Reproducibility of *a posteriori* dietary patterns across time and studies: a scoping review[a]

Valeria Edefonti[1], Roberta De Vito[2,3], Andrea Salvatori[1], Francesca Bravi[1], Linia Patel[1], Michela Dalmartello[1] and Monica Ferraroni[1]

Review Articles

[1] Branch of Medical Statistics, Biometry and Epidemiology ”G. A. Maccacaro”, Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milano, Italy;

[2] Department of Biostatistics, Brown University, Providence, RI, USA;

[3] Data Science Initiative, Brown University, Providence, RI, USA;

Corresponding Author: Valeria Edefonti, Branch of Medical Statistics, Biometry and Epidemiology ”G. A. Maccacaro”, Department of Clinical Sciences and Community Health, Università degli Studi di Milano, via Venezian 1, 20133 Milano, Italy; telephone: 0039 02-

[a] Supplemental Materials and Methods, Supplemental Tables 1, 2, and 3, and Supplemental Figure 1 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/advances

50320853; fax: 0039 02-50320866; email: valeria.edefonti@unimi.it.

Word count: 6724 words.

Number of figures: 1 figure.

Number of tables: 4 tables.

Running title: consistency of dietary patterns in time or studies

List of abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; CA: cluster analysis; CFA: confirmatory factor analysis; DP: dietary pattern; EFA: exploratory factor analysis; EPIC: European Prospective Investigation into Cancer and Nutrition; FFQ: food-frequency questionnaire; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NHS: Nurses' Health Study; PCA: principal component analysis; PCFA: principal component factor analysis; SMC: Swedish Mammography Cohort; SWS: Southampton Women's Survey

List of financial support: Valeria Edefonti was supported by Università degli Studi di Milano 'Young Investigator Grant Program 2017'.

Conflicts of interest: The authors have declared no conflicts of interest.

Abstract - 299 words

Few papers have considered if *a posteriori* dietary patterns (DPs) are generalizable across different centers or studies, or if they are consistently seen over time. To date, no systematic search of the literature on these topics has been carried out.

A scoping review was conducted through a systematic search on the PubMed database. In the current paper, we included the 34 articles examining the extent to which *a posteriori* DPs were consistently seen: 1. across centers from the same study or across different studies potentially representing different populations or countries (here indicated as cross-study reproducibility); and 2. over longer time periods (i.e., ≥2 years) (here indicated as stability over time).

Selected articles (published in 1981–2019, 32% from 2010 onwards) were based on observational studies, mostly from Europe and North America. Five articles were based on children and/or adolescents and 14 papers included adults (2 men; 12 women, of which 3 on pregnant women). *A posteriori* DPs were mostly derived (32 papers) with principal component or factor analyses.

Among the 9 articles assessing DP reproducibility across studies (number of centers/studies: 2-27, median: 3), 5 provided a formal assessment using statistical methods (4 index-based approaches of different complexity, 1 statistical model). A median of 4 DPs was reproduced across centers/studies (range: 1-7). Among the 25 articles assessing DP

stability over time (number of time-occasions: 2-6, median: 3), 19 provided a formal assessment with statistical methods (17 index-based and/or test-based approaches, 1 statistical model, 1 with both strategies). The number and composition of DPs remained mostly stable over time.

Based on the limited evidence collected, most identified DPs showed good reproducibility across studies and stability over time. However, when present within the single studies, the criteria for the formal assessment of cross-study reproducibility or stability over time were generally very basic.

***Keywords (5-10)*:**

a posteriori dietary patterns; cluster analysis; consistency of dietary patterns; cross-study reproducibility of dietary patterns; factor analysis; generalizability of dietary patterns; reproducibility of dietary patterns; reproducibility of dietary patterns across studies; stability of dietary patterns over time.

Introduction

Over the last twenty years, the analysis of dietary patterns (DPs)[b] has provided a complementary strategy to the traditional single-food or single-nutrient approach. Use of dietary patterns captures the intrinsic complexity of diet, the potential synergistic effects between its different components as well as the variability in DPs existing within and between populations (1).

The *a posteriori* (or empirically derived) DPs are obtained from the application of multivariate statistics [e.g., principal component analysis (PCA), exploratory factor analysis (EFA), or cluster analysis (CA)] to the available dietary data (2). Therefore, a meaningful set of *a posteriori* DPs synthesizes the different aspects of the *actual* dietary behavior, as measured at a single time-point reflecting recent dietary habits of a population. Compared to the *a priori* DPs (i.e., comparing subjects' diet against evidence-based benchmark diets) or to the mixed-type reduced rank regression (i.e., using *a priori* knowledge on a set of response variables whose variation has to be maximized within a PCA-like multivariate approach to regression) (3), the *a posteriori* DPs are less prone to be generalized to different populations or over time. Indeed, actual DPs reflect the food supply, geography/climate, socio-economic

[b] ALSPAC: Avon Longitudinal Study of Parents and Children; CA: cluster analysis; CFA: confirmatory factor analysis; DP: dietary pattern; EFA: exploratory factor analysis; EPIC: European Prospective Investigation into Cancer and Nutrition; FFQ: food-frequency questionnaire; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NHS: Nurses' Health Study; PCA: principal component analysis; PCFA: principal component factor analysis; SMC: Swedish Mammography Cohort; SWS: Southampton Women's Survey

status, ethnicity, religion, impact of media and society, changes in policy that affect dietary habits, etc. (4). In combination with biological mechanisms, these latent factors are responsible for any differences in both the number and structure of DPs identified across populations and also over time.

Given the considerable body of evidence on the topic, the time is now ripe to summarize evidence on the specific dimensions of generalizability of *a posteriori* dietary patterns, including their reproducibility and validity. In the absence of a consensus on these definitions, we have initiated the first scoping review on reproducibility and validity of *a posteriori* DPs. After clarifying basic terminology and the use of terms in nutritional epidemiology (**Supplemental Table 1** and **Supplemental Figure 1**), evidence was summarized into two papers. The current review examined the extent to which similar DPs are consistently seen: 1. across centers from the same study or across different studies potentially representing different populations or countries (here indicated as cross-study reproducibility); and 2. over longer time periods (i.e., ≥2 years) (here indicated as stability over time). A recently published companion paper has synthesized evidence on other forms of reproducibility [e.g., across different statistical solutions or in a short-term period (i.e., <2 years)], relative validity, and construct validity of *a posteriori* DPs (5) (see Supplemental Table 1 and Supplemental Figure 1 for additional definitions).

Besides providing a summary of the existing literature, we have focused the two reviews on statistical methods for the assessment of generalizability of *a posteriori* DPs. While real-life

factors are the main drivers of this issue, from the statistical standpoint, the assessment of generalizability is fraught with difficulties that should be clarified to distinguish true differences in time or space from artifacts or noise. Firstly, results depend on subjective decisions (e.g., data preprocessing or not, multivariate statistical approach to use, algorithm to carry out the analysis, number of DPs to retain) taken during the DP identification process within the single studies. However, some pioneer papers adopting a standardized approach to DP identification across studies (6-8) have already shown that 2 to 4 DPs were consistently identified across similar cohorts in Europe. Similarly, in the assessment of stability of DPs over time, the use of the same statistical approach to DP identification has allowed attributing any differences (including those from artifacts of subjective decisions) to true differences. This consistency in the statistical approach has already contributed to identifying sets of reproducible DPs across multiple administrations of the same dietary assessment tool up to 6-7 years of follow-up (e.g., (9, 10)).

Secondly, evaluations of generalizability of *a posteriori* DPs should be based on ad-hoc statistical methods tailored to disentangle the true differences in time or populations from time-specific or study-specific effects or simpler artifacts. A few novel methods have been proposed for the assessment of reproducibility of *a posteriori* DPs across studies (8, 11-14), including the use of the congruence coefficient for factor loading comparison. Despite the several challenges to confront with - including individual and population-specific dimensions of stability (e.g. (15, 16)) as well as transitions of target populations to a later stage in life

(e.g., (16-18)), fewer research efforts have been focused on methods for the assessment of DP stability over time.

To compensate for these issues, more recent evaluations of generalizability of DPs over time and/or across studies are more likely to be sound and fair. Indeed, since the early 2000s, some researchers have investigated the effect of single subjective decisions in performing PCA and EFA (e.g., (19-21)). Particularly, confirmatory factor analysis (CFA) has been more often proposed in the validation of sensible (possibly, EFA-based) constructs representing correlation structures among food groups and among DPs (e.g. (22, 23)). These examples indicated to us that a scoping review on reproducibility and validity of *a posteriori* DPs would have been feasible.

The current paper has two aims: 1. summarizing the evidence on reproducibility of *a posteriori* DPs across studies and their stability over time; 2. providing a focus on statistical methods to assess reproducibility of DPs across studies and their stability over time.

Materials and methods

Literature search strategy

A scoping review was conducted using a systematic search of the literature through MEDLINE via PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) to identify all the articles on reproducibility and validity of *a posteriori* DPs, based on the following string: *"(reproducibility or validity) and dietary pattern*"*. The guidelines from the Preferred Reporting Items for

147 Systematic Reviews and Meta-Analyses (PRISMA) group were followed (24). The search
148 was restricted to human studies reported in the English language and published up to
149 January 11, 2019. Two authors (MD and VE) independently screened titles and then
150 abstracts and retrieved the potentially relevant articles. The reference lists of the identified
151 articles and other systematic reviews based on similar topics were also scanned.
152 Discrepancies were resolved by involving a third researcher (MF).

153 ***Inclusion and exclusion criteria***

154 Articles were included or excluded based on the following criteria.

155 *A posteriori dietary patterns*

156 We focused our scoping review on *a posteriori* DPs. However, in the absence of previously
157 published reviews on this topic, we preferred not to add the term *"a posteriori"* to our search
158 string. Therefore, we further had to exclude papers presenting reproducibility or validity of *a*
159 *priori* DPs only, or applying reduced rank regression, or treelet transform.

160 *Reproducibility and validity of a posteriori dietary patterns*

161 In the current review, we summarized evidence on cross-study reproducibility of DPs
162 (including both reproducibility across centers from a multicentric study and reproducibility
163 across different studies), and stability of DPs over time. Supplemental Table 1 and
164 Supplemental Figure 1 provide an overview of the general terminology used in this review
165 and of its use in nutritional epidemiology. Definition and use of terms introduced in our earlier
166 review (5) (i.e. reproducibility across different statistical methods, short-term reproducibility,

relative validity, and construct validity) were also presented within the **Supplementary Material**. We also chose not to exclude studies on the basis of their quality, because of the lack of previous evidence on reproducibility and/or validity of DPs.

Stability of dietary patterns over time: possible forms of assessment

Table 1 provides a detailed description of the different levels of analysis available within an assessment of stability of DPs. In detail, when the primary research question is to target potential transitions of subjects from one DP to another DP over time (individual-level stability analysis), the most straightforward approach is to apply a CA and to track changes by calculating the percentages of transitioners (or stable eaters) across successive time-points. When the primary aim is to describe potential changes over time in the covariance structure among dietary items within a population (population-level stability analysis), the most suitable approach is to apply PCA/EFA; changes can be tracked through the monitoring of the following aspects (in order of importance): 1. number of identified DPs: are there DPs gained or lost?; 2. percentage of explained variance by each DP: do stable DPs show similar percentages over time?; 3. DP composition: are similar DPs characterized by the same relevant food groups or nutrients? Or are factor loadings similar or congruent over time?; 4. DP scores: do the mean DP scores change (e.g., increase or decrease following some path) over time? Additional levels of complexity may arise when important changes in the life-course (e.g., from childhood to adolescence, or before and after pregnancy) happen within the period of observation. Within these designs, secular trends can be tracked

identifying parallel sub-cohorts of different ages at baseline and comparing DPs derived on the sub-cohorts considered at the same age-period.

Data extraction

Quantitative and qualitative data were extracted from the selected studies for in-depth review by 3 independent researchers (LP, MD, and VE); any discrepancies were resolved after consultation with a fourth author (MF) to maintain consistency. Information extracted included the following: 1. general characteristics of the studies (first author, year of publication of the article, country, and study name); 2. study design and characteristics (type of design, data collection, study location, number and age of the participants, and years of follow-up); 3. dietary assessment tools used; 4. dietary pattern identification method; 5. dietary pattern name and composition; 6. statistical methods used for the assessment of reproducibility of DPs; and 7. main results on DP reproducibility.

Results

Study selection process

Figure 1 shows the flowchart of the study selection process carried out within the systematic search of the literature supporting this scoping review. From the PubMed database literature search, we identified 218 articles, of which 181 remained for detailed evaluation after the search was limited to human studies and articles written in the English language. Thirty-five review articles were removed, and 124 original research articles were also not included

because they met the exclusion criteria. The most frequent reasons for exclusion were previously described in detail in the companion review (5). Forty-two additional articles were identified from manual searches of reference lists of selected original and review articles. Thus, 64 articles were included in our scoping review. Of these, the 34 articles that focused on stability of DPs over time and on their reproducibility across studies were included in this review, whereas the 38 articles on reproducibility, relative and construct validity of DPs were included in the companion paper (5). Eight papers (6, 9, 10, 22, 23, 25-27) were common to both reviews.

Main characteristics of the included studies

General characteristics and study design information from the 34 papers on stability and cross-study reproducibility of DPs (6-12, 15-18, 22, 23, 25-45) are presented in **Table 2**. The selected papers were published between 1981 and 2019, with 32% of them published from 2010 onwards; the studies were mostly carried out across Europe and North America. Several articles were based on the same studies, including (but not limited to) those from the Swedish Mammography Cohort (SMC) (6, 7, 9, 22, 23), the Avon Longitudinal Study of Parents and Children (ALSPAC) (17, 18, 39, 40), and the Nurses' Health Study (NHS) I and II (35, 36, 38, 42). All the articles were based on observational studies, including 1 case-control (32), 24 cohort (6-10, 15-18, 22, 23, 26, 28, 30, 31, 33, 35-40, 42, 45) and 2 cross-sectional (43, 44) studies; in addition, there were 3 multiple administrations of the same survey (27, 34, 41), 1 validation study of the SMC food-frequency questionnaire (FFQ) (25),

and 3 papers including studies with different designs (11, 12, 29). Two articles included adult men only (33, 45), 12 included adult women only (9, 11, 12, 15, 22, 23, 28, 30, 35, 36, 38, 40), with 3 studies based on pregnant women (15, 30, 40); five papers considered the recruitment of children and/or adolescents (16-18, 31, 39). With a few exceptions (16, 18, 30, 37, 43, 44), dietary information was collected with a FFQ. The FFQs were self-administered (except for the Southampton Women's Survey (SWS) (15, 28)); the reference period of assessment was generally of 1 year, except for diet during pregnancy (15, 28) or the high school period (36, 38). The number of food items inquired in the FFQs ranged from 26 (27, 34) to 276 (6), with a median value of 111.5 items. When more than 1 FFQ administration was available from cohort studies, the time-interval between successive administrations could be fixed or variable [range of the minimum distance between dietary data used for DP identification: 1 month (during pregnancy) (30) - 7 years (37)]. The reproducibility and/or relative validity of the FFQs were assessed within 1 validation study included in the review (25); in addition, 20 articles reported information on FFQ reproducibility and/or relative validity (6-12, 15, 22, 23, 26, 29, 31-33, 35, 36, 38, 42, 45). Dietary patterns were based on data collected through a dietary record and/or a recall of 24 or 48 hours in 6 articles (16, 18, 30, 37, 43, 44).

