



Advances in dietary pattern analysis in nutritional epidemiology

Christina-Alexandra Schulz¹ · Kolade Oluwagbemigun¹ · Ute Nöthlings¹

Received: 13 August 2020 / Accepted: 22 March 2021 / Published online: 25 April 2021
© The Author(s) 2021

Abstract

Background and Purpose It used to be a common practice in the field of nutritional epidemiology to analyze separate nutrients, foods, or food groups. However, in reality, nutrients and foods are consumed in combination. The introduction of dietary patterns (DP) and their analysis has revolutionized this field, making it possible to take into account the synergistic effects of foods and to account for the complex interaction among nutrients and foods. Three approaches of DP analysis exist: (1) the hypothesis-based approach (based on prior knowledge regarding the current understanding of dietary components and their health relation), (2) the exploratory approach (solely relying on dietary intake data), and (3) the hybrid approach (a combination of both approaches). During the recent past, complementary approaches for DP analysis have emerged both conceptually and methodologically.

Method We have summarized the recent developments that include incorporating the Treelet transformation method as a complementary exploratory approach in a narrative review.

Results Uses, peculiarities, strengths, limitations, and scope of recent developments in DP analysis are outlined. Next, the narrative review gives an overview of the literature that takes into account potential relevant dietary-related factors, specifically the metabolome and the gut microbiome in DP analysis. Then the review deals with the aspect of data processing that is needed prior to DP analysis, particularly when dietary data arise from assessment methods other than the long-established food frequency questionnaire. Lastly, potential opportunities for upcoming DP analysis are summarized in the outlook.

Conclusion Biological factors like the metabolome and the microbiome are crucial to understand diet-disease relationships. Therefore, the inclusion of these factors in DP analysis might provide deeper insights.

Keywords Dietary pattern · Hypothesis based · Exploratory · Gut microbiome · Metabolome

Introduction

Dietary pattern (DP) analysis has long been established in nutritional epidemiological research [1, 2]. DPs are examined to characterize dietary behavior achieving a resemblance of real life, in which foods (and nutrients) are eaten in combination. They offer a close relation to dietary recommendations on the food level and account for the potential interplay of nutrients in whole diets, which generally is not reflected in the single food group or nutrient approaches. DP,

covering the complexity of dietary intakes, are then most often investigated with respect to their relation to disease risk and contribute to the generation of evidence for disease prevention [3].

The approaches in dietary pattern analysis can be grouped into three categories [3–6]: the hypothesis-driven approaches, the exploratory approaches and hybrid approaches of the two. The hypothesis-driven approach relies on prior information based on current knowledge regarding defined dietary components and their relation to health promotion and/or to major diet-related-diseases. Hypothesis-driven DPs arise from allocating points to the predefined dietary components using a scoring system. The scores, also called dietary indices, reflect diet quality or adherence to national dietary guidelines (Healthy Eating Index, etc.), or are based on hypotheses on healthful dietary habits such as the Dietary Approaches to Stop Hypertension (DASH) diet [7], or the Mediterranean (MED) diet [8].

Christina-Alexandra Schulz and Kolade Oluwagbemigun contributed equally to this work.

✉ Ute Nöthlings
noethlings@uni-bonn.de

¹ Department of Nutrition and Food Sciences, Nutritional Epidemiology, University of Bonn, Endenicher Allee 19b, 53115 Bonn, Germany

The exploratory approaches derive DPs without any hypotheses. Exploratory approaches traditionally derive DPs solely based on the underlying dietary data whereby a larger set of dietary variables is aggregated and reduced to form a smaller set of variables determining the DPs. The DPs describe the variation in dietary intake in a population based on correlations between nutrients, food items or food groups as reported in the respective dietary assessment instruments [3]. This approach usually employs simple to extensive statistical analysis to achieve its aim. The most widely adopted methods are cluster analysis and principal component analysis (PCA) [9] and the most often derived exploratory patterns in Western populations are commonly given simplistic interpretation of “Western DP” or “Prudent DP” [9]. Typically, a Western DP is characterized by greater intakes of white bread, red meat, processed meat, potatoes, and high-fat dairy products. In a Prudent DP, greater amounts of fruits, vegetables, whole grains, poultry and fish are consumed [9].

Since 2003, the hybrid methods that help to explain the relationship of diet to health via intermediate factors such as the reduced rank regression (RRR) have more widely been used [10, 11]. RRR takes into account prior knowledge about variables potentially relevant for the pathophysiological consequences of dietary intake, and thus takes a priori knowledge into account. With respect to the grouping of food items, however, it is exploratory by design, making it a hybrid approach.

Nutritional epidemiological research questions and methodology are evolving and high-dimensional biomarker data are increasingly integrated. Given these expansions, the scope of traditional DP approaches may be limited. Indeed, this includes both conceptual and statistical limitations. Since DPs could be a reflection of broader lifestyle patterns beyond diets [12], incorporating potentially relevant dietary-related factors such as sociodemographic, socioeconomic as well as genes and biological factors that are not assumed to be dependent on dietary intake may be necessary. Recently, a DP was generated to account for the contributions of sex, race/ethnicity, and body mass index [13]. To extend these thoughts, hypothesis-driven dietary patterns lately have also expanded by taking environmental aspects into account [14–16]. Moreover, non-traditional biological factors such as the metabolome and the gut microbiome, which are closely associated with dietary intake, have only limitedly been accounted for in DP analysis. Improvements have also been made in exploratory methods that derive DPs solely based on reported dietary intake, because the most commonly used exploratory method, the conventional PCA involves several but crucial subjective decisions, and the resulting DPs are sometimes challenging to interpret. Some of the novel exploratory methods that addresses these limitations include Treelet transform analysis [17] and the Gaussian

graphical models [18]. Similarly, the RRR, for instance, can handle multiple response variables, thus implying the use of a number of metabolites (derived from metabolomics) or taxa as response variables. But, RRR being linear regression model based is not suitable when the functional relationship to model between dietary variables and the dependent variable is nonlinear [19]. Besides, RRR essentially treats dietary data as independent variables.

A crucial step in utilizing dietary data for DPs, regardless of the aforementioned approaches per se, is dietary data preprocessing. The extent of the data preprocessing depends on the assessment instrument used to generate the dietary data and on the statistical methods to be applied. The food frequency questionnaire (FFQs) is the conventional dietary assessment instrument in large epidemiological studies, and data from FFQs have been traditionally used to derive DP. However, prone to bias, nowadays the use of multiple short-term instruments potentially complemented by an FFQ have been recommended [20], which leaves more granular dietary data to reflect dietary intake. Besides, nutritional epidemiological research has evolved to incorporate novel tools to assess dietary intake [21]. These novel tools may have an impact on the form and utility of dietary intake in DP analysis. Therefore, these advances call for both conceptual and statistical adaptations and advances in DP analysis.

The current review aims to pick up some of the recent developments and strategic thoughts into dietary pattern analysis. Particularly, we attempt (1) to give an overview of complementary or novel hypothesis-driven and exploratory approaches to derive DPs, and highlight their uses, peculiarities, strengths, limitations, and scope; (2) to review the literature that takes into account potential relevant dietary-related factors, specifically the metabolome and the gut microbiome in dietary pattern analysis; and (3) to give an overview of the literature using dietary data from dietary assessment methods other than FFQ to compose DPs.

Recent advances in hypothesis-driven dietary pattern analysis

The centerpiece of a priori index is the underlying information used to derive the DP. Healthful dietary habits are usually based on an interpretation of the overall literature relating diet to health [9]. The most commonly known indices take into account either cultural aspects or a certain health condition. Thus, existing hypothesis-driven indices are based on either food-based dietary guidelines, traditional eating habits of a specific geographical region, disease prevention, or the broader lifestyle. An overview of selected dietary indices and their rationalities, components and scoring systems is given in Table 1.

Table 1 Selected examples of dietary indices

Rationale	Score	Hypothesis based on	Dietary components, recommended intake, and scoring system
Food-based dietary guidelines	Healthy Eating Index (HEI-2015) [25]	2015–2002 Dietary Guidelines for Americans	<p>5/10* points each if complying to the recommended servings below</p> <ul style="list-style-type: none"> Total fruits: ≥ 0.8 c equivalents/1000 kcal Whole fruits: ≥ 0.4 c equivalents/1000 kcal Total vegetables: ≥ 1.1 c equivalents/1000 kcal Greens and beans: ≥ 0.2 c equivalents/1000 kcal Whole grains*: ≥ 1.5 oz equivalents/1000 kcal Dairy*: ≥ 1.3 c equivalents/1000 kcal Total protein foods: ≥ 2.5 oz equivalents/1000 kcal Seafood and Plant Protein: ≥ 0.8 oz equivalents Fatty acids*: (PUFAs + MUFAs)/SFAs ≥ 2.5 Refined grains*: ≤ 1.8 oz equivalents/1000 kcal Sodium*: ≤ 1.1 g/1000 kcal Added sugar*: ≤ 6.5% of energy Saturated Fats*: ≤ 8% of energy <p>TOTAL HEI-2015 score = 0–100</p>
Region origin	Mediterranean (MED) [8]	Adherence to the Mediterranean diet	<p>5 points each if complying with the recommended servings below</p> <ul style="list-style-type: none"> Non-refined Grains > 4/d Vegetables > 4/d Potatoes > 2/d Fruits > 3/d Full-fat dairy ≤ 10/wk Red meat ≤ 1/wk Fish > 6/wk Poultry ≤ 3/wk Legumes, nuts and beans > 6/wk Olive oil ≥ 1/d Alcohol < 300 mL/d but > 0 <p>TOTAL MedDiet score = 0 to 55</p>
Disease related	Dietary Approaches to Stop Hypertension (DASH) [7]	Recommended diet to prevent and treat high blood pressure [Joint National Committee on Prevention 2003] based on DASH-trials [41, 42]	<p>1 point each if complying with the recommended servings below</p> <ul style="list-style-type: none"> Total grains ≥ 7/d Vegetables ≥ 4/d Fruits ≥ 4/d Dairy ≥ 2/d Meat, poultry and fish ≤ 2/d Nuts, seeds and legumes ≥ 4/wk Total fat ≤ 27% of kcal Saturated Fat ≤ 6% of kcal Sweets ≤ 5/wk Sodium ≤ 2400 mg/d <p>TOTAL DASH score = 0 to 10</p>

