6')*100
                     post power1 (`power')

                    }

    else {
          local power=0
          post power1 (`power')

          }
```

```
                if r(nreject) > 0 {

                                forvalues i=1(1)74 {

                                                if dummy1[`i']==1  {

                                                                                local  R1 = `R1'+1
                                                                        }

                                                 }
                                  if `R1'==0 {
                                              local fdr=0
                                              post fdr1 (`fdr')
                                             }

                                  else {
                                        local fdr=((`R1' -`S1')/`R1')*100
                                        post fdr1 (`fdr')

                                        }

                                  }
                     else {
                           local fdr=0
                           post fdr1 (`fdr')

                          }
                    }

postclose power1
postclose fdr1

use "C:\Phd\power1.dta" ,replace
tab power1
sort power1
sum power1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
```

```
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" power1[int(`lowerp')] ","power1[int(`upperp')] ")"

use "C:\Phd\fdr1.dta" ,replace
sort fdr1
sum fdr1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" fdr1[int(`lowerp')] ","fdr1[int(`upperp')] ")"

end

local i=log(1.5)
mybootunmix 10 1200 -1.74 `i' 5 10 0.3 0.2 79
local i=log(1.5)
mybootunmix  10000 2400 -1.74 `i' 5 10 0.3 0.2 79
local i=log(1.5)
mybootunmix  10000 300 -1.74 `i' 5 10 0.3 0.2 79
local i=log(1.5)
mybootunmix  10000 600 -1.74 `i' 5 10 0.3 0.2 79
local i=log(1.5)
mybootunmix  10000 4800 -1.74 `i' 5 10 0.3 0.2 79

*** Estimating Power and FDR of PCA  ***

program drop mybootpcaneg
program mybootpcaneg, rclass

display " Monte Carlo "
display " ------------------------------ "
display " Varying Sample size  "
display " ------------------------------ "
display "  effect size is `4' , constant=`3' , Simes procedure"
display " ------------------------------ "
display " FDR at  " `8'*100 "%"  " level , " " odds ratio is " exp(`4') ", cut off value=`7'"
```

```
display " ------------------------------- "
display " `6' random selected foods , `5' principal components , sample size= `2' , number of replications=`1' , food items=`9’
"
display " ------------------------------- "

clear
quietly set memory 400m
set matsize 500
quietly postutil clear
postfile power power using "C:\Phd\Phd\Phd\Phd\power.dta" , every (1) replace
postfile fdr fdr using "C:\Phd\Phd\Phd\Phd\fdr.dta" ,every (1) replace

forvalues i=1(1)`1' {

                    local S = 0
                    local R =0

                    quietly use "C:\Phd\Phd\Phd\Phd\number1.dta" ,replace
                    quietly sample `6', count

                     forvalues j=1(1)`6' {

                                        local k`j' = number[`j']

                                        }

                    quietly use "C:\Phd\echrsbootstm.dta" ,replace
                    quietly bsample `2'

                                         local conf " port`k1' "

                    local v=(`6'/2)

                    forvalues w=2(1)`v' {

                                   local conf  "`conf' +  port`k`w''"

                                   }
```

```
 local z=`v'+1

 forvalues w=`z'(1)`6' {

                        local conf  "`conf' - port`k`w'"

                        }

quietly gen p=1 /(1+exp(-(`3'+`4'*(`conf'))))

 gen uniform=uniform()

 quietly gen disease = 0
 quietly replace disease = 1  if uniform <= p
 quietly pca port6-port79, comp(`5')
 quietly rotate, varimax
 quietly predict pc1-pc`5'

 quietly corr   port`k1' port`k2' port`k3' port`k4' port`k5' port`k6' port`k7' port`k8' port`k9' port`k10'
pc1-pc`5'


 quietly matrix A=r(C)
 quietly corr port6-port79 pc1-pc`5'
 quietly matrix A1=r(C)
 postfile multipc1 coef pvalue using "C:\Phd\Phd\Phd\Phd\multipc1.dta" ,replace

 forvalues i=1(1)`5' {

                      quietly xi: logit disease pc`i' alkcal
                      local b`i'=_b[pc`i']
                      local z`i'=_b[pc`i']/_se[pc`i']
                      local pval`i'= 2*normal(-abs(`z`i''))
                      post multipc1 (`b`i'') (`pval`i'')

                      }

  postclose multipc1
```

```
quietly use "C:\Phd\Phd\Phd\Phd\multipc1.dta" ,replace

quietly multproc, pvalue(pvalue) reject(dummy) pcor(0.05)

if r(nreject)> 0 {

                  forvalues j=1(1)`6' {

                                      local w=0
                                      forvalues i=1(1)`5' {

                                       if  abs(A[`6'+`i',`j']) >=`7' & dummy[`i']==1 & `w'<1 {

                                                                                    local S = `S'+ 1


                                                                                                  }
                                                                  }
                                     }

                   local power=(`S'/`6')*100
                   post power (`power')

                  }

 else {

       local S=0
       local power=0
       post power (`power')

       }

  if r(nreject)> 0 {

                    forvalues j=1(1)74 {
                                        local l=0
```

```
                                forvalues i=1(1)`5' {


                  if  abs(A1[74 +`i',`j']) >= `7'  & dummy[`i']==1 & `l'<1 {

                                                                          local R = `R'+ 1

            local l=1


                                                                          }

                                                          }
                                    }

                    if `R'==0 {

                              local fdr=0
                              post fdr (`fdr')

                             }


                      else {

                            local fdr=((`R' -`S')/`R')*100
                            post fdr (`fdr')

                            }
                   }

       else {
              local fdr=0
              post fdr (`fdr')

            }
}
```

```
postclose power
postclose fdr

use "C:\Phd\Phd\Phd\Phd\power.dta" ,replace
sort power
sum power
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*((1-0.05)/2)
display " 95% CI (" power[int(`lowerp')] ","power[int(`upperp')] ")"

use "C:\Phd\Phd\Phd\Phd\fdr.dta" ,replace
sort fdr
sum fdr
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" fdr[int(`lowerp')] ","fdr[int(`upperp')] ")"

end

local i=log(1.5)
mybootpcaneg  10 1200 -1.74 `i' 2 10 0.3 0.2
local i=log(1.5)
mybootpcaneg  10000 300 -1.74 `i' 2 10 0.3 0.2
local i=log(1.5)
mybootpcaneg  10000 4800 -1.74 `i' 2 10 0.3 0.2

local i=log(1.5)
mybootpcaneg  10000 1200 -1.74 `i' 5 10 0.3 0.2
local i=log(1.5)
mybootpcaneg  10000 300 -1.74 `i' 5 10 0.3 0.2
local i=log(1.5)
mybootpcaneg  10000 4800 -1.74 `i' 5 10 0.3 0.2
```

```
local i=log(1.5)
mybootpcaneg  10000 1200 -1.74 `i' 10 10 0.3 0.2
local i=log(1.5)
mybootpcaneg  10000 300 -1.74 `i' 10 10 0.3 0.2
local i=log(1.5)
mybootpcaneg  10000 4800 -1.74 `i' 10 10 0.3 0.2

*** Estimating power and FDR of ESFA adjusting for propensity scores ***

program drop mybootadjneg
program mybootadjneg , rclass

display " Monte Carlo"
display " ------------------------------- "
display " Varying Effect Size "
display " ------------------------------- "
display "  effect size is `4' , constant=`3' , Simes procedure"
display " ------------------------------- "
display " FDR at  " `8'*100 "%"  " level , " " odds ratio is " exp(`4') ", cut off value=`7'"
display " ------------------------------- "
display " `6' random selected foods , `5' principal components , sample size= `2' , number of replications=`1', food items= `9'
"
display " ------------------------------- "

clear
quietly set memory 64m
set matsize 500
quietly postutil clear
postfile power1 power1 using "C:\Phd\Phd\Phd\Phd\power1.dta" ,every (1)replace
postfile fdr1 fdr1 using "C:\Phd\Phd\Phd\Phd\fdr1.dta" ,every(1) replace

                                        forvalues l=6(1)`9' {
                                        local conf " "
                                        forvalues j=6(1)`9' {

                                        if `j'!=`l' {

                                        local conf "`conf' port`j'"
```

```
                                        }
                                        local ps`l' "`conf'"
                                        }
                                        }
forvalues i=1(1)`1' {

                    local S1=0
                    local R1=0


                    quietly use "C:\Phd\Phd\Phd\Phd\number1.dta" ,replace

                    quietly sample `6', count

                    forvalues j=1(1)`6' {

                                        local k`j' = number[`j']

                                        }


                    quietly use "C:\Phd\Phd\Phd\Phd\echrsbootstm.dta" ,replace

                    quietly bsample `2'

                                         local conf "port`k1'"

                    local v=(`6'/2)

                    forvalues w=2(1)`v' {

                                  local conf  "`conf' +  port`k`w''"

                                  }

                    local z=`v'+1
```

```
 forvalues w=`z'(1)`6' {

                        local conf  "`conf' - port`k`w''"

                        }

quietly gen p=1 /(1+exp(-(`3'+`4'*(`conf'))))

 gen uniform=uniform()

 quietly gen disease = 0
 quietly replace disease = 1  if uniform <= p

 postfile multisf1  coef pvalue using "C:\Phd\Phd\Phd\Phd\multisf1.dta" ,replace

 forvalues k=6(1)`9' {

                                      quietly regress  port`k' `ps`k''

                                      quietly predict ps`k'

                                      quietly xi:logit disease port`k' ps`k' alkcal

                                      local port " "

                                      if e(df_m) > 2 {


                                                  local bs=_b[port`k']
                                                  local zs=_b[port`k']/_se[port`k']
                                                  local pvals=2*normal(-abs(`zs'))
                                                  post multisf1 (`bs') (`pvals')


                                                  }

                                             else  {
```

```
                                                    local k=1
                                                    local bs=0
                                                    local pvals=1

                                                    post multisf1 (`bs') (`pvals')

                                                }
                                }


postclose multisf1

quietly use "C:\Phd\Phd\Phd\Phd\multisf1.dta" ,replace

quietly multproc, pvalue(pvalue) method(simes) reject(dummy1) puncor(`8')

if r(nreject) > 0 {

                    forvalues i=1(1)`6' {

                                        local v=(`k`i''-5)
                                        if dummy1[`v']==1 {

                                                            local  S1 = `S1'+1


                                                        }
                                         }

                     local power=(`S1'/`6')*100
                     post power1 (`power')

                    }

    else {
          local power=0
```

```
          post power1 (`power')

          }

if r(nreject) > 0 {

                    forvalues i=1(1)74 {


                                        if dummy1[`i']==1  {

                                                            local  R1 = `R1'+1

                                                             }
                                         }


                      if `R1'==0 {
                                  local fdr=0
                                  post fdr1 (`fdr')
                                 }

                      else {
                            local fdr=((`R1' -`S1')/`R1')*100
                            post fdr1 (`fdr')

                            }

                      }


   else {
         local fdr=0
         post fdr1 (`fdr')

        }
```

```
                    }

postclose power1
postclose fdr1

use "C:\Phd\Phd\Phd\Phd\power1.dta" ,replace
tab power1
sort power1
sum power1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" power1[int(`lowerp')] ","power1[int(`upperp')] ")"

use "C:\Phd\Phd\Phd\Phd\fdr1.dta" ,replace
sort fdr1
sum fdr1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" fdr1[int(`lowerp')] ","fdr1[int(`upperp')] ")"
end
local i=log(1.5)
mybootadjneg 10 300 -1.74 `i' 5 4 0.3 0.2 79
local i=log(1.5)
mybootadjneg  10000 1200 -1.74 `i' 5 4 0.3 0.2 79
local i=log(1.5)
mybootadjneg  10000 4800 -1.74 `i' 5 4 0.3 0.2 79

*** Estimating Power and FDR of ESFA adjusted for 5 principal components ***

program drop mybootadjpcneg
program mybootadjpcneg , rclass

display " Bootstrap"
display " ------------------------------ "
```

```
display " Varying Effect Size "
display " ------------------------------ "
display " Modelling Disease with main effects only, effect size is `4' , constant=`3' , Simes procedure"
display " ------------------------------ "
display " FDR at  " `8'*100 "%"  " level , " " odds ratio is " exp(`4') ", cut off value=`7'"
display " ------------------------------ "
display " `6' random selected foods , `5' principal components , sample size= `2' , number of replications=`1' , food items = `9'
"
display " ------------------------------ "

clear
quietly set memory 64m
set matsize 500
quietly postutil clear
postfile power1 power1 using "C:\Phd\power1.dta" ,every (1)replace
postfile fdr1 fdr1 using "C:\Phd\fdr1.dta" ,every(1) replace
local f=`9'+ 5

forvalues i=1(1)`1' {

                    local S1=0
                    local R1=0

                    quietly use "C:\Phd\number1.dta" ,replace
                    quietly sample `6', count

                    forvalues j=1(1)`6' {

                                        local k`j' = number[`j']

                                        }

                    quietly use "C:\Phd\echrsbootstm.dta" ,replace
                    quietly bsample `2'

                                         local conf " port`k1'"

                    local v=(`6'/2)
```

```
 forvalues w=2(1)`v' {

                local conf  "`conf' + port`k`w'"

                }

 local z=`v'+1

 forvalues w=`z'(1)`6' {

                       local conf  "`conf' + port`k`w'"

                       }


quietly gen p=1 /(1+exp(-(`3'-`4'*(`conf'))))


 gen uniform=uniform()

 quietly gen disease = 0
 quietly replace disease = 1  if uniform <= p

 quietly pca port6-port`f', comp(`5')

 quietly rotate, varimax
 quietly predict pc1-pc`5'

 local conf1 " "

 forvalues j=1(1)`5' {


                    local conf1 "`conf1' pc`j'"

                    }
```

```
  postfile multisf1  coef pvalue using "C:\Phd\multisf1.dta" ,replace

  forvalues k=6(1)`f' {

                                        quietly xi:logit disease port`k' `conf1' alkcal

                                        if e(df_m) > (`5'+1) {


                                                local bs=_b[port`k']
                                                local zs=_b[port`k']/_se[port`k']
                                                local pvals=2*normal(-abs(`zs'))
                                                post multisf1 (`bs') (`pvals')

                                                     }

                                              else  {

                                                local bs=0
                                                local pvals=1

                                                post multisf1 (`bs') (`pvals')

                                             }

                                }
postclose multisf1

quietly use "C:\Phd\multisf1.dta" ,replace
quietly multproc, pvalue(pvalue) method(simes) reject(dummy1) puncor(`8')
quietly tab dummy

if r(nreject) > 0 {

                    forvalues i=1(1)`6' {

                                        local v=(`k`i''-5)
```

```
                                        if dummy1[`v']==1 {

                                                            local  S1 = `S1'+1

                                                           }
                                         }


                       local power=(`S1'/`6')*100
                       post power1 (`power')

                      }

    else {
          local power=0

          post power1 (`power')

          }

if r(nreject) > 0 {

                    forvalues i=1(1)`9' {

                                        if dummy1[`i']==1  {

                                                                local  R1 = `R1'+1

                                                             }

                                         }

                       if `R1'==0 {
                                   local fdr=0
                                   post fdr1 (`fdr')
                                  }

                       else {
```

```
                                        local fdr=(`R1' -`S1')/`R1'
                                        post fdr1 (`fdr')

                                        }

                                    }


                        else {
                              local fdr=0
                              post fdr1 (`fdr')

                             }
                       }

postclose power1
postclose fdr1

use "C:\Phd\power1.dta" ,replace
tab power1
sort power1
sum power1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" power1[int(`lowerp')] ","power1[int(`upperp')] ")"

use "C:\Phd\fdr1.dta" ,replace
sort fdr1
sum fdr1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*(1-(0.05/2))
display " 95% CI (" fdr1[int(`lowerp')] ","fdr1[int(`upperp')] ")"
```

```
end

local i=log(1.5)
mybootadjpcneg 10000 1200 -1.74 `i' 5 10 0.3 0.2 74
local i=log(1.5)
mybootadjpcneg 10000 300 -1.74 `i' 5 10 0.3 0.2 74
local i=log(1.5)
mybootadjpcneg 10000 4800 -1.74 `i' 5 10 0.3 0.2 74