Irrespective of the dietary assessment tool used, the number of food groups defined from the available food items ranged from 15 (43, 44) to 152 (31), with a median value of 37 food groups included in the statistical analysis.

Tables 2 and **3** present details on the DP identification process, on the methods for the assessment of DP reproducibility and validity, and on the results of the assessment. Details on DP composition are presented in **Supplemental Tables 2** and **3**. Among the 34 articles included, 32 performed PCA, EFA or CFA and 2 performed CA (10, 18).

Cross-study reproducibility of dietary patterns

Table 3 concerned the 9 articles on cross-study reproducibility of *a posteriori* DPs. All the papers applied PCA or EFA, and one (26) added a CFA to validate results from a previous EFA. The number of involved centers or studies ranged from 2 (12, 43) to 27 (8), with a median of 3 centers/studies included per article.

Identification of dietary patterns across centers or studies

In the easiest set-up (6, 7, 43, 44), separate PCA/EFAs were carried out for each available study/center following the same approach and results were further explored for potential similarities. Within the European Prospective Investigation into Cancer and Nutrition (EPIC) (8), an “overall PCA” (based on the merged data matrix) was compared with the separate center-specific PCAs using the Krzanowski’s index, which measures the proportion of variance captured by the center-specific DPs which is also captured by the overall-PCA-based DPs. A similar approach was used in a study from the US (26) to assess the importance of population subgroups of interest (i.e., region, sex, and race) in identifying separate sets of DPs.

Another 2 companion papers from Spain formally explored: 1. the cross-study reproducibility

of PCA-based DPs in 2 different samples extracted from similar Spanish populations (12), and 2. the applicability of three “internal” DPs derived from the previous Spanish case-control study (12) to independent (“external”) populations with similar characteristics from France, the United States and Sweden (as identified by a bibliographic search of the literature on the association between DPs and breast cancer) (11). The former paper (12) applied a bootstrap-based approach to compare results from separate study-specific PCAs based on the same food-grouping scheme. The latter paper (11) proposed to reconstruct the “external” DP scores as linear combinations of the published DP loadings and consumption of the published food groups, as re-calculated on the dietary data from the Spanish study. Similarly, the authors re-calculated the “external” DP loadings as based on the reference set of Spanish food groups to allow for direct comparison between loadings (11).

Finally, when individual-level data were available from studies of the same collaborative project, multi-study factor analysis was proposed in one paper (32) to extend standard maximum-likelihood EFA and allowed for a partial sharing of EFA-based DPs across studies. Some DPs were derived to be common across all the studies; in addition to them, each study may express extra study-specific DPs. The number of shared and study-specific DPs was identified using a combination of standard criteria for EFA and information criteria for model selection (32).

The number of described DPs ranged from 2 (7) to 8 (44), with a median of 4 DPs per article;

two articles (6, 7) reported the presence of additional population-specific DPs not described in detail (Supplemental Table 2).

Assessment of cross-study reproducibility of dietary patterns

Four papers (6, 7, 43, 44) did not formally assess cross-study reproducibility and concluded that the study-specific sets of PCA/EFA-based DPs were qualitatively similar based on loadings and percentages of explained variances. A formal assessment was carried out in the remaining 5 papers (8, 11, 12, 26, 32). Congruence coefficients between factor loadings and correlation coefficients between factor scores were used in 3 papers (11, 12, 26), whereas the other 2 papers used the Krzanowski's index (8) and multi-study factor analysis (13), respectively. The aim of the analyses was also different across the 5 papers. In 2 articles (8, 26) the statistical analysis was meant to support an overall PCA/EFA model where the single centers/studies were merged in one database. Another 2 studies (11, 12) were aimed at testing the extent to which *a posteriori* DPs are generalizable within and between countries. One paper (32) was in between the 2 approaches as it was focused on an assessment of cross-study reproducibility in an international context as in (11), however the availability of consortia data allowed to fit a statistical model that accounted simultaneously for common and study-specific DPs.

Summary of the evidence on cross-study reproducibility of dietary patterns

No matter of the statistical approach used, the number of DPs reproduced across the studies ranged from 1 (12) to 7 (43), with a median value of 4 common DPs identified. In addition,

2 papers (6, 32) described 1 (32) and 4 (6) DPs that were reproducible among subsets of the included studies. Among the reproducible DPs, most studies identified variants of a Western-like DP (6-8, 11, 12, 26, 32) and/or a Prudent-like DP (6-8, 11, 26, 32, 43, 44); in addition, some papers identified a variant of a *Fat-* or *Condiment-based* DP (8, 11, 26, 32, 43, 44), whereas another paper added to its reproducible set of DPs a *Traditional (Southern)* and *Alcohol/Salads* DP across 8 US regions (26).

Stability of dietary patterns over time

Table 4 presents details on stability of DPs over time (9, 10, 15-18, 22, 23, 25, 27-31, 33-42, 45). With the exception of 2 papers applying CA (10, 18), all the articles derived DPs from PCA or principal component factor analysis (PCFA) or EFA; 4 articles additionally derived DPs with CFA (9, 22, 23, 27). Time-points when DPs were identified ranged from 2 (9, 22, 23, 25, 27-29, 31, 35, 39-41) to 6 (30), with a median of 3 time-occasions included in the stability analysis.

Identification of dietary patterns over multiple time-occasions

With the exception of a single paper (27), DPs were separately identified at each time-point following the same standardized approach across time-occasions. While most of the papers simply proposed separate time-specific statistical analyses (9, 16, 17, 22, 23, 29-31, 33-36, 38, 39, 42, 45), a few others proposed either applied (15, 25, 28, 40, 41) or simplified (37) scores to harmonize PCA- or EFA-based DPs derived at different time-points. As opposed to standard or “natural” scores, applied scores were calculated at a later time-point

327 combining loadings from a PCA/EFA at a previous (analysis at 2 time-points) or reference
328 time-point (analysis at 3 or more time-points) with dietary information at the current time-
329 point (40); at a fixed time-point, simplified scores (46) were calculated as an unweighted
330 sum of dominant food groups, where only the sign (and not the value) of the loading is used.
331 To further improve comparability of DPs at different time-points, the paper by Togo et al. (27)
332 used a mean-structure CFA model that allowed the jointly modelling of dietary data at the 2
333 time-points within a formal statistical approach that explored trends in (potentially correlated)
334 DP scores across time.
335 The number of described DPs ranged from 2 to 6, with 11 of the articles naming and
336 describing 2 DPs; however, in 5 articles (9, 15, 16, 22, 23), the authors reported additional
337 DPs not common to all time-points and/or not relevant/interpretable (Supplemental Table 3).
338 The described DPs were generally similar across time-points in terms of factor loadings and
339 percentages of explained variance; their names reflected these similarities. Some variation
340 in DP composition was reported, either leading to a change in the DP name across time-
341 points (29) or not (16, 22, 30, 31, 36, 40). Additional DPs were identified at earlier (17) and/or
342 later time-points (17, 25, 29, 31); some other DPs were lost at later time-points (17, 30, 40)
343 (Supplemental Table 3).
344 *Assessment of stability over time: dietary patterns and their relevant food groups*
345 Six articles (22, 29, 31, 35, 39, 42) did not formally assess stability of DPs over time; except
346 for one DP in 2 studies (22, 29), the main conclusion from these papers was that the time-

specific sets of PCA/EFA-based DPs were qualitatively similar based on loadings and percentages of explained variances.

A formal assessment of DP stability was carried out in the remaining papers. The number of criteria used to assess stability ranged from 1 to 5, with a median value of 2 criteria under consideration. Intra-class (25), Spearman (9, 15-17, 28, 36), or Pearson (23, 33, 38, 40, 45) correlation coefficients between factor scores and congruence coefficients between factor loadings (30) were the most used criteria across papers. Four articles considered the change in mean factor scores over the period and assessed stability with a paired t-test or within a regression model (17, 34, 40, 41). The Bland-Altman method, with 95% limits of agreement, was presented in 4 papers (15, 17, 28, 40). Proportions of subjects classified into the same, adjacent, or opposite category of factor scores over subsequent time-occasions and/or corresponding Kappa coefficient were used in 5 articles applying PCA/EFA (16, 17, 25, 37, 40); similarly, when CA was applied, transitions of individuals between DPs over time were described as proportions of stable eaters or transitioners across time-occasions in 2 papers (10, 18), also combined with a sequence index plot to illustrate graphically changes in cluster membership (18).

In addition to these standard approaches, the assessment of stability over time of DPs might include a detailed analysis of trend of consumption of the most relevant food groups within each DP. Among possible approaches to assess differences in food group consumption within each DP, authors modelled the number of relevant food groups (37), the mean intake

of relevant food groups (9, 10, 23, 30), or the mean change in relevant food group intakes (10, 18) across time-occasions. One paper (10) stratified the analysis of trends of consumption by stable eaters or not.

Finally, when a CFA was carried out together with EFA, it was possible to assess DP stability within a more refined model where changes in the time-specific covariance matrices were assessed (9) or changes were directly modeled within a mean-structure factor analysis model (27).

Summary of the evidence on stability over time: dietary patterns

Besides the weak evidence from the 6 articles (22, 29, 31, 35, 39, 42) based on a qualitative assessment, a summary of the evidence from papers formally evaluating DP stability was provided below. In addition to (31, 39), the stability of DPs from childhood onwards was formally evaluated in 3 papers (16-18), with 2 of them exploring the issue in subjects that moved from childhood to adolescence (18) or from childhood/adolescence to adulthood (16). The main conclusions were the following ones: 1. During childhood, the identified DPs were very stable, with the highest agreement found between successive waves (4 and 7 years; 7 and 9 years) and for the *Health-conscious* DP (17); 2. From childhood to adolescence, the number of children remaining in the same cluster across time-occasions was still reasonably high, with the greatest stability found for the *Healthy* cluster (33% of subjects in the same cluster at all 3 ages) (18); 3. From childhood/adolescence to adulthood (~20 year period), both the correlation coefficients between time-specific scores and the proportion of subjects

remaining in the extreme quintiles over time pointed to DP stability, with the highest stability found for the uppermost quintile category of subjects and for the 15-18 years old subjects at baseline (16).

Two papers (36, 38) explored the stability of DP from the high-school period to adulthood based on the NHS II. Women between 34 and 53 years were asked to fill in a reproducible and valid FFQ tailored to the high-school period. The comparison of the high-school DPs with those derived in successive waves during the next 10 years provided correlation coefficients between 0.30 and 0.40, with better results for the *Prudent* DP (36, 38).

In addition, 3 articles assessed the stability of DPs around the pregnancy period (15, 30) and up to 4 years of the child (40). Results suggested high stability of DPs identified within this timeframe. Exceptions were the following ones: 1. a *High-energy* DP was significantly increased in late pregnancy, as compared with before or early pregnancy, and had wider limits of agreement than a *Prudent* DP (15); 2. at 4 years of the child, women had a significantly lower score on a *Health-conscious* DP (40).

Finally, 11 papers (9, 10, 23, 25, 27, 28, 33, 34, 37, 41, 45) assessed the stability of DPs identified in successive waves on adult men and/or women. Three of them (25, 27, 34) showed instability over time for most or all the identified DPs. In detail, at 12 years from the validation study of the Teheran Lipid and Glucose Study, the *Iranian traditional* DP was found to be unreproducible according to all criteria, whereas quintile categories of the *Western* DP showed poor agreement over time (25). Going from the 1982-84 to the 1987-

88 survey of the Danish MONItoring of trends and determinants in CArdiovascular Disease (MONICA) study (27), increasing mean scores were found for the *Green* DP, but the *Traditional* (in men) and the *Sweet-Traditional* (in women) DPs showed decreased mean scores, within an overall mean-structure CFA model. However, while going from the 1982-84 to the 1991-92 survey of the Danish MONICA study, both men and women showed the same trend of increasing consumption of *Coarse Bread*, *Pasta and Rice*, and *Baked Goods and Sweets* DPs at the expense of a decrease in mean intakes of the *Meat, Potatoes, and Fats* DP and the *Breakfast* DP (34). In one pioneering paper that compared 2 consecutive US surveys (41), 2 [*Component 1 (high in fruit and vegetables)* and *Component 4 (high in sugary foods)*] out of the 4 identified DPs increased over time more than it would have been expected for the 7-year advance in age.

A weaker form of instability over time concerned single DPs within a set of substantially stable DPs. This issue was evident across the '80s and '90s for the *Meat, Potatoes and Sweet Foods* DP in 36-years old females from the United Kingdom over 17 years of follow-up (37) and for the *Western/Swedish* DP in 52-years old females from Sweden over 9 years of follow-up (23). Finally, several studies (9, 10, 28, 33, 45) showed good stability of all DPs found during adulthood.

When identified (SMC study) (9, 23), the *Alcohol* DP showed the best reproducibility; however, the more refined analysis of changes in the time-specific covariances matrices revealed instability after 7 years in one (9) of the papers. With 2 exceptions (23, 37), the

Western-like (e.g., *Western*, *High-energy, Low-fiber Bread*, *Meat, Potatoes and Sweet Foods,* and *Western/Swedish*) and the Prudent-like (e.g., *Prudent*, *High-fiber Bread*, *Healthy*, and *Fruit, Vegetables and Dairy*) DPs generally showed a similar and moderate stability over time. Traditional-like DPs (e.g., *Iranian traditional*, *Sweet-traditional*, and *Traditional*) were less likely to be stable over time (18, 25, 27, 40).

In addition, most of the papers with 3 or more measurement occasions (i.e., (16, 17, 33, 37, 45)) showed that the agreement was higher when the DPs were identified on data from consecutive, as compared to non-consecutive, waves.

Finally, the use of applied versus natural scores in PCA was formally explored in 2 papers (15, 40). The former paper suggested similar ranges of correlation coefficients for natural and applied scores (15), whereas the latter paper provided inconclusive results (40).