Table 1 (continued)

Rationale	Score	Hypothesis based on	Dietary components, recommended intake, and scoring system
Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) [43]		Cross-method of the MED and DASH dietary pattern with an emphasis on linking the dietary components and servings to neuroprotection and dementia prevention	<p>1 point each if complying with the recommended servings below</p> <p>Whole grains $\geq 3/d$</p> <p>Green leafy vegetables $\geq 6/wk$</p> <p>Other vegetables $\geq 1/d$</p> <p>Berries $\geq 2/wk$</p> <p>Red meat and products $< 4/wk$</p> <p>Fish $\geq 1/wk$</p> <p>Poultry $\geq 2/wk$</p> <p>Beans $> 3/wk$</p> <p>Nuts $\geq 5/wk$</p> <p>Fast/fried food $< 1/wk$</p> <p>Olive oil primary oil</p> <p>Butter, margarine $< 1 T/d$</p> <p>Cheese $< 1/wk$</p> <p>Pastries, sweets $< 5/wk$</p> <p>Alcohol/wine $1/d$</p> <p>Total MIND score = 0–15</p>
Lifestyle focused	Low-Risk Lifestyle Behaviors score [129]	Relationship between lifestyle behaviors on various health outcomes including cardiovascular disease, diabetes, all-cause mortality, and mortality from cancer [129]	<p>1 point each if complying with the recommendation below</p> <p>Never smoked or less than 100 cigarettes smoked during life time</p> <p>Healthy diet = top 40% of the Healthy Eating Index of the population [129]</p> <p>Adequate physical activity = being moderately or vigorously active, 5 and/or 3 times per week</p> <p>Moderate alcohol consumption = men ≤ 2 drinks/day and women ≤ 1 drinks/day, respectively, but for both more than 0 drinks per month</p> <p>Total low-risk lifestyle behaviors score = 0–4</p>

* indicates the following components where 10 points apply (i.e. Whole grain*; Dairy*; Fatty acids*; Refined Grain*; Sodium*; Added sugar*; and Saturated Fats*)
d day, *g* grams, *mg* milligrams, *wk* week, *PUFAs* polyunsaturated fatty acids, *MUFAs* monounsaturated fatty acids, *SFAs* saturated fatty acids

Food-based dietary guidelines focused

Various dietary indices have been introduced to assess diet quality [1, 2, 22] or adherence to dietary recommendations [1]. The Healthy Eating Index (HEI), developed by Kenneth et al. to assess adherence to the US Dietary Guidelines [23], was probably one of the first indices to measure how well national dietary recommendations were met. Since its creation in 1995, the HEI has been updated in correspondence to the most current version of the Dietary Guidelines for Americans at any time (i.e. HEI-2010 [24], HEI-2015 [25]). Moreover, an alternative HEI (aHEI) has been developed. As an alternative, this DP is based on foods and also nutrients that have been consistently associated with lower chronic disease risk [26]. Furthermore, the Dietary Guidelines for Americans Adherence Index (DGAI) has also been developed in reference to the Dietary Guidelines for Americans [27]. The concept of creating dietary indices based on national recommendations has been transferred to several dietary guidelines (e.g. Danish, French, Swedish and German recommendations) [28–31].

Regional origin

Knowledge about differences in death rates across populations from different geographical regions relating to dietary intake, in particular mono- and saturated fat consumption, goes back to the “seven countries study” by Ancel Keys [32]. The traditional Mediterranean diet, a DP found in Italy and Greece in the 1950s and 1960s, was found to be associated with a lower risk of coronary heart disease [32]. Since there was a great interest in investigating DP from different geographical regions and indices have been developed which capture geographical eating patterns. Well-known examples are the MED [33, 34] or Nordic [35] dietary indices, which focus on the traditional diet consumed in the Mediterranean and the Nordic countries. In addition, an alternative MED index (aMED) exists, which reflects an adaptation of the traditional MED [33] to non-Mediterranean countries [36]. Furthermore, some modified versions of the MED have been developed to adapt the MED to the countries such as Lebanon [37] and Japan [38], which differ in food culture compared to Greece where the MED was developed [33].

Disease relationship oriented

The World Cancer Research Fund (WCRF) and the American Institute of Cancer Research (AICR) have released eight recommendations (plus two special recommendations) on diet, physical activity, and weight management for cancer prevention [39]. A composite score of these recommendations has been found to potentially lower the risk for most types of cancers [40] as well as death from respiratory

diseases, death from circulatory diseases and, death from cancer [16]. In addition, the DASH DP has been associated with health benefits particularly with lower blood pressure [41, 42]. Of note, also a cross-method of the Mediterranean diet and the DASH diet, called MIND (Mediterranean-DASH Intervention for Neurodegenerative Delay) exists, which emphasizes the dietary components and servings linked to neuroprotection and dementia prevention [43].

Lifestyle oriented

In line with the WCRF/AICR score [16], which includes several lifestyle factors in addition to dietary factors, some more recently developed scores that, in addition to diet, incorporate several modifiable lifestyle factors such as smoking, obesity and physical activity exist [44, 45]. This development goes back to the insight, that dietary behavior can also be seen as an element of lifestyles and links to further behavioral factors which have been shown to be associated with disease risk and prevention. A recent meta-analysis of 22 prospective observational studies reported a 66% reduced cardiovascular risk with adherence to several lifestyle behaviors [46]. This indicates there is a growing interest and relevance in the analysis of patterns going beyond the inclusion of dietary factors only.

Complementary exploratory approaches in dietary pattern analysis

Exploratory DP analysis has been extensively reviewed in several pieces of literatures such as Moeller et al. [3]. Thus, in the current work, we focus on methods complementary to traditional ones, which have been developed in the last 15 years. The fact that the widely used exploratory method, the PCA has some shortcomings that make interpretation difficult, primarily necessitated the development of these complementary methods. While variants of PCA, such as sparse PCA that addressed some of the limitations of conventional PCA have been developed, other methods have emerged. These complementary approaches which can be broadly classified into latent class and clustering techniques have long found extensive application in a broad range of research fields, albeit they are new to DP analysis. The commonly used complementary approaches are presented below.

Treelet transform analysis (TT) is a latent class method that can be viewed as an amalgamation of PCA and hierarchical clustering analysis [17]. It provides the opportunity to describe DPs in multivariate dietary data with substantive meaning and interpretation. TT produces sparse components in an elegant and simple fashion by combining ideas from PCA and hierarchical clustering analysis. It also provides a concise visual representation of loading sparsity patterns

and the general dependency structure of the data. Due to its sparsity (interpretability), a trade-off with explained variance of DPs becomes inevitable. Moreover, unlike PCA that relies on rules of thumb and ad hoc thinking for determining the number of DPs to retain, TT has a fast and efficient algorithm for determining the optimal number of DPs to be retained as well as assessing the stability of DPs. TT has been used to derive DP at the level of nutrients [47], food items [48] and food groups [49, 50]. The five DPs obtained from the intake into 39 food groups in a group of older German Adults were “alcohol and red meat”, “bread, margarine, and processed meat”, “fruiting vegetables and vegetable oils”, “tea and miscellaneous”, and “pasta and rice, and sauce” [50].

Another complementary method, the sparse latent factor models (SLFM) that are based on a Bayesian modeling framework aims to provide parsimonious relations between high-dimensional variables and latent factors by forcing less influential associations to have a zero association in the model [51, 52]. The only available literature on the application of sparse latent factor models in DP analysis was a recent paper on the intake of 102 food items in a group of young American adults [13]. The obtained seven DP were “fruit, nuts, several vegetables, fish and poultry”, “red and processed meat, potatoes, fried foods, pasta dishes, and breads”, “a cluster of snacks and sweets”, “a group of Hispanic foods”, “meat and dietary alternatives”, “alcoholic beverages”, and “cereal with milk” [13].

Gaussian graphical models (GGM) is an important exploratory analysis that identifies the conditional independence structure in the data set by assessing pairwise correlation between two variables controlling for other variables [18, 53]. GGM also introduces sparsity in generating DPs. A study that used GGM for DP analysis on 49 food groups identified clusters of red meat and cooked vegetables, dairy products, sweet foods, and fruits and vegetables [54].

The random forest with classification tree analysis (RF-CTA) is a clustering-based method. Random forest uses an algorithm to create multiple classification or regression trees based on a dependent variable of interest [55]. Subsequently, the classification tree analysis identifies mutually exclusive subgroups of a population that share common characteristics that influence the dependent variable of interest [56]. This method has been used to identify DPs comprising soy products, wine, fast foods, and French fries [57].

Independent component analysis (ICA) is another latent class method that finds a linear representation of non-Gaussian data so that the components are statistically independent or as independent as possible [58]. To our knowledge, ICA has not yet been applied in DP analysis; however, a report from the Nordic Research Council suggests that ICA would be very relevant because the DP that would be obtained will be independent of one another [19].