*** Estimating Power and FDR of ESFA adjusting for statistical significant foods ***

program drop mybootadjf
program mybootadjf , rclass

display " Monte Carlo"
display " ------------------------------ "
display " Varying Effect Size "
display " ------------------------------ "
display " Modelling Disease with main effects only, effect size is `4' , constant=`3' , Simes procedure"
display " ------------------------------ "
display " FDR at  " `8'*100 "%"  " level , " " odds ratio is " exp(`4') ", cut off value=`7'"
display " ------------------------------ "
display " `6' random selected foods , `5' principal components , sample size= `2' , number of replications=`1' , food items = `9'
"
display " ------------------------------ "

clear
quietly set memory 64m
set matsize 500
quietly postutil clear

postfile power1 power1 using "C:\Phd\power1.dta" ,every (1) replace
postfile fdr1 fdr1 using "C:\Phd\fdr1.dta" ,every (1) replace

local f=`9'+ 5

forvalues i=1(1)`1' {
```

```
local S1=0
local R1=0

quietly use "C:\Phd\number1.dta" ,replace
quietly sample `6', count

forvalues j=1(1)`6' {

                    local k`j' = number[`j']

                    }


quietly use "C:\Phd\echrsbootstm.dta" ,replace


quietly bsample `2'

                     local conf " port`k1'"


local v=(`6'/2)

forvalues w=2(1)`v' {

                local conf  "`conf' -  port`k`w''"

                }

local z=`v'+1

forvalues w=`z'(1)`6' {

                      local conf  "`conf' + port`k`w''"

                      }
```

```
quietly gen p=1 /(1+exp(-(`3'+`4'*(`conf'))))

gen uniform=uniform()

quietly gen disease = 0
quietly replace disease = 1  if uniform <= p

 postfile multisf1  coef pvalue using "C:\Phd\multisf1.dta" ,replace

 forvalues k=6(1)`f' {

                                        quietly xi:logit disease port`k' alkcal

                                        if e(df_m) > 1 {


                                                    local bs=_b[port`k']
                                                    local zs=_b[port`k']/_se[port`k']
                                                    local pvals=2*normal(-abs(`zs'))
                                                    post multisf1 (`bs') (`pvals')


                                                         }

                                               else  {

                                                    local bs=0
                                                    local pvals=1
                                                    post multisf1 (`bs') (`pvals')

                                                   }

                                 }


postclose multisf1
```

```
preserve

quietly use "C:\Phd\multisf1.dta" ,replace

quietly multproc, pvalue(pvalue) method(simes) reject(dummy1) puncor(`8')

quietly tab dummy1

local conf " "

local w=0

if r(nreject) > 0 {

                    forvalues i=1(1)`9' {

                                        if dummy1[`i']==1  {

                                                            local w=`i'+5

                                                            local conf "`conf' port`w'"

                                                        }
                                                                                    }

                         }
 restore

 postfile multisf1_a  coef pvalue using "C:\Phd\multisf1_a.dta" ,replace

  forvalues k=6(1)`f' {

                                       quietly xi:logit disease port`k' `conf' alkcal

                                                    local bs=_b[port`k']
                                                    local zs=_b[port`k']/_se[port`k']
                                                    local pvals=2*normal(-abs(`zs'))
```

```
                                                post multisf1_a (`bs') (`pvals')

                                        else {
                                                local bs=0
                                                local pvals=1

                                                post multisf1_a (`bs') (`pvals')


                                            }
                            }
postclose multisf1_a

quietly use "C:\Phd\multisf1_a.dta" ,replace

quietly multproc, pvalue(pvalue) method(simes) reject(dummy1) puncor(`8')

quietly tab dummy1

if r(nreject) > 0 {

                    forvalues i=1(1)`6' {


                                    local v=(`k`i''-5)

                                    if dummy1[`v']==1 {

                                                    local  S1 = `S1'+1


                                                   }

                                     }


                     local power=(`S1'/`6')*100
                     post power1 (`power')
```

```
                    }

    else {
          local power=0

          post power1 (`power')

          }

if r(nreject) > 0 {

                    forvalues i=1(1)`9' {

                                        if dummy1[`i']==1  {

                                                                local  R1 = `R1'+1

                                                              }
                                         }


                     if `R1'==0 {
                                 local fdr=0
                                 post fdr1 (`fdr')
                                }

                     else {
                           local fdr=((`R1' -`S1')/`R1')*100
                           post fdr1 (`fdr')

                           }

                     }

   else {
         local fdr=0
         post fdr1 (`fdr')
```

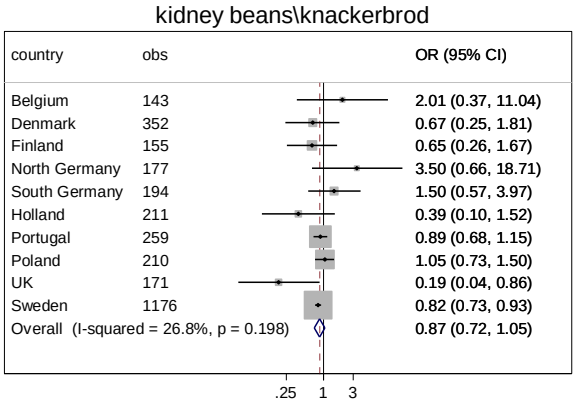
```
                            }

dis `power'
dis `fdr'
                    }

postclose power1
postclose fdr1

use "C:\Phd\power1.dta" ,replace
sort power1
sum power1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*((1-0.05)/2)
display " 95% CI (" power1[int(`lowerp')] ","power1[int(`upperp')] ")"

use "C:\Phd\fdr1.dta" ,replace
sort fdr1
sum fdr1
local seboot= r(sd)/sqrt(`1')
display "standard error is   " `seboot'
local lowerp=`1'*(0.05/2)
local upperp=`1'*((1-0.05)/2)
display " 95% CI (" fdr1[int(`lowerp')] ","fdr1[int(`upperp')] ")"

end

local i=log(1.5)
mybootadjf 10000 1200 -1.74 `i' 5 10 0.3 0.2 74
local i=log(1.5)
mybootadjf 10000 300 -1.74 `i' 5 10 0.3 0.2 74
local i=log(1.5)
mybootadjf 10000 4800 -1.74 `i' 5 10 0.3 0.2 74
```

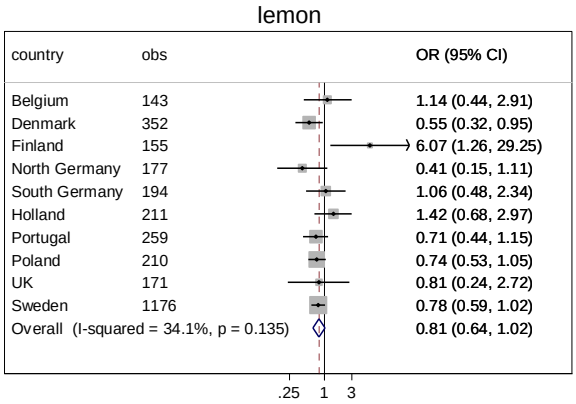
```
*** Simulating disease from "Western" dietary pattern derived from the UK ECHRS II dataset instead of randomly selected foods
(rest of the code remains the same ***

                                    local k1=13
                                     local k2=18
                                     local k3=21
                                     local k4=25
                                     local k5=30
                                     local k6=31
                                     local k7=37
                                     local k8=41
                                     local k9=52
                                     local k10=75

quietly gen p1=1 /(1+exp(-(`3'+`4'*(port`k1'+
port`k2'+port`k3'+port`k4'+port`k5'+port`k6'+port`k7'+port`k8'+port`k9'+port`k10'))))

                        gen uniform1=uniform()

                        quietly gen disease1 = 0
                        quietly replace disease1 = 1  if uniform1 <= p1

*** Simulating disease from "Western" dietary pattern derived from the UK ECHRS dataset instead of randomly selected foods (rest
of the code remains the same ***

                                     local k1=6
                                    local k2=28
                                    local k3=31
                                    local k4=34
                                    local k5=112
                                    local k6=113
                                    local k7=114
                                    local k8=115
                                    local k9=119
                                    local k10=121
                                    local k11=122
```

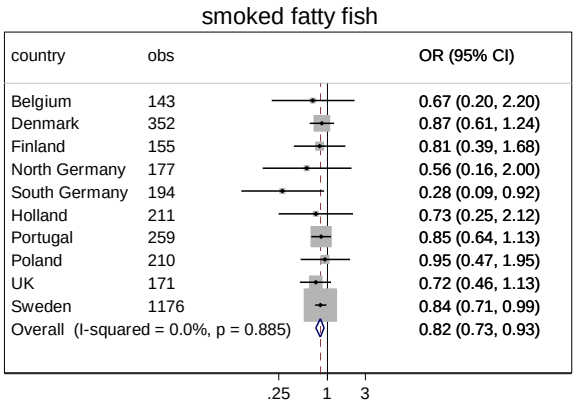
```
local k12=123
local k13=124
local k14=125
local k15=129
local k16=131
local k17=139
local k18=140
local k19=157
local k20=165
local k21=166
local k22=177
local k23=178
local k24=183
local k25=184
local k26=186
local k27=189
local k28=190
local k29=194
local k30=215
```

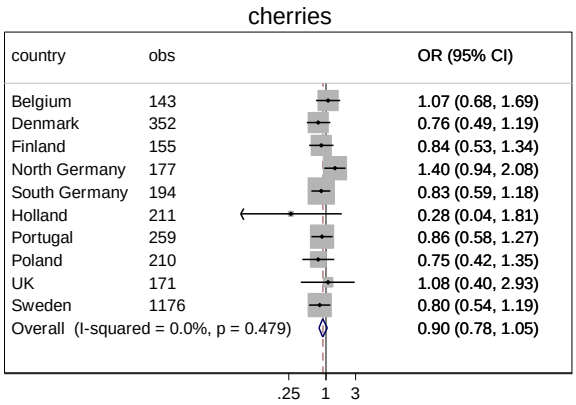
```
quietly gen p2=1 /(1+exp(-(`3'-`4'*(port`k1'+ port`k2'+
port`k3'+port`k4'+port`k5'+port`k6'+port`k7'+port`k8'+port`k9'+port`k10' +port`k11'+ port`k12'+
port`k13'+port`k14'+port`k15'+port`k16'+port`k17'+port`k18'+port`k19'+port`k20'+port`k21'+ port`k22'+
port`k23'+port`k24'+port`k25'+port`k26'+port`k27'+port`k28'+port`k29'+port`k30'))))