Summary of the evidence on stability over time: relevant food groups within dietary patterns

The analysis of trends of consumption of relevant food groups within each DP (9, 10, 18, 23, 30, 37) supported or further strengthened results on DP stability over time. When the DPs were stable (9, 10, 30), no material differences in mean consumption of relevant food groups were found in one paper (30) or less than a half of them underwent significant changes (9, 10). When one DP was not stable over time (23, 37), the mean intakes (23) (or the number (37)) of relevant food groups changed over time, and this also had an impact on relevant food groups for other DPs over time; a change might also occur in the number of relevant food groups that characterized stable DPs over time, reflecting an increasing variety in

447 consumption over time within the same DP (37). When moving from childhood to
448 adolescence, the mean amount of food groups consumed generally increased over time,
449 but the foods in each cluster with higher- and lower-than average consumptions were similar
450 at each age (18).
451
452 **Conclusions**
453 The present scoping review provides a preliminary summary of the current results on
454 reproducibility of *a posteriori* DPs across studies and over long time-periods. The evidence
455 collected is still limited, with only 9 papers identified on cross-study reproducibility. In
456 addition, only 55% (cross-study reproducibility) and 76% (stability over time) of the papers
457 adopted a formal statistical approach, which, however, relied on elementary statistics (i.e.,
458 correlation coefficients) in most of the cases and on a statistical model in 3 papers only.
459 Based on the evidence collected, most identified DPs (in particular, *Alcohol*, *Prudent*, and
460 *Western* DPs) showed good reproducibility across studies and stability over time.
461 The assessment of cross-study reproducibility has gained recent attention in the literature
462 (8, 11, 12, 26, 32), after some sparse pioneering attempts in the ’80 (43, 44) and ‘00s (6, 7).
463 Recent papers (8, 11, 12, 26, 32) have definitely confirmed the merits of the assessment of
464 cross-study reproducibility of PCA/EFA-based DPs. Besides having found a high
465 congruence between apparently similar pairs of DPs in terms of food composition and
466 association with cancer risk, some novelties in methods (11, 12, 32) have been introduced.

These include multi-study factor analysis (13) [when individual-level dietary data are available, see the corresponding R package “MSFA” (13) from GitHub] and the approach by Castello and colleagues (11, 12) [when published factor-loading matrices and food-grouping schemes are available, see the Supplementary Material of (11)]. Moreover, following two papers identified in the current review (26, 30), Castello and colleagues (11, 12) popularized the use of the congruence coefficient between factor loadings to assess DP similarity. In addition to set-up specific cut-offs to identify DP similarity or equivalence, they showed that the congruence coefficient outperforms the correlation coefficient between factor scores and overcomes the misuse of its statistical significance.

Although the assessment of cross-study reproducibility has undergone a major improvement in statistical methods, researchers have still to face with the interpretation of similarities and differences across centers/studies: which latent factors (e.g., climate, influence of media or society, or food supply) are responsible for the identification of DPs in a country, but not in another one, or for the different variants of similar DPs across countries? For example, given the same climate and food supply, groups with different age, religion, ethic or socio-economic background may show different versions of a similar DP (4). Similarly, sources of beneficial or detrimental nutrients differ across populations or subpopulations with varied age, ethnic, or socio-economic background. For example, in 10 case-control studies from the International Head and Neck Cancer Epidemiology consortium (47), we have shown that the primary sources of vitamin C were different across countries: within the European

studies subjects mainly derived natural vitamin C from citrus fruits, kiwi, tomatoes, green salad, and apples/pears, whereas, in the US studies, fruit juices and potatoes were relevant contributors too. Within countries, sources were different in (otherwise comparable) populations from urban or rural areas (e.g., miso-soup in the rural, vegetables and green tea in a more industrialized area from Japan), among young people or Blacks from the US, where fortified drinks and southern greens were the major contributors of vitamin C, respectively. Besides the complexity of DP analysis, these considerations suggest the importance of working at a subpopulation level and the need for statistical criteria assessing similarity of subpopulation-specific DPs to allow for the merging of data from different subpopulations.

The assessment of stability of *a posteriori* DPs over time has been traditionally considered in cohort and survey studies over the last 30 years, to identify the more appropriate timeframe for scheduling successive dietary information queries. This justifies why we have found 25 relevant papers, as compared to the 9 on cross-study reproducibility, in this systematic review.

The analysis of DP stability can be very complicated. For example, research can target the individual- and/or the population-specific levels of stability and can assess stability of the identified DPs and/or the relevant food groups. Also, the stability of DPs identified across different life-course periods can be the focus of the research (e.g. (15, 16, 31)). Even when considering adults only, differences in the study designs arose from subjects' age at baseline,

the time-intervals between successive waves and the maximum time-interval between the first and the last wave considered. In addition, the statistical methods used for the assessment of DP stability differ markedly across papers: 25% of them did not use any statistical procedure (but simply inspected the factor-loading matrices over time), whereas 50% considered 2 criteria.

Within this complicated scenario, we can only comment on some preliminary results. Firstly, the closer the examined waves of dietary information collection are, the better is the stability of the identified DPs. This conclusion is very well supported, without any restriction on the statistical approach used for the analysis. When the dietary assessment tool, subject's life, and the DP identification process are stable over successive administrations, DP instabilities are either unexpected or due to essential and timely modifications of diet-related policies (e.g., the ban on trans fats) which lead to changes in behavior and food product development and marketing (4). Secondly, in the 75% of the papers, the number of identified DPs and the percentage of explained variance were substantially stable over time. We can conclude that, to date, overall dietary habits have been generally expressed in a stable number of constructs over time, with a few new or lost DPs over 10 or 20 years. Also, the ability of the identified DPs to capture the overall variance did not change over time, although the relative importance of the single DPs (in terms of percentage of the total variance explained) may vary. Thirdly, within an identified DP, the correlation structure among food groups is still stable over time, although changes in relevant food groups have been reported in more

refined statistical analyses. Dietary patterns are more likely to evolve, rather than disappear or emerge as brand-new ones. This conclusion may reflect the combination of several aspects. Among the most relevant ones, we mention early-life experiences with various tastes and flavors and parental feeding practices, which tend to persist over the lifespan (48): however, later food choices could be influenced by media/society or ageing. At a population-level, several other factors may influence the potential evolution of DPs over time, including changes in food supply (e.g., preferences for ethnic foods) as well as in nutrition-related policies. For example, we might hypothesize that the ban on trans fats will favor a change in the DP structure of those putative DP named *Snacks,* or *Sweets,* or *Desserts* (based on bakery products, baked goods, commercially fried foods, and spreads, which are likely to contain trans fats) in favor of similar processed foods made with unhydrogenated oils.

Evidence from the current review is still too limited to provide a firm conclusion on the most suitable timeframe to administrate successive dietary assessment tools within longitudinal studies or repeated surveys. In the absence of major life changes in the target population, DPs still show a good stability within 6 - 7 years after the previous dietary assessment; however, within a more refined statistical model, marked signs of instability were found after the same number of years for one (at 6 years) or two (at 7 years) DPs, but not for the last DP identified on the same dataset (9). Thus, scheduling successive administrations of the dietary assessment tool every 4 years, like in the NHS II, and updating the Dietary

Guidelines for Americans every 5 years are recommended strategies to monitoring DPs at their maximum potential stability over time.

Similarly, the current review does not provide clear insights into the question about some types of DPs being more stable than others. Except for the well-characterized and stable *Alcohol* DP (based on beer, liquors, and wine) in the Swedish SMC study, the Prudent-like and the Western-like DPs show similar and acceptable levels of stability. Nonetheless, we notice a general tendency of the Western-like DPs (mainly based on meat, processed meat, potatoes, and sometimes on fats, sweets, or grains) in the European studies (9, 23, 27, 34, 37) to show decreasing mean scores and/or decreasing intakes of relevant food groups. The same trajectory was not evident for their American counterparts (33, 45), although the analyses were based on weaker criteria.

Another major limitation of our review is that we did not summarize information on the potential association between changes of DPs (across studies or over time) and changes in disease occurrence. From a public health perspective, a common or stable DP is more critical to preserve if it protects against the risk of major chronic diseases, whereas the loss of previously identified DPs may derive from successful public health campaigns to discourage unhealthy dietary behaviors, like the ban on trans fats.

Future efforts should be directed on defining the generalizability of *a posteriori* DPs within a statistical model where time or study variables are explicitly modelled and the selection of the type and number of DPs to retain at each measurement occasion is carried out

borrowing information across any levels of the analysis. The use of multi-study factor analysis (13) in nutritional epidemiology (32) has provided an example of a fruitful application of a novel statistical modelling strategy to tackle cross-study reproducibility of *a posteriori* DPs. Similarly, multilevel latent class analysis (49) may offer insights in cross-study reproducibility, and latent class transition models (i.e., latent Markov models) (50) can offer a natural framework to track changes of DPs over time. These possibilities rely not only on statistical skills, but also on an effort of integration of study protocols and data. As far as studies are conceived as isolate attempts of knowledge, any assessment of reproducibility will likely end up into a unified but distorted combination of results from separate studies with their own decisions and limitations. In the short-term, as researchers, we can at least contribute to spread out a general culture of reproducibility by assessing reproducibility of DPs according to a series of different criteria, although based on elementary statistics.

In conclusion, preliminary evidence from the first scoping review on the topic suggests that most identified DPs showed good reproducibility across studies and stability over time. This evidence is based on a qualitative assessment of DP similarities across measurement occasions in ~50% of the papers on cross-study reproducibility and 25% of papers on stability over time. Our focus on statistical methods for the assessment of DP reproducibility and stability provides crucial suggestions for researchers who approach these novel aspects and they thus may contribute to spread out the importance of reproducible messages in nutritional epidemiology.

588 **Acknowledgements**

589 VE and MF designed research; VE and MD collected the relevant papers and selected those
590 to be included in the systematic review; MD and LP prepared the first draft of Table 2 and of
591 part of Tables 3 and 4; VE, RDV, and MF completed and refined Tables 3 and 4; AS and FB
592 revised all the tables and checked their consistency with the text; AS prepared Supplemental
593 Figure 1; AS prepared Table 1; MD prepared Figure 1; VE wrote the paper and had primary
594 responsibility for final content. All authors read and approved the final manuscript.

References

1. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Current opinion in lipidology 2002;13(1):3-9.
2. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev 2004;62(5):177-203.
3. Weikert C, Schulze MB. Evaluating dietary patterns: the role of reduced rank regression. Curr Opin Clin Nutr Metab Care 2016;19(5):341-6. doi: 10.1097/MCO.0000000000000308.
4. Tucker KL. Dietary patterns, approaches, and multicultural perspective. Appl Physiol Nutr Metab 2010;35(2):211-8. doi: 10.1139/H10-010.
5. Edefonti V, De Vito R, Dalmartello M, Patel L, Salvatori A, Ferraroni M. Reproducibility and Validity of A Posteriori Dietary Patterns: A Systematic Review. Adv Nutr 2019. doi: 10.1093/advances/nmz097.
6. Balder HF, Virtanen M, Brants HA, Krogh V, Dixon LB, Tan F, Mannisto S, Bellocco R, Pietinen P, Wolk A, et al. Common and country-specific dietary patterns in four European cohort studies. J Nutr 2003;133(12):4246-51. doi: 10.1093/jn/133.12.4246.
7. Mannisto S, Dixon LB, Balder HF, Virtanen MJ, Krogh V, Khani BR, Berrino F, van den Brandt PA, Hartman AM, Pietinen P, et al. Dietary patterns and breast cancer risk: results from three cohort studies in the DIETSCAN project. Cancer causes &

control : CCC 2005;16(6):725-33. doi: 10.1007/s10552-005-1763-7.

8. Moskal A, Pisa PT, Ferrari P, Byrnes G, Freisling H, Boutron-Ruault MC, Cadeau C, Nailler L, Wendt A, Kuhn T, et al. Nutrient patterns and their food sources in an International Study Setting: report from the EPIC study. PLoS One 2014;9(6):e98647. doi: 10.1371/journal.pone.0098647.

9. Weismayer C, Anderson JG, Wolk A. Changes in the stability of dietary patterns in a study of middle-aged Swedish women. J Nutr 2006;136(6):1582-7.

10. Dekker LH, Boer JM, Stricker MD, Busschers WB, Snijder MB, Nicolaou M, Verschuren WM. Dietary patterns within a population are more reproducible than those of individuals. J Nutr 2013;143(11):1728-35. doi: 10.3945/jn.113.177477.

11. Castello A, Buijsse B, Martin M, Ruiz A, Casas AM, Baena-Canada JM, Pastor-Barriuso R, Antolin S, Ramos M, Munoz M, et al. Evaluating the Applicability of Data-Driven Dietary Patterns to Independent Samples with a Focus on Measurement Tools for Pattern Similarity. J Acad Nutr Diet 2016;116(12):1914-24 e6. doi: 10.1016/j.jand.2016.05.008.

12. Castello A, Lope V, Vioque J, Santamarina C, Pedraz-Pingarron C, Abad S, Ederra M, Salas-Trejo D, Vidal C, Sanchez-Contador C, et al. Reproducibility of data-driven dietary patterns in two groups of adult Spanish women from different studies. Br J Nutr 2016;116(4):734-42. doi: 10.1017/S000711451600252X.

13. De Vito R, Bellio R, Trippa L, Parmigiani G. Multi-study factor analysis. Biometrics

2019;75(1):337-46. doi: 10.1111/biom.12974.

14. Murakami K, Shinozaki N, Fujiwara A, Yuan X, Hashimoto A, Fujihashi H, Wang HC, Livingstone MBE, Sasaki S. A Systematic Review of Principal Component Analysis-Derived Dietary Patterns in Japanese Adults: Are Major Dietary Patterns Reproducible Within a Country? Adv Nutr 2019;10(2):237-49. doi: 10.1093/advances/nmy079.

15. Crozier SR, Robinson SM, Godfrey KM, Cooper C, Inskip HM. Women's dietary patterns change little from before to during pregnancy. J Nutr 2009;139(10):1956-63. doi: 10.3945/jn.109.109579.

16. Mikkila V, Rasanen L, Raitakari OT, Pietinen P, Viikari J. Consistent dietary patterns identified from childhood to adulthood: the cardiovascular risk in Young Finns Study. Br J Nutr 2005;93(6):923-31.

17. Northstone K, Emmett PM. Are dietary patterns stable throughout early and mid-childhood? A birth cohort study. Br J Nutr 2008;100(5):1069-76. doi: 10.1017/S0007114508968264.

18. Northstone K, Smith AD, Newby PK, Emmett PM. Longitudinal comparisons of dietary patterns derived by cluster analysis in 7- to 13-year-old children. Br J Nutr 2013;109(11):2050-8. doi: 10.1017/S0007114512004072.

19. Northstone K, Ness AR, Emmett PM, Rogers IS. Adjusting for energy intake in dietary pattern investigations using principal components analysis. Eur J Clin Nutr

2008;62(7):931-8. doi: 10.1038/sj.ejcn.1602789.

20. Castro MA, Baltar VT, Selem SS, Marchioni DM, Fisberg RM. Empirically derived dietary patterns: interpretability and construct validity according to different factor rotation methods. Cad Saude Publica 2015;31(2):298-310.

21. Varraso R, Garcia-Aymerich J, Monier F, Le Moual N, De Batlle J, Miranda G, Pison C, Romieu I, Kauffmann F, Maccario J. Assessment of dietary patterns in nutritional epidemiology: principal component analysis compared with confirmatory factor analysis. Am J Clin Nutr 2012;96(5):1079-92. doi: 10.3945/ajcn.112.038109.

22. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. J Nutr 2006;136(10):2580-7. doi: 10.1093/jn/136.10.2580.

23. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Long-term stability of food patterns identified by use of factor analysis among Swedish women. J Nutr 2006;136(3):626-33. doi: 10.1093/jn/136.3.626.

24. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, PRISMA-P. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1. doi: 10.1186/2046-4053-4-1.