Preprocessing of input dietary variables such as dealing with missing variables and standardization (scaling or centering) are issues that should be addressed before using some of these approaches. Methods to handle missing dietary variables include multiple imputations [59] and mixed models [60]. Standardization is encouraged because it ensures that mean and the variance of each dietary variable have no influence on the obtained DPs [3]. It should also be noted that since exploratory DPs are data-driven, DPs might not be reproducible across populations or even within the same population over time [3]. Additionally, to ensure widespread applicability, it is necessary that these approaches are available across the well-known statistical software packages. More importantly, statistical principles of these approaches should be well understood since the default settings may not always give optimal solutions. Obviously, there are advantages and disadvantages of these complementary methods with their respective practical worth and repercussions. Table 2 provides an overview of the features/aims, peculiarities, strengths, and limitations of TT, SLFM, GGM, and RF-CTA in DPs analysis.

The metabolome and the gut microbiome in dietary pattern analysis

The metabolome and gut microbiome are expected to unlock the ‘black box’ in chronic disease epidemiology [61–63]. Although their causal relationship with most health conditions has yet to be confirmed, the metabolome and the gut microbiome are likely to improve our understanding of the link between DPs and health conditions. Therefore, there is a need to integrate the metabolome and the gut microbiome in DP analysis. A number of epidemiological studies associating DPs with the metabolome and the gut microbiome, particularly bacterial taxonomic composition appeared in the literature over the last ten years [50, 64–86]. These primary studies were retrieved from the PubMed (MEDLINE) database (terms: dietary patterns/indices, gut microbiome, and metabolome, and limited to articles published in English). We reviewed the references of retrieved studies for additional studies. These studies include those associating hypothesis-driven and exploratory DPs with single metabolites and metabolite patterns or scores as well as hypothesis-driven and exploratory DPs with single metabolites and metabolite patterns or scores. A summary of these studies is provided in Table 3.

Evidence exists regarding the relation of Mediterranean DP with metabolites and microbiome. As opposed to the metabolome, findings from studies relating DPs, particularly exploratory DPs to the microbiome composition are few. While this may indicate that the interest in the role of the microbiome in DP is just emerging, it also suggests DPs

Table 2 Selected complementary exploratory dietary pattern analysis approaches

Exploratory dietary pattern approach	Features/aims	Peculiarity	Strengths	Limitations	References
Treelet transform analysis (TT)	TT is a cross-method of principal component analysis and hierarchical clustering analysis	<ol style="list-style-type: none"> 1. Forces sparse DPs 2. Produces cluster tree that reflects local dependency relationships of the dietary variables as well as a coordinate system for the dietary data at each level of the cluster tree 3. Cluster tree level (cut-level) that must be chosen influences both the sparsity and the composition of DPs 4. Implemented as a Mata function in STATA statistical software 	<ol style="list-style-type: none"> 1. Uses the correlation or covariance matrix of dietary intake to derive DPs 2. TT procedure enables the selection of the number of DPs and a cut-level by a data-driven (cross-validation) method 3. Several functions to aid in model selection and output analysis 4. Stability assessment analysis by subsampling approach 5. Sparsity is optimal because a dietary variable is loaded on only one DP 6. DPs are potentially simpler to interpret 7. The latent variable is numerical so DP scores (continuous values) can be assigned to study participants 	DPs are not independent	[17, 49, 50]
Sparse latent factor models (SLFM)	SLFM provide parsimonious relations between high-dimensional variables and latent factors by forcing less influential associations to have a zero association in the model	<ol style="list-style-type: none"> 1. Uses the posterior inclusion probability obtained from the sparse latent factor analysis to determine food and DP pairs that have non-zero associations 2. A model-based solution for dealing with missing dietary data 3. Sparsity of DPs: sparsity-inducing priors is an integral part of SLFM 4. Implemented in software for Bayesian factor regression models 	<ol style="list-style-type: none"> 1. Interpretability of DPs is achieved by its sparsity and inclusion threshold of dietary variables showing a significant and a posterior inclusion probability > 0.95 for a given DP 2. Several approaches to infuse biological or non-dietary variables 	Sparsity is not optimal because a dietary intake variable may be loaded on one or a few DPs	[13, 51]

Table 2 (continued)

Exploratory dietary pattern approach	Features/aims	Peculiarity	Strengths	Limitations	References
Gaussian graphical models (GGM)	GGM identifies conditional independence structure in dietary variables by assessing pairwise correlation between two variables, while controlling for effects of the other variables in the network model	<ol style="list-style-type: none"> 1. Selects a model on rank-based dietary variable 2. Estimates a single DP as a unique solution for the estimated model 3. Sparsity of DPs depends on a regularization parameter that can be derived using different criteria 4. Implemented in R statistical software 	<ol style="list-style-type: none"> 1. It can also be used for ordinal data or data comprising a mixture of categorical and continuous dietary variables 2. Stability assessment analysis 	<p>Requires Gaussian-distributed data</p> <p>The latent variable is categorical so DP scores (continuous values) cannot be assigned to study participants</p> <p>Spurious relationships may affect network interpretation</p> <p>A single DP may not fully exploit the multivariate dietary data sets</p>	[18, 53, 54]
Random forest with classification tree analysis (RF-CTA)	RF-CTA belongs to the family of classification trees. This approach generates many decision trees that are constructed using different sets of randomly selected dietary variables	<ol style="list-style-type: none"> 1. Intuitive approach: easy to implement and understand 2. Produces decision trees that can be visualized 3. Implemented in R and other statistical software 	<ol style="list-style-type: none"> 1. It is a highly accurate classifier which produces an internal unbiased estimation of the generalization error 2. The model is robust to outliers 3. No standardization or scaling required prior to the analysis 4. Ability to handle missing values 	<p>The latent variable is categorical so DP scores (continuous values) cannot be assigned to study participants</p> <p>Less transparent</p> <p>Highly data dependent</p>	[55–57]

shape the metabolome more than the microbiome. Nevertheless, the fact that recent studies have shown that the gut microbiome plays a role in the occurrence of several health conditions [63, 87], gut microbiome composition improves the prediction of glucose response [88], and the fact that a prudent DP was associated with a lower risk of *Fusobacterium nucleatum*-positive colorectal cancer [89] strengthens the relevance of the gut microbiome in DP analysis.

Irrespective of the DP approach or the number of metabolome or microbiome variables that were investigated, the approaches in the aforementioned studies can be broadly classified into two. The first is the two-stage approach where DPs are generated first and then they are related to single (or groups of) metabolites or single (or groups of) taxa in regression models. The second is a one-stage approach where metabolite-related or taxa-related DPs are generated using multivariable regression-based methods such as RRR. These two approaches are conceptually similar. The first approach has been applied to both the metabolome and the microbiome. On the other hand, the second approach using RRR has only been applied to the metabolome to generate metabolite-related DPs for seven classes of chemically similar metabolites [66], isoleucine, leucine, and valine [79], and six islet autoimmunity-associated metabolites [82].

The main goal of RRR is to generate DPs that capture as much variation in predicted response variables which are usually disease-related nutrients or biomarkers as such the focus is on the response variables. Therefore, RRR-generated DPs may not be diet behaviorally meaningful. This method would therefore be suitable when the goal of DP analysis is against the backdrop of exploring the effect of the combination of foods on health conditions that is mediated through the appropriately identified components of the metabolome and/or the gut microbiome. These components may be defined a priori [66, 79] or identified a posteriori [82]. As it has been explored for the metabolome, studies should use RRR or related methods to create DPs explaining disease or health outcomes microbiome signatures such as specific taxa or functions, broader microbiome compositions or functions, and microbiome diversity. The gut microbiome with its extensive and distinct metabolic repertoire complements the metabolic activity of the host enzymes by contributing microbial metabolites to the metabolome, either diet-derived or from other sources [90]. Therefore, a pragmatic approach would be to generate DPs of disease-related microbial metabolites.

If a DP is related to a particular metabolite or taxon, there is a possibility that it is also related to other closely or distantly related metabolite(s) or taxa. Hence, an approach that maps the association of DPs with aggregated correlated metabolites or taxa would potentially yield higher statistical power. The study by Oluwagbemigun and colleagues underscores the importance of this approach where

Table 3 Studies associating dietary patterns with the metabolome and the gut microbiome

	Single metabolites	Metabolite patterns or scores	Single bacterial taxa	Population groups of bacterial taxa
Hypothesis-driven dietary patterns (DPs)	Four DPs related to the ratio of 363 metabolites [64], Two DPs related to 4369 metabolites [68], Four DPs related to 116 metabolites [71], Mediterranean DP associated with three short-chain fatty acids [72], four diet quality indexes associated with 1316 metabolites [74], four DPs related to several metabolites [76], Patterns of meat and seafood consumption related to 42 metabolites [77], Mediterranean DP associated with 1165 metabolites [78], Mediterranean DP associated with 59 metabolites [81], Mediterranean DP associated with 175 metabolites [84]	Mediterranean DP associated with a metabolite score [84]	Three DPs related to 26 bacteria [70], Mediterranean diet associated with several bacteria [72], Mediterranean DP and Healthy Food DP related to several taxa [73], Two DPs related to several bacterial taxa [87]	Three DPs associated with enterotypes [81], unprocessed foods and processed food groups associated with phylogenetically grouped bacterial taxa [86], Two DPs related to two enterotypes [87]
Exploratory DPs	Seven PCA-derived DPs related to ratio of 363 metabolites [64], Three Cluster analysis-derived DPs related to plasma fatty acid profiles and metabolomic data [65], Seven PCA-derived DPs related to 163 metabolites [67], sparse PCA-derived meat and vegetable DPs associated with 130 metabolites [81]	Five TT-derived DPs related to eight TT-derived metabolite patterns [50], DPs explaining a maximum variation in the concentration of the seven classes of chemically similar metabolites were derived by RRR [66], Two PCA-derived DPs correlated with two PCA-derived metabolite patterns [74], RRR-derived DP associated with branched chain amino acids [79], RRR-derived DP associated with 853 metabolites [82]	Three cluster analysis-derived DPs related to seven bacteria [69], factor analysis-derived DPs associated with several bacterial taxa [85], Three PCA-derived DP associated with 24 bacterial taxa [86] two DPs were related to several bacterial taxa [89]	Five TT-derived DPs related to seven TT-derived groups of bacteria [50]

structured patterns of correlations present in each of the dietary, metabolome, and microbiome compositional domains were identified using TT and possible DP-metabolite patterns and DP-microbiome patterns associations were fitted in linear regression models [50]. This strategy is appealing because it fully explores the complex multivariate structure of the metabolome and microbiome data sets and ensures that grouping of metabolites is beyond chemical similarities or grouping of bacteria is beyond taxonomic relationships. This approach may be sufficient to reveal the potential relationships between few sets of dietary, metabolome, and microbiome variables that can be explored further. For instance, the single diets, metabolites or taxa contributing to each pattern can be used to generate new metabolome- or microbiome-related DPs using RRR or more parsimonious

methods such as the partial least squares regression. Unlike the RRR, partial least squares regression (PLS) would generate DPs that have maximum covariance of both the dietary intake predictor variables and the metabolome and microbiome response variables.