gen uniform2=uniform()
quietly gen disease2 = 0
quietly replace disease2 = 1  if uniform2 <= p2
```

III. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and asthma: results of meta-analyses. OR; odds ratio.

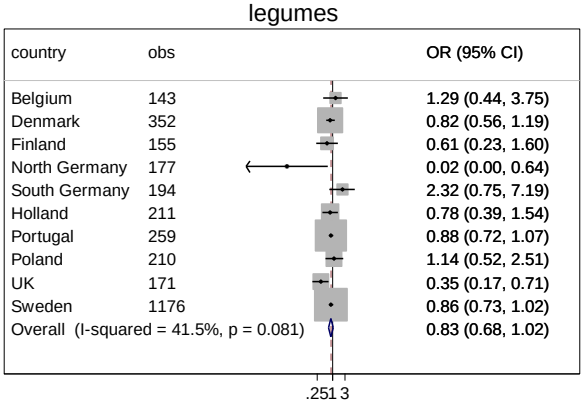




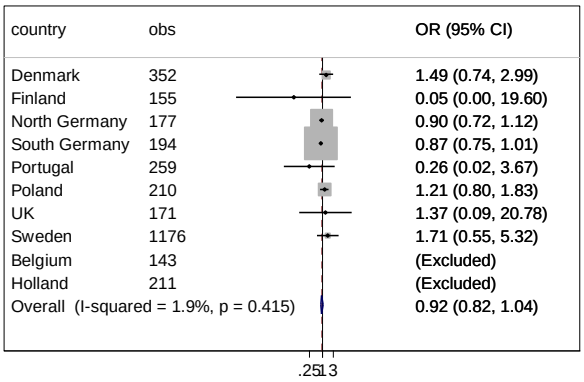




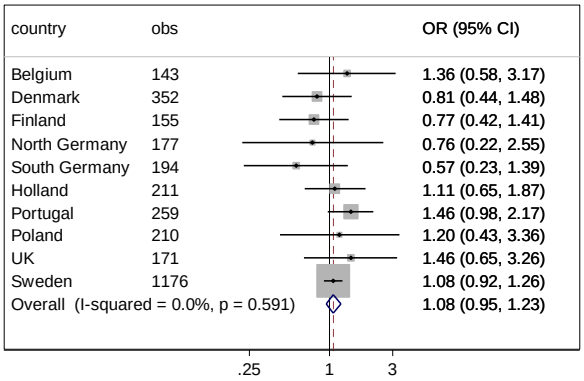
legumes



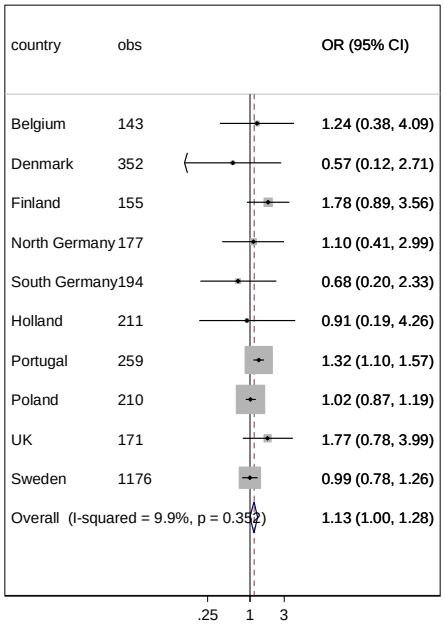
condensed milk



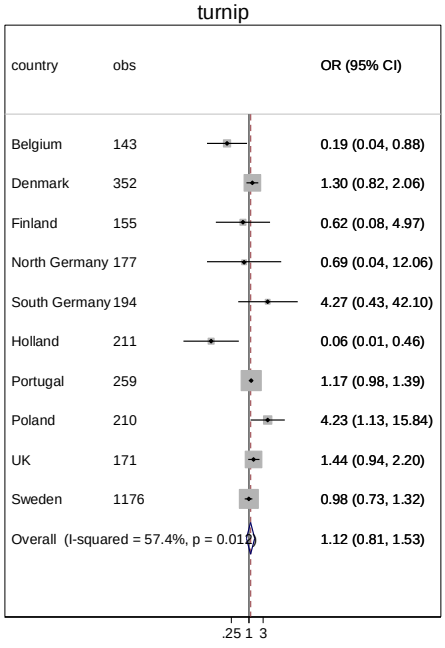
thin biscuits



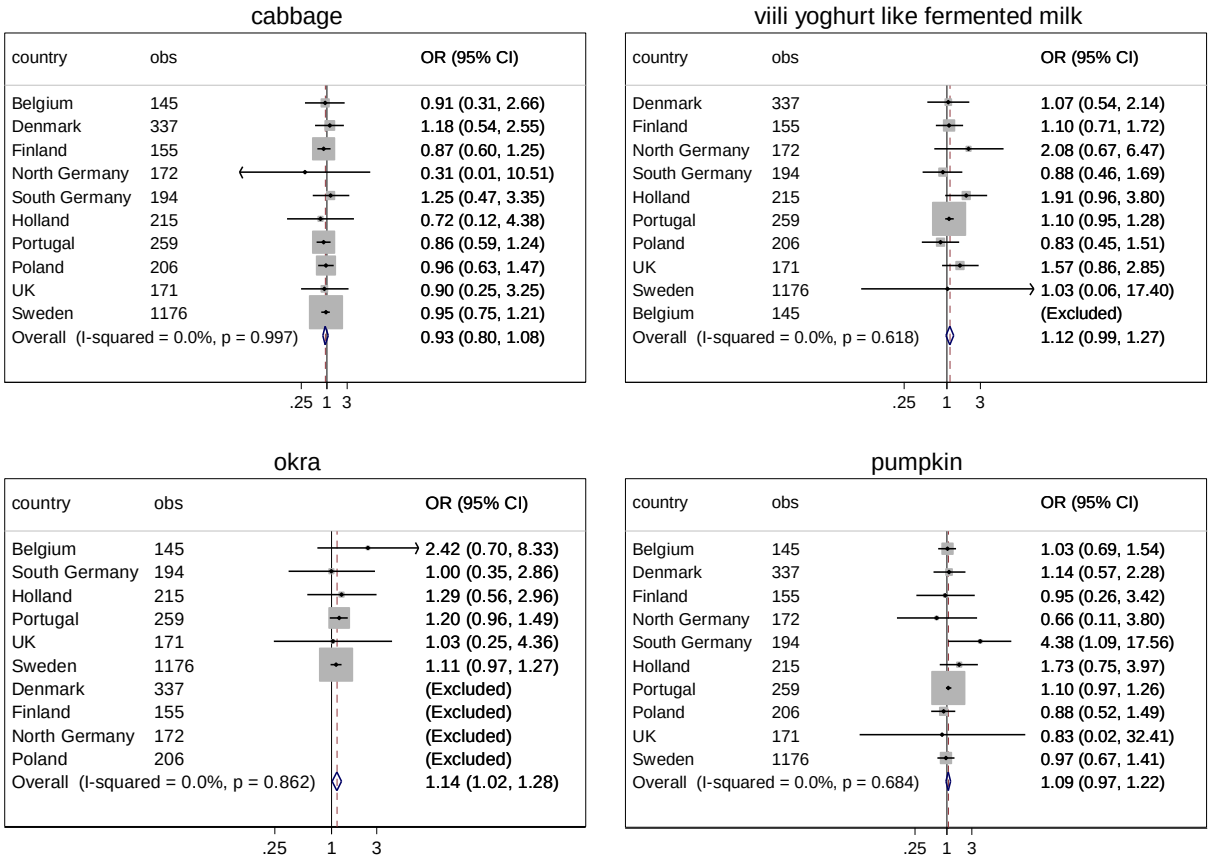
couscous



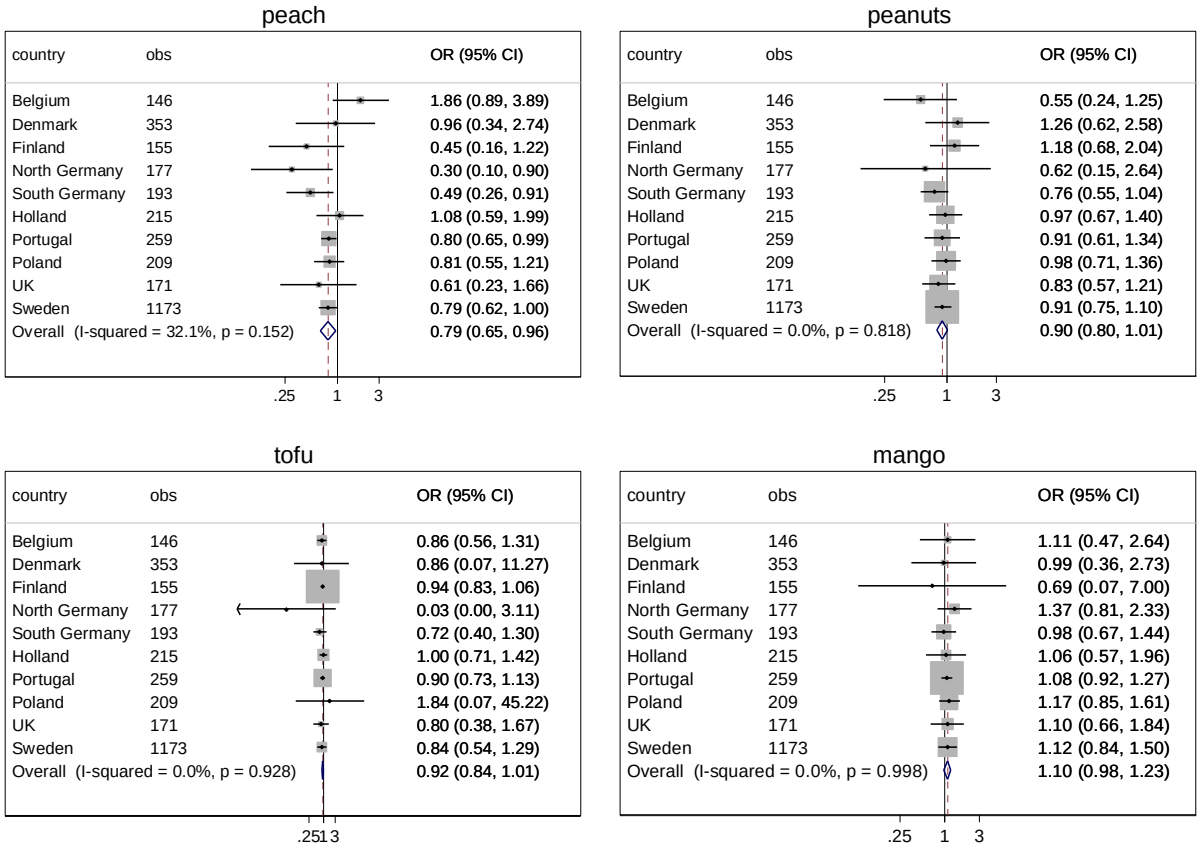
turnip



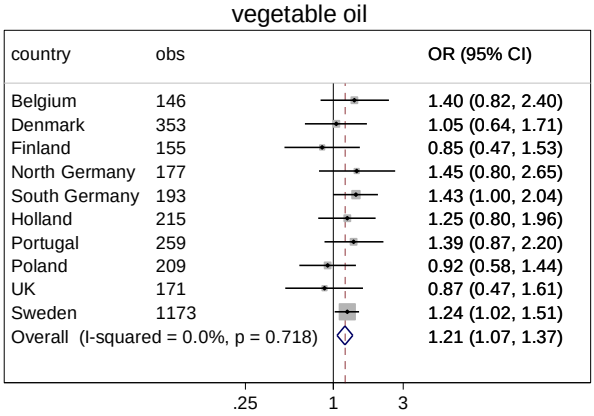
IV. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and chronic sinusitis: results of meta-analyses. OR; odds ratio.



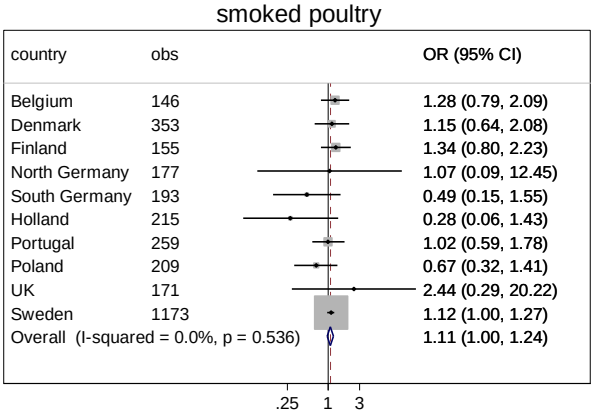
V. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and Allergic Rhinitis: results of meta-analyses. OR; odds ratio.



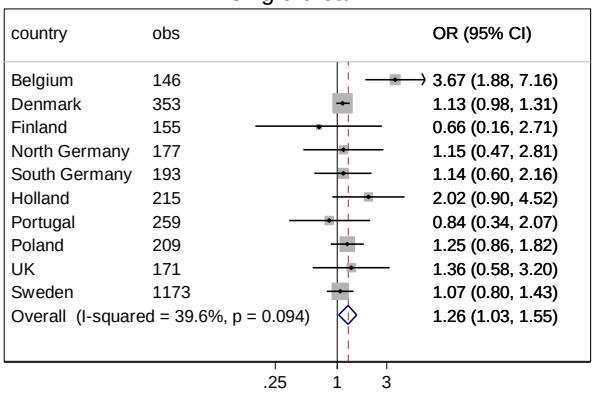
vegetable oil



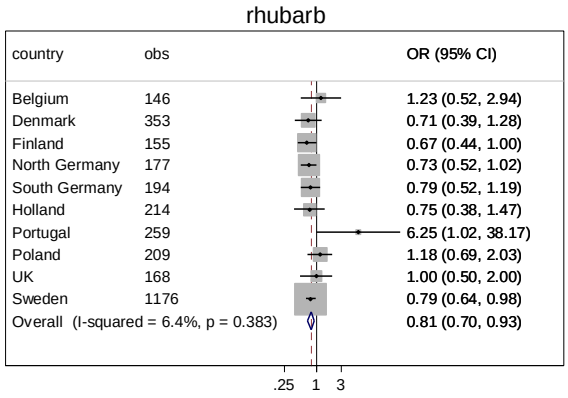
smoked poultry

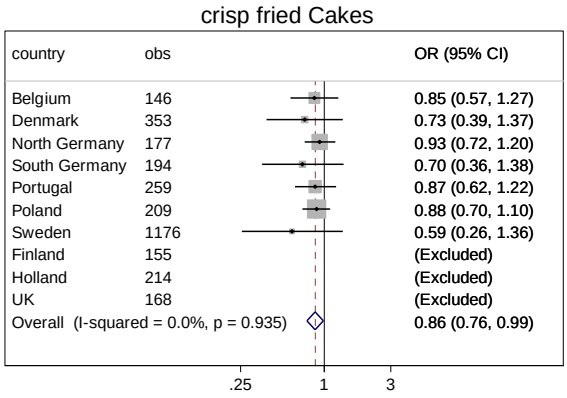


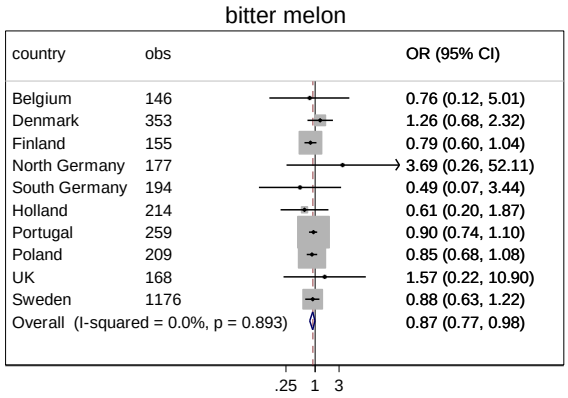
single cream

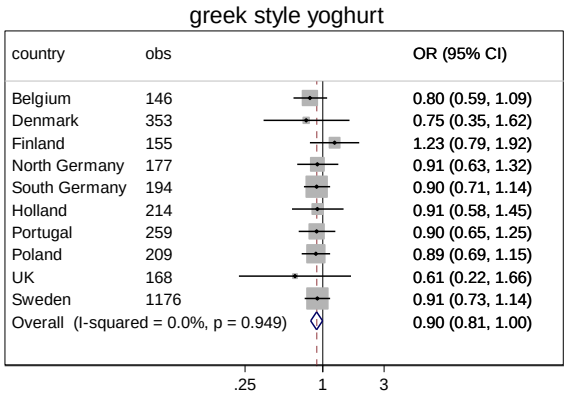


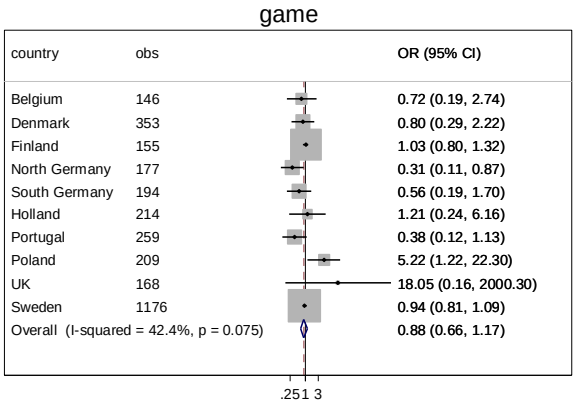
VI. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and eczema: results of meta-analyses. OR; odds ratio

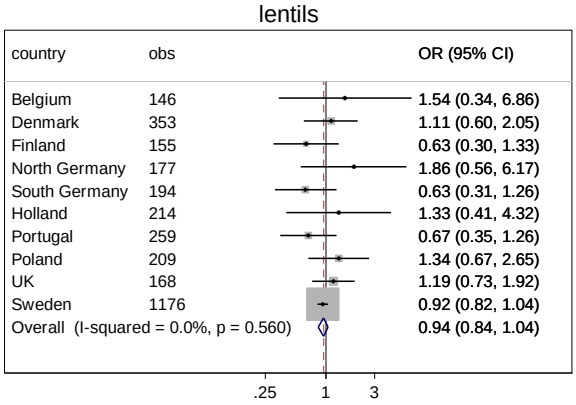


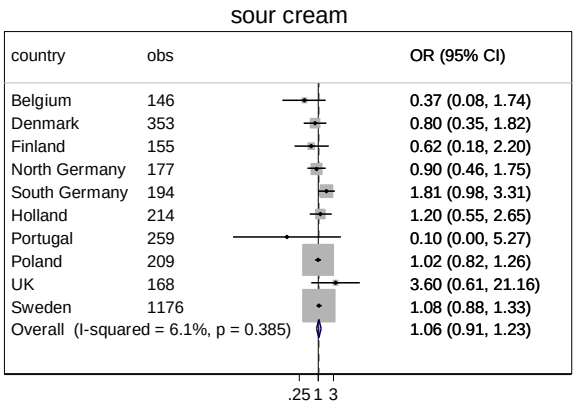




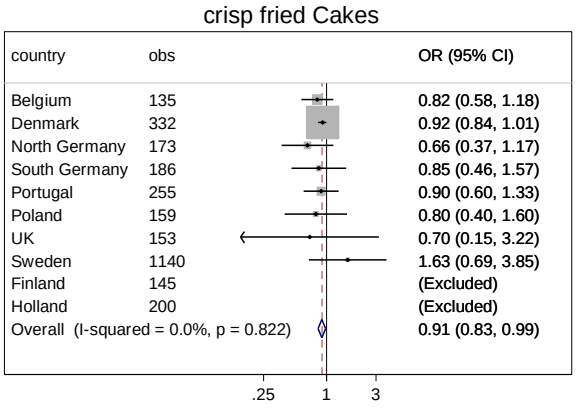


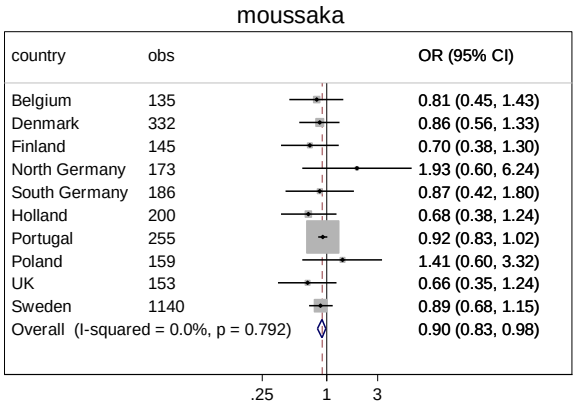

game
country
obs
OR (95% CI)
Belgium 146 0.72 (0.19, 2.74)
Denmark 353 0.80 (0.29, 2.22)
Finland 155 1.03 (0.80, 1.32)
North Germany 177 0.31 (0.11, 0.87)
South Germany 194 0.56 (0.19, 1.70)
Holland 214 1.21 (0.24, 6.16)
Portugal 259 0.38 (0.12, 1.13)
Poland 209 5.22 (1.22, 22.30)
UK 168 18.05 (0.16, 2000.30)
Sweden 1176 0.94 (0.81, 1.09)
Overall (I-squared = 42.4%, p = 0.075) 0.88 (0.66, 1.17)
.25 1 3

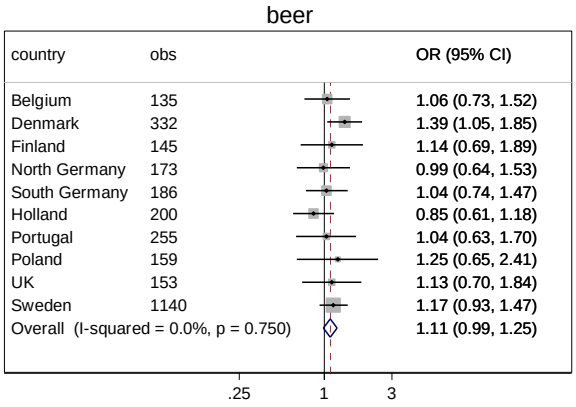

lentils
country
obs
OR (95% CI)
Belgium 146 1.54 (0.34, 6.86)
Denmark 353 1.11 (0.60, 2.05)
Finland 155 0.63 (0.30, 1.33)
North Germany 177 1.86 (0.56, 6.17)
South Germany 194 0.63 (0.31, 1.26)
Holland 214 1.33 (0.41, 4.32)
Portugal 259 0.67 (0.35, 1.26)
Poland 209 1.34 (0.67, 2.65)
UK 168 1.19 (0.73, 1.92)
Sweden 1176 0.92 (0.82, 1.04)
Overall (I-squared = 0.0%, p = 0.560) 0.94 (0.84, 1.04)
.25 1 3

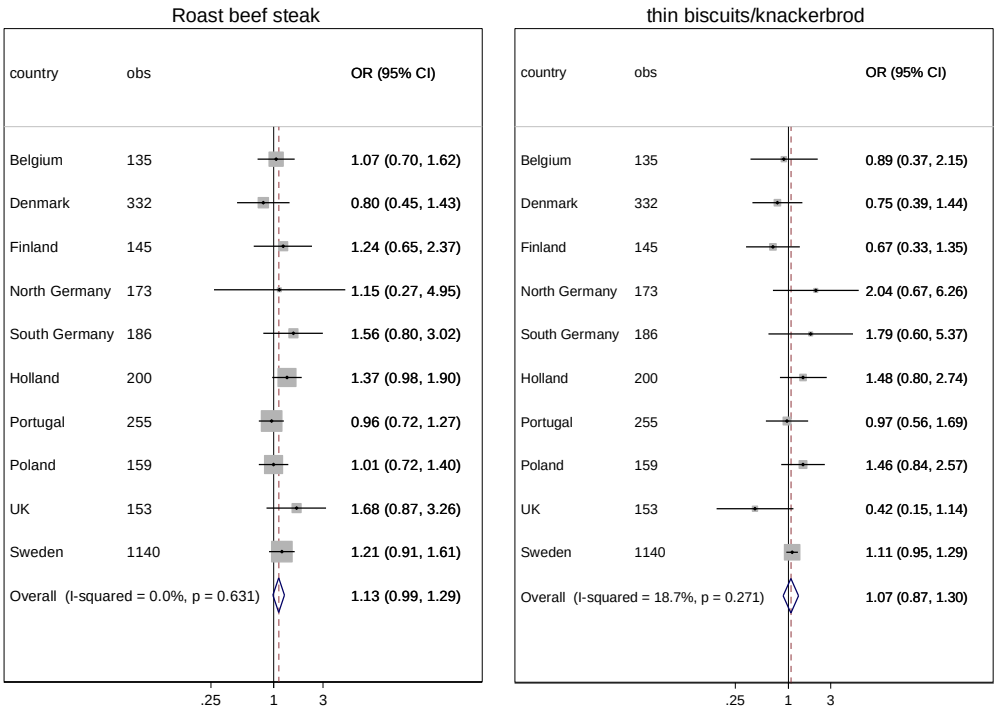

sour cream
country
obs
OR (95% CI)
Belgium 146 0.37 (0.08, 1.74)
Denmark 353 0.80 (0.35, 1.82)
Finland 155 0.62 (0.18, 2.20)
North Germany 177 0.90 (0.46, 1.75)
South Germany 194 1.81 (0.98, 3.31)
Holland 214 1.20 (0.55, 2.65)
Portugal 259 0.10 (0.00, 5.27)
Poland 209 1.02 (0.82, 1.26)
UK 168 3.60 (0.61, 21.16)
Sweden 1176 1.08 (0.88, 1.33)
Overall (I-squared = 6.1%, p = 0.385) 1.06 (0.91, 1.23)
.25 1 3

VII. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and atopy: results of meta-analyses. OR; odds ratio.








Roast beef steak
country
obs
OR (95% CI)
Belgium 135 1.07 (0.70, 1.62)
Denmark 332 0.80 (0.45, 1.43)
Finland 145 1.24 (0.65, 2.37)
North Germany 173 1.15 (0.27, 4.95)
South Germany 186 1.56 (0.80, 3.02)
Holland 200 1.37 (0.98, 1.90)
Portugal 255 0.96 (0.72, 1.27)
Poland 159 1.01 (0.72, 1.40)
UK 153 1.68 (0.87, 3.26)
Sweden 1140 1.21 (0.91, 1.61)
Overall (I-squared = 0.0%, p = 0.631) 1.13 (0.99, 1.29)
.25
1
3
thin biscuits/knackerbrod
country
obs
OR (95% CI)
Belgium 135 0.89 (0.37, 2.15)
Denmark 332 0.75 (0.39, 1.44)
Finland 145 0.67 (0.33, 1.35)
North Germany 173 2.04 (0.67, 6.26)
South Germany 186 1.79 (0.60, 5.37)
Holland 200 1.48 (0.80, 2.74)
Portugal 255 0.97 (0.56, 1.69)
Poland 159 1.46 (0.84, 2.57)
UK 153 0.42 (0.15, 1.14)
Sweden 1140 1.11 (0.95, 1.29)
Overall (I-squared = 18.7%, p = 0.271) 1.07 (0.87, 1.30)
.25
1
3

IX. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and respiratory and allergic outcomes: results of a random coefficient logistic model. OR; odds ratio.

ESFA (2nd step)* FDR=5% all P-values<0.001									
Asthma		**Chronic Sinusitis**		**Allergic Rhinitis**		**Eczema**		**Atopy**	
food	**OR**	**food**	**OR**	**food**	**OR**		**OR**		**OR**
kidney (black beans)	0.85	cabbage	0.87	peach	0.77	rhubarb	0.84	crisp fried cakes	0.86
lemon	0.85	Viili yogurt like fermented milk	1.11	peanuts	0.90	crisp fried cakes	0.84	moussaka	0.88
cured or smoked fatty fish	0.85	okra	1.12	tofu	0.90	bitter melon	0.88	beer	1.10
cherries	0.86	pumpkin	1.14	mango	1.10	Greek style yogurt	0.89	thin biscuits (knackerb rod)	1.13
legumes, any	0.89			vegetable oil	1.17	game, other	0.92	Hot cold Roast beef/ boiled beef	1.13
condensed milk	0.88			any smoked/ cured poultry	1.09	lentils	0.93		
thin biscuits (knackerbr od)	1.09			single cream	1.24	Sour cream	1.15		
turnip	1.13								
couscous	1.17								

*Adjusted for age, sex, body mass index, smoking status and all the foods that were significant at the univariate analysis (1st step) ***Bold type indicates foods which were statistically significant associated with eczema across 3 or more of the different procedures being used

X. Description of mean and median intake (grams/per day) for each food item in our study

food item	mean (min-max)	median (p25-p75)	percentage of individuals with non zero consumption of the specific food item
wholemeal bread	145.90(0.00 -350.00)	137.50(25.00-175.00)	88.82
white bread	126.54(0.00 -658.00)	47.00(23.50-141.00)	80.45
rye bread	53.84(0.00 -350.00)	12.50(0.00 -75.00)	60.25
kneipp bread	8.99 (0.00 -560.00)	0.00 (0.00 -0.00)	8.80
nan paratha bread	9.27 (0.00 -2100.00)	0.00 (0.00 -0.00)	8.50
chappati bread	3.78 (0.00 -770.00)	0.00 (0.00 -0.00)	6.61
wheat yeast rusks	7.96 (0.00 -420.00)	0.00 (0.00 -0.00)	19.35
breakfast cereals	98.57(0.00 -560.00)	40.00(0.00 -220.00)	66.54
couscous	25.93(0.00 -2100.00)	0.00 (0.00 -0.00)	24.54
total pasta	378.14(0.00 -3220.00)	230.00(115.00-690.00)	89.44
refined pasta	285.53(0.00 -3220.00)	230.00(115.00-230.00)	84.03
wholemeal pasta	105.44(0.00 -3220.00)	0.00 (0.00 -115.00)	40.34
filled pasta	50.11(0.00 -3220.00)	0.00 (0.00 -115.00)	30.34
noodles	55.45(0.00 -3220.00)	0.00 (0.00 -115.00)	30.43
any cakes	113.57(0.00 -840.00)	60.00(30.00-180.00)	85.96
cakes	46.25(0.00 -840.00)	30.00(0.00 -60.00)	71.49
danish pastries	44.21(0.00 -1540.00)	0.00 (0.00 -55.00)	45.06
sweet rolls	5.53 (0.00 -630.00)	0.00 (0.00 -0.00)	11.53
muffins	8.47 (0.00 -840.00)	0.00 (0.00 -0.00)	21.86
doughnuts/buns	34.13(0.00 -1050.00)	0.00 (0.00 -37.50)	44.97
Puddings/ deserts	47.96(0.00 -2100.00)	0.00 (0.00 -75.00)	37.70
custard cream	13.24(0.00 -840.00)	0.00 (0.00 -30.00)	28.12
greek cakes	1.75 (0.00 -1400.00)	0.00 (0.00 -0.00)	2.15
pancakes	32.26(0.00 -770.00)	0.00 (0.00 -55.00)	49.53
Italian biscuits	1.16 (0.00 -280.00)	0.00 (0.00 -0.00)	7.79
plain biscuits	15.89(0.00 -280.00)	10.00(0.00 -10.00)	56.86
crisp fried cakes	0.49 (0.00 -140.00)	0.00 (0.00 -0.00)	3.10
thin biscuits	4.92 (0.00 -56.00)	2.00 (0.00 -4.00)	57.84
sweet biscuits	3.86 (0.00 -98.00)	0.00 (0.00 -3.50)	43.34
rice	137.32(0.00 -1400.00)	100.00(50.00-300.00)	88.47
white rice	116.88(0.00 -1400.00)	50.00(50.00-100.00)	82.40
brown wholemeal pasta	32.78(0.00 -700.00)	0.00 (0.00 -50.00)	35.13
rice noodles	9.13 (0.00 -1400.00)	0.00 (0.00 -0.00)	11.50
table sugar	15.49(0.00 -70.00)	2.50 (0.00 -27.50)	54.06
jam	23.10(0.00 -210.00)	7.50 (0.00 -15.00)	67.77
marmelade	10.13(0.00 -210.00)	0.00 (0.00 -7.50)	37.28
honey	7.27 (0.00 -112.00)	0.00 (0.00 -4.00)	45.62

syrup spreads	1.21 (0.00 -70.00)	0.00 (0.00 -0.00)	12.22
apple spreads	1.24 (0.00 -140.00)	0.00 (0.00 -0.00)	8.67
total sweets	37.97(0.00 -280.00)	20.00(0.00 -60.00)	73.97
Boiled sweets	3.29 (0.00 -70.00)	0.00 (0.00 -2.50)	46.04
candies	12.30(0.00 -168.00)	6.00 (0.00 -12.00)	61.00
cereal bars	8.14 (0.00 -420.00)	0.00 (0.00 -0.00)	24.47
halva	0.39 (0.00 -140.00)	0.00 (0.00 -0.00)	5.12
water ice	12.52(0.00 -1050.00)	0.00 (0.00 -0.00)	18.70
any chocolates	12.99(0.00 -98.00)	7.00 (3.50 -21.00)	84.75
chocolates bars	31.13(0.00 -700.00)	25.00(0.00 -25.00)	51.45
plain chocolate	36.45(0.00 -350.00)	12.50(12.50-75.00)	78.92
vegetable oil	31.39(0.00 -154.00)	33.00(5.50 -60.50)	77.71
sunflower oil	13.55(0.00 -154.00)	5.50 (0.00 -11.00)	53.73
olive oil	32.50(0.00 -154.00)	33.00(5.50 -60.50)	81.33
any margarine	47.21(0.00 -140.00)	30.00(0.00 -70.00)	66.89
low fat margarine	19.58(0.00 -140.00)	0.00 (0.00 -10.00)	35.58
half fat margarine	15.96(0.00 -140.00)	0.00 (0.00 -5.00)	30.24
normal margarine	12.36(0.00 -140.00)	0.00 (0.00 -10.00)	36.69
blended spreads	18.20(0.00 -140.00)	0.00 (0.00 -10.00)	32.16
soya based spreads	1.88 (0.00 -140.00)	0.00 (0.00 -0.00)	5.05
any butter	29.09(0.00 -140.00)	5.00 (0.00 -55.00)	60.83
low fat butter	5.34 (0.00 -140.00)	0.00 (0.00 -0.00)	14.79
half fat butter	4.82 (0.00 -140.00)	0.00 (0.00 -0.00)	12.90
full fat butter	21.39(0.00 -140.00)	5.00 (0.00 -30.00)	50.31
lard	0.09 (0.00 -14.00)	0.00 (0.00 -0.00)	6.71
any nuts	14.05(0.00 -182.00)	6.50 (0.00 -13.00)	67.12