25. Asghari G, Rezazadeh A, Hosseini-Esfahani F, Mehrabi Y, Mirmiran P, Azizi F. Reliability, comparative validity and stability of dietary patterns derived from an FFQ

in the Tehran Lipid and Glucose Study. Br J Nutr 2012;108(6):1109-17. doi: 10.1017/S0007114511006313.

26. Judd SE, Letter AJ, Shikany JM, Roth DL, Newby PK. 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 2014;1:29. doi: 10.3389/fnut.2014.00029.

27. Togo P, Osler M, Sorensen TI, Heitmann BL. A longitudinal study of food intake patterns and obesity in adult Danish men and women. Int J Obes Relat Metab Disord 2004;28(4):583-93. doi: 10.1038/sj.ijo.0802598.

28. Borland SE, Robinson SM, Crozier SR, Inskip HM, Group SWSS. Stability of dietary patterns in young women over a 2-year period. Eur J Clin Nutr 2008;62(1):119-26. doi: 10.1038/sj.ejcn.1602684.

29. Chen Z, Wang PP, Shi L, Zhu Y, Liu L, Gao Z, Woodrow J, Roebothan B. Comparison in dietary patterns derived for the Canadian Newfoundland and Labrador population through two time-separated studies. Nutr J 2015;14:75. doi: 10.1186/s12937-015-0064-6.

30. Cuco G, Fernandez-Ballart J, Sala J, Viladrich C, Iranzo R, Vila J, Arija V. Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. Eur J Clin Nutr 2006;60(3):364-71. doi: 10.1038/sj.ejcn.1602324.

31. Cutler GJ, Flood A, Hannan P, Neumark-Sztainer D. Major patterns of dietary intake

in adolescents and their stability over time. J Nutr 2009;139(2):323-8. doi: 10.3945/jn.108.090928.

32. De Vito R, Lee YCA, Parpinel M, Serraino D, Olshan AF, Zevallos JP, Levi F, Zhang ZF, Morgenstern H, Garavello W, et al. Shared and study-specific dietary patterns and head and neck cancer risk in an international consortium. Epidemiology 2019;30(1):93-102. doi: 10.1097/EDE.0000000000000902.

33. Fung TT, Rimm EB, Spiegelman D, Rifai N, Tofler GH, Willett WC, Hu FB. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. Am J Clin Nutr 2001;73(1):61-7.

34. Gerdes LU, Bronnum-Hansen H, Osler M, Madsen M, Jorgensen T, Schroll M. Trends in lifestyle coronary risk factors in the Danish MONICA population 1982-1992. Public Health 2002;116(2):81-8. doi: 10.1038/sj.ph.1900824.

35. Lopez-Garcia E, Schulze MB, Fung TT, Meigs JB, Rifai N, Manson JE, Hu FB. Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. Am J Clin Nutr 2004;80(4):1029-35. doi: 10.1093/ajcn/80.4.1029.

36. Malik VS, Fung TT, van Dam RM, Rimm EB, Rosner B, Hu FB. Dietary patterns during adolescence and risk of type 2 diabetes in middle-aged women. Diabetes Care 2012;35(1):12-8. doi: 10.2337/dc11-0386.

37. Mishra GD, McNaughton SA, Bramwell GD, Wadsworth ME. Longitudinal changes in

dietary patterns during adult life. Br J Nutr 2006;96(4):735-44.

38. Nimptsch K, Malik VS, Fung TT, Pischon T, Hu FB, Willett WC, Fuchs CS, Ogino S, Chan AT, Giovannucci E, et al. Dietary patterns during high school and risk of colorectal adenoma in a cohort of middle-aged women. International journal of cancer 2014;134(10):2458-67. doi: 10.1002/ijc.28578.

39. Northstone K, Emmett P. Multivariate analysis of diet in children at four and seven years of age and associations with socio-demographic characteristics. Eur J Clin Nutr 2005;59(6):751-60. doi: 10.1038/sj.ejcn.1602136.

40. Northstone K, Emmett PM. A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis. Br J Nutr 2008;99(5):1099-106. doi: 10.1017/S0007114507842802.

41. Prevost AT, Whichelow MJ, Cox BD. Longitudinal dietary changes between 1984-5 and 1991-2 in British adults: association with socio-demographic, lifestyle and health factors. Br J Nutr 1997;78(6):873-88.

42. Schulze MB, Fung TT, Manson JE, Willett WC, Hu FB. Dietary patterns and changes in body weight in women. Obesity (Silver Spring) 2006;14(8):1444-53. doi: 10.1038/oby.2006.164.

43. Schwerin HS, Stanton JL, Riley AM, Jr., Schaefer AE, Leveille GA, Elliott JG, Warwick KM, Brett BE. Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. Am J Clin Nutr 1981;34(4):568-80. doi: 10.1093/ajcn/34.4.568.

44. Schwerin HS, Stanton JL, Smith JL, Riley AM, Jr., Brett BE. Food, eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. Am J Clin Nutr 1982;35(5 Suppl):1319-25. doi: 10.1093/ajcn/35.5.1319.

45. van Dam RM, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Dietary patterns and risk for type 2 diabetes mellitus in U.S. men. Ann Intern Med 2002;136(3):201-9.

46. Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. Br J Nutr 2003;89(3):409-19. doi: 10.1079/BJN2002778.

47. Edefonti V, Hashibe M, Parpinel M, Turati F, Serraino D, Matsuo K, Olshan AF, Zevallos JP, Winn DM, Moysich K, et al. Natural vitamin C intake and the risk of head and neck cancer: A pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. International journal of cancer 2015;137(2):448-62. doi: 10.1002/ijc.29388.

48. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors Influencing Children's Eating Behaviours. Nutrients 2018;10(6). doi: 10.3390/nu10060706.

49. Vermunt JK. Multilevel latent class models. Sociological methodology 2003;33(1):213-39.

50. Sotres-Alvarez D, Herring AH, Siega-Riz AM. Latent transition models to study

women's changing of dietary patterns from pregnancy to 1 year postpartum. Am J Epidemiol 2013;177(8):852-61. doi: 10.1093/aje/kws303.

51. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 2008;31(4):466-75. doi: 10.1002/uog.5256.

52. Ryman TK, Boyer BB, Hopkins S, Philip J, O'Brien D, Thummel K, Austin MA. Characterising the reproducibility and reliability of dietary patterns among Yup'ik Alaska Native people. Br J Nutr 2015;113(4):634-43. doi: 10.1017/S0007114514003596.

53. Willett W. Nutritional epidemiology: Oxford University Press, 2012.

54. Beck KL, Kruger R, Conlon CA, Heath AL, Coad J, Matthys C, Jones B, Stonehouse W. The relative validity and reproducibility of an iron food frequency questionnaire for identifying iron-related dietary patterns in young women. J Acad Nutr Diet 2012;112(8):1177-87. doi: 10.1016/j.jand.2012.05.012.

55. Ambrosini GL, O'Sullivan TA, de Klerk NH, Mori TA, Beilin LJ, Oddy WH. Relative validity of adolescent dietary patterns: a comparison of a FFQ and 3 d food record. Br J Nutr 2011;105(4):625-33. doi: 10.1017/S0007114510004137.

56. Bountziouka V, Tzavelas G, Polychronopoulos E, Constantinidis TC, Panagiotakos DB. Validity of dietary patterns derived in nutrition surveys using a priori and a posteriori multivariate statistical methods. Int J Food Sci Nutr 2011;62(6):617-27. doi:

10.3109/09637486.2011.561783.

57. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr 1999;69(2):243-9.

58. Strauss ME. Introduction to the special section on construct validity of psychological tests: 50 years after Cronbach and Meehl (1955). Psychol Assess 2005;17(4):395. doi: 10.1037/1040-3590.17.4.395.

Table 1. Dimensions of stability according to possible levels of analysis[1]

Level	Methods[2]	Forms of stability[2]
Individual-level: Are single subjects stable eaters over time or do they change their DPs?	Dietary patterns CA	Dietary patterns - Percentages of stable eaters or transitioners - Ranking of clusters with the higher stability
	Relevant food groups ANOVA for testing differences in the mean intakes across clusters	Relevant food groups Lower-than- or higher-than-average consumption of food groups within clusters of subjects

Population-level: - Are DPs stable within a target population? - Is there a change in individuals' life-course in the period under examination? - If yes, is the entire population experiencing a change in the life-course? - Are there parallel sub-cohorts of different ages that get older, to assess «secular trends»?	Dietary patterns PCA/EFA with potential CFA on EFA-based results	Dietary patterns - Number of identified DPs over time: are there DPs gained or lost during the period? - Percentage of explained variance of single DPs: are percentages similar over time for stable DPs? - DP composition: are factor-loading matrices similar over time? - DP scores: do mean scores from similar DPs change over time? Do quantile categories assigned to the same subject change over time?
	Relevant food groups MANOVA or ANOVA for testing differences in mean intakes or changes over time for EFA- or CFA-based relevant food groups	Relevant food groups - Number of relevant food groups within a DP: is the number of food groups increasing or decreasing consistently over time? - Food group intakes within a DP: do mean intakes from the same relevant food group change over time?

[1]ABBREVIATIONS: ANOVA: analysis of variance; CA: cluster analysis; CFA: confirmatory factor analysis; DP: dietary pattern; EFA: exploratory factor analysis; MANOVA: multivariate analysis of variance; PCA: principal component analysis

[2] Methods for the assessment of stability over time can target dietary patterns directly as well as the relevant food-groups defining these dietary patterns; likewise, stability can be inspected at the dietary pattern level or at the relevant food group one.

Table 2. Basic characteristics of observational studies on cross-study reproducibility and stability over time of a posteriori dietary patterns[1]

Reference	Location and Study	Study Design	Participants (n)	Age (y)[b]	Follow-up (y)	Questionnaire
Asghari, 2012 (25)	Iran TLGS	TLGS: cohort study on urban residents in Teheran in 1999 - 2001; Validation study of the TLGS FFQ based on a random sample of participants who were proportionately distributed across 5 10-y age intervals and sexes plus extra wave of the cohort study with FFQ administration	132 (89 completed FFQ3)	35.6 ± 16.8 (20 - 70)	8, until 2011 (baseline: 1999-2001)	FFQ (based on a Willett format): 1 y; SA; reproducibility and validity to be assessed in this study, but validity granted for the analysis of stability over time; 168 FI; 12 24HRs: collected monthly on 2 formal weekend days and 10 week days; FFQ1: completed 1 month before collection of the first 24HRs; FFQ2: completed 1 month after the last 24HR, 14 months between FFQ1 and FFQ2; FFQ3: completed at the end of the follow-up; 19 FG common to all dietary sources

Balder, 2003 (6)	Netherlands, Sweden, Finland, and Italy DIETSCAN Project (NLSC, SMC, ATBC, ORDET)	Parallel analysis of 4 prospective cohort studies according to the same strategy (no pooled analysis); NLSC (random subcohort of): population-based cohort of Ms and Fs from Dutch municipalities that began in 1986; SMC: population-based cohort of Fs based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; ATBC: randomized placebo-controlled intervention study conducted among M smokers who lived in southwestern Finland (1985-1988); ORDET: cohort study of Italian healthy volunteer Fs from the province of Varese, northern Italy (1987-1992)	NLSC: 3123 Ms and Fs (1598 Fs and 1525 Ms); SMC: 61,469 Fs; ATBC: 27,111 Ms; all numbers referred to subjects with complete dietary data; ORDET: 9208 Fs	NLSC: at baseline 61.4 ± 4.2 for Ms and ± 4.3 for Fs (55 - 69); SMC: at baseline 53.7 ± 9.7 (40 - 74); ATBC: at baseline 57.7 ± 5.1 (50 - 69); ORDET: at baseline 48 ± 8.5 (35 - 69)	7 for NLSC (baseline: 1986), 13 for SMC (baseline: 1987-1990), and NA for ATBC (baseline: 1985 - 1988, intervention ended in 1993 after 5-8 ys, follow-up later on), 9 for ORDET (baseline: 1987-1992)	4 different but validated FFQs: NLSC-FFQ: 1 y; SA; NA reproducibility but valid; 150 FI (51 FG, but final number equal to 49); SMC-FFQ: 6 months; SA; NA reproducibility but valid; 67 FI (51 FG, but final number equal to 42); ATBC-FFQ: 1 y; SA; reproducible and valid; 276 FI (51 FG, but smaller final number of FG); ORDET-FFQ: 1 y; SA; reproducible and valid; 107 FI (51 FG, but final number equal to 32)
Borland, 2008 (28)	UK SWS	SWS: prospective study including Fs from the general population living in the western part of Southampton; subset of Fs interviewed 2 ys later at the same time of the y as the first interview (1998, November, 13 - December, 22) from the cohort of 6129 SWS nonpregnant Fs; a subset of 29 diet changers out of all included in a separate analysis	94 non-pregnant Fs	at baseline (20 - 34)	2 ys (baseline: 1998)	FFQ: 3 months; IA; 100 FI (49 FG); NA reproducibility and validity; FFQ administered 2 times, at baseline and after 2 ys

Castello, 2016 (12)	Spain EpiGEICAM, DDM-Spain	EpiGEICAM: case-control study on F breast cancer based on 14 Spanish provinces (2006-2011); DDM-Spain: cross-sectional study based on a random sample of Fs from 7 screening centers (minimum 500 from each center) (2007-2008)	EpiGEICAM: 973 healthy Fs; DDM-Spain: 3550 Fs	EpiGEICAM: 50.63 ± 9.47 (22 - 71); DDM-Spain: 56.20 ± 5.46 (45 - 69)	Not applicable	EpiGEICAM: FFQ: 5 ys; NA SA; based on a validated FFQ; 117 FI (26 FG); DDM-Spain: FFQ: 1 y; IA; based on a validated FFQ; 99 FI (all in common with EpiGEICAM FFQ) (26 FG)
Castello, 2016 (11)	Spain EpiGEICAM	EpiGEICAM: case-control study on F breast cancer based on 14 Spanish provinces (2006-2011); selection of 3 studies (Bessaud et al, Adebamowo et al, and Terry et al) from a systematic review of the literature on DPs and breast cancer	EpiGEICAM: 973 case-control pairs of Fs (1946 Fs in total)	EpiGEICAM: 50.63 ± 9.47 (22 - 71); other studies: NA	Not applicable	EpiGEICAM: FFQ: 5 ys; NA SA; based on a validated FFQ; 117 FI (26 FG); other studies: FFQs described in the paper
Chen, 2015 (29)	Canada CCS, FFQVP	Two time-separated studies (over a decade) in the Newfoundland and Labrador, including non-istituzionalized adult residents; CCS: case-control study with a frequency matching on age (5 ys) and sex (2001-2005) - controls only from the CCS study; FFQVP: validation study conducted with a stratified random digit dialing with proportional allocation (2011 - 2012)	CCS: 554 controls; FFQVP: 192	CCS: 58.7 ± 7.7 (35 - 70) (20 - 74 in all CCS cases and controls); FFQVP: 56.2 ± 8.7 (35 - 70)	Not applicable in either studies	Modified FFQ based on an Hawaii FFQ: 2 ys; SA; 169 FI (39 FG); valid; same FFQ administered in both studies