In future, DP analysis that intends to incorporate the metabolome and/or microbiome should employ novel methods that can accommodate all high-dimensional predictor and response variables and simultaneously explore their associations and dependency patterns. Example of these novel methods, which have been applied in our research fields, includes the sparse PLS [91] and the sparse RRR [92]. The foregoing suggests that the additional information offered by the metabolome and the gut microbiome would be more exploited in exploratory DP when compared

to hypothesis-driven DP analysis. Indeed, using a large part of the metabolome and the gut microbiome data and metabolotyping [93] and enterotyping [94] populations into generally valid and biologically meaningful homogeneous subgroups would facilitate the integration of the metabolome and the gut microbiome in DP analysis.

Despite the optimism and the advantages of incorporating the metabolome and microbiome in DP analysis, their huge biological complexities and the methodological variation associated with generating metabolome and microbiome data suggest that translating metabolome- or microbiome-related DPs into practice may be challenging.

Implication of dietary assessment methods for dietary pattern analysis

In the past, dietary data used for DP analysis was most commonly assessed by long-term dietary assessment instrument, FFQs [49, 95–99]. FFQs include informative food items consumed regularly, contributing substantially to the nutrient(s) of interest and ability to rank individuals according to their intake [9]. Different procedures for selecting informative food items have been described [9]. Fewer studies have used short-term instruments such as dietary recalls and dietary records [100–103]. All assessment methods do have their strengths and limitations. Thus, more recently it has been discussed that combining short-term and long-term dietary assessment instruments is a superior approach to derive usual dietary intake estimates in large-scale studies [20]. Whereas the combination of these instruments is promising, some challenges for DP which will be addressed below cannot be excluded.

Prior to DP analysis, food items are usually grouped into food groups. The main purpose of such grouping is to reduce the complexity of the obtained data. The number of food groups composed reported in the literature range from 25 [104] to 99 [102]. Most commonly the groups are formed according to their macronutrient composition (e.g., fat or fiber content), nutritional similarities, or culinary use [105–108]. Further rationalities for grouping food items include related food and drinks [102, 109], nutrient profile and consumption in the cohort under investigation [110],

and their hypothesized contribution to diet–disease relationships [100, 111]. Additionally, some foods are oftentimes grouped into a separate group such as eggs, butter, margarine, pizza, soup, coffee, tea, or garlic [105, 108, 112]. Allocating a minimum of 5–10 participants per food group was another approach mentioned in the literature [113] based on the recommendation by Kass and Tinsley [114]. Decisions of aggregating food items may also be based on expert knowledge and experience. The extent to which different aggregation of food items from the same population and assessment instrument affect DP solutions are questions that should be considered in future studies.

At first, it needs to be considered that the number of food items reported in different populations may vary immensely. This may be on the one hand due to the variety of the diet consumed by a population under investigation, but also due to the applied dietary assessment instruments. Applying the instruments simultaneously over a similar period may be beneficial. FFQs are designed to capture the habitual dietary intake and include a list of defined food and drinks. Over a specific period in the past (commonly the previous year), the participants are asked to indicate their typical frequency of consumption. Depending on the purpose of the FFQ, these questionnaires include up to a few hundred selected food items, commonly ranging from, e.g. 20 items in “screeners” to 200 items for fully comprehensive lists intended to assess the total diet [115] (Table 4). Whereas, data from dietary recalls and records capture either dietary intake from the recent past (usually previous 24 h or 48 h) or are completed by participants at the time of consumption throughout a specific period of time (often 3–7 days). Portion sizes may be either weighted or estimated (non-weighted), brands (if relevant), and preparation (e.g. cooking method, the addition of fat, recipe ingredients, etc.) may also be reported [115]. Both of the latter tools are open ended in contrast to the predefined list of food items captured via FFQs. Thus, by the usage of a dietary recall and dietary record, the number of items reported in a study population may easily add up to a couple of thousand items (e.g. from the literature > 3500 [102], 4850 [116], and 7444 [117]).

Thus, the more the detailed information arises from dietary recalls and records the more the effort might be required by the investigators to cluster the food items into food groups

Table 4 Characteristics of dietary assessment tool

	Characteristics	Number of food items
FFQ	Retrospective, usually concerning the previous year, both quantitative and semi-quantitative versions exist	List of predefined food items and food groups; usually < 300
Diet recall	Retrospective, usually previous 24 h or 48 h	Individually based the food intake recorded by the participant; open-ended; > 1000 items
Dietary record	Filled in during actual food intake, both weighted and unweighted	Individually based the food intake recorded by the participant; open-ended; > 1000 items

prior to analysis, given that much more items appear in the database which need to be sorted into groups. Still, FFQ-derived DP has been shown to be in general comparable with those derived from dietary record data in both children [118] and adults [108, 119]. Having access to much more detailed dietary data such as multiple FFQs, more recalls or recording days and variety of foods and the extent of their processing however may potentially offer the opportunity that more detailed food groups can be formed and thus more in-depth hypotheses can be tested. It has been discussed for instance that the preparation method of meat may play a role in the disease relationship [120–122] or that fermented dairy products may have a different disease association than non-fermented dairy products [123, 124].

Summary and implications for future research

This review summarizes the most recent advances in DP analysis, which has been an established analytic approach in nutritional epidemiology. DP analysis has advanced the field of nutritional epidemiology by combining nutrients and foods and thus taking into account their synergistic effects [1, 2]. The three approaches, (a) hypothesis based, (b) data-driven, and (c) a hybrid approach of both, have been widely used during the last two decades. A relationship between diets and the metabolome and gut microbiome is strongly hypothesized. Recent developments in biomedical sciences, particularly advances in omics technologies, have made the quantification of several metabolites and the composition of a large number of gut bacteria feasible. This has made the inclusion of metabolome and gut microbiome in DP analysis relevant, necessary, and important, and calls for adequate statistical approaches. Lastly, the advancements in assessments of dietary data were also picked up in this review. The availability of more detailed data will allow defining food groups customized to the scope of specific research questions.

The strength of this review is that it provides a state of knowledge of DP analysis with a focus on novel approaches. To our knowledge, at writing, this review was the first to cover the potential inclusion of the metabolome and the gut microbiome in DP analysis. The major weakness of this review article is that some relevant studies might have not been evaluated. However, this review is comprehensive, each aspect was objectively introduced, and our inferences mostly adhere to the cited evidence.

The perspective of nutritional epidemiology has nowadays become more comprehensive and studies have a more holistic approach beyond the sole focus on the associations of dietary intake and disease risk. This widening of perspective addresses the linking of additional aspects related to diets in

exposure definitions on the one hand, and the disentangling of underlying mechanisms between dietary intake and disease risk on the other hand. The latter is of importance for the general mechanistic understanding, but might also support the development of dietary disease prevention recommendations stratified for population groups. DP analysis can be instrumental in both directions. Advanced statistical approaches in DP analysis enable the incorporation of metabolome and gut microbiome to foster a thorough understanding of disease risk mechanisms and serve potential stratifications with respect to susceptible population groups. Finally, DP analysis should be adapted to the advancements in dietary assessment methods with respect to detail and granularity. Lately, DP very recently started to take effects beyond health outcomes into account and implications in relation to sustainability and environmental impact have come into focus. Yet, only very few studies have addressed sustainability aspects in exploring DP to-date [14, 15, 125–128]. Thus, there is a strong need to investigate the broader impact of diets on the environment and link these to disease prevention, for which DP analysis has the means to respectively expand its scope and close the gap of analyses of sustainable DPs with chronic disease risk.

Considering the above-mentioned advances in DP analysis would offer progressive insights into diet–disease relationships.

Conclusion

DP analysis has revolutionized the field of nutritional epidemiology long ago making it possible to take into account that foods are consumed in combination and thus may have synergistic effects. In the recent past, new aspects relevant for DP analysis have emerged both conceptually as well as methodologically. Thus, future DP analysis should take additional aspects such as diet sustainability and environmental impact into account, to advance the capture of diet as a whole. Additionally, incorporating the metabolome and microbiome in DP analysis might help to reveal necessary insights into the diet–disease relationships helping targeted disease prevention through dietary efforts in the long run.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported by Metabolic HEALTH through nutrition, microbiota and tryptophan bioMARKers in the frame of the EU Joint Programming Initiative “A Healthy Diet for a Healthy Life” (ERA HDHL-HEALTHMARK) through the Federal Ministry of Education and Research (BMBF), Germany (Grant number: 01EA1705A), and is adjunct to the work in Diet Body Brain (Competence Cluster in Nutrition Research) funded by the German Federal Ministry of Education and Research (BMBF) (Grant number: 01EA1410A).