peanuts	4.92 (0.00 -140.00)	0.00 (0.00 -5.00)	49.23
cashew nuts	4.85 (0.00 -140.00)	0.00 (0.00 -5.00)	44.22
nut based spreads	2.23 (0.00 -140.00)	0.00 (0.00 -0.00)	15.71
any legumes	128.11(0.00 -1680.00)	60.00(0.00 -120.00)	70.38
kidney	35.81(0.00 -1260.00)	0.00 (0.00 -45.00)	46.53
lentils	28.05(0.00 -1680.00)	0.00 (0.00 -60.00)	28.64
chickpeas	30.44(0.00 -1260.00)	0.00 (0.00 -45.00)	38.94
clusterbeans	4.74 (0.00 -630.00)	0.00 (0.00 -0.00)	6.88
frenchbeans	41.28(0.00 -1260.00)	45.00(0.00 -45.00)	52.43
favabeans	16.69(0.00 -1260.00)	0.00 (0.00 -0.00)	23.26
soyabeans	14.01(0.00 -1260.00)	0.00 (0.00 -0.00)	13.59
any leafy vegetables	475.33(0.00 -1120.00)	440.00(240.00-560.00)	93.29
lettuce	247.03(0.00 -1120.00)	240.00(80.00-440.00)	92.54
spinach	57.26(0.00 -1260.00)	45.00(0.00 -45.00)	67.45
chard	26.40(0.00 -1120.00)	0.00 (0.00 -0.00)	22.61
fenugreek	1.90 (0.00 -440.00)	0.00 (0.00 -0.00)	2.44
wildgreens	13.22(0.00 -1260.00)	0.00 (0.00 -0.00)	14.66
okra	1.19 (0.00 -330.00)	0.00 (0.00 -0.00)	2.35

caper	12.52(0.00 -560.00)	0.00 (0.00 -0.00)	21.60
tomato	332.51(0.00 -1190.00)	255.00(85.00-467.50)	94.75
aubergine	44.91(0.00 -1820.00)	0.00 (0.00 -65.00)	39.95
courgette	58.86(0.00 -1400.00)	50.00(0.00 -50.00)	55.65
sweet pepper	309.04(0.00 -2240.00)	160.00(80.00-480.00)	87.29
cucumber	66.78(0.00 -322.00)	69.00(11.50-69.00)	89.87
bitter mellon	8.86 (0.00 -1120.00)	0.00 (0.00 -0.00)	6.06
carrot	150.07(0.00 -840.00)	60.00(30.00-180.00)	92.15
parsnip	20.98(0.00 -910.00)	0.00 (0.00 -32.50)	30.27
turnsip	16.31(0.00 -770.00)	0.00 (0.00 -27.50)	28.25
artichoke	4.11 (0.00 -275.00)	0.00 (0.00 -0.00)	14.34
radish	3.97 (0.00 -112.00)	0.00 (0.00 -4.00)	39.04
beetroot	22.66(0.00 -560.00)	20.00(0.00 -20.00)	56.99
celery	23.36(0.00 -700.00)	0.00 (0.00 -25.00)	42.10
coleslaw	20.59(0.00 -980.00)	0.00 (0.00 -35.00)	30.34
sweetcorn	56.50(0.00 -1190.00)	42.50(0.00 -42.50)	59.34
asparagus	45.92(0.00 -875.00)	0.00 (0.00 -62.50)	48.39
herbs	58.17(0.00 -420.00)	30.00(15.00-90.00)	81.36
leek	90.94(0.00 -1260.00)	45.00(45.00-90.00)	79.18
white mushroom	68.63(0.00 -1120.00)	40.00(40.00-80.00)	77.78
onion	215.40(0.00 -980.00)	210.00(70.00-385.00)	93.97
garlic	6.74 (0.00 -42.00)	3.00 (1.50 -9.00)	85.04
cauliflower	71.01(0.00 -1260.00)	45.00(45.00-90.00)	75.69
pumpkin	20.03(0.00 -1330.00)	0.00 (0.00 -0.00)	19.26
brussel spouts	29.25(0.00 -1260.00)	0.00 (0.00 -45.00)	39.95
broccoli	78.99(0.00 -1190.00)	42.50(42.50-85.00)	77.84
cabbage	81.93(0.00 -1330.00)	47.50(0.00 -95.00)	68.13
stuffed vegetables	12.35(0.00 -1260.00)	0.00 (0.00 -0.00)	19.81
pickled vegetables	63.89(0.00 -1260.00)	45.00(0.00 -45.00)	55.78
ginger	25.05(0.00 -840.00)	0.00 (0.00 -30.00)	33.37
potatoes	523.57(0.00 -2240.00)	480.00(160.00-880.00)	93.68
mashed potato	48.85(0.00 -840.00)	30.00(30.00-60.00)	75.40
casserole	68.76(0.00 -1190.00)	42.50(0.00 -85.00)	74.36
chips/fries	89.84(0.00 -2310.00)	82.50(0.00 -82.50)	61.55
salads	24.89(0.00 -1190.00)	0.00 (0.00 -42.50)	45.16
potato dumplings	5.77 (0.00 -560.00)	0.00 (0.00 -0.00)	23.33
potato tortilla	8.65 (0.00 -1400.00)	0.00 (0.00 -0.00)	13.95
sweet potato	21.00(0.00 -1120.00)	0.00 (0.00 -0.00)	17.24
apple	266.79(0.00 -1050.00)	225.00(37.50-412.50)	88.69
pear	234.89(0.00 -2100.00)	75.00(75.00-450.00)	74.88
banana	252.55(0.00 -1400.00)	100.00(50.00-300.00)	86.41
peach	105.27(0.00 -1540.00)	55.00(0.00 -110.00)	58.85
avocado	48.65(0.00 -1260.00)	0.00 (0.00 -45.00)	45.49
cherry	51.76(0.00 -1330.00)	0.00 (0.00 -47.50)	37.54

rhubarb	29.36(0.00 -980.00)	0.00 (0.00 -70.00)	26.91
forest fruits	195.60(0.00 -9100.00)	70.00(0.00 -140.00)	73.64
melon watermelon	149.04(0.00 -2800.00)	100.00(0.00 -100.00)	64.78
grape	119.30(0.00 -1400.00)	50.00(50.00-100.00)	79.93
mango	12.64(0.00 -560.00)	0.00 (0.00 -20.00)	31.54
apricot	12.60(0.00 -560.00)	0.00 (0.00 -20.00)	27.11
nectarine	74.96(0.00 -1260.00)	45.00(0.00 -90.00)	57.51
plum	41.23(0.00 -770.00)	27.50(0.00 -27.50)	50.90
squeezed fruit	57.69(0.00 -700.00)	25.00(0.00 -50.00)	50.24
pineapple	43.98(0.00 -1120.00)	40.00(0.00 -40.00)	55.69
kiwi	44.52(0.00 -840.00)	30.00(0.00 -30.00)	57.71
lemon	9.50 (0.00 -140.00)	5.00 (0.00 -10.00)	57.84
orange	267.32(0.00 -2240.00)	80.00(80.00-480.00)	79.80
mandarine/tangarine	169.04(0.00 -1400.00)	50.00(50.00-300.00)	78.27
grapefruit	30.59(0.00 -1120.00)	0.00 (0.00 -40.00)	29.68
canned fruits	20.11(0.00 -1260.00)	0.00 (0.00 -45.00)	29.16
raisins	21.46(0.00 -420.00)	0.00 (0.00 -15.00)	43.70
fig	12.17(0.00 -770.00)	0.00 (0.00 -0.00)	20.95
prune	13.16(0.00 -840.00)	0.00 (0.00 -0.00)	19.22
olive	62.09(0.00 -980.00)	35.00(0.00 -70.00)	58.36
juice with sugar	130.43(0.00 -2240.00)	0.00 (0.00 -80.00)	40.14
juice without sugar	207.81(0.00 -2240.00)	80.00(0.00 -160.00)	49.98
soft drinks	378.23(0.00 -3500.00)	0.00 (0.00 -250.00)	48.39
tap water	1863.43(0.00 -2800.00)	2800.00(600.00-2800.00)	82.83
mineral water	873.34(0.00 -2800.00)	200.00(100.00-1400.00)	75.76
soda with sugar	107.23(0.00 -2800.00)	0.00 (0.00 -100.00)	26.95
soda without sugar	138.96(0.00 -2800.00)	0.00 (0.00 -100.00)	25.42
black tea	653.13(0.00 -2660.00)	95.00(0.00 -1330.00)	59.89
coffee	1597.57(0.00 -2660.00)	1330.00(190.00-2660.00)	79.60
greek coffee	51.31(0.00 -2660.00)	0.00 (0.00 -0.00)	6.00
herbal tea	416.33(0.00 -2660.00)	95.00(0.00 -570.00)	53.08
beer	384.34(0.00 -3990.00)	142.50(0.00 -285.00)	61.23
any wine	189.03(0.00 -1750.00)	62.50(0.00 -375.00)	65.92
red wine	164.81(0.00 -1750.00)	62.50(0.00 -125.00)	61.81
white wine	80.70(0.00 -1750.00)	62.50(0.00 -62.50)	50.86
rose wine	26.52(0.00 -1750.00)	0.00 (0.00 -0.00)	21.28
fortified wines	8.12 (0.00 -700.00)	0.00 (0.00 -0.00)	18.44
spirits	20.81(0.00 -700.00)	0.00 (0.00 -25.00)	37.21
any red meat	220.74(0.00 -1050.00)	225.00(75.00-225.00)	88.33
roast beef steak	123.96(0.00 -1680.00)	60.00(60.00-120.00)	76.31
beef burger	55.76(0.00 -1680.00)	60.00(0.00 -60.00)	56.83
minced beef meat	156.10(0.00 -1960.00)	140.00(70.00-140.00)	86.18
meat stew	71.10(0.00 -1960.00)	70.00(0.00 -70.00)	61.62
pork steak	82.35(0.00 -1260.00)	45.00(45.00-90.00)	80.42

meat pies	24.36(0.00 -1960.00)	0.00 (0.00 -0.00)	23.10
sausages	42.80(0.00 -840.00)	30.00(0.00 -30.00)	56.34
meat spreads	2.76 (0.00 -140.00)	0.00 (0.00 -0.00)	12.58
veal	27.57(0.00 -630.00)	0.00 (0.00 -45.00)	34.51
small game	18.32(0.00 -2450.00)	0.00 (0.00 -0.00)	14.73
other game	23.18(0.00 -1225.00)	0.00 (0.00 -0.00)	17.07
lamb	19.62(0.00 -630.00)	0.00 (0.00 -45.00)	33.24
cured pork	42.00(0.00 -630.00)	22.50(0.00 -45.00)	52.10
Salami/gammon/ham	13.56(0.00 -168.00)	6.00 (0.00 -12.00)	71.36
frankfurter	82.43(0.00 -1960.00)	70.00(0.00 -70.00)	57.80
bacon cubes/bacon	21.79(0.00 -560.00)	20.00(0.00 -20.00)	65.53
smoked lamb	1.36 (0.00 -330.00)	0.00 (0.00 -0.00)	3.32
smoked game	3.27 (0.00 -980.00)	0.00 (0.00 -0.00)	6.45
any poultry	145.16(0.00 -1400.00)	100.00(50.00-300.00)	86.25
chicken boiled roast	122.34(0.00 -1400.00)	100.00(50.00-100.00)	90.91
chicken in stews	45.87(0.00 -1260.00)	45.00(0.00 -45.00)	56.08
turkey roasted boile	28.95(0.00 -1260.00)	0.00 (0.00 -45.00)	34.54
stews breadcrumbs	10.81(0.00 -770.00)	0.00 (0.00 -0.00)	13.36
smoked poultry	14.04(0.00 -980.00)	0.00 (0.00 -0.00)	18.74
liver pates	23.17(0.00 -560.00)	0.00 (0.00 -20.00)	45.06
other offal	1.30 (0.00 -137.50)	0.00 (0.00 -0.00)	8.86
fresh fatty fish	121.63(0.00 -1680.00)	60.00(60.00-120.00)	83.84
fresh white fish	81.25(0.00 -1540.00)	55.00(0.00 -110.00)	72.21
other fish seafood	67.54(0.00 -2240.00)	0.00 (0.00 -80.00)	41.41
fresh crustaceans mollucks	12.68(0.00 -280.00)	0.00 (0.00 -20.00)	48.19
smoked fatty fish	27.51(0.00 -490.00)	35.00(0.00 -35.00)	54.38
smoked white fish	12.70(0.00 -525.00)	0.00 (0.00 -0.00)	19.32
tinned fatty fish	29.07(0.00 -980.00)	35.00(0.00 -35.00)	52.79
tinned crustaceans mollucks	6.18 (0.00 -275.00)	0.00 (0.00 -0.00)	20.30
All eggs	94.87(0.00 -840.00)	60.00(30.00-180.00)	89.44
egg based dishes	68.21(0.00 -1540.00)	55.00(0.00 -110.00)	68.23
egg based desserts	35.84(0.00 -1680.00)	0.00 (0.00 -60.00)	35.39
total milk	1117.37(0.00 -2800.00)	1100.00(100.00-1400.00)	81.04
sour milk	42.22(0.00 -420.00)	0.00 (0.00 -30.00)	46.33
full fat milk	176.51(0.00 -2800.00)	0.00 (0.00 -100.00)	25.25
semi skimmed milk	592.44(0.00 -2800.00)	100.00(0.00 -1100.00)	56.24
skimmed milk	311.02(0.00 -2800.00)	0.00 (0.00 -100.00)	28.35
condensed milk	16.96(0.00 -700.00)	0.00 (0.00 -0.00)	9.29
total yoghurt	290.42(0.00 -1750.00)	125.00(0.00 -375.00)	70.67
greek style yogurt	59.17(0.00 -1750.00)	0.00 (0.00 -62.50)	31.96
fromage frais	56.28(0.00 -1190.00)	0.00 (0.00 -42.50)	42.49
soya milk	22.16(0.00 -1750.00)	0.00 (0.00 -0.00)	7.04
viili	35.49(0.00 -1750.00)	0.00 (0.00 -0.00)	10.39
tofu	5.02 (0.00 -1120.00)	0.00 (0.00 -0.00)	6.03

any cheese	192.25(0.00 -560.00)	120.00(40.00-280.00)	91.63
hard cheeses	43.11(0.00 -560.00)	20.00(0.00 -40.00)	60.31
soft cheeses	46.03(0.00 -560.00)	20.00(0.00 -40.00)	73.25