Crozier, 2009 (15)	UK SWS	SWS: prospective cohort study including Fs from the general population living in the western part of Southampton (1998-2002)	2270 (early pregnancy) and 2649 (late pregnancy) from a cohort of 12572 nonpregnant Fs; 2057 Fs with complete information at the 3 time-points of interest used for the stability analysis	at baseline (20 - 34)	before pregnancy (median time to conception: 1,8 ys from initial interview) - late pregnancy (34 weeks of gestation) (baseline: 1998-2002)	FFQ: 3 months; IA; 98 FI (48 FG); valid; FFQ administered at 3 time-points, before pregnancy, in early pregnancy, and late pregnancy
Cucò, 2006 (30)	Spain NA	Longitudinal cohort study based in the city of Reus, Spain, including healthy Fs volunteers who have planned and completed a pregnancy and had complete dietary information at all assessment occasions (1992-1996)	80 Fs	mean: 29 (at baseline 18 - 35, final range: 24 - 35)	last preconception visit (1-3 menstrual cycles) to weeks 6, 10, 26, and 38 of pregnancy, and 6 months postpartum (baseline: 1992-1996)	1 7-consecutive day DR at each time-point; check with trained interviewers; 22 FG common to all time-points

Cutler, 2009 (31)	USA (Minnesota) EAT	EAT: cohort study of ethnically diverse youth from Minnesota schools during early and middle adolescence; EAT-I (Time 1) and EAT-II (Time 2)	Time 1: 4746; Time 2: 2516	Time 1: at baseline (12-13: early adolescence or middle school [younger cohort], and 15 - 16: middle adolescence or high school [older cohort]); Time 2: same students 5 ys later	5 ys (Time 2: 2003 - 2004) (baseline: Time 1: 1998-1999)	YAO-FFQ, based on the NHS FFQ: NA reference period; NA SA; reproducible and valid in children and adolescents 9-18 ys old; 152 FI (152 FG); pretested in a low-income, ethnically diverse middle school population with good results for comprehension
De Vito, 2019 (32)	USA, Italy, and Switzerland INHANCE	INHANCE: consortium of case-control studies on head and neck cancer; subsample of 7 case-control studies providing information on a common set of 23 nutrients derived from study-specific FFQs. North Carolina (2002-2006) (2002-2006); Milan (2006-2009) (2006-2009); New York MSKCC (1992-1994); Los Angeles (1999-2004);	10,668 (3844 cases; 6824 controls)	NA, but adults	Not applicable	5 study-specific FFQs, as the European studies [Italy Multicenter, Switzerland, and Milan (2006-2009)] shared the same FFQ; 1 y for the 4 US studies and 2 ys for the 3 European studies; IA for 3 studies and SA for 4 studies; either reproducible and valid or based on previously validated FFQs; number of FI varying

		Switzerland (1991-1997); Italy Multicenter (1990-1999)				from 72 to 138 (23 common nutrients)
Dekker, 2013 (10)	Netherlands Doetinchem Cohort Study	3 successive surveys (surveys 2, 3, and 4, after the first one) within the same population-based cohort study including at baseline an age- and sex-stratified random sample of residents from Doetinchem town (1987-1991; follow-up available for 2/3 of the original random sample by design	4007 subjects with information available for the 3 rounds. In detail: 6113 (survey 2); 4916 (survey 3); 4520 (survey 4)	(~47 - 66)	6 ys (survey 2: 1993-1997), 11 ys (survey 3: 1998-2002), 16 ys (survey 4: 2003-2007) after the first survey, so 10-y follow-up from survey 2 to survey 4 (baseline: 1987-1991)	FFQ: 1 y; NA SA; reproducible and valid; 178 FI (32 FG)
Fung, 2001 (33)	USA HPFS	HPFS: prospective cohort study of US M health professionals started in 1986; random sample from the 18,255 subjects of the HPF Study recruited between 1993 and 1994 who volunteered to provide blood sample	466 Ms	at baseline (40 - 75)	1990 and 1994 waves (baseline: 1986)	FFQ: 1 y; SA; reproducible and valid; 131 FI (42 FG)

Gerdes, 2002 (34)	Denmark MONICA	Three consecutive surveys from MONICA project, including at baseline (DAN-MONICA I, 1982-1984) a random sample of Danish citizens who lived in the western part of the Copenhagen County and had 30, 40, 50, and 60 ys at baseline and further re-examined in two successive surveys (DAN-MONICA II and DAN-MONICA III)	3317 Fs (1822 + 737 + 778) and 3378 Ms (1876 + 725 + 777)	at baseline 30, or 40, or 50, or 60	1982 - 1984 (baseline, DAN-MONICA I) - 1986-1987 (DAN-MONICA II) and 1991-1992 (DAN-MONICA III)	FFQ: 1 y; NA SA; NA reproducibility and validity; 26 FI (23 FI, with 3 excluded, no FG built)
Judd, 2014 (26)	USA REGARDS	Population-based cohort study including a random sample of black and white individuals and designed to oversample black participants and people residing in the stroke belt, a US region at particularly high risk for stroke (8 US states) (2003-2007)	21,636	> 45	No follow-up	FFQ: 1 y; SA; NA reproducibility, but valid; 107 FI (58 FG, but final analysis on 56 FG due to low communalities and zero consumption)
Lopez-Garcia, 2004 (35)	USA NHS	NHS: prospective cohort study of US F registered nurses started in 1976; sample of Fs who were selected as control subjects for a nested case-control study on diabetes and that did not have cardiovascular disease, cancer, or diabetes mellitus at the time of blood drawing	732 Fs	at blood drawing mean: 56 (43 – 69) (1989 - 1990)	1986 and 1990 waves (baseline: 1976)	FFQ; 1 y; SA; reproducible and valid; administered two times in 1986 and 1990; 116 FI (37 FG)

Malik, 2012 (36)	USA NHS II	NHS II: prospective cohort study of US F registered nurses started in 1989; sample of Fs who returned a FFQ on high school diet in 1998 and did not have confirmed diabetes/history of diabetes/gestational diabetes, cancer, or cardiovascular disease	37,038 Fs	at baseline in 1989 (24 - 44), in 1997 at high school FFQ completion (34 - 53)	1997 - 2005 (baseline: 1989)	HS-FFQ: high school period; SA; reproducible and valid; 124 FI (37 FG); NHS II FFQ: 1 y; SA; reproducible and valid; 131 FI (40 FG); NHS II administered 4 times to assess adult diet (in 1991, 1995, 1999, and 2003)
Mannisto, 2005 (7)	Netherlands, Sweden, and Italy DIETSCAN Projec (NLSC, SMC, ATBC, ORDET)	Parallel analysis of 3 prospective cohort studies according to the same strategy (no pooled analysis); NLSC (random subcohort of): population-based cohort of Ms and Fs from Dutch municipalies that began in 1986; SMC: population-based cohort of Fs based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; ORDET: cohort study of Italian healthy volunteer Fs from the province of Varese, northern Italy (1987-1992)	NLSC: 1598 Fs; SMC: 61,463 Fs; ORDET: 10,788 Fs	NLSC: 61.4 ± 4.3 at baseline (55 - 69); SMC: 53.7 ± 9.7 at baseline (40 - 74); ORDET: 48 ± 8.5 at baseline (34 - 70)	7 ys for NLSC (baseline: 1986), and 13 ys for SMC (baseline: 1987-1990), 9 ys for ORDET (baseline: 1987-1992)	3 different but validated FFQs: NLSC-FFQ: 1 y; SA; NA reproducibility but valid; 150 FI (51 FG, but final number equal to 49); SMC-FFQ: 6 months; SA; NA reproducibility but valid; 67 FI (51 FG, but final number equal to 42); ORDET-FFQ: 1 y; SA; reproducible and valid; 107 FI (51 FG, but final number equal to 31)
Mikkila, 2005 (16)	Finland Cardiovascular Risk in Young Finns Study	Cardiovascular Risk in Young Finns Study: multicenter prospective cohort study of children, adolescents, and young adults started in 1980 in Finland; random sample of 50% of the	1768 subjects in 1980, 1200 in 1986, and 1037 in 2001, giving a total of 1037	at baseline (3 - 18), in 2001 (24 - 39)	1980 (baseline) - 2001, with a first wave of follow-up in 1986	1 48HR for each time-point (in 1980, 1986, and 2001); different number of recorded FI for each time-point (23 FG)

		participants who had dietary information and was followed at two-time points	subjects with complete information at the 3 time-points			
Mishra, 2006 (37)	UK Medical Research Council National Survey of Health and Development (1946 British Birth Cohort)	1946 British Birth Cohort: longitudinal cohort study based on a social class stratified, random sample of 5362 singleton births in England, Scotland or Wales during the first week of March, 1946, with 21 occasions for collecting information throughout the life-course until published paper; data from interviews at 3 time-points in 1982, 1989, and 1999	1265 subjects with dietary information at the 3 time-points	36 in 1982, 43 in 1989, 53 in 1999	1946 (baseline) - 1999	1 5-day DR completed between spring and autumn for each time-point in 1982, 1989, and 1999; different number of recorded FI for each DR (126 FG)
Moskal, 2014 (8)	Europe EPIC	EPIC: cohort study on healthy Ms and Fs from 23 centers representing 10 European countries, including a Calibration Study based on a random sample of 5-12% subjects from each EPIC center	477,312 (including 34,436 from the Calibration Study with 24HR)	at baseline (35 - 70)	1992 - 1998 (for FFQ); 1995 - 2000 (for 24HR)	Country-specific dietary questionnaries, mostly FFQs; NA reference period; SA; valid; NA FI (23 nutrients); 1 24HR recall via face-to-face interview to describe the identified DPs

Newby, 2006 (23)	Sweden SMC	SMC: population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; subsample of SMC including healthy Fs at baseline with complete information on FFQ1 and FFQ2	33,840 Fs	mean: 52 at baseline (all Fs born between 1914 and 1948)	from 1987 - 1990 (baseline) to 1997 - onwards	FFQ1 (1987 - 1990): 6 months; SA; reproducible and valid; 67 FI (29 FG); FFQ2 (1997): 1 y; SA; based on the 1987 reproducible and valid FFQ; 97 FI (32 FG); mean time interval between FFQs: 8.8 ys
Newby, 2006 (22)	Sweden SMC	SMC: population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; subsample of SMC including healthy Fs at baseline with complete information on FFQ1 and FFQ2	33,840 Fs	mean: 52 at baseline (all Fs born between 1914 and 1948)	from 1987 - 1990 (baseline) to 1997, 9 ys of follow-up	FFQ1 (1987 - 1990): 6 months; SA; reproducible and valid; 67 FI (29 FG); FFQ2 (1997): 1 y; SA; based on the 1987 reproducible and valid FFQ; 97 FI (32 FG)
Nimptsch, 2014 (38)	USA NHS II	NHS II: prospective cohort study of US F registered nurses started in 1989; sample of Fs who returned a FFQ on high school diet in 1998, underwent at least 1 lower bowel endoscopy between 1998 and 2007, and had no history of cancer, colorectal adenomas, hyperplastic polyps	17,221 Fs	at baseline in 1989 (24 - 42), in 1997 at high school FFQ completion (34 - 51)	1997 - 2007 (baseline: 1989)	HS-FFQ: high school period (1960 - 1980); SA; reproducible and valid; 124 FI (37 FG); NHS II FFQ: 1 y; SA; reproducible and valid; 131 FI (40 FG); NHS II administered 5 times to assess adult diet (in 1991, 1995, 1999, 2003, and 2007)

Northstone, 2005 (39)	UK ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant Fs residents in the former Avon Health Authority with expected delivery date between 1st April 1991 - 31st December 1992; subset of ALSPAC study including 4- and 7-ys old children (2 waves)	9550 and 8286 children at 4 and 7 ys, respectively	4 and 7	2 waves for the children (4 and 7 ys of age) (baseline: 1991-1992)	FFQ adapted from the one used to assess maternal diet at 32 weeks of pregnancy; NA reference period; SA, completed by the mother/main carer; NA reproducibility and validity; 90 FI (57 FG)
Northstone, 2013 (18)	UK ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant Fs residents in the former Avon Health Authority with expected delivery date between 1st April 1991 - 31st December 1992; subset of ALSPAC study including 7-, 10-, and 13-ys old children (3 waves)	7285, 7473, and 6105 children, at 7, 10, and 13 ys, respectively	~7, 10, and 13	3 waves for the children (7, 10, and 13 ys of age) (baseline: 1991-1992)	1 3-day DR for each time-point, including 2 weekdays and 1 weekend; at 7 ys care-giver completion, at 10 and 13 ys, child completion; 62 FG at each time point
Northstone, 2008 (17)	UK ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant Fs residents in the former Avon Health Authority with expected delivery date between 1st April 1991 - 31st December 1992; subset of ALSPAC study including 3-, 4-, 7-, and 9-ys old children (4 waves)	10139, 9550, 8286, and 8010 children, at 3, 4, 7, and 9 ys, respectively; 6177 children with information at 4 time points	~3, 4, 7, and 9	4 waves for the children (3, 4, 7, and 9 ys of age) (baseline: 1991-1992)	Slightly different FFQs adapted from the one used to assess maternal diet at 32 weeks of pregnancy; NA reference period; SA, completed by the mother/main carer; NA reproducibility and validity; NA FI, increasing number for increasing study wave number;

			for stability analysis			34, 35, 41, and 41 FG at 3-, 4-, 7-, and 9-ys data
Northstone, 2008 (40)	UK ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant Fs residents in the former Avon Health Authority with expected delivery date between 1st April 1991 - 31st December 1992; subset of ALSPAC study including Fs during pregnancy and 4 ys after delivery (2 waves)	12053 and 9504 Fs pregnant at baseline and at 4 ys of the child, respectively; 8953 Fs with complete information at both time-points	NA, but pregnant Fs	4 ys (47 months post birth) (baseline: 1991-1992)	Slightly different FFQs with extra information added in the second FFQ; NA reference period; SA; NA reproducibility and validity; NA FI (44 FG at pregnancy assessment and 52 FG at 4-ys wave)
Prevost, 1997 (41)	UK HALS	Two consecutive surveys (1984 - 1985: HALS1, 1991 - 1992: HALS2); HALS1: random stratified sample of adults resident in England, Scotland, and Wales	HALS1: 9003; HALS2: 5352 from HALS1, still alive and able to participate	(18 - 74)	1991 - 1992 (HALS2) (baseline: 1984-1985, HALS1)	FFQ: NA reference period; NA SA; 39 FI (39 FG); NA reproducibility and validity; FFQ administered 2 times, at baseline and at follow-up