Declarations

Conflict of interest The authors have no conflict of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 13(1):3–9. <https://doi.org/10.1097/00041433-200202000-00002>
- Kant AK (1996) Indexes of overall diet quality: a review. *J Am Diet Assoc* 96(8):785–791. [https://doi.org/10.1016/S0002-8223\(96\)00217-9](https://doi.org/10.1016/S0002-8223(96)00217-9)
- Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM (2007) Dietary patterns: challenges and opportunities in dietary patterns research an experimental biology workshop, April 1, 2006. *J Am Diet Assoc* 107(7):1233–1239. <https://doi.org/10.1016/j.jada.2007.03.014>
- Ocke MC (2013) Evaluation of methodologies for assessing the overall diet: dietary quality scores and dietary pattern analysis. *Proc Nutr Soc* 72(2):191–199. <https://doi.org/10.1017/S0029665113000013>
- Jannasch F, Riordan F, Andersen LF, Schulze MB (2018) Exploratory dietary patterns: a systematic review of methods applied in pan-European studies and of validation studies. *Br J Nutr* 120(6):601–611. <https://doi.org/10.1017/S0007114518001800>
- Schulze MB, Martinez-Gonzalez MA, Fung TT, Lichtenstein AH, Forouhi NG (2018) Food based dietary patterns and chronic disease prevention. *BMJ* 361:k2396. <https://doi.org/10.1136/bmj.k2396>
- Epstein DE, Sherwood A, Smith PJ, Craighead L, Caccia C, Lin PH, Babyak MA, Johnson JJ, Hinderliter A, Blumenthal JA (2012) Determinants and consequences of adherence to the dietary approaches to stop hypertension diet in African-American and white adults with high blood pressure: results from the ENCORE trial. *J Acad Nutr Diet* 112(11):1763–1773. <https://doi.org/10.1016/j.jand.2012.07.007>
- Panagiotakos DB, Pitsavos C, Arvaniti F, Stefanadis C (2007) Adherence to the Mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the MedDietScore. *Prev Med* 44(4):335–340. <https://doi.org/10.1016/j.ypmed.2006.12.009>
- Willett W (2012) *Nutritional Epidemiology*. Print ISBN-13:9780199754038
- Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H (2004) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* 159(10):935–944. <https://doi.org/10.1093/aje/kwh134>
- Barbarekko J, Koch M, Schulze MB, Nothlings U (2013) Dietary pattern analysis and biomarkers of low-grade inflammation: a systematic literature review. *Nutr Rev* 71(8):511–527. <https://doi.org/10.1111/nure.12035>
- Martinez ME, Marshall JR, Sechrest L (1998) Invited commentary: factor analysis and the search for objectivity. *Am J Epidemiol* 148(1):17–19. <https://doi.org/10.1093/oxfordjournals.aje.a009552>
- Joo J, Williamson SA, Vazquez AI, Fernandez JR, Bray MS (2018) advanced dietary patterns analysis using sparse latent factor models in young adults. *J Nutr* 148(12):1984–1992. <https://doi.org/10.1093/jn/nxy188>
- Nelson ME, Hamm MW, Hu FB, Abrams SA, Griffin TS (2016) Alignment of healthy dietary patterns and environmental sustainability: a systematic review. *Adv Nutr* 7(6):1005–1025. <https://doi.org/10.3945/an.116.012567>
- Fresan U, Martinez-Gonzalez MA, Sabate J, Bes-Rastrollo M (2019) Global sustainability (health, environment and monetary costs) of three dietary patterns: results from a Spanish cohort (the SUN project). *BMJ Open* 9(2):e021541. <https://doi.org/10.1136/bmjopen-2018-021541>
- Vergnaud AC, Romaguera D, Peeters PH, van Gils CH, Chan DS, Romieu I, Freisling H, Ferrari P, Clavel-Chapelon F, Fagherazzi G, Dartois L, Li K, Tikk K, Bergmann MM, Boeing H, Tjonneland A, Olsen A, Overvad K, Dahm CC, Redondo ML, Agudo A, Sanchez MJ, Amiano P, Chirlaque MD, Ardanaz E, Khaw KT, Wareham NJ, Crowe F, Trichopoulou A, Orfanos P, Trichopoulos D, Masala G, Sieri S, Tumino R, Vineis P, Panico S, Bueno-de-Mesquita HB, Ros MM, May A, Wirfalt E, Sonestedt E, Johansson I, Hallmans G, Lund E, Weiderpass E, Parr CL, Riboli E, Norat T (2013) Adherence to the World Cancer Research Fund/ American Institute for Cancer Research guidelines and risk of death in Europe: results from the European Prospective Investigation into Nutrition and Cancer cohort study 1,4. *Am J Clin Nutr* 97(5):1107–1120. <https://doi.org/10.3945/ajcn.112.049569>
- Gorst-Rasmussen A (2012) tt: Treelet transform with Stata. *Stata J* 12(1):130–146
- Villers F, Schaeffer B, Bertin C, Huet S (2008) Assessing the validity domains of graphical Gaussian models in order to infer relationships among components of complex biological systems. *Stat Appl Genet Mol Biol*. <https://doi.org/10.2202/1544-6115.1371>
- Edberg A, Freyhult E, Sand S, Fagt S, Knudsen VK, Andersen LF, Lindroos AK, Soeria-Atmadja D, Gustafsson MG, Hammerling U (2013) Discovering and characterisation of dietary patterns in two Nordic countries. Nordic Council, Copenhagen, Denmark. <https://doi.org/10.6027/TN2013-548>
- Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, Kipnis V (2012) Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol* 175(4):340–347. <https://doi.org/10.1093/aje/kwr317>
- Conrad J, Nothlings U (2017) Innovative approaches to estimate individual usual dietary intake in large-scale epidemiological studies. *Proc Nutr Soc* 76(3):213–219. <https://doi.org/10.1017/S0029665116003025>
- Burggraf C, Teuber R, Brosig S, Meier T (2018) Review of a priori dietary quality indices in relation to their construction criteria. *Nutr Rev* 76(10):747–764. <https://doi.org/10.1093/nutrit/nuy027>
- Kennedy ET, Ohls J, Carlson S, Fleming K (1995) The Healthy Eating Index: design and applications. *J Am Diet Assoc* 95(10):1103–1108. [https://doi.org/10.1016/S0002-8223\(95\)00300-2](https://doi.org/10.1016/S0002-8223(95)00300-2)
- Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HA, Kuczynski KJ, Kahle LL, Krebs-Smith SM (2013) Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet* 113(4):569–580. <https://doi.org/10.1016/j.jand.2012.12.016>
- Krebs-Smith SM, Pannucci TE, Subar AF, Kirkpatrick SI, Lerman JL, Tooze JA, Wilson MM, Reedy J (2018) Update of the Healthy Eating Index: HEI-2015. *J Acad Nutr Diet* 118(9):1591–1602. <https://doi.org/10.1016/j.jand.2018.05.021>