semi hard cheeses	130.42(0.00 -560.00)	120.00(20.00-220.00)	80.55
cottage cheese	25.74(0.00 -560.00)	0.00 (0.00 -20.00)	40.47
greek cheeses	3.55 (0.00 -560.00)	0.00 (0.00 -0.00)	6.91
fresh cheeses	29.05(0.00 -560.00)	20.00(0.00 -40.00)	66.05
ice cream	50.17(0.00 -1050.00)	37.50(0.00 -37.50)	72.01
single cream	4.78 (0.00 -210.00)	0.00 (0.00 -7.50)	26.49
creme fraiche	8.64 (0.00 -210.00)	7.50 (0.00 -7.50)	56.47
sour cream	4.94 (0.00 -105.00)	0.00 (0.00 -7.50)	40.47
Double clotted cream	5.07 (0.00 -210.00)	0.00 (0.00 -7.50)	35.78
dressing sauces	12.68(0.00 -280.00)	0.00 (0.00 -10.00)	48.65
mayonnaise	9.05 (0.00 -210.00)	7.50 (0.00 -7.50)	55.07
white sauce	4.44 (0.00 -105.00)	0.00 (0.00 -7.50)	42.82
ketchup	12.80(0.00 -210.00)	7.50 (0.00 -15.00)	66.47
fresh vegetable	270.74(0.00 -3080.00)	110.00(0.00 -220.00)	69.60
fresh meat	55.13(0.00 -1540.00)	0.00 (0.00 -110.00)	33.07
pizza	45.61(0.00 -770.00)	55.00(0.00 -55.00)	67.25
moussaka	6.12 (0.00 -660.00)	0.00 (0.00 -0.00)	9.51

XI. Description of mean and median intake (grams per day) for each food item that was identified as significant from adjusted ESFA (FDR =5%) at 2nd step by each country.

Food item	mean	median	number of individuals with non zero consumption of the specific food item	area
any legumes	95.75(0.00 -840.00)	60.00(60.00-120.00)	110	Belgium
any legumes	87.29(0.00 -840.00)	60.00(0.00 -60.00)	209	Denmark
any legumes	72.77(0.00 -840.00)	60.00(0.00 -60.00)	84	Finaland
any legumes	57.97(0.00 -660.00)	60.00(0.00 -60.00)	122	North Germany
any legumes	69.28(0.00 -840.00)	60.00(0.00 -60.00)	139	South Germany
any legumes	112.47(0.00 -840.00)	60.00(60.00-120.00)	172	Holland
any legumes	294.21(0.00 -1680.00)	120.00(60.00-360.00)	215	Portugal
any legumes	116.57(0.00 -1680.00)	60.00(60.00-120.00)	183	Poland
any legumes	99.65(0.00 -1680.00)	60.00(0.00 -120.00)	89	UK
any legumes	144.44(0.00 -1680.00)	60.00(0.00 -120.00)	837	Sweden
beer	539.74(0.00 -3990.00)	142.50(0.00 -855.00)	92	Belgium
beer	516.06(0.00 -3990.00)	142.50(0.00 -855.00)	253	Denmark
beer	340.16(0.00 -3990.00)	142.50(0.00 -285.00)	85	Finaland
beer	402.54(0.00 -3990.00)	142.50(0.00 -285.00)	98	North Germany
beer	405.46(0.00 -3990.00)	142.50(0.00 -285.00)	114	South Germany
beer	503.06(0.00 -3990.00)	142.50(0.00 -285.00)	114	Holland
beer	237.68(0.00 -3990.00)	0.00 (0.00 -285.00)	119	Portugal
beer	308.75(0.00 -3990.00)	142.50(0.00 -285.00)	130	Poland
beer	412.50(0.00 -3990.00)	0.00 (0.00 -285.00)	82	UK
beer	344.98(0.00 -3990.00)	142.50(0.00 -285.00)	792	Sweden
bitter mellon	1.10 (0.00 -80.00)	0.00 (0.00 -0.00)	2	Belgium
bitter mellon	10.73(0.00 -440.00)	0.00 (0.00 -0.00)	37	Denmark
bitter mellon	22.71(0.00 -1120.00)	0.00 (0.00 -0.00)	13	Finaland
bitter mellon	4.52 (0.00 -560.00)	0.00 (0.00 -0.00)	6	North Germany
bitter mellon	2.27 (0.00 -240.00)	0.00 (0.00 -0.00)	5	South Germany
bitter mellon	3.72 (0.00 -240.00)	0.00 (0.00 -0.00)	14	Holland
bitter mellon	34.90(0.00 -1120.00)	0.00 (0.00 -0.00)	58	Portugal
bitter mellon	22.29(0.00 -1120.00)	0.00 (0.00 -0.00)	14	Poland

bitter mellon	4.21 (0.00 -240.00)	0.00 (0.00 -0.00)	8	UK
bitter mellon	2.65 (0.00 -560.00)	0.00 (0.00 -0.00)	29	Sweden
cabbage	52.38(0.00 -522.50)	47.50(0.00 -47.50)	105	Belgium
cabbage	57.03(0.00 -665.00)	47.50(0.00 -47.50)	236	Denmark
cabbage	106.03(0.00 -1330.00)	47.50(0.00 -95.00)	109	Finaland
cabbage	46.16(0.00 -522.50)	47.50(0.00 -47.50)	125	North Germany
cabbage	49.21(0.00 -665.00)	47.50(0.00 -47.50)	131	South Germany
cabbage	49.71(0.00 -285.00)	47.50(0.00 -47.50)	156	Holland
cabbage	131.50(0.00 -1330.00)	47.50(0.00 -95.00)	153	Portugal
cabbage	152.23(0.00 -1330.00)	95.00(47.50-95.00)	199	Poland
cabbage	90.83(0.00 -522.50)	47.50(47.50-95.00)	133	UK
cabbage	81.83(0.00 -1330.00)	47.50(0.00 -95.00)	744	Sweden
cherry	74.83(0.00 -1330.00)	47.50(0.00 -47.50)	75	Belgium
cherry	42.00(0.00 -1330.00)	0.00 (0.00 -47.50)	94	Denmark
cherry	21.76(0.00 -1045.00)	0.00 (0.00 -0.00)	33	Finaland
cherry	76.75(0.00 -1330.00)	47.50(0.00 -95.00)	92	North Germany
cherry	64.88(0.00 -1330.00)	0.00 (0.00 -47.50)	90	South Germany
cherry	39.77(0.00 -1330.00)	0.00 (0.00 -47.50)	103	Holland
cherry	129.48(0.00 -1330.00)	47.50(0.00 -95.00)	157	Portugal
cherry	146.57(0.00 -1330.00)	47.50(0.00 -285.00)	141	Poland
cherry	41.94(0.00 -665.00)	0.00 (0.00 -47.50)	80	UK
cherry	19.43(0.00 -1330.00)	0.00 (0.00 -0.00)	287	Sweden
condensed milk	0.51 (0.00 -25.00)	0.00 (0.00 -0.00)	3	Belgium
condensed milk	3.04 (0.00 -350.00)	0.00 (0.00 -0.00)	7	Denmark
condensed milk	1.77 (0.00 -150.00)	0.00 (0.00 -0.00)	5	Finaland
condensed milk	85.73(0.00 -700.00)	0.00 (0.00 -25.00)	52	North Germany
condensed milk	135.31(0.00 -700.00)	0.00 (0.00 -275.00)	78	South Germany
condensed milk	4.65 (0.00 -350.00)	0.00 (0.00 -0.00)	4	Holland
condensed milk	5.02 (0.00 -50.00)	0.00 (0.00 -0.00)	49	Portugal
condensed milk	24.88(0.00 -700.00)	0.00 (0.00 -0.00)	50	Poland
condensed milk	2.78 (0.00 -150.00)	0.00 (0.00 -0.00)	12	UK
condensed milk	0.85 (0.00 -350.00)	0.00 (0.00 -0.00)	25	Sweden
couscous	24.66(0.00 -150.00)	0.00 (0.00 -75.00)	43	Belgium
couscous	6.78 (0.00 -450.00)	0.00 (0.00 -0.00)	26	Denmark
couscous	28.55(0.00 -450.00)	0.00 (0.00 -75.00)	41	Finaland
couscous	21.61(0.00 -150.00)	0.00 (0.00 -75.00)	46	North Germany
couscous	11.98(0.00 -150.00)	0.00 (0.00 -0.00)	29	South Germany
couscous	13.26(0.00 -150.00)	0.00 (0.00 -0.00)	34	Holland

couscous	29.54(0.00 -1050.00)	0.00 (0.00 -0.00)	32	Portugal
couscous	46.43(0.00 -2100.00)	0.00 (0.00 -75.00)	67	Poland
couscous	25.88(0.00 -450.00)	0.00 (0.00 -75.00)	43	UK
couscous	32.33(0.00 -450.00)	0.00 (0.00 -75.00)	392	Sweden
crisp fried Cakes	0.68 (0.00 -60.00)	0.00 (0.00 -0.00)	3	Belgium
crisp fried Cakes	0.45 (0.00 -140.00)	0.00 (0.00 -0.00)	3	Denmark
crisp fried Cakes	0.00 (0.00 -0.00)	0.00 (0.00 -0.00)	0	Finaland
crisp fried Cakes	1.13 (0.00 -60.00)	0.00 (0.00 -0.00)	14	North Germany
crisp fried Cakes	0.57 (0.00 -20.00)	0.00 (0.00 -0.00)	10	South Germany
crisp fried Cakes	0.05 (0.00 -10.00)	0.00 (0.00 -0.00)	1	Holland
crisp fried Cakes	1.24 (0.00 -20.00)	0.00 (0.00 -0.00)	27	Portugal
crisp fried Cakes	2.00 (0.00 -110.00)	0.00 (0.00 -0.00)	27	Poland
crisp fried Cakes	0.53 (0.00 -60.00)	0.00 (0.00 -0.00)	3	UK
crisp fried Cakes	0.08 (0.00 -20.00)	0.00 (0.00 -0.00)	7	Sweden
greek style yogurt	56.93(0.00 -1750.00)	0.00 (0.00 -0.00)	29	Belgium
greek style yogurt	34.07(0.00 -875.00)	0.00 (0.00 -62.50)	112	Denmark
greek style yogurt	37.50(0.00 -875.00)	0.00 (0.00 -0.00)	33	Finaland
greek style yogurt	66.74(0.00 -875.00)	0.00 (0.00 -62.50)	45	North Germany
greek style yogurt	61.21(0.00 -1750.00)	0.00 (0.00 -0.00)	44	South Germany
greek style yogurt	53.49(0.00 -875.00)	0.00 (0.00 -62.50)	56	Holland
greek style yogurt	40.06(0.00 -1750.00)	0.00 (0.00 -0.00)	25	Portugal
greek style yogurt	71.73(0.00 -1750.00)	0.00 (0.00 -62.50)	70	Poland
greek style yogurt	73.46(0.00 -875.00)	0.00 (0.00 -62.50)	71	UK
greek style yogurt	69.30(0.00 -1750.00)	0.00 (0.00 -62.50)	496	Sweden
kidney	11.71(0.00 -90.00)	0.00 (0.00 -0.00)	31	Belgium
kidney	10.55(0.00 -270.00)	0.00 (0.00 -0.00)	73	Denmark
kidney	22.94(0.00 -270.00)	0.00 (0.00 -45.00)	58	Finaland
kidney	14.24(0.00 -90.00)	0.00 (0.00 -45.00)	52	North Germany
kidney	25.52(0.00 -270.00)	0.00 (0.00 -45.00)	85	South Germany
kidney	29.72(0.00 -270.00)	45.00(0.00 -45.00)	114	Holland
kidney	81.49(0.00 -630.00)	45.00(45.00-90.00)	209	Portugal
kidney	45.64(0.00 -630.00)	45.00(0.00 -45.00)	132	Poland
kidney	32.11(0.00 -270.00)	45.00(0.00 -45.00)	86	UK
kidney	42.90(0.00 -1260.00)	22.50(0.00 -45.00)	588	Sweden
lemon	7.36 (0.00 -70.00)	0.00 (0.00 -5.00)	69	Belgium
lemon	7.58 (0.00 -70.00)	5.00 (0.00 -5.00)	188	Denmark
lemon	5.26 (0.00 -140.00)	0.00 (0.00 -5.00)	72	Finaland
lemon	8.95 (0.00 -70.00)	5.00 (0.00 -10.00)	101	North Germany
lemon	5.72 (0.00 -140.00)	0.00 (0.00 -5.00)	90	South

				Germany
lemon	5.28 (0.00 -70.00)	0.00 (0.00 -5.00)	96	Holland
lemon	8.98 (0.00 -140.00)	0.00 (0.00 -10.00)	128	Portugal
lemon	35.88(0.00 -140.00)	10.00(5.00 -55.00)	181	Poland
lemon	7.40 (0.00 -70.00)	5.00 (0.00 -10.00)	89	UK
lemon	8.09 (0.00 -140.00)	5.00 (0.00 -10.00)	761	Sweden
lentils	18.08(0.00 -120.00)	0.00 (0.00 -0.00)	35	Belgium
lentils	6.10 (0.00 -360.00)	0.00 (0.00 -0.00)	24	Denmark
lentils	11.61(0.00 -120.00)	0.00 (0.00 -0.00)	27	Finaland
lentils	34.92(0.00 -120.00)	60.00(0.00 -60.00)	96	North Germany
lentils	35.88(0.00 -360.00)	0.00 (0.00 -60.00)	96	South Germany
lentils	9.49 (0.00 -360.00)	0.00 (0.00 -0.00)	28	Holland
lentils	13.20(0.00 -360.00)	0.00 (0.00 -0.00)	40	Portugal
lentils	10.00(0.00 -360.00)	0.00 (0.00 -0.00)	18	Poland
lentils	40.70(0.00 -360.00)	0.00 (0.00 -60.00)	59	UK
lentils	43.78(0.00 -1680.00)	0.00 (0.00 -60.00)	456	Sweden
mango	10.41(0.00 -120.00)	0.00 (0.00 -20.00)	51	Belgium
mango	5.37 (0.00 -120.00)	0.00 (0.00 -0.00)	69	Denmark
mango	3.61 (0.00 -40.00)	0.00 (0.00 -0.00)	27	Finaland
mango	17.74(0.00 -280.00)	0.00 (0.00 -20.00)	70	North Germany