Schulze, 2006 (42)	USA NHS II	NHS II: prospective cohort study of US F registered nurses started in 1989; sample of Fs who returned 3 plausible FFQs and did not have history of diabetes, cancer, cardiovascular disease, or were pregnant at FFQ administration time	51,67	at baseline (24 - 44), in 1991 (26 - 46)	1991 - 1999 (baseline: 1989)	NHS II FFQ: 1 y; SA; reproducible and valid; 133 FI (39 FG); NHS II administered 3 times to assess adult diet (in 1991, 1995, and 1999)
Schwerin, 1981 (43)	USA Ten-State Nutrition Survey (Ten-State), HANES I	Merging of 2 cross-sectional studies; Ten-State (1968 -1970): sample disproportionately poor, with few young adults, and a disproportionate number of Blacks and Spanish Americans from geographically scattered states; subjects are provided with detailed information from special clinics; HANES I (1971 - 1974): broad-based national sample including all age groups between 1 and 74 ys	Ten-State: 11,337; HANES I: 20,749	(1 - 74)	No follow-up	1 24HR (15 FG) for both surveys

Schwerin, 1982 (44)	USA Ten-State Nutrition Survey (Ten-State), HANES I, NFCS	Merging of 3 cross-sectional studies; Ten-State (1968 -1970): sample disproportionately poor, with few young adults, and a disproportionate number of Blacks and Spanish Americans from geographically scattered states; subjects are provided with detailed information from special clinics; HANES I (1971 - 1974): broad-based national sample including all age groups between 1 and 74 ys; NFCS (1977 - 1978): representative sample of US population	Ten-State: 11,337; HANES I: 20,749; NFCS: 28,030	(1 - 74)	No follow-up	1 24HR (15 FG) for all 3 surveys, plus for NFCS 2-day DR; for NCFS, combination of information from 24HR and 2-days DR into a 3-days food consumption in grams
Togo, 2004 (27)	Denmark MONICA	Three consecutive surveys from MONICA project, including at baseline (M-82) a random sample of Danish citizens who lived in the western part of the Copenhagen County and had 30, 40, 50, and 60 ys at baseline (1982-1984) and further rexamined in two successive surveys (M-87, M-93)	2436 subjects participating in all 3 surveys, including 1806 subjects in M-82	30, or 40, or 50, or 60 at baseline	at 5 ys (1987 - 1988, M-87) and 11 ys (1993 - 1994, M-93)	FFQ: 1 y; NA SA; NA reproducibility and validity; 26 FI (21 FG)
van Dam, 2002 (45)	USA HPFS	HPFS: prospective cohort study of US M health professionals started in 1986; all Ms without diagnosed diabetes,	42,504 Ms	at baseline in 1986 (40 - 75)	1986 - 1998	FFQ; 1 y; SA; reproducible and valid; 131 FI (37 FG)

		cardiovascular disease, or cancer at baseline				
Weismayer, 2006 (9)	Sweden SMC	SMC: population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; subsample of SMC including 4 randomly selected subsamples of 1000 Fs each (giving a total of 4000 Fs), who completed 2 identical FFQs, to avoid survey learning effects	3606 Fs (871, 864, 887, and 967, at 4, 5, 6, 7 ys after baseline, respectively)	(49 - 70)	4, 5, 6, 7 ys after baseline (1987-1990) depending of the subsample	FFQ (1987 - 1990): 6 months; SA; reproducible and valid; 67 FI (25 FG); FFQ completed at baseline and after 4, 5, 6 or 7 ys depending of the subsample

[1]ABBREVIATIONS: 24HR: 24 hours recall; 48HR: 48 hours recall; ALSPAC: Avon Longitudinal Study of Parents and Children; ATBC: Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; DDM-Spain: Determinantes de la Densidad Mamográfica en España; DIETSCAN: CCS: Case-Control Study, here intended as the full name of one of the included studies and not as the case-control study design; DIETary patternS and CANcer in four European countries project; DR: dietary record; DP: dietary pattern; EAT: Eating Among Teens; EPIC: European Prospective Investigation into Cancer and Nutrition; EpiGEICAM: Grupo Español de investigación en Cáncer de Mama; F: female; FFQ: food-frequency questionnaire; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1/2/3; FFQVP: Food-Frequency Questionnaire Validation Project; FG: food groups; FI: food items; HALS: Health and Lifestyle Survey; HANES: Health and Nutrition Examination Survey; HS: High School; HPFS: Health Professionals Follow-up Study; IA: interviewer-administered; INHANCE: International Head and Neck Cancer Epidemiology consortium; M: male; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; MSKCC: Memorial Sloan Kettering Cancer Center; NA: not available; NFCS: Nationwide Food Consumption Survey; NHS: Nurses' Health Study; NLSC: Netherlands Cohort Study on diet and cancer; ORDET: Ormoni e Dieta nella Eziologia dei Tumori in Italy; REGARDS: Reasons for Geographic and Racial Differences in Stroke; SA: self-administered; SD: standard deviation; SMC: Swedish Mammography Cohort; SWS: Southampton Women's Survey; TLGS: Teheran Lipid and Glucose Study; y: year; YAO: Youth Adolescent Questionnaire

[b] Values are means ± SDs (ranges)

Table 3. Cross-study reproducibility of a posteriori dietary patterns[1]

Reference	Location and Study	Dietary pattern identification methods	Expl. Var % (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main Results
Balder, 2003 (6)	Netherlands, Sweden, Finland, and Italy DIETSCAN (NLCS, SMC, ATBC, ORDET)	Separate EFAs on each of the 4 studies: standardization and separate analysis by sex; within each study, sensitivity analyses assessing the effect of: 1. untransformed vs. dichotomized variables (for FG with >75% of nonusers); 2. unadjusted vs energy-adjusted variables using residual method; 3. solutions with 2-6 factors; 4. split-half analysis using the procrustes rotation to compare different solutions; Scree test to assess the final number of factors to retain in a range from 2 to 6 factors; Varimax rotation; Loading >= |0.35|cut-off	NLCS: 23 (5) for Ms, 23.2 (5) for Fs; SMC: 21.8 (4); ATBC: 20.3 (3); ORDET: 28.5 (4); final results based on unadjusted variables for energy	Internal reproducibility: see (5) for details; Cross-study reproducibility: no formal assessment	Internal reproducibility: see (5) for details; Cross-study reproducibility: Two of the identified DPs were qualitatively similar across studies and between Ms and Fs

Castello, 2016 (12)	Spain EpiGEICAM, DDM-Spain	Separate PCAs on EpiGEICAM and DDM studies: PCA on EpiGEICAM data: PCA on controls only; EIG>1; No rotation; Loading >= |0.30| cut-off; PCA on DDM data: separate PCAs on 5000 replicates of the DDM-Spain study within bootstrap estimation with selection of the 3 DPs that were more similar to those from EpiGEICAM study; PCA on controls only; EIG>1; No rotation; Loading >= |0.30| cut-off	37 (3) with PCA on EpiGEICAM data	Cross-study reproducibility: CC (95% percentile CI) between factor loadings (with values of 0.85-0.94 indicate fair similarity and values >=0.95 indicate 2 DPs were equivalent); Spearman correlation coefficient (Corr) (95% percentile CI) between factor scores (considering any significant correlation as being indicative of DP similarity)	Cross-study reproducibility: satisfactory reproducibility of WESTERN DP, but not of PRUDENT and MEDITERRANEAN DPs (WESTERN DPs: CC=0.90 (95% CI: 0.58-0.95), Corr=0.92 (95% CI 0.55-0.98); PRUDENT: CC=0.76 (95% CI 0.40-0.84), Corr=0.83 (95% CI 0.47-0.91); MEDITERRANEAN: CC = 0.77 (95% CI 0.65-0.83), Corr = 0.74 (95 %CI 0.63-0.79)); had we considered any significant correlation as being indicative of similarity, all DPs from the EpiGEICAM data were reproducible in the DDM-Spain study

Castello, 2016 (11)	Spain EpiGEICAM	PCA on EpiGEICAM study: PCA on controls only; EIG>1; No rotation; Loading >= |0.30|cut-off; food consumption information from EpiGEICAM study grouped into FG proposed in 3 other papers (Bessaoud et al, Adebamowo et al, and Terry et al) and factor scores calculated with loadings from the original papers and FG defined as in the original papers but recalculated on EpiGEICAM data; factor loadings recalculated using the definition of FG from (10)	37 (3) with PCA on EpiGEICAM data	Cross-study reproducibility: CC (95% percentile CI) between factor loadings (with values of 0.85-0.94 indicate fair similarity and values >=0.95 indicate 2 DPs were equivalent); Spearman correlation coefficient (Corr) (95% percentile CI) between factor scores (considering any significant correlation as being indicative of DP similarity	Cross-study reproducibility: 5 of the 6 reconstructed DPs showed high CC (>0.9) to their corresponding DP derived on the EpiGEICAM study data (CC(Castello-WESTERN, Bessaoud-WESTERN)=0.82, Corr(Castello-WESTERN, Bessaoud-WESTERN)=0.57; CC(Castello-WESTERN, Adebamowo-WESTERN)=0.92, Corr(Castello-WESTERN, Adebamowo-WESTERN)=0.83; CC(Castello-WESTERN, Terry-WESTERN)=0.94, Corr(Castello-WESTERN, Terry-WESTERN)=0.85; CC(Castello-PRUDENT, Bassaoud-MEDITERRANEAN)=0.86, Corr(Castello-PRUDENT, Bassaoud-MEDITERRANEAN)=0.67; CC(Castello-MEDITERRANEAN, Bassaoud-MEDITERRANEAN)=0.95, Corr(Castello-MEDITERRANEAN, Bassaoud-MEDITERRANEAN)=0.85; CC(Castello-PRUDENT, Adebamowo-PRUDENT)=0.95, Corr(Castello-PRUDENT, Adebamowo-PRUDENT)=0.85; CC(Castello-MEDITERRANEAN, Adebamowo-PRUDENT)=0.88, Corr(Castello-MEDITERRANEAN, Adebamowo-PRUDENT)=0.73; CC(Castello-PRUDENT, Terry-HEALTHY)=0.95, Corr(Castello-PRUDENT, Terry-HEALTHY)=0.89; CC(Castello-MEDITERRANEAN, Terry-HEALTHY)=0.77, Corr(Castello-MEDITERRANEAN, Terry-HEALTHY)=0.52); some smaller CC between comparable DPs depended on lack of FG in the original studies

De Vito, 2019 (32)	USA, Italy, and Switzerland INHANCE	Multi-study factor analysis on the merged dataset including the 7 studies: within-study logtransformation (base e) and standardization; controls-only analysis; identification of shared (among all studies) and (potential) study-specific dietary patterns within an integrated statistical model based on the maximum likelihood approach; number of factors to retain chosen according to a combination of standard techniques for FA, including Horn's parallel analysis, Cattell's scree plot, and the Steiger's RMSEA index, for the best number of total factors allowed, and to Akaike Information Criterion, for the number of shared factors; Varimax rotation on the shared factor loading matrix; Loading >= |0.60|cut-off for the	75-81 (3 common DPs shared among all the studies plus 1 additional study-specific DP for each of the 4 US studies)	Cross-study reproducibility: multi-study factor analysis	Cross-study reproducibility: Study populations from Italy, Switzerland, and the United States shared 3 reproducible DPs characterized by consumption of animal products and cereals, vitamin-rich foods, and fats, respectively; each of the American studies was characterized by a somewhat similar additional DP, which opposed calcium and niacin as dominant nutrients

		shared (rotated) factors and loading >= |0.25|cut-off for the study-specific (unrotated) factors; robustness analyses and stratified multi-study factor analysis by sex			

Judd, 2014 (26)	USA REGARDS	EFA on the first split-sample, CFA on the second split-sample, and final PCA on the whole sample as far as the model is correctly identified: EFA: 3 separate PCAs by population subgroups [region (southeastern US stroke belt/non-belt), sex (male/female), and race (black/white)] to identify the optimal number of factors in a range from 3 to 6 factors; EIG>1.5, Scree test, interpretability of results from stratified PCAs; Varimax rotation; Descriptive labelling; CFA: Loading > |0.20| cut-off on EFA results; No different correlation structures specified; RMSEA and CFI	NA (5)	Cross-study reproducibility: CC determined for each stratification pair for each of the factor number solutions ("excellent" when the smallest coefficient was >0.8, "good"; between 0.65 and 0.8, "acceptable" between 0.5 and 0.65, and "poor" <0.5); Validity: CFA	Cross-study reproducibility: PCA stratified by region of residence on the first half-sample: excellent CC for the 4- and 5-factor solutions, and acceptable CC for the 3- and 6-factor solutions; PCA stratified by gender: good CC for the 5- and 6-factor solutions and poor CC for the 3- and 4-factor solutions; PCA stratified by race: acceptable CC in the 5-factor solution, but poor CC for the other 3; the 5-factor solution had an acceptable CC in all stratified analyses and it was interpretable, so this was the final model selected for CFA; CFA on the second half-sample using the 5-factor solution: very good results, even when removing FG with low factor loadings (RMSEA values below 0.05)
Mannisto, 2005 (7)	Netherlands, Sweden, and Italy DIETSCAN	Separate PCFAs on each of the 3 studies: Scree test; Varimax rotation; Loading >= |0.35|cut-off	NLCS: 23.2 (5); ORDET: 29 (4); SMC: 21.8 (4)	Cross-study reproducibility: no formal assessment	Cross-study reproducibility: both the identified DPs remained quite consistent across cohort studies

	(NLCS, SMC, ATBC, ORDET)				
Moskal, 2014 (8)	Europe EPIC	Overall PCA on combined but country-specific questionnaire intakes and separate PCAs by center: logtransformation (base e) and energy adjustment with energy density method (based on alcohol-free energy) but no adjustment for center; Separate analysis by sex; PCA on covariance matrix; Scree-plot, interpretability; Varimax rotation; Loading > |0.45|cut-off	Overall PCA: 67 (4)	Cross-study reproducibility: Krzanowski's index, B_k, which measures the proportion of variance captured by k center-specific PCs which is also captured by overall PCA	Cross-study reproducibility: More than 75% of the variance that would be captured by center-specific PCs was captured by the PCs from the overall PCA (B_j>0.76 for all j>=2, B_2>0.85 for 23 of 27 centers); retaining 4 or more PCs was sufficient to capture at least 80% of variance in any center (B_j>0.80 for all j>=4); differences between sexes in each center were small when k>2
Schwerin, 1981 (43)	USA Ten-State Nutrition Survey (Ten-State), HANES I	Separate PCAs on the 2 surveys: standardization; EIG>1; Varimax rotation; Alphanumeric labelling; assignment algorithm of subjects based on the highest factor score; (probably) applied scores on HANES I data based on Ten-State DP loadings in the final solution	55.3 (7)	Cross-study reproducibility**:** no formal assessment	Cross-study reproducibility: the identified DPs were similar in the 2 surveys in terms of FG consumed

Schwerin, 1982 (44)	USA Ten-State Nutrition Survey (Ten-State), HANES I, NFCS	Separate PCAs on the 3 surveys: standardization; EIG>1; Varimax rotation with Kaiser normalization; Alphanumeric labelling	NA (6, or 7, or 8)	Cross-study reproducibility: no formal assessment	Cross-study reproducibility: 4 **of** the identified DPs remained quite consistent across studies that covered a decade