26. McCullough ML, Willett WC (2006) Evaluating adherence to recommended diets in adults: the Alternate Healthy Eating Index. *Public Health Nutr* 9(1A):152–157. <https://doi.org/10.1079/phn2005938>
27. Fogli-Cawley JJ, Dwyer JT, Saltzman E, McCullough ML, Troy LM, Jacques PF (2006) The 2005 dietary guidelines for Americans Adherence Index: development and application. *J Nutr* 136(11):2908–2915. <https://doi.org/10.1093/jn/136.11.2908>
28. Knudsen VK, Fagt S, Trolle E, Matthiessen J, Groth MV, Biloft-Jensen A, Sorensen MR, Pedersen AN (2012) Evaluation of dietary intake in Danish adults by means of an index based on food-based dietary guidelines. *Food Nutr Res*. <https://doi.org/10.3402/fnr.v56i0.17129>
29. Chaltiel D, Adjibade M, Deschamps V, Touvier M, Hercberg S, Julia C, Kesse-Guyot E (2019) Programme National Nutrition Sante—guidelines score 2 (PNNS-GS2): development and validation of a diet quality score reflecting the 2017 French dietary guidelines. *Br J Nutr* 122(3):331–342. <https://doi.org/10.1017/S0007114519001181>
30. Drake I, Gullberg B, Ericson U, Sonestedt E, Nilsson J, Wallstrom P, Hedblad B, Wirfalt E (2011) Development of a diet quality index assessing adherence to the Swedish nutrition recommendations and dietary guidelines in the Malmo Diet and Cancer cohort. *Public Health Nutr* 14(5):835–845. <https://doi.org/10.1017/S1368980010003848>
31. von Ruben AIA, Boeing H, Flthkötter M (2009) Die Bewertung der Lebensmittelaufnahme mittels eines, Healthy Eating Index‘ (HEI-EPIC). *Ernährungs Umschau* 56:450–456
32. Keys A, Menotti A, Karvonen MJ, Aravanis C, Blackburn H, Buzina R, Djordjevic BS, Dontas AS, Fidanza F, Keys MH et al (1986) The diet and 15-year death rate in the seven countries study. *Am J Epidemiol* 124(6):903–915. <https://doi.org/10.1093/oxfordjournals.aje.a114480>
33. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D (2003) Adherence to a Mediterranean diet and survival in a Greek population. *N Engl J Med* 348(26):2599–2608. <https://doi.org/10.1056/NEJMoa025039>
34. Bach-Faig A, Berry EM, Lairon D, Reguant J, Trichopoulou A, Dernini S, Medina FX, Battino M, Belahsen R, Miranda G, Serra-Majem L, Mediterranean Diet Foundation Expert G (2011) Mediterranean diet pyramid today. Science and cultural updates. *Public Health Nutr* 14(12A):2274–2284. <https://doi.org/10.1017/S1368980011002515>
35. Roswall N, Eriksson U, Sandin S, Lof M, Olsen A, Skeie G, Adami HO, Weiderpass E (2015) Adherence to the healthy Nordic food index, dietary composition, and lifestyle among Swedish women. *Food Nutr Res* 59:26336. <https://doi.org/10.3402/fnr.v59.26336>
36. Fung TT, Rexrode KM, Mantzoros CS, Manson JE, Willett WC, Hu FB (2009) Mediterranean diet and incidence of and mortality from coronary heart disease and stroke in women. *Circulation* 119(8):1093–1100. <https://doi.org/10.1161/CIRCULATIONAHA.108.816736>
37. Naja F, Hwalla N, Itani L, Baalbaki S, Sibai A, Nasreddine L (2015) A novel Mediterranean diet index from Lebanon: comparison with Europe. *Eur J Nutr* 54(8):1229–1243. <https://doi.org/10.1007/s00394-014-0801-1>
38. Kanauchi M, Kanauchi K (2016) Development of a Mediterranean diet score adapted to Japan and its relation to obesity risk. *Food Nutr Res* 60:32172. <https://doi.org/10.3402/fnr.v60.32172>
39. Wiseman M (2008) The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *Proc Nutr Soc* 67(3):253–256. <https://doi.org/10.1017/S002966510800712X>
40. Romaguera D, Vergnaud AC, Peeters PH, van Gils CH, Chan DS, Ferrari P, Romieu I, Jenab M, Slimani N, Clavel-Chapelon F, Fagherazzi G, Perquier F, Kaaks R, Teucher B, Boeing H, von Rusten A, Tjonneland A, Olsen A, Dahm CC, Overvad K, Quiros JR, Gonzalez CA, Sanchez MJ, Navarro C, Barricarte A, Dorransoro M, Khaw KT, Wareham NJ, Crowe FL, Key TJ, Trichopoulou A, Lagiou P, Bamia C, Masala G, Vineis P, Tumino R, Sieri S, Panico S, May AM, Bueno-de-Mesquita HB, Buchner FL, Wirfalt E, Manjer J, Johansson I, Hallmans G, Skeie G, Benjaminsen Borch K, Parr CL, Riboli E, Norat T (2012) Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. *Am J Clin Nutr* 96(1):150–163. <https://doi.org/10.3945/ajcn.111.031674>
41. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH, Karanja N (1997) A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *N Engl J Med* 336(16):1117–1124. <https://doi.org/10.1056/NEJM199704173361601>
42. Sacks FM, Svetkey LP, Vollmer WM, Appel LJ, Bray GA, Harsha D, Obarzanek E, Conlin PR, Miller ER 3rd, Simons-Morton DG, Karanja N, Lin PH, Group DA-SCR (2001) Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. *N Engl J Med* 344(1):3–10. <https://doi.org/10.1056/NEJM200101043440101>
43. Morris MC, Tangney CC, Wang Y, Sacks FM, Bennett DA, Aggarwal NT (2015) MIND diet associated with reduced incidence of Alzheimer’s disease. *Alzheimer’s Dement* 11(9):1007–1014. <https://doi.org/10.1016/j.jalz.2014.11.009>
44. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, Fuster V, Boerwinkle E, Melander O, Orho-Melander M, Ridker PM, Kathiresan S (2016) Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med* 375(24):2349–2358. <https://doi.org/10.1056/NEJMoa1605086>
45. Lourida I, Hannon E, Littlejohns TJ, Langa KM, Hypponen E, Kuzma E, Llewellyn DJ (2019) Association of lifestyle and genetic risk with incidence of dementia. *JAMA*. <https://doi.org/10.1001/jama.2019.9879>
46. Barbaresco J, Rienks J, Nothlings U (2018) Lifestyle indices and cardiovascular disease risk: a meta-analysis. *Am J Prev Med* 55(4):555–564. <https://doi.org/10.1016/j.amepre.2018.04.046>
47. Assi N, Moskal A, Slimani N, Viallon V, Chajes V, Freisling H, Monni SK, Forster J, Weiderpass E, Lujan-Barroso L, Amiano P, Ardanaz E, Molina-Montes E, Salmeron D, Quiros JR, Olsen A, Tjonneland A, Dahm CC, Overvad K, Dossus L, Fournier A, Baglietto L, Fortner RT, Kaaks R, Trichopoulou A, Bamia C, Orfanos P, De Magistris MS, Masala G, Agnoli C, Ricceri F, Tumino R, Bueno de Mesquita HB, Bakker MF, Peeters PH, Skeie G, Braaten T, Winkvist A, Johansson I, Khaw KT, Wareham NJ, Key T, Travis R, Schmidt JA, Merritt MA, Riboli E, Romieu I, Ferrari P (2016) A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor-defined breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr* 19(2):242–254. <https://doi.org/10.1017/S1368980015000294>
48. Schoenaker DA, Dobson AJ, Soedamah-Muthu SS, Mishra GD (2013) Factor analysis is more appropriate to identify overall dietary patterns associated with diabetes when compared with treelet transform analysis. *J Nutr* 143(3):392–398. <https://doi.org/10.3945/jn.112.169011>
49. Gorst-Rasmussen A, Dahm CC, Dethlefsen C, Scheike T, Overvad K (2011) Exploring dietary patterns by using the treelet

- transform. *Am J Epidemiol* 173(10):1097–1104. <https://doi.org/10.1093/aje/kwr060>
50. Oluwagbemigun K, Foerster J, Watkins C, Fouhy F, Stanton C, Bergmann MM, Boeing H, Nothlings U (2020) Dietary patterns are associated with serum metabolite patterns and their association is influenced by gut bacteria among older German adults. *J Nutr* 150(1):149–158. <https://doi.org/10.1093/jn/nxz194>
 51. Carvalho CM, Chang J, Lucas JE, Nevins JR, Wang Q, West M (2008) High-dimensional sparse factor modeling: applications in gene expression genomics. *J Am Stat Assoc* 103(484):1438–1456. <https://doi.org/10.1198/016214508000000869>
 52. West M (2003) Bayesian factor regression models in the “Large p, Small n” paradigm. *Bayesian statistics* 7:723–732
 53. Hevey D (2018) Network analysis: a brief overview and tutorial. *Health Psychol Behav Med* 6(1):301–328
 54. Iqbal K, Buijsse B, Wirth J, Schulze MB, Floegel A, Boeing H (2016) Gaussian graphical models identify networks of dietary intake in a German adult population. *J Nutr* 146(3):646–652. <https://doi.org/10.3945/jn.115.221135>
 55. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
 56. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W (2003) Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med* 26(3):172–181. https://doi.org/10.1207/S15324796ABM2603_02
 57. Biesbroek S, van der AD, Broens MC, Beulens JW, Verschuren WM, van der Schouw YT, Boer JM, (2015) Identifying cardiovascular risk factor-related dietary patterns with reduced rank regression and random forest in the EPIC-NL cohort. *Am J Clin Nutr* 102(1):146–154. <https://doi.org/10.3945/ajcn.114.092288>
 58. Hyvarinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13(4–5):411–430. [https://doi.org/10.1016/s0893-6080\(00\)00026-5](https://doi.org/10.1016/s0893-6080(00)00026-5)
 59. Fraser GE, Yan R, Butler TL, Jaceldo-Siegl K, Beeson WL, Chan J (2009) Missing data in a long food frequency questionnaire: are imputed zeroes correct? *Epidemiology* 20(2):289–294. <https://doi.org/10.1097/EDE.0b013e31819642c4>
 60. Stevens J, Ou FS, Truesdale KP, Zeng D, Vaughn AE, Pratt C, Ward DS (2015) A suggested approach for imputation of missing dietary data for young children in daycare. *Food Nutr Res* 59:28626. <https://doi.org/10.3402/fnr.v59.28626>
 61. Bictash M, Ebbels TM, Chan Q, Loo RL, Yap IK, Brown II, de Iorio M, Daviglus ML, Holmes E, Stamler J, Nicholson JK, Elliott P (2010) Opening up the “Black Box”: metabolic phenotyping and metabolome-wide association studies in epidemiology. *J Clin Epidemiol* 63(9):970–979. <https://doi.org/10.1016/j.jclinepi.2009.10.001>
 62. Hu FB (2011) Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology. *Clin Chem* 57(9):1224–1226. <https://doi.org/10.1373/clinchem.2011.167056>
 63. Kho ZY, Lal SK (2018) The Human Gut Microbiome - A Potential Controller of Wellness and Disease. *Front Microbiol* 9:1835. <https://doi.org/10.3389/fmicb.2018.01835>
 64. Altmaier E, Kastenmuller G, Romisch-Margl W, Thorand B, Weinberger KM, Illig T, Adamski J, Doring A, Suhre K (2011) Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol* 26(2):145–156. <https://doi.org/10.1007/s10654-010-9524-7>
 65. O’Sullivan A, Gibney MJ, Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* 93(2):314–321. <https://doi.org/10.3945/ajcn.110.000950>
 66. Floegel A, von Ruesten A, Drogan D, Schulze MB, Prehn C, Adamski J, Pischon T, Boeing H (2013) Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr* 67(10):1100–1108. <https://doi.org/10.1038/ejcn.2013.147>
 67. Menni C, Zhai G, Macgregor A, Prehn C, Romisch-Margl W, Suhre K, Adamski J, Cassidy A, Illig T, Spector TD, Valdes AM (2013) Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* 9(2):506–514. <https://doi.org/10.1007/s11306-012-0469-6>
 68. Andersen MB, Rinnan A, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, Astrup A, Dragsted LO (2014) Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *J Proteome Res* 13(3):1405–1418. <https://doi.org/10.1021/pr400964s>
 69. Kong LC, Holmes BA, Cotillard A, Habi-Rachedi F, Brazeilles R, Gougis S, Gausseres N, Cani PD, Fellahi S, Bastard JP, Kennedy SP, Dore J, Ehrlich SD, Zucker JD, Rizkalla SW, Clement K (2014) Dietary patterns differently associate with inflammation and gut microbiota in overweight and obese subjects. *PLoS ONE* 9(10):e109434. <https://doi.org/10.1371/journal.pone.0109434>
 70. Matijasic BB, Obermajer T, Lipoglavsek L, Grabnar I, Avgustin G, Rogelj I (2014) Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *Eur J Nutr* 53(4):1051–1064. <https://doi.org/10.1007/s00394-013-0607-6>
 71. Schmidt JA, Rinaldi S, Ferrari P, Carayol M, Achaintre D, Scalbert A, Cross AJ, Gunter MJ, Fensom GK, Appleby PN, Key TJ, Travis RC (2015) Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *Am J Clin Nutr* 102(6):1518–1526. <https://doi.org/10.3945/ajcn.115.111989>
 72. Gutierrez-Diaz I, Fernandez-Navarro T, Sanchez B, Margolles A, Gonzalez S (2016) Mediterranean diet and faecal microbiota: a transversal study. *Food Funct* 7(5):2347–2356. <https://doi.org/10.1039/c6fo00105j>
 73. De Filippis F, Pellegrini N, Vannini L, Jeffery IB, La Storia A, Laghi L, Serrazanetti DI, Di Cagno R, Ferrocino I, Lazzi C, Turrioni S, Cocolin L, Brigidi P, Neviani A, Gobbetti M, O’Toole PW, Ercolini D (2016) High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut* 65(11):1812–1821. <https://doi.org/10.1136/gutjnl-2015-309957>
 74. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, Albanes D, Gu F, Kontto J, Lassale C, Liao LM, Mannisto S, Mondul AM, Weinstein SJ, Irwin ML, Mayne ST, Stolzenberg-Solomon R (2017) Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr* 105(2):450–465. <https://doi.org/10.3945/ajcn.116.144501>
 75. Guasch-Ferre M, Bhupathiraju SN, Hu FB (2018) Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin Chem* 64(1):82–98. <https://doi.org/10.1373/clinchem.2017.272344>
 76. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, Hansen T, Beckmann M, Pedersen O, Elliott P, Stamler J, Nicholson JK, Draper J, Mathers JC, Holmes E, Frost G (2017) Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol* 5(3):184–195. [https://doi.org/10.1016/S2213-8587\(16\)30419-3](https://doi.org/10.1016/S2213-8587(16)30419-3)
 77. Lu Y, Zou L, Su J, Tai ES, Whitton C, Dam RMV, Ong CN (2017) Meat and seafood consumption in relation to plasma metabolic profiles in a Chinese population: a combined untargeted and targeted metabolomics study. *Nutrients*. <https://doi.org/10.3390/nu9070683>
 78. Almanza-Aguilera E, Urpi-Sarda M, Llorach R, Vazquez-Fresno R, Garcia-Aloy M, Carmona F, Sanchez A, Madrid-Gambin