mango	8.25 (0.00 -280.00)	0.00 (0.00 -0.00)	44	South Germany
mango	11.16(0.00 -220.00)	0.00 (0.00 -20.00)	83	Holland
mango	34.05(0.00 -560.00)	20.00(0.00 -40.00)	132	Portugal
mango	14.57(0.00 -560.00)	0.00 (0.00 -0.00)	43	Poland
mango	19.42(0.00 -560.00)	0.00 (0.00 -20.00)	75	UK
mango	10.48(0.00 -560.00)	0.00 (0.00 -20.00)	374	Sweden
moussaka	10.27(0.00 -120.00)	0.00 (0.00 -0.00)	24	Belgium
moussaka	6.27 (0.00 -60.00)	0.00 (0.00 -0.00)	37	Denmark
moussaka	7.74 (0.00 -60.00)	0.00 (0.00 -0.00)	20	Finaland
moussaka	2.03 (0.00 -60.00)	0.00 (0.00 -0.00)	6	North Germany
moussaka	3.09 (0.00 -120.00)	0.00 (0.00 -0.00)	9	South Germany
moussaka	5.58 (0.00 -120.00)	0.00 (0.00 -0.00)	19	Holland
moussaka	4.40 (0.00 -660.00)	0.00 (0.00 -0.00)	8	Portugal
moussaka	1.43 (0.00 -60.00)	0.00 (0.00 -0.00)	5	Poland
moussaka	8.42 (0.00 -120.00)	0.00 (0.00 -0.00)	22	UK
moussaka	7.45 (0.00 -120.00)	0.00 (0.00 -0.00)	142	Sweden
okra	1.23 (0.00 -30.00)	0.00 (0.00 -0.00)	6	Belgium
okra	0.17 (0.00 -30.00)	0.00 (0.00 -0.00)	2	Denmark
okra	0.00 (0.00 -0.00)	0.00 (0.00 -0.00)	0	Finaland
okra	0.00 (0.00 -0.00)	0.00 (0.00 -0.00)	0	North

				Germany
okra	0.62 (0.00 -30.00)	0.00 (0.00 -0.00)	4	South Germany
okra	1.12 (0.00 -30.00)	0.00 (0.00 -0.00)	8	Holland
okra	3.71 (0.00 -330.00)	0.00 (0.00 -0.00)	14	Portugal
okra	0.43 (0.00 -60.00)	0.00 (0.00 -0.00)	2	Poland
okra	1.93 (0.00 -60.00)	0.00 (0.00 -0.00)	10	UK
okra	1.40 (0.00 -330.00)	0.00 (0.00 -0.00)	26	Sweden
other game	10.19(0.00 -175.00)	0.00 (0.00 -0.00)	16	Belgium
other game	6.43 (0.00 -175.00)	0.00 (0.00 -0.00)	25	Denmark
other game	27.10(0.00 -525.00)	0.00 (0.00 -0.00)	32	Finaland
other game	10.88(0.00 -175.00)	0.00 (0.00 -0.00)	20	North Germany
other game	7.67 (0.00 -87.50)	0.00 (0.00 -0.00)	17	South Germany
other game	9.36 (0.00 -87.50)	0.00 (0.00 -0.00)	23	Holland
other game	5.74 (0.00 -525.00)	0.00 (0.00 -0.00)	10	Portugal
other game	3.75 (0.00 -175.00)	0.00 (0.00 -0.00)	7	Poland
other game	2.05 (0.00 -87.50)	0.00 (0.00 -0.00)	4	UK
other game	46.65(0.00 -1225.00)	0.00 (0.00 -87.50)	370	Sweden
peach	105.48(0.00 -770.00)	55.00(0.00 -110.00)	97	Belgium
peach	60.13(0.00 -770.00)	55.00(0.00 -55.00)	182	Denmark
peach	42.94(0.00 -605.00)	0.00 (0.00 -55.00)	74	Finaland
peach	87.32(0.00 -770.00)	55.00(0.00 -110.00)	104	North Germany
peach	119.92(0.00 -1540.00)	55.00(0.00 -110.00)	134	South Germany
peach	59.09(0.00 -1540.00)	55.00(0.00 -55.00)	113	Holland
peach	337.64(0.00 -1540.00)	110.00(55.00-605.00)	228	Portugal
peach	144.83(0.00 -1540.00)	55.00(55.00-110.00)	163	Poland
peach	98.42(0.00 -770.00)	55.00(0.00 -110.00)	108	UK
peach	78.52(0.00 -1540.00)	55.00(0.00 -55.00)	603	Sweden
peanuts	3.08 (0.00 -30.00)	0.00 (0.00 -5.00)	57	Belgium
peanuts	3.21 (0.00 -70.00)	0.00 (0.00 -5.00)	122	Denmark
peanuts	3.35 (0.00 -70.00)	0.00 (0.00 -5.00)	54	Finaland
peanuts	2.85 (0.00 -30.00)	0.00 (0.00 -5.00)	74	North Germany
peanuts	5.28 (0.00 -140.00)	5.00 (0.00 -5.00)	100	South Germany
peanuts	8.49 (0.00 -70.00)	5.00 (0.00 -10.00)	140	Holland
peanuts	4.54 (0.00 -55.00)	5.00 (0.00 -5.00)	133	Portugal
peanuts	4.74 (0.00 -140.00)	5.00 (0.00 -5.00)	106	Poland
peanuts	6.35 (0.00 -70.00)	5.00 (0.00 -5.00)	86	UK
peanuts	5.37 (0.00 -140.00)	5.00 (0.00 -5.00)	639	Sweden
pumpkin	43.60(0.00 -665.00)	47.50(0.00 -47.50)	85	Belgium
pumpkin	9.80 (0.00 -285.00)	0.00 (0.00 -0.00)	47	Denmark

pumpkin	21.45(0.00 -522.50)	0.00 (0.00 -0.00)	37	Finaland
pumpkin	16.10(0.00 -285.00)	0.00 (0.00 -0.00)	43	North Germany
pumpkin	9.79 (0.00 -522.50)	0.00 (0.00 -0.00)	25	South Germany
pumpkin	14.80(0.00 -285.00)	0.00 (0.00 -0.00)	50	Holland
pumpkin	107.47(0.00 -1330.00)	47.50(0.00 -95.00)	156	Portugal
pumpkin	26.01(0.00 -1330.00)	0.00 (0.00 -0.00)	47	Poland
pumpkin	10.83(0.00 -285.00)	0.00 (0.00 -0.00)	29	UK
pumpkin	4.24 (0.00 -665.00)	0.00 (0.00 -0.00)	72	Sweden
rhubarb	28.77(0.00 -980.00)	0.00 (0.00 -70.00)	38	Belgium
rhubarb	27.88(0.00 -420.00)	0.00 (0.00 -70.00)	103	Denmark
rhubarb	33.87(0.00 -770.00)	0.00 (0.00 -70.00)	51	Finaland
rhubarb	41.92(0.00 -980.00)	0.00 (0.00 -70.00)	53	North Germany
rhubarb	39.33(0.00 -980.00)	0.00 (0.00 -0.00)	41	South Germany
rhubarb	20.84(0.00 -420.00)	0.00 (0.00 -0.00)	52	Holland
rhubarb	1.89 (0.00 -140.00)	0.00 (0.00 -0.00)	5	Portugal
rhubarb	32.67(0.00 -980.00)	0.00 (0.00 -0.00)	44	Poland
rhubarb	35.61(0.00 -770.00)	0.00 (0.00 -70.00)	64	UK
rhubarb	31.85(0.00 -980.00)	0.00 (0.00 -70.00)	375	Sweden
roast beef steak	156.99(0.00 -840.00)	60.00(0.00 -120.00)	109	Belgium
roast beef steak	93.73(0.00 -840.00)	60.00(60.00-120.00)	289	Denmark
roast beef steak	110.32(0.00 -360.00)	60.00(0.00 -120.00)	114	Finaland
roast beef steak	44.75(0.00 -360.00)	0.00 (0.00 -60.00)	87	North Germany
roast beef steak	71.13(0.00 -840.00)	60.00(0.00 -60.00)	123	South Germany
roast beef steak	237.77(0.00 -840.00)	120.00(60.00-360.00)	197	Holland
roast beef steak	259.00(0.00 -1680.00)	120.00(60.00-360.00)	221	Portugal
roast beef steak	158.86(0.00 -1680.00)	60.00(60.00-120.00)	176	Poland
roast beef steak	121.40(0.00 -840.00)	60.00(60.00-120.00)	134	UK
roast beef steak	95.00(0.00 -840.00)	60.00(60.00-120.00)	892	Sweden
single cream	7.14 (0.00 -105.00)	7.50 (0.00 -7.50)	80	Belgium
single cream	6.86 (0.00 -210.00)	0.00 (0.00 -0.00)	40	Denmark
single cream	2.03 (0.00 -45.00)	0.00 (0.00 -0.00)	27	Finaland
single cream	5.81 (0.00 -82.50)	7.50 (0.00 -7.50)	92	North Germany
single cream	6.57 (0.00 -45.00)	7.50 (0.00 -7.50)	116	South Germany
single cream	6.49 (0.00 -210.00)	0.00 (0.00 -7.50)	99	Holland
single cream	2.58 (0.00 -45.00)	0.00 (0.00 -7.50)	70	Portugal
single cream	16.07(0.00 -210.00)	7.50 (0.00 -15.00)	142	Poland
single cream	3.11 (0.00 -45.00)	0.00 (0.00 -7.50)	49	UK
single cream	2.17 (0.00 -210.00)	0.00 (0.00 -0.00)	98	Sweden

smoked fatty fish	30.21(0.00 -210.00)	35.00(0.00 -35.00)	102	Belgium
smoked fatty fish	23.53(0.00 -385.00)	0.00 (0.00 -35.00)	160	Denmark
smoked fatty fish	30.94(0.00 -385.00)	35.00(0.00 -35.00)	94	Finaland
smoked fatty fish	16.41(0.00 -70.00)	0.00 (0.00 -35.00)	72	North Germany
smoked fatty fish	17.32(0.00 -490.00)	0.00 (0.00 -35.00)	76	South Germany
smoked fatty fish	25.56(0.00 -210.00)	35.00(0.00 -35.00)	123	Holland
smoked fatty fish	25.81(0.00 -385.00)	0.00 (0.00 -35.00)	105	Portugal
smoked fatty fish	32.67(0.00 -385.00)	35.00(0.00 -35.00)	130	Poland
smoked fatty fish	20.06(0.00 -210.00)	0.00 (0.00 -35.00)	74	UK
smoked fatty fish	32.17(0.00 -490.00)	35.00(0.00 -35.00)	733	Sweden
smoked poultry	21.34(0.00 -490.00)	0.00 (0.00 -35.00)	47	Belgium
smoked poultry	4.94 (0.00 -490.00)	0.00 (0.00 -0.00)	34	Denmark
smoked poultry	15.13(0.00 -980.00)	0.00 (0.00 -0.00)	16	Finaland
smoked poultry	3.36 (0.00 -35.00)	0.00 (0.00 -0.00)	17	North Germany
smoked poultry	7.58 (0.00 -210.00)	0.00 (0.00 -0.00)	24	South Germany
smoked poultry	3.09 (0.00 -70.00)	0.00 (0.00 -0.00)	16	Holland
smoked poultry	9.19 (0.00 -210.00)	0.00 (0.00 -0.00)	35	Portugal
smoked poultry	12.33(0.00 -210.00)	0.00 (0.00 -35.00)	55	Poland
smoked poultry	4.09 (0.00 -210.00)	0.00 (0.00 -0.00)	13	UK
smoked poultry	23.21(0.00 -980.00)	0.00 (0.00 -35.00)	318	Sweden
sour cream	0.87 (0.00 -15.00)	0.00 (0.00 -0.00)	15	Belgium
sour cream	1.74 (0.00 -45.00)	0.00 (0.00 -0.00)	63	Denmark
sour cream	3.39 (0.00 -45.00)	0.00 (0.00 -7.50)	62	Finaland
sour cream	4.66 (0.00 -82.50)	0.00 (0.00 -7.50)	70	North Germany
sour cream	4.21 (0.00 -105.00)	0.00 (0.00 -7.50)	71	South Germany
sour cream	4.01 (0.00 -45.00)	0.00 (0.00 -7.50)	80	Holland
sour cream	0.14 (0.00 -7.50)	0.00 (0.00 -0.00)	5	Portugal
sour cream	11.64(0.00 -105.00)	7.50 (0.00 -15.00)	116	Poland
sour cream	1.40 (0.00 -15.00)	0.00 (0.00 -0.00)	30	UK
sour cream	7.32 (0.00 -105.00)	7.50 (0.00 -7.50)	730	Sweden
thin biscuits	2.18 (0.00 -28.00)	0.00 (0.00 -2.00)	50	Belgium
thin biscuits	2.14 (0.00 -28.00)	0.00 (0.00 -2.00)	152	Denmark
thin biscuits	3.94 (0.00 -28.00)	2.00 (0.00 -4.00)	78	Finaland
thin biscuits	2.10 (0.00 -28.00)	0.00 (0.00 -2.00)	81	North Germany
thin biscuits	2.08 (0.00 -28.00)	0.00 (0.00 -2.00)	90	South Germany
thin biscuits	5.17 (0.00 -28.00)	2.00 (0.00 -4.00)	137	Holland
thin biscuits	2.41 (0.00 -56.00)	0.00 (0.00 -2.00)	93	Portugal
thin biscuits	2.00 (0.00 -56.00)	0.00 (0.00 -2.00)	99	Poland

thin biscuits	3.30 (0.00 -28.00)	2.00 (0.00 -4.00)	97	UK
thin biscuits	8.37 (0.00 -56.00)	2.00 (2.00 -12.00)	898	Sweden
tofu	10.68(0.00 -560.00)	0.00 (0.00 -0.00)	9	Belgium
tofu	1.24 (0.00 -240.00)	0.00 (0.00 -0.00)	5	Denmark
tofu	16.77(0.00 -1120.00)	0.00 (0.00 -0.00)	18	Finaland
tofu	2.03 (0.00 -80.00)	0.00 (0.00 -0.00)	7	North Germany
tofu	2.27 (0.00 -240.00)	0.00 (0.00 -0.00)	5	South Germany
tofu	8.56 (0.00 -240.00)	0.00 (0.00 -0.00)	27	Holland
tofu	4.32 (0.00 -560.00)	0.00 (0.00 -0.00)	12	Portugal
tofu	3.24 (0.00 -240.00)	0.00 (0.00 -0.00)	10	Poland
tofu	4.44 (0.00 -240.00)	0.00 (0.00 -0.00)	13	UK
tofu	4.73 (0.00 -1120.00)	0.00 (0.00 -0.00)	79	Sweden