[1]ABBREVIATIONS: ATBC: Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; CA: cluster analysis; CC: congruence coefficient; CFA: confirmatory factor analysis; CFI: comparative fit index; CI: confidence interval; DDM-Spain: Determinantes de la Densidad Mamográfica en España; DIETSCAN: DIETary patternS and CANcer in four European countries project; DP: dietary pattern; EFA: exploratory factor analysis; EIG: Eigenvalue; EPIC: European Prospective Investigation into Cancer and Nutrition; EpiGEICAM: Grupo Español de investigación en Cáncer de Mama; F: female; FA: factor analysis; FG: food groups; HANES: Health and Nutrition Examination Survey; INHANCE: International Head and Neck Cancer Epidemiology consortium; M: male; NA: not available; NFCS: Nationwide Food Consumption Survey; NLCS: Netherlands Cohort Study on diet and cancer; ORDET: Ormoni e Dieta nella Eziologia dei Tumori in Italy; PC: principal component; PCA: principal component analysis; PCFA: principal component factor analysis; REGARDS: Reasons for Geographic and Racial Differences in Stroke; RMSEA: root mean square error of approximation; SMC: Swedish Mammography Cohort; y: year

Table 4. Stability over time of a posteriori dietary patterns[1]

Reference	Location and Study	Dietary pattern identification methods	Expl. Var % (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main Results
Asghari, 2012 (25)	Iran TLGS	Separate PCFAs on FFQ1, FFQ2, FFQ3, and m24HRs: Scree test and interpretability; Varimax rotation; Descriptive labelling; Applied scores from previous EFAs to data from FFQ3 were reported but their use was not clear	27.4 (2) with FFQ1 data, 31.6 (2) with FFQ2 data, 39.0 (3) with FFQ3 data, and 32.0 (2) with m24HR data	Reproducibility: see (5) for details; Relative validity: see (5) for details; Stability over time: intra-class correlation coefficient between continuous scores from FFQ2 and FFQ3 data, weighted kappa coefficient and proportions of subjects at the same quintile, adjacent quintile and opposite quintile when comparing quintiles classification of factor scores between baseline and follow-up data	Reproducibility: see (5) for details; Relative validity: see (5) for details; Stability over time: intra-class coefficients between FFQ2- and FFQ3-based scores equal to -0.09 (P=0.653) for the IRANIAN TRADITIONAL and 0.49 (P<0.001) for the WESTERN DPs; percentage of subjects at the same quintile higher for the WESTERN DP VS. the IRANIAN TRADITIONAL DP (27.1% vs. 20.2%); proportion of individuals at the opposite quintile reversed (35.8% vs. 41.5%); weighted kappa coefficient: 0.09 (95% CI: -0.05, 0.23) for the IRANIAN TRADITIONAL and 0.20 (95% CI: 0.05, 0.34) for the WESTERN DP

Borland, 2008 (28)	UK SWS	Separate PCAs at baseline and at follow-up: Interpretability; NA varimax rotation; Descriptive labelling; Applied scores calculated with loadings from the PCA on the whole cohort with complete FFQ (6125 subjects); Scores expressed in units of SD at initial visit (scores at both time-points divided by the SD of the scores at initial visit)	NA (2)	Stability over time: Spearman correlation coefficient between DP scores at 2 time-points; Bland-Altman method	Stability over time: Reasonable Spearman correlation coefficients (on the overall sample of 94 Fs: 0.81 for PRUDENT DP and 0.64 for the HIGH-ENERGY DP; higher correlations among the no major change group than in the diet changers group for both DPs); Bland-Altman method: average change (repeat - initial visit) equal to 0.13 SD for the PRUDENT DP score and equal to -0.01 SD for the HIGH-ENERGY DP; wider LOA for the HIGH-ENERGY than for the PRUDENT DP; narrower LOA in the no major change group than in the diet changers group for both DPs
Chen, 2015 (29)	Canada CCS, FFQVP	Separate EFAs in the 2 studies: EIG>1.5, Scree test, >50% variance explained by a factor, interpretability; Varimax rotation; Loading >|0.35|cut-off for CCS and >|0.5|cut-off for FFQVP study	54 (3) for the CCS study and 63 (4) for the FFQVP study	Stability over time: no formal assessment	Stability over time: The DPs of the Newfoundland and Labrador adult population have remained reasonably stable over almost a decade, although the PLANT-BASED DP derived from CCS study was a combination of the VEGETABLES/FRUITS DP and the GRAINS DP in the FFQVP study

Crozier, 2009 (15)	UK SWS	Separate PCAs at 3 time-points: standardization; NA criteria for choosing the number of factors; NA rotation; Descriptive labelling; Natural scores calculated with the factor loadings derived at each time-point; Applied scores calculated at a follow-up time with loadings obtained from PCA at the baseline time-point	14.5 (2) before pregnancy, 14.2 (2) in early pregnancy, and 14.5 (2) in later pregnancy	Stability over time: Spearman correlation coefficient between pairs of DP scores across the 3 time-points; Bland-Altman method; formal comparison between natural and applied scores	Stability over time: The identified DPs were strikingly similar at all 3 time-points in terms of factor loadings and explained variances; high Spearman correlation coefficients for both natural and applied DP scores before pregnancy and during early pregnancy and late pregnancy (natural scores with range: 0.51 - 0.81, applied scores with range: 0.52 - 0.80); Bland-Altman method: minimal change in PRUDENT DP score in early (–0.01 SD; P = 0.35) and late (–0.03 SD; P = 0.11) pregnancy compared with before pregnancy; no overall change in HIGH-ENERGY DP score in early pregnancy compared with before pregnancy (0.01 SD; P = 0.49), but a small significant increase in late pregnancy compared with before pregnancy (0.07 SD; P = 0.0002); narrower LOA for the PRUDENT score than the HIGH-ENERGY DP score
Cucò, 2006 (30)	Spain NA	Separate PCFAs at each of the 6 time-points: EIG>1, Scree test, interpretability; No rotation; Descriptive labelling starting from a |0.20| cut-off	21.48 (2) at preconception, 20.91 (2) at 6th week, 21.64 (2) at 10th week, 24.23 (2) at 26th week, 24.21 (2) at 38th week, and 12.79 (1) at 6th month of the child	Stability over time: CC between loadings from similar DPs across different available time-points; MANOVA for the analysis of consumption trend of dominant FG for each DP using standardized consumptions	Stability over time: coefficients of congruence: for the SWEETENED BEVERAGES AND SUGARS DP, quite high coefficients, ranging between 0.39 and 0.88 in absolute values, with high coefficients also between pregnancy and post-partum periods; for the VEGETABLES AND MEAT DP, high coefficients of congruence, ranging between 0.30 and 0.79 in absolute values; analysis of trend in dominant FG: no significant differences in the standardized mean consumption of dominant FG for both DPs

Cutler, 2009 (31)	USA (Minnesota) EAT	Separate PCFAs by cohort (older/younger) and sex (boys/girls) based on responses at Time 1 and responses at Time 2: standardization and energy-density transformation; EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling	NA (4 at Time 1, 4 or 5 at Time 2, depending on subgroup)	Stability over time: stability between DPs at Time 1 and Time 2 not formally assessed; secular trends (examined comparing DPs of middle adolescents at Time 1 (older cohort) with DPs in middle adolescents at Time 2 (younger cohort)) not formally assessed	Stability over time: The same set of 4 DPs found in boys and girls in early and middle adolescence was relatively stable over a 5-y time-period; when examining age-matched secular trends in middle adolescents at Time 1 and Time 2, almost identical DPs 5 ys apart were identified, except for the FAST FOOD DP that emerged in the middle adolescent boys at Time 2

Dekker, 2013 (10)	Netherlands Doetinchem Cohort Study	Separate CAs at each of the 3 surveys: percentage energy contributed variables (nutrient density); k-means algorithm; Bootstrap and internal cluster validity indexes (Calinski-Harabasz index, Davies-Bouldin index, and prediction-strength method) to assess the optimal number of clusters to retain between 2 and 6 clusters; Labelling based on FG that contributed the highest percentage of total energy compared with other DPs within the same survey (>= 40% higher energy indicated an important FG); Robustness analysis with partitioning around medoids method	Not applicable, 2-cluster solution chosen according to Jaccard similarity and internal cluster validity indexes	Reproducibility: see (5) for details; Stability over time: 1. stability of DPs over time in terms of contribution of a FG to total energy between the 2 clusters within the same survey (t- test, 99% CI, highly important FG were those with >1.4 time the percentage of total energy contributed for one compared to the other cluster by any FG) and comparison of the differences across surveys with a 5% cut-off; 2. Transitions of individuals between DPs over time: proportion of stable eaters (those assigned to the same cluster)	Reproducibility: see (5) for details; Stability over time: 1. stability of DPs over time in terms of contribution of a FG to total energy: the 2 DPs were similar in all 3 surveys in terms of percentages of total energy contributed by relevant FG within each survey, although with small differences in FG composition across surveys (i.e. soft drinks with sugar and high-fiber cereals); the 2 DPs retained their relative difference in FG intake at each of the surveys, with FG relative intakes in each DP not changing >5% per survey; low-fiber bread was the only exception, with relative differences being equal to -7.06, -13.1, and -4.56 percentage of total energy contributed in survey 2, 3, and 4 respectively, so 2 changes on 3 were >5%; 2. Transitions of individuals between DPs over time: 30.7% of the 4007 subjects with complete FFQ information were stable eaters assigned to HIGH-FIBER BREAD DP in all 3 surveys and 11.1% were stable eaters assigned to LOW-FIBER BREAD DP in all 3 surveys, giving a total of 41.8%; when comparing survey 2 and 4 on the the longest time frame (10 ys), 57.8% of participants assigned to HIGH-FIBER BREAD DP in both surveys, 15.2% assigned to LOW-FIBER BREAD DP at both surveys, 18.7% went from the HIGH- to LOW-FIBER BREAD DP, and 9.6% went from the LOW- to HIGH-FIBER BREAD DP; among stable eaters over time, no significant differences in percentage of energy intake contributed by important FG was found during the 10-y period; transitioners had higher relative differences in percentage of energy intake for important FG than stable eaters (0.27-3.01 as compared to 0.86-1.88)

				and transitioners (those assigned to different clusters) in all 3 surveys and in survey 2 and 4 (over the higher 10-y period); relative change in mean percentage of total energy a specific FG contributed from survey 2 to survey 4 between individuals with stable and unstable behavior	

Fung, 2001 (33)	USA HPFS	Separate PCFAs at the 3 time-points (in 1986, 1990, and 1994): NA criteria for choosing the number of factors; Varimax rotation; Descriptive labelling	NA (2)	Stability over time: Pearson correlation coefficient between scores from similar DPs across time-points	Stability over time: The 2 identified DPs were qualitatively similar across time; Pearson correlation coefficient between 1986 and 1990 equal to 0.65 for PRUDENT and 0.70 for WESTERN DP; Pearson correlation coefficient between 1990 and 1994 equal to 0.67 for PRUDENT and 0.69 for WESTERN DP; Pearson correlation coefficient between 1986 and 1994 equal to 0.58 for both PRUDENT and WESTERN DPs
Gerdes, 2002 (34)	Denmark MONICA	Separate PCFAs at each of the 3 surveys: separate analyses by sex and age group; Scree test, interpretability; Varimax rotation; Descriptive labelling	45 (6) with single survey data	Stability over time: trends in mean DP scores with pooled and age-specific data from linear regression models including time per age interaction term	Stability over time: Profound changes happened in the period, with coarse bread, rice and pasta much more frequently chosen at the expense of traditional Danish main meals; DP scores showed both variance heterogeneity and heterogeneity in trends across age groups; for Ms, COARSE BREAD and PASTA AND RICE DPs both increased 7 (95% CI: 6 - 8) $*10^{-2}$ points per y, i.e. about 0.7 SDs per 10 ys, BAKED GOODS AND SWEETS score increased 4 (95% CI: 3 - 5) $*10^{-2}$ points per y, FRUIT AND VEGETABLES DP score did not change, MEAT, POTATOES AND FAT score declined 4 (95% CI: 3 - 5) $*10^{-2}$ points per y, and BREAKFAST declined 2 (95% CI: 1 - 3) $*10^{-2}$ points per y; for Fs, survey-specific levels differed from the findings in Ms, notably for COARSE BREAD, FRUIT AND VEGETABLES and MEAT, POTATOES AND FAT, but showed the same trends: COARSE BREAD and PASTA AND RICE DP scores increased 6 (95% CI: 5 - 7) $*10^{-2}$ and 8 (95% CI: 7 - 9) $*10^{-2}$ points per y, respectively, BAKED GOODS AND SWEETS score increased 3 (95% CI: 2 - 4) $*10^{-2}$ points per y, FRUIT AND VEGETABLES score remained constant, MEAT, POTATOES AND FAT score declined 6 (95% CI: 5 - 7) $*10^{-2}$ points per y and BREAKFAST score declined 3 (95% CI: 2 - 4) $*10^{-2}$ points per y

Lopez-Garcia, 2004 (35)	USA NHS	Separate PCFAs on FFQ in 1986 and 1990 and average consumption across FFQ data: EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling	NA (2)	Stability over time: no formal assessment	Stability over time: The 2 major DPs were qualitatively similar across time
Malik, 2012 (36)	USA NHS II	Separate EFAs at the 5 time-points (during high school and in adulthood in 1991, 1995, 1999, and 2003): EIG>1, Scree test, interpretability; Varimax rotation; Loading >=|0.30|cut-off; Adjustment of DP scores by total energy with residual method	NA (2)	Stability over time: Spearman correlation coefficient between scores from similar DPs obtained during high school and in adulthood (cumulative updated average)	Stability over time: The 2 identified DPs were qualitatively similar across time; Spearman correlation between high school and adult DP scores equal to 0.49 for PRUDENT and 0.40 for WESTERN DP

Mikkila, 2005 (16)	Finland Cardiovascular Risk in Young Finns Study	Separate PCFAs at each of the 3 time-points (in 1980, 1986, and 2001): EIG>1, Scree test, interpretability; Varimax rotation; Alphanumeric labelling; Adjustment of DP scores by total energy with residual method	18 (2) with 1980 data, 21 (2) with 1986 data, and 17 (2) with 2001 data	Stability over time: Spearman correlation coefficient between scores from similar DPs in 1980 and 2001; Tracking analysis (cross-classification): proportion of subjects originally in the lowest or highest quintile of factor scores who remained in the same category over 6 (from 1980 to 1986) or 21 (from 1980 to 2001) ys, separately for those who were children (3 to 12 ys old) and adolescents (15 to 18 ys old) at the beginning of the study	Stability over time: The 2 identified DPs were qualitatively similar across time, over a 21-y period; Spearman correlation coefficient between factor scores in 1980 and 2001 were equal to 0.32 for PATTERN 1 and 0.38 for PATTERN 2; Tracking analysis: the proportion of subjects in the lowest or highest quintile of pattern scores remaining in the same quintile after 6 and 21 ys was 1.5 to 2 times the expected in both DPs if no stability is assumed; tracking was stronger among 15-18 ys-old subjects at baseline with 30–42% and 27–41% of subjects originally belonging to the extreme quintile of the energy-adjusted DP scores persisted in the same quintile 6 and 21 ys later, respectively; highest stability found in the uppermost quintile in both DPs