- F, Estruch R, Corella D, Andres-Lacueva C (2017) Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a (1)H-NMR-based untargeted metabolomics approach. *J Nutr Biochem* 48:36–43. <https://doi.org/10.1016/j.jnutbio.2017.06.001>
79. Merz B, Frommherz L, Rist MJ, Kulling SE, Bub A, Watzl B (2018) Dietary pattern and plasma BCAA-variations in healthy men and women-results from the KarMeN study. *Nutrients*. <https://doi.org/10.3390/nu10050623>
 80. Wei R, Ross AB, Su M, Wang J, Guiraud SP, Draper CF, Beaumont M, Jia W, Martin FP (2018) Metabotypes related to meat and vegetable intake reflect microbial, lipid and amino acid metabolism in healthy people. *Mol Nutr Food Res* 62(21):e1800583. <https://doi.org/10.1002/mnfr.201800583>
 81. Zhang C, Bjorkman A, Cai K, Liu G, Wang C, Li Y, Xia H, Sun L, Kristiansen K, Wang J, Han J, Hammarstrom L, Pan-Hammarstrom Q (2018) Impact of a 3-months vegetarian diet on the gut microbiota and immune repertoire. *Front Immunol* 9:908. <https://doi.org/10.3389/fimmu.2018.00908>
 82. Johnson RK, Vanderlinden L, DeFelice BC, Kechris K, Uusitalo U, Fiehn O, Sontag M, Crume T, Beyerlein A, Lernmark A, Toppari J, Ziegler AG, She JX, Hagopian W, Rewers M, Akolkar B, Krischer J, Virtanen SM, Norris JM, Group TS (2019) Metabolite-related dietary patterns and the development of islet autoimmunity. *Sci Rep* 9(1):14819. <https://doi.org/10.1038/s41598-019-51251-4>
 83. Macias S, Kirma J, Yilmaz A, Moore SE, McKinley MC, McKeown PP, Woodside JV, Graham SF, Green BD (2019) Application of (1)H-NMR metabolomics for the discovery of blood plasma biomarkers of a Mediterranean diet. *Metabolites*. <https://doi.org/10.3390/metabo9100201>
 84. Tong TYN, Koulman A, Griffin JL, Wareham NJ, Forouhi NG, Imamura F (2019) A combination of metabolites predicts adherence to the Mediterranean Diet pattern and its associations with insulin sensitivity and lipid homeostasis in the general population: the Fenland study, United Kingdom. *J Nutr*. <https://doi.org/10.1093/jn/nxz263>
 85. Shikany JM, Demmer RT, Johnson AJ, Fino NF, Meyer K, Ensrud KE, Lane NE, Orwoll ES, Kado DM, Zmuda JM, Langsetmo L, Osteoporotic Fractures in Men Research G (2019) Association of dietary patterns with the gut microbiota in older, community-dwelling men. *Am J Clin Nutr* 110(4):1003–1014. <https://doi.org/10.1093/ajcn/nqz174>
 86. Matsuyama M, Morrison M, Cao KL, Pruihl S, Davies PSW, Wall C, Lovell A, Hill RJ (2019) Dietary intake influences gut microbiota development of healthy Australian children from the age of one to two years. *Sci Rep* 9(1):12476. <https://doi.org/10.1038/s41598-019-48658-4>
 87. Durack J, Lynch SV (2019) The gut microbiome: Relationships with disease and opportunities for therapy. *J Exp Med* 216(1):20–40. <https://doi.org/10.1084/jem.20180448>
 88. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosover N, Rein M, Zilberman-Schapira G, Dohnalova L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E (2015) Personalized nutrition by prediction of glycemic responses. *Cell* 163(5):1079–1094. <https://doi.org/10.1016/j.cell.2015.11.001>
 89. Mehta RS, Nishihara R, Cao Y, Song M, Mima K, Qian ZR, Nowak JA, Kosumi K, Hamada T, Masugi Y, Bullman S, Drew DA, Kostic AD, Fung TT, Garrett WS, Huttenhower C, Wu K, Meyerhardt JA, Zhang X, Willett WC, Giovannucci EL, Fuchs CS, Chan AT, Ogino S (2017) Association of dietary patterns with risk of colorectal cancer subtypes classified by fusobacterium nucleatum in tumor tissue. *JAMA Oncol* 3(7):921–927. <https://doi.org/10.1001/jamaoncol.2016.6374>
 90. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, Tuohy K (2018) Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr* 57(1):1–24. <https://doi.org/10.1007/s00394-017-1445-8>
 91. Chun H, Keles S (2010) Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Series B Stat Methodol* 72(1):3–25. <https://doi.org/10.1111/j.1467-9868.2009.00723.x>
 92. Chen L, Huang JZ (2012) Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *J Am Stat Assoc* 107(500):1533–1545
 93. Riedl A, Gieger C, Hauner H, Daniel H, Linseisen J (2017) Metabotyping and its application in targeted nutrition: an overview. *Br J Nutr* 117(12):1631–1644. <https://doi.org/10.1017/S0007114517001611>
 94. Christensen L, Roager HM, Astrup A, Hjorth MF (2018) Microbial enterotypes in personalized nutrition and obesity management. *Am J Clin Nutr* 108(4):645–651. <https://doi.org/10.1093/ajcn/nqy175>
 95. Schulze MB, Manson JE, Willett WC, Hu FB (2003) Processed meat intake and incidence of type 2 diabetes in younger and middle-aged women. *Diabetologia* 46(11):1465–1473. <https://doi.org/10.1007/s00125-003-1220-7>
 96. Schulze MB, Hoffmann K, Kroke A, Boeing H (2003) Risk of hypertension among women in the EPIC-Potsdam Study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. *Am J Epidemiol* 158(4):365–373. <https://doi.org/10.1093/aje/kwg156>
 97. Reedy J, Mitrou PN, Krebs-Smith SM, Wirfalt E, Flood A, Kipnis V, Leitzmann M, Mouw T, Hollenbeck A, Schatzkin A, Subar AF (2008) Index-based dietary patterns and risk of colorectal cancer: the NIH-AARP Diet and Health Study. *Am J Epidemiol* 168(1):38–48. <https://doi.org/10.1093/aje/kwn097>
 98. Barbaresko J, Siegert S, Koch M, Aits I, Lieb W, Nikolaus S, Laudes M, Jacobs G, Nothlings U (2014) Comparison of two exploratory dietary patterns in association with the metabolic syndrome in a Northern German population. *Br J Nutr* 112(8):1364–1372. <https://doi.org/10.1017/S0007114514002098>
 99. Judd SE, Letter AJ, Shikany JM, Roth DL, Newby PK (2014) Dietary patterns derived using exploratory and confirmatory factor analysis are stable and generalizable across race, region, and gender subgroups in the REGARDS study. *Front Nutr* 1:29. <https://doi.org/10.3389/fnut.2014.00029>
 100. Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN (1998) Eating patterns and risk of colon cancer. *Am J Epidemiol* 148(1):4–16. <https://doi.org/10.1093/aje/148.1.4-a>
 101. Barker ME, McClean SI, Thompson KA, Reid NG (1990) Dietary behaviours and sociocultural demographics in Northern Ireland. *Br J Nutr* 64(2):319–329. <https://doi.org/10.1079/bjn19900034>
 102. Hamer M, McNaughton SA, Bates CJ, Mishra GD (2010) Dietary patterns, assessed from a weighed food record, and survival among elderly participants from the United Kingdom. *Eur J Clin Nutr* 64(8):853–861. <https://doi.org/10.1038/ejcn.2010.93>
 103. Diethelm K, Gunther AL, Schulze MB, Standl M, Heinrich J, Buyken AE (2014) Prospective relevance of dietary patterns at the beginning and during the course of primary school to the development of body composition. *Br J Nutr* 111(8):1488–1498. <https://doi.org/10.1017/S0007114513004017>
 104. Naja F, Jomaa L, Itani L, Zidek J, El Labban S, Sibai AM, Hwalla N (2018) Environmental footprints of food consumption and dietary patterns among Lebanese adults: a cross-sectional study. *Nutr J* 17(1):85. <https://doi.org/10.1186/s12937-018-0393-3>
 105. Newby PK, Muller D, Hallfrisch J, Qiao N, Andres R, Tucker KL (2003) Dietary patterns and changes in body mass index and