turnip	9.04 (0.00 -165.00)	0.00 (0.00 -27.50)	38	Belgium
turnip	4.97 (0.00 -385.00)	0.00 (0.00 -0.00)	36	Denmark
turnip	7.10 (0.00 -55.00)	0.00 (0.00 -0.00)	38	Finaland
turnip	5.59 (0.00 -165.00)	0.00 (0.00 -0.00)	31	North Germany
turnip	2.84 (0.00 -55.00)	0.00 (0.00 -0.00)	18	South Germany
turnip	5.50 (0.00 -302.50)	0.00 (0.00 -0.00)	31	Holland
turnip	66.47(0.00 -770.00)	27.50(0.00 -55.00)	151	Portugal
turnip	9.43 (0.00 -770.00)	0.00 (0.00 -0.00)	29	Poland
turnip	40.69(0.00 -302.50)	27.50(0.00 -55.00)	107	UK
turnip	14.29(0.00 -385.00)	0.00 (0.00 -27.50)	388	Sweden
vegetable oil	25.84(0.00 -154.00)	5.50 (0.00 -33.00)	93	Belgium
vegetable oil	31.26(0.00 -154.00)	33.00(5.50 -60.50)	282	Denmark
vegetable oil	21.25(0.00 -77.00)	5.50 (0.00 -33.00)	104	Finaland
vegetable oil	32.44(0.00 -154.00)	33.00(5.50 -60.50)	149	North Germany
vegetable oil	42.72(0.00 -154.00)	33.00(11.00-60.50)	179	South Germany
vegetable oil	31.13(0.00 -154.00)	11.00(0.00 -60.50)	153	Holland
vegetable oil	15.40(0.00 -154.00)	5.50 (0.00 -11.00)	149	Portugal
vegetable oil	30.75(0.00 -154.00)	33.00(5.50 -60.50)	176	Poland
vegetable oil	16.79(0.00 -154.00)	5.50 (0.00 -33.00)	99	UK
vegetable oil	37.24(0.00 -154.00)	33.00(5.50 -60.50)	1001	Sweden
viili like fermented yoghurt	0.86 (0.00 -125.00)	0.00 (0.00 -0.00)	1	Belgium
viili like fermented yoghurt	3.88 (0.00 -875.00)	0.00 (0.00 -0.00)	8	Denmark
viili like fermented yoghurt	89.11(0.00 -875.00)	0.00 (0.00 -62.50)	67	Finaland
viili like fermented yoghurt	9.53 (0.00 -687.50)	0.00 (0.00 -0.00)	15	North Germany
viili like fermented	18.36(0.00 -875.00)	0.00 (0.00 -0.00)	29	South

yoghurt				Germany
viili like fermented yoghurt	18.90(0.00 -1750.00)	0.00 (0.00 -0.00)	9	Holland
viili like fermented yoghurt	285.23(0.00 -1750.00)	62.50(0.00 -375.00)	147	Portugal
viili like fermented yoghurt	33.93(0.00 -1750.00)	0.00 (0.00 -0.00)	24	Poland
viili like fermented yoghurt	7.68 (0.00 -875.00)	0.00 (0.00 -0.00)	3	UK
viili like fermented yoghurt	1.33 (0.00 -375.00)	0.00 (0.00 -0.00)	16	Sweden

XII. Associations between dietary patterns and respiratory outcomes by smoking status, body mass index and gender: results of meta-analyses. OR; odds ratio.

respiratory outcomes	Dietary pattenr	odds ratio (95%CI)	isquare (p-value for heterogeneity)	Strata
asthma	fruit, fish and vegetables	1.12(0.92,1.35)	43.84(0.066)	never smoker
asthma	meat, potatoes and sweets	0.99(0.88,1.11)	0.00(0.758)	never smoker
chronic sinusitis	fruit, fish and vegetables	1.07(0.89,1.27)	6.47(0.382)	never smoker
chronic sinusitis	meat, potatoes and sweets	1.03(0.85,1.25)	24.68(0.216)	never smoker
allergic rhinitis	fruit, fish and vegetables	0.99(0.79,1.25)	63.24(0.004)	never smoker
allergic rhinitis	meat, potatoes and sweets	0.99(0.85,1.14)	25.21(0.211)	never smoker
eczema	fruit, fish and vegetables	1.13(0.92,1.39)	57.10(0.013)	never smoker
eczema	meat, potatoes and sweets	0.93(0.78,1.11)	40.97(0.084)	never smoker
asthma	fruit, fish and vegetables	1.15(0.95,1.38)	11.64(0.336)	ever smoker
asthma	meat, potatoes and sweets	0.88(0.71,1.07)	15.74(0.298)	ever smoker
chronic sinusitis	fruit, fish and vegetables	1.19(0.92,1.54)	37.74(0.107)	ever smoker
chronic sinusitis	meat, potatoes and sweets	1.21(0.93,1.57)	36.15(0.119)	ever smoker
allergic rhinitis	fruit, fish and vegetables	0.96(0.76,1.22)	36.62(0.115)	ever smoker
allergic rhinitis	**meat, potatoes and sweets**	**1.28(1.07,1.55)**	**15.31(0.302)**	**ever smoker**
eczema	fruit, fish and vegetables	1.35(0.99,1.83)	64.16(0.003)	ever smoker
eczema	meat, potatoes and sweets	1.07(0.83,1.38)	55.56(0.016)	ever smoker
asthma	fruit, fish and vegetables	0.93(0.56,1.55)	73.33(0.000)	current smoker
asthma	meat, potatoes and sweets	1.00(0.67,1.47)	46.01(0.054)	current smoker
chronic sinusitis	fruit, fish and vegetables	0.85(0.66,1.08)	10.12(0.351)	current smoker
chronic sinusitis	meat, potatoes and sweets	1.40(0.95,2.06)	44.31(0.064)	current smoker

allergic rhinitis	fruit, fish and vegetables	1.12(0.80,1.55)	59.87(0.008)	current smoker
allergic rhinitis	meat, potatoes and sweets	1.20(0.96,1.51)	0.00(0.536)	current smoker
eczema	fruit, fish and vegetables	1.01(0.73,1.38)	40.60(0.087)	current smoker
eczema	**meat, potatoes and sweets**	**1.31(1.04,1.65)**	**0.00(0.852)**	**current smoker**
asthma	fruit, fish and vegetables	1.10(0.94,1.27)	7.85(0.370)	female
asthma	meat, potatoes and sweets	1.02(0.87,1.19)	0.00(0.736)	female
chronic sinusitis	fruit, fish and vegetables	0.97(0.76,1.25)	43.69(0.067)	female
chronic sinusitis	meat, potatoes and sweets	1.13(0.88,1.45)	48.51(0.042)	female
allergic rhinitis	fruit, fish and vegetables	1.11(0.84,1.47)	65.22(0.002)	female
allergic rhinitis	meat, potatoes and sweets	1.00(0.87,1.15)	0.00(0.471)	female
eczema	fruit, fish and vegetables	1.06(0.85,1.33)	54.14(0.020)	female
eczema	meat, potatoes and sweets	0.89(0.73,1.08)	42.20(0.076)	female
asthma	fruit, fish and vegetables	1.10(0.95,1.26)	22.41(0.237)	male
asthma	meat, potatoes and sweets	0.90(0.81,1.01)	0.00(0.568)	male
chronic sinusitis	fruit, fish and vegetables	1.08(0.95,1.23)	0.00(0.575)	male
chronic sinusitis	meat, potatoes and sweets	1.08(0.95,1.22)	0.00(0.468)	male
allergic rhinitis	fruit, fish and vegetables	0.87(0.69,1.08)	68.01(0.001)	male
allergic rhinitis	meat, potatoes and sweets	1.10(0.98,1.24)	8.37(0.365)	male
eczema	fruit, fish and vegetables	1.17(0.97,1.42)	59.37(0.008)	male
eczema	meat, potatoes and sweets	1.09(0.93,1.27)	41.14(0.083)	male
asthma	fruit, fish and vegetables	1.12(0.97,1.31)	5.87(0.387)	non-obese
asthma	meat, potatoes and sweets	0.90(0.75,1.08)	26.80(0.197)	non-obese
chronic sinusitis	fruit, fish and vegetables	1.13(0.97,1.32)	6.58(0.381)	non-obese
chronic sinusitis	meat, potatoes and sweets	1.00(0.80,1.25)	48.02(0.044)	non-obese
allergic rhinitis	fruit, fish and vegetables	0.92(0.67,1.26)	75.87(0.000)	non-obese

allergic rhinitis	meat, potatoes and sweets	1.04(0.79,1.38)	68.90(0.001)	non-obese
eczema	fruit, fish and vegetables	1.24(1.00,1.55)	45.12(0.059)	non-obese
eczema	meat, potatoes and sweets	0.94(0.75,1.18)	56.51(0.014)	non-obese
asthma	fruit, fish and vegetables	1.05(0.94,1.17)	0.00(0.812)	obese
asthma	meat, potatoes and sweets	1.05(0.92,1.19)	0.00(0.771)	obese
chronic sinusitis	fruit, fish and vegetables	1.04(0.82,1.31)	50.90(0.032)	obese
chronic sinusitis	meat, potatoes and sweets	1.05(0.83,1.32)	45.71(0.056)	obese
allergic rhinitis	fruit, fish and vegetables	0.97(0.84,1.12)	21.95(0.241)	obese
allergic rhinitis	meat, potatoes and sweets	1.09(0.97,1.23)	0.00(0.480)	obese
eczema	fruit, fish and vegetables	1.04(0.87,1.24)	46.01(0.054)	obese
eczema	meat, potatoes and sweets	1.01(0.88,1.15)	15.51(0.300)	obese

XIII. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and respiratory outcomes results of random intercept models; stratified by smoking status.

strata	Food item	Respiratory outcome	odds ratio	p-value
non smoker	kindey and black beans	asthma	0.96	0.562
ever smoker	kindey and black beans	asthma	0.63	0.007
current smoker	kindey and black beans	asthma	0.91	0.060
non smoker	lemon	asthma	0.78	0.017
ever smoker	lemon	asthma	0.87	0.177
current smoker	lemon	asthma	0.85	0.017
non smoker	**smoked fatty fish**	**asthma**	**0.89**	**0.021**
ever smoker	**smoked fatty fish**	**asthma**	**0.77**	**0.004**
current smoker	**smoked fatty fish**	**asthma**	**0.84**	**0.018**
non smoker	cherry	asthma	0.85	0.000
ever smoker	cherry	asthma	0.84	0.084
current smoker	cherry	asthma	0.88	0.198
non smoker	any legumes	asthma	1.00	0.911
ever smoker	any legumes	asthma	0.74	0.021
current smoker	any legumes	asthma	0.89	0.366
non smoker	**condensed milk**	**asthma**	**0.85**	**0.014**
ever smoker	**condensed milk**	**asthma**	**0.92**	**0.016**
current smoker	**condensed milk**	**asthma**	**0.88**	**0.019**
non smoker	thin biscuits	asthma	1.10	0.000
ever smoker	thin biscuits	asthma	1.11	0.047
current smoker	thin biscuits	asthma	1.03	0.670
non smoker	turnip	asthma	1.14	0.069
ever smoker	turnip	asthma	1.19	0.087
current smoker	turnip	asthma	1.01	0.947
non smoker	couscous	asthma	0.97	0.568
ever smoker	couscous	asthma	1.31	0.000
current smoker	couscous	asthma	0.96	0.849
non smoker	cabbage	chronic sinusitis	0.91	0.026
ever smoker	cabbage	chronic sinusitis	0.82	0.151
current smoker	cabbage	chronic sinusitis	0.85	0.213
non smoker	viili or fermented yogurt	chronic sinusitis	1.21	0.063
ever smoker	viili or fermented yogurt	chronic sinusitis	1.20	0.010
current smoker	viili or fermented yogurt	chronic sinusitis	1.12	0.208
non smoker	okra	chronic sinusitis	1.34	0.000
ever smoker	okra	chronic sinusitis	1.16	0.024
current smoker	okra	chronic sinusitis	1.06	0.243
non smoker	pumpkin	chronic sinusitis	1.20	0.001

ever smoker	pumpkin	chronic sinusitis	1.21	0.036
current smoker	pumpkin	chronic sinusitis	1.02	0.928
non smoker	**peach**	**allergic rhinitis**	**0.87**	**0.055**
ever smoker	**peach**	**allergic rhinitis**	**0.81**	**0.001**
current smoker	**peach**	**allergic rhinitis**	**0.60**	**0.000**
non smoker	peanuts	allergic rhinitis	0.77	0.338
ever smoker	peanuts	allergic rhinitis	0.97	0.682
current smoker	peanuts	allergic rhinitis	0.95	0.574
non smoker	tofu	allergic rhinitis	0.80	0.013
ever smoker	tofu	allergic rhinitis	0.99	0.805
current smoker	tofu	allergic rhinitis	1.15	0.233
non smoker	mango	allergic rhinitis	1.05	0.472
ever smoker	mango	allergic rhinitis	1.07	0.450
current smoker	mango	allergic rhinitis	1.35	0.049
non smoker	vegetable oil	allergic rhinitis	1.21	0.002
ever smoker	vegetable oil	allergic rhinitis	1.10	0.144
current smoker	vegetable oil	allergic rhinitis	1.14	0.149
non smoker	smoked poultry	allergic rhinitis	1.17	0.000
ever smoker	smoked poultry	allergic rhinitis	0.91	0.078
current smoker	smoked poultry	allergic rhinitis	1.17	0.000
non smoker	single cream	allergic rhinitis	1.28	0.000
ever smoker	single cream	allergic rhinitis	1.22	0.201
current smoker	single cream	allergic rhinitis	1.31	0.139
non smoker	**rhubarb**	**eczema**	**0.86**	**0.005**
ever smoker	**rhubarb**	**eczema**	**0.85**	**0.008**
current smoker	**rhubarb**	**eczema**	**0.69**	**0.045**
non smoker	crisp fried Cakes	eczema	0.86	0.105
ever smoker	crisp fried Cakes	eczema	0.82	0.024
current smoker	crisp fried Cakes	eczema	0.76	0.060
non smoker	bitter mellon	eczema	0.78	0.001
ever smoker	bitter mellon	eczema	0.95	0.564
current smoker	bitter mellon	eczema	0.96	0.563
non smoker	greek style yogurt	eczema	0.79	0.003
ever smoker	greek style yogurt	eczema	0.85	0.000
current smoker	greek style yogurt	eczema	1.26	0.212
non smoker	other game	eczema	0.96	0.025
ever smoker	other game	eczema	1.12	0.184
current smoker	other game	eczema	0.76	0.005
non smoker	**lentils**	**eczema**	**1.06**	**0.049**
ever smoker	**lentils**	**eczema**	**0.92**	**0.000**
current smoker	**lentils**	**eczema**	**0.76**	**0.000**
non smoker	sour cream	eczema	1.22	0.019
ever smoker	sour cream	eczema	1.10	0.080
current smoker	sour cream	eczema	0.83	0.292

non smoker	crisp fried Cakes	atopy	0.63	0.006
ever smoker	crisp fried Cakes	atopy	1.01	0.858
current smoker	crisp fried Cakes	atopy	0.92	0.278
non smoker	moussaka	atopy	0.92	0.212
ever smoker	moussaka	atopy	0.97	0.789
current smoker	moussaka	atopy	0.53	0.000
non smoker	beer	atopy	1.16	0.148
ever smoker	beer	atopy	1.08	0.388
current smoker	beer	atopy	1.14	0.119
non smoker	**thin biscuits**	**atopy**	**1.23**	**0.000**
ever smoker	**thin biscuits**	**atopy**	**0.89**	**0.021**
current smoker	**thin biscuits**	**atopy**	**1.19**	**0.001**
non smoker	roast beef steak	atopy	1.24	0.011
ever smoker	roast beef steak	atopy	1.11	0.174
current smoker	roast beef steak	atopy	1.07	0.350

XIV. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and respiratory outcomes; results of random intercept models; stratified by gender.

strata	food item	respiratory outcome	odds ratio	pvalue
male	**kidney and black beans**	**asthma**	**0.85**	**0.003**
female	**kidney and black beans**	**asthma**	**0.79**	**0.004**
male	lemon	asthma	0.87	0.068
female	lemon	asthma	0.84	0.000
male	**smoked fatty fish**	**asthma**	**0.82**	**0.000**
female	**smoked fatty fish**	**asthma**	**0.87**	**0.000**
male	cherry	asthma	0.71	0.160
female	cherry	asthma	0.89	0.000
male	any legumes	asthma	0.80	0.000
female	any legumes	asthma	0.95	0.096
male	**condensed milk**	**asthma**	**0.80**	**0.024**
female	**condensed milk**	**asthma**	**0.93**	**0.026**
male	thin biscuits	asthma	1.20	0.000
female	thin biscuits	asthma	1.01	0.747
male	turnsip	asthma	1.09	0.380
female	turnsip	asthma	1.15	0.056
male	couscous	asthma	1.19	0.000
female	couscous	asthma	1.12	0.128
male	**cabbage**	**chronic sinusitis**	**0.69**	**0.022**
female	**cabbage**	**chronic sinusitis**	**0.90**	**0.051**
male	viili and fermented yogurt	chronic sinusitis	0.94	0.476
female	viili and fermented yogurt	chronic sinusitis	1.30	0.000
male	okra	chronic sinusitis	1.01	0.756
female	okra	chronic sinusitis	1.19	0.154
male	pumpkin	chronic sinusitis	1.29	0.063
female	pumpkin	chronic sinusitis	1.16	0.000
male	peach	allergic rhinitis	0.87	0.183
female	peach	allergic rhinitis	0.81	0.001
male	peanuts	allergic rhinitis	0.90	0.079
female	peanuts	allergic rhinitis	0.91	0.390
male	**tofu**	**allergic rhinitis**	**0.91**	**0.023**
female	**tofu**	**allergic rhinitis**	**0.81**	**0.016**
male	mango	allergic rhinitis	1.15	0.391
female	mango	allergic rhinitis	1.10	0.002
male	vegetable oil	allergic rhinitis	1.22	0.000
female	vegetable oil	allergic rhinitis	1.09	0.134

male	smoked poultry	allergic rhinitis	1.08	0.038
female	smoked poultry	allergic rhinitis	1.12	0.000
male	single cream	allergic rhinitis	1.15	0.201
female	single cream	allergic rhinitis	1.46	0.013
male	rhubarb	eczema	0.86	0.127
female	rhubarb	eczema	0.84	0.000
male	**crisp fried Cakes**	**eczema**	**0.81**	**0.036**
female	**crisp fried Cakes**	**eczema**	**0.86**	**0.040**
male	bitter mellon	eczema	0.86	0.176
female	bitter mellon	eczema	0.90	0.004
male	greek style yogurt	eczema	1.03	0.802
female	greek style yogurt	eczema	0.82	0.001
male	other game	eczema	1.07	0.115
female	other game	eczema	0.88	0.000
male	lentils	eczema	0.89	0.000
female	lentils	eczema	1.01	0.495
male	sour cream	eczema	1.07	0.171
female	sour cream	eczema	1.20	0.004
male	crisp fried Cakes	atopy	0.97	0.619
female	crisp fried Cakes	atopy	0.72	0.031
male	**moussaka**	**atopy**	**0.89**	**0.004**
female	**moussaka**	**atopy**	**0.88**	**0.006**
male	**beer**	**atopy**	**1.09**	**0.002**
female	**beer**	**atopy**	**1.32**	**0.038**
male	**thin biscuits**	**atopy**	**1.08**	**0.051**
female	**thin biscuits**	**atopy**	**1.14**	**0.001**
male	**roast beef steak**	**atopy**	**1.16**	**0.025**
female	**roast beef steak**	**atopy**	**1.15**	**0.034**

XV. Associations between specific food items that were declared significant from adjusted ESFA (FDR =5%) and respiratory outcomes; results of random intercept models; stratified by body mass index.

strata	respiratory outcome	food item	odds ratio	pvalue
non-obese	**asthma**	**kidney and black beans**	**0.79**	**0.000**
obese	**asthma**	**kidney and black beans**	**0.83**	**0.000**
non-obese	**asthma**	**lemon**	**0.84**	**0.046**
obese	**asthma**	**lemon**	**0.85**	**0.000**
non-obese	asthma	smoked fatty fish	0.98	0.781
obese	asthma	smoked fatty fish	0.78	0.000
non-obese	asthma	cherry	0.86	0.314
obese	asthma	cherry	0.83	0.004
non-obese	asthma	any legumes	0.88	0.251
obese	asthma	any legumes	0.85	0.012
non-obese	asthma	condensed milk	0.97	0.725
obese	asthma	condensed milk	0.85	0.071
non-obese	**asthma**	**thin biscuits**	**1.05**	**0.003**
obese	**asthma**	**thin biscuits**	**1.16**	**0.000**
non-obese	asthma	turnsip	1.07	0.354
obese	asthma	turnsip	1.20	0.083
non-obese	asthma	couscous	1.08	0.446
obese	asthma	couscous	1.20	0.000
non-obese	chronic sinusitis	cabbage	0.86	0.517
obese	chronic sinusitis	cabbage	0.90	0.308
non-obese	chronic sinusitis	viili and fermented yogurt	1.37	0.000
obese	chronic sinusitis	viili and fermented yogurt	1.01	0.871
non-obese	chronic sinusitis	okra	1.21	0.001
obese	chronic sinusitis	okra	1.06	0.096
non-obese	**chronic sinusitis**	**pumpkin**	**1.16**	**0.000**
obese	**chronic sinusitis**	**pumpkin**	**1.19**	**0.006**
non-obese	allergic rhinitis	peach	0.69	0.038
obese	allergic rhinitis	peach	0.93	0.455
non-obese	allergic rhinitis	peanuts	0.94	0.259
obese	allergic rhinitis	peanuts	0.89	0.001
non-obese	allergic rhinitis	tofu	0.96	0.441
obese	allergic rhinitis	tofu	0.79	0.302
non-obese	allergic rhinitis	mango	1.27	0.198
obese	allergic rhinitis	mango	0.99	0.950
non-obese	allergic rhinitis	vegetable oil	1.19	0.001
obese	allergic rhinitis	vegetable oil	1.12	0.290
non-obese	allergic rhinitis	smoked poultry	0.97	0.441
obese	allergic rhinitis	smoked poultry	1.24	0.019
non-obese	allergic rhinitis	single cream	1.22	0.150

obese	allergic rhinitis	single cream	1.32	0.000
non-obese	eczema	rhubarb	0.69	0.000
obese	eczema	rhubarb	0.94	0.270
non-obese	eczema	crisp fried Cakes	0.84	0.003
obese	eczema	crisp fried Cakes	0.86	0.108
non-obese	**eczema**	**bitter mellon**	**0.88**	**0.048**
obese	**eczema**	**bitter mellon**	**0.89**	**0.000**
non-obese	eczema	greek style yogurt	0.89	0.041
obese	eczema	greek style yogurt	0.90	0.077
non-obese	eczema	other game	0.89	0.001
obese	eczema	other game	1.05	0.076
non-obese	**eczema**	**lentils**	**0.93**	**0.000**
obese	**eczema**	**lentils**	**0.95**	**0.047**
non-obese	eczema	sour cream	1.09	0.161
obese	eczema	sour cream	1.18	0.000
non-obese	atopy	crisp fried Cakes	0.76	0.027
obese	atopy	crisp fried Cakes	0.93	0.209
non-obese	atopy	moussaka	0.85	0.026
obese	atopy	moussaka	0.91	0.074
non-obese	atopy	beer	1.09	0.116
obese	atopy	beer	1.12	0.084
non-obese	atopy	thin biscuits	1.31	0.000
obese	atopy	thin biscuits	0.96	0.290
non-obese	atopy	roast beef steak	1.26	0.010
obese	atopy	roast beef steak	1.10	0.165