- waist circumference in adults. *Am J Clin Nutr* 77(6):1417–1425. <https://doi.org/10.1093/ajcn/77.6.1417>
106. Chen Z, Wang PP, Shi L, Zhu Y, Liu L, Gao Z, Woodrow J, Roebothan B (2015) Comparison in dietary patterns derived for the Canadian Newfoundland and Labrador population through two time-separated studies. *Nutr J* 14:75. <https://doi.org/10.1186/s12937-015-0064-6>
 107. Schulz M, Hoffmann K, Weikert C, Nothlings U, Schulze MB, Boeing H (2008) Identification of a dietary pattern characterized by high-fat food choices associated with increased risk of breast cancer: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. *Br J Nutr* 100(5):942–946. <https://doi.org/10.1017/S0007114508966149>
 108. Khani BR, Ye W, Terry P, Wolk A (2004) Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *J Nutr* 134(6):1541–1545. <https://doi.org/10.1093/jn/134.6.1541>
 109. Mishra GD, Prynne CJ, Paul AA, Greenberg DC, Bolton-Smith C (2004) The impact of inter-generational social and regional circumstances on dietary intake patterns of British adults: results from the 1946 British Birth Cohort. *Public Health Nutr* 7(6):737–744
 110. Drake I, Sonestedt E, Ericson U, Wallstrom P, Orho-Melander M (2018) A Western dietary pattern is prospectively associated with cardio-metabolic traits and incidence of the metabolic syndrome. *Br J Nutr* 119(10):1168–1176. <https://doi.org/10.1017/S000711451800079X>
 111. Appannah G, Pot GK, O’Sullivan TA, Oddy WH, Jebb SA, Ambrosini GL (2014) The reliability of an adolescent dietary pattern identified using reduced-rank regression: comparison of a FFQ and 3 d food record. *Br J Nutr* 112(4):609–615. <https://doi.org/10.1017/S0007114514001111>
 112. Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC (2000) Prospective study of major dietary patterns and risk of coronary heart disease in men. *Am J Clin Nutr* 72(4):912–921. <https://doi.org/10.1093/ajcn/72.4.912>
 113. Schrijvers JK, McNaughton SA, Beck KL, Kruger R (2016) Exploring the dietary patterns of young New Zealand women and associations with BMI and body fat. *Nutrients*. <https://doi.org/10.3390/nu8080450>
 114. Kass RA, Tinsley H (1979) Factor analysis. *J Leis Res* 11:120–138
 115. Dao MC, Subar AF, Warthon-Medina M, Cade JE, Burrows T, Golley RK, Forouhi NG, Pearce M, Holmes BA (2019) Dietary assessment toolkits: an overview. *Public Health Nutr* 22(3):404–418. <https://doi.org/10.1017/S1368980018002951>
 116. Pawellek I, Grote V, Theurich M, Closa-Monasterolo R, Stolarczyk A, Verduci E, Xhonneux A, Koletzko B (2017) Factors associated with sugar intake and sugar sources in European children from 1 to 8 years of age. *Eur J Clin Nutr* 71(1):25–32. <https://doi.org/10.1038/ejcn.2016.206>
 117. Luque V, Escribano J, Closa-Monasterolo R, Zaragoza-Jordana M, Ferre N, Grote V, Koletzko B, Totzauer M, Verduci E, ReDonigi A, Gruszfeld D, Socha P, Rousseaux D, Moretti M, Oddy W, Ambrosini GL (2018) Unhealthy dietary patterns established in infancy track to mid-childhood: the EU Childhood Obesity Project. *J Nutr* 148(5):752–759. <https://doi.org/10.1093/jn/nxy025>
 118. Korkalo L, Vepsalainen H, Ray C, Skaffari E, Lehto R, Hauta-Alus HH, Nissinen K, Meinila J, Roos E, Erkkola M (2019) Parents’ reports of preschoolers’ diets: relative validity of a food frequency questionnaire and dietary patterns. *Nutrients*. <https://doi.org/10.3390/nu11010159>
 119. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* 69(2):243–249. <https://doi.org/10.1093/ajcn/69.2.243>
 120. John EM, Stern MC, Sinha R, Koo J (2011) Meat consumption, cooking practices, meat mutagens, and risk of prostate cancer. *Nutr Cancer* 63(4):525–537. <https://doi.org/10.1080/01635581.2011.539311>
 121. Liu G, Zong G, Wu K, Hu Y, Li Y, Willett WC, Eisenberg DM, Hu FB, Sun Q (2018) Meat cooking methods and risk of type 2 diabetes: results from three prospective cohort studies. *Diabetes Care* 41(5):1049–1060. <https://doi.org/10.2337/dc17-1992>
 122. Tabatabaei SM, Fritschi L, Knuiaman MW, Boyle T, Iacopetta BJ, Platell C, Heyworth JS (2011) Meat consumption and cooking practices and the risk of colorectal cancer. *Eur J Clin Nutr* 65(6):668–675. <https://doi.org/10.1038/ejcn.2011.17>
 123. Sonestedt E, Wirfalt E, Wallstrom P, Gullberg B, Orho-Melander M, Hedblad B (2011) Dairy products and its association with incidence of cardiovascular disease: the Malmo diet and cancer cohort. *Eur J Epidemiol* 26(8):609–618. <https://doi.org/10.1007/s10654-011-9589-y>
 124. Tholstrup T (2006) Dairy products and cardiovascular disease. *Curr Opin Lipidol* 17(1):1–10. <https://doi.org/10.1097/01.mol.0000199813.08602.58>
 125. Willett W, Rockstrom J, Loken B, Springmann M, Lang T, Vermeulen S, Garnett T, Tilman D, DeClerck F, Wood A, Jonell M, Clark M, Gordon LJ, Fanzo J, Hawkes C, Zurayk R, Rivera JA, De Vries W, Majele Sibanda L, Afshin A, Chaudhary A, Herrero M, Agustina R, Branca F, Lartey A, Fan S, Crona B, Fox E, Bignet V, Troell M, Lindahl T, Singh S, Cornell SE, Srinath Reddy K, Narain S, Nishtar S, Murray CJL (2019) Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. *Lancet* 393(10170):447–492. [https://doi.org/10.1016/S0140-6736\(18\)31788-4](https://doi.org/10.1016/S0140-6736(18)31788-4)
 126. Seconda L, Egnell M, Julia C, Touvier M, Herberg S, Pointereau P, Lairon D, Alles B, Kesse-Guyot E (2019) Association between sustainable dietary patterns and body weight, overweight, and obesity risk in the NutriNet-Sante prospective cohort. *Am J Clin Nutr*. <https://doi.org/10.1093/ajcn/nqz259>
 127. Seconda L, Baudry J, Pointereau P, Lacour C, Langevin B, Herberg S, Lairon D, Alles B, Kesse-Guyot E (2019) Development and validation of an individual sustainable diet index in the NutriNet-Sante study cohort. *Br J Nutr* 121(10):1166–1177. <https://doi.org/10.1017/S0007114519000369>
 128. Fresan U, Martinez-Gonzalez MA, Sabate J, Bes-Rastrollo M (2018) The Mediterranean diet, an environmentally friendly option: evidence from the Seguimiento Universidad de Navarra (SUN) cohort. *Public Health Nutr* 21(8):1573–1582. <https://doi.org/10.1017/S1368980017003986>
 129. Ford ES, Zhao G, Tsai J, Li C (2011) Low-risk lifestyle behaviors and all-cause mortality: findings from the National Health and Nutrition Examination Survey III Mortality Study. *Am J Public Health* 101(10):1922–1929. <https://doi.org/10.2105/AJPH.2011.300167>