Mishra, 2006 (37)	UK Medical Research Council National Survey of Health and Development (1946 British Birth Cohort)	Separate EFAs at the 3 time-points (in adulthood in 1982, 1989, and 1999) on binary data (non-consumption/consumption): separate analyses by sex; EIG>1, Scree test, interpretability, root mean square residual; Varimax rotation; Loading >=|0.25|cut-off; Simplified DP scores to calculate individual DP scores in 1982 (36 ys) and 1989 (43 ys) based on EFA performed in 1999 (53 ys)	In 1999, 18.9 (3) among Fs and 17.4 (2) among Ms; in 1982 and 1989: NA (3 for Fs and 2 for Ms)	Stability over time: Number of FG consumed over time for each DP; Weighted kappa coefficient (95% CI) between thirds of DP scores between 1982 and 1989, between 1982 and 1999, and between 1989 and 1999	Stability over time: The identified DPs were similar over time among Ms and Fs; Number of FG consumed over time: for Fs, increased number of FG consumed in the ETHNIC FOOD AND ALCOHOL and FRUIT, VEGETABLES AND DAIRY DPs, and a decrease in MEAT, POTATOES AND SWEET FOODS DP; for Ms, number of FG consumed from both DPs increased significantly over time; fair-to-moderate values of kappa coefficient, except for MEAT, POTATOES AND SWEET FOODS DP, which showed poor agreement in Fs across time

Newby, 2006 (23)	Sweden SMC	Separate PCFAs at each of the 2 time-points: Scree test, interpretability; Varimax rotation; Descriptive labelling; Separate CFAs at each time point: Loading >=|0.15|cut-off based on loadings >=|0.20|cut-off from EFA results and a priori knowledge	PCFA: 35.4 (6) with FFQ1 (1987) data, 32.4 (6) with FFQ2 (1997) data; CFA: No model selection	Validity: CFA; Stability over time: mean and SD intakes of CFA-based FG at both time points and Spearman correlation coefficient between CFA-based FG; Pearson correlation coefficient between DP scores at 2 time-points; Pearson correlation coefficient between DP scores from PCFA and CFA at fixed time-point	Validity: CFA, but no goodness of fit assessment or formal comparison with EFA; Stability over time: intakes of vegetables, fruit, seafood, refined grains, soda, sugary foods, and sweet baked goods increased over the time period, whereas intakes of meat and whole grains decreased over the time period; Spearman correlation coefficient between CFA-based FG ranged from 0.23 to 0.70 (all P<0.0001); Pearson correlation coefficient between DP scores in 1987 and 1997 ranged from 0.27 (WESTERN/SWEDISH DP) to 0.54 (ALCOHOL DP) for CFA-based DPs (all P<0.0001) and were similar for PCFA-based DPs; Pearson correlation coefficient between DP scores from PCFA and CFA at fixed time-point were >=0.90 (all P<0.0001)
Newby, 2006 (22)	Sweden SMC	Separate PCFAs at each of the 2 time-points: Scree test, interpretability; Varimax rotation; Descriptive labelling; Separate CFAs at each time point: Loading >=|0.15|cut-off based on loadings >=|0.20|cut-off	PCFA: 35.4 (6) with FFQ1 (1987) data, 32.4 (6) with FFQ2 (1997) data; CFA: No model selection	Validity: CFA; Stability over time: no formal assessment	Validity: CFA, but no goodness of fit assessment or formal comparison with EFA; Stability over time: Similar FG and factor loadings for each DP were seen in 1987 and 1997; some variation was observed for HEALTHY DP

		from EFA results and a priori knowledge			
Nimptsch, 2014 (38)	USA NHS II	Separate EFAs at the 5 time-points (during high school and in adulthood in 1991, 1995, 1999 and 2003): EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling; Adjustment of DP scores by total energy with residual method	NA (2)	Stability over time: Pearson correlation coefficient between scores from similar DPs obtained during high school and in adulthood (cumulative updated average)	Stability over time: The 2 identified DPs were qualitatively similar across time; Spearman correlation between high school and adult DP scores equal to 0.48 for PRUDENT and 0.39 for WESTERN DP
Northstone, 2005 (39)	UK ALSPAC	Separate PCAs on 4- and 7-ys data: standardization; Scree test, interpretability; Varimax rotation; Loading >|0.3| cut-off	17.7 (3) with 4-ys old children data and 18.3 (3) with 7-ys old children data	Stability over time: no formal assessment	Stability over time: The 3 DPs were similar at both time points in terms of loadings and explained variances

Northstone, 2013 (18)	UK ALSPAC	Separate CAs at each of the 3 time-points: standardization with division by the range; k-means algorithm run 100 times with different starting positions to find the solution with the smallest sum of squares differences; internal stability testing of the final solution; number of clusters ranging from 2 to 6; outliers removed from the analysis at each time-point; 62 FG based on average consumption at each time-occasion	Not applicable, 4-cluster solution chosen at each-time point according to internal stability measures based on split-half technique performed 5 times (number of children allocated to a different cluster) and interpretability of the results	Stability over time: changes in mean scores of relevant FG characterizing the cluster; cross-tabulation of cluster solutions at different ages and proportion of subjects who remained in the same cluster between each pair of ages; sequence index plot to illustrate changes in cluster membership over time	Stability over time: 1. Internal stability based on 5 sets of split-sample testing: 4-cluster solution is the most stable, with <10% misclassified children at each time-point; 2. Changes in mean consumption for relevant FG: mean amount of FG consumed within each cluster differed between ages, generally increasing as the children got older, although the patterns of foods consumed and the foods in each cluster with higher- and lower-than-average consumptions were similar at each age; 3. Cross-tabulation of subjects at different ages: reasonably high number of children remaining in the same cluster at different ages (50 and 43% of children in the HEALTHY and PROCESSED clusters, respectively, at age 7 ys were in the same clusters at age 13 ys; proportion of children who stayed in the same cluster at all 3 ages equal to 20%; for individual clusters, the greatest stability was seen for the HEALTHY cluster at 33%, with the PROCESSED cluster second at 22%; less stable results for TRADITIONAL and PACKED LUNCH clusters, with 25 – 34% remaining in those clusters over time); 4. Sequence index plot: the most consistent cluster membership over time was for the HEALTHY cluster, followed by the PROCESSED cluster

Northstone, 2008 (17)	UK ALSPAC	Separate PCAs on 3, 4, 7, and 9 ys data on subjects available at each time-point and on subjects with information at 4 time-points: standardization; Scree test, interpretability; Varimax rotation; Loading >|0.3| cut-off	23.4 (4) with 3-ys old children data, 17.7 (3) with 4-ys old children data, 18.1 (3) with 7-ys old children data, and 19.2 (3) with 9-ys old children data	Stability over time: Stability assessed for the PROCESSED, TRADITIONAL, and HEALTH CONSCIOUS DPs: Spearman correlation coefficient between DP scores at each time-point, paired t-test for the change in mean DP scores between periods of questioning; Bland-Altman method and LOA (95% CI) across time-points using z-scores of each DP score with mean and SD depending on the comparison under consideration; Cross-classification using quintiles; weighted kappa coefficient to compare scores	Stability over time: High Spearman correlation coefficients between the same DP score at each pair of time-points (range: 0.35 - 0.69, all P<0.001), but little Spearman correlation coefficients between different DPs across time, except for the HEALTH CONSCIOUS/VEGETARIAN (9 ys only) negatively correlated with the TRADITIONAL DP at previous time points (but no significant p-values); paired t-test on mean differences in DP scores across time: consistent increase in the mean PROCESSED DP scores at the later ages compared to 3 ys old (all P<0.001), but no differences for the other DPs; 95% LOA for the adjusted scores: widest LOA for all pairings between 3 and 9 ys old data, narrowest LOA between 4 and 7 ys old data, narrowest LOA for the HEALTH CONSCIOUS between both 3 and 4 ys of age and 7 ys of age; weighted kappa coefficient: reasonable level of agreement between categorized scores from each time point (range: 0.25 - 0.47), with higher levels between 4 and 7 ys of age and 7 and 9 ys of age

				between each pair of time-points	

Northstone, 2008 (40)	UK ALSPAC	Separate PCAs on pregnancy data and on 4-ys data: standardization; Scree test, interpretability; Varimax rotation; Loading >|0.3| cut-off; Natural and applied scores at 4 ys with applied scores calculated with loadings obtained from pregnancy data PCA	31.3 (5) with pregnancy data and 25.1 (4) with 4-y follow-up data	Stability over time: Stability assessed for the HEALTH CONSCIOUS, PROCESSED, CONFECTIONARY, and VEGETARIAN DPs: Pearson correlation coefficient between scores from similar DPs obtained at pregnancy and at 4-y follow-up using both natural and applied scores; paired t-test to assess the change in mean scores over the 4-y period between questioning; Bland-Altman method and LOA (95% CI) between scores at the 2 time points; cross-tabulation between pregnancy score quintiles and the	Stability over time: Similar Pearson correlation coefficients across DPs for the natural and applied scores, although slightly larger using the applied method; paired t-test: considerably lower 4-y applied scores on average as compared to corresponding natural mean scores from the separate PCA at 4-ys follow-up, but SDs were much larger with applied scores; Fs decreased their scores on the HEALTH-CONSCIOUS DP over time (mean difference -0.075 and -0.284; P<0.0001, with natural and applied scores, respectively), but results for natural and applied scores were inconsistent in sign and/or statistical significance for the other DPs; Bland-Altman method: LOA were wider for applied scores; weighted kappa coefficient: reasonable level of agreement (0.267 < kappa < 0.306) between categorized scores from pregnancy and 4-y natural scores; weighted kappa coefficient generally higher when comparing pregnancy and 4-y applied scores; cross-classification: agreement was slightly better for the applied score of the HEALTH-CONSCIOUS DP compared to the 4-y natural score, but this was not true for the PROCESSED DP where the applied score was much less stable than the natural score

				2 (natural and applied) sets of 4-y score quintiles; weighted kappa coefficient (95% CI) on quintile of factor scores across time (pregnancy vs. 4-y; pregnancy vs. applied 4-y; 4-y vs. applied 4-y)	

Prevost, 1997 (41)	UK HALS	Separate PCAs on HALS1 (previous publication), HALS2 (current publication) and PCA on the merged dataset including subjects from HALS1 and HALS2: Scree test, Chi square test of isotropic variation; No rotation; Loading >|0.3|cut-off; In the final analysis, HALS2 scores calculated with loadings from PCA on HALS1 data, as factor loadings from HALS2 were identical to those originally derived from the full sample and to the HALS2 subset at HALS1	NA (4)	Stability over time: graphical representation of unadjusted mean DP scores for HALS1 and HALS2 by 7-y age groups at HALS1 (separately for Ms and Fs); unadjusted changes (HALS2 score - HALS1 score) in mean DP scores, with corresponding F test	Stability over time**:** Marked stability of DPs, in terms of variety of foods consumed, from the 1984-1985 survey to the 1991-1992 survey; graphical representation: COMPONENT 1 (HIGH IN FRUIT AND VEGETABLES, LOW IN FAT): the scores had risen by HALS2, in each age group, considerably more than would be expected for the 7-y advance in age, with the greatest increase in scores occurring in the youngest subjects (P for interaction between survey indicator and age at survey < 0.05); COMPONENT 2 (HIGH IN ENERGY-DENSE FOODS): HALS2 scores were all less than would have been expected for the 7-y advance in age and the score decreases were not uniform across the age groups, but were smaller in the older subjects (P for interaction < 0.05); COMPONENT 3 (HIGH IN CONVENIENCE FOODS): in Ms (except those aged 67-73 ys) and Fs aged 39 ys and over at HALS1, the scores had decreased in each age group, but the changes were small and less than expected just for the 7-y advance in age; in the younger Fs there was an increase in score by HALS2, contrary to the expected age trend; COMPONENT 4 (HIGH IN SUGARY FOODS, LOW IN VEGETABLES): same behavior of DP scores for both Ms and Fs at HALS1 and HALS2 (high in youth and older age, and low in middle age), but HALS2 scores were all higher, and higher than would have been expected for the 7-y advance in age; unadjusted mean scores increased significantly for Ms and Fs on COMPONENT 1 and 4 and fell significantly on COMPONENT 2 (men and Fs) and 3 (men only) (P<0.001)

Schulze, 2006 (42)	USA NHS II	Separate PCFAs at each of the 3 time-points (in adulthood in 1991, 1995, and 1999): EIG>1, Scree test, interpretability; Varimax rotation; Loading >=|0.30|cut-off; Adjustment of DP scores by total energy with residual method	NA (2)	Stability over time: no formal assessment	Stability over time: The 2 identified DPs were qualitatively similar across time

Togo, 2004	Denmark MONICA	EFA: on a subsample of the M-82 data (who filled a DR too); Separate analyses by sex; Scree test, interpretability; Varimax rotation; Descriptive labelling; CFA: Loading >=|0.30|cut-off on EFA results; CFA: 3-factor model with correlated factors; CFA performed on M-82 data (all M-82 participants) and on the subgroup including M-82-87 data; to include diet information at 5-y follow-up, CFA performed as a mean-structure factor analysis with group mean factor scores at baseline equal to 0 (but free to be estimated at M-87) and fixed loadings and factor-factor correlations over time; minimization	EFA: 30.5 (3) among Ms; 23.8 (3) among Fs; CFA: 3-factor model with correlated factors separately for Ms and Fs applied for the baseline cross-sectional analysis and as a mean-structure factor analysis	Validity**:** CFA at baseline; Stability over time: CFA as mean-structure factor analysis on the subgroup with data at both time points (M82-87)	Validity: CFA, but no goodness of fit assessment or formal comparison with EFA; Stability over time: CFA: by design, high correlations between corresponding DP scores at both time points (range: 0.88 - 0.95); between M-82 and M-87, the GREEN DP score mean increased to 0.30 for Ms and to 0.24 for Fs, the TRADITIONAL (men) and the SWEET-TRADITIONAL (women) DPs decreased to -0.27 and -0.18, and the SWEET DP (men) was virtually unchanged

		technique to calculate factor scores			

van Dam, 2002	USA HPFS	Separate PCFAs at each of the 3 time-points (1986, 1990, and 1994): EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling; Robustness analyses to assess the effect of number of factors retained, estimation method, and type of rotation	NA (2)	Stability over time: Pearson correlation coefficient between scores from similar DPs across time-points	Stability over time: The 2 major DPs were qualitatively similar across time; Pearson correlation between the PRUDENT DP score was 0.59 between 1986 and 1990, 0.60 between 1990 and 1994, and 0.55 between 1986 and 1994; for the WESTERN DP scores, the Pearson correlation was 0.69 between 1986 and 1990, 0.72 between 1990 and 1994, and 0.64 between 1986 and 1994
Weismayer, 2006 (9)	Sweden SMC	Separate EFAs at baseline and at follow-up for each of the 4 subgroups: Scree test, interpretability; Varimax rotation; Descriptive labelling; Separate CFAs at baseline and at follow-up for each of the 4 subgroups: Loading >=|0.20| cut-off on EFA results	EFA: NA (3); CFA: No model selection	Validity: CFA; Stability over time: 1. Spearman correlation coefficient between baseline and follow-up scores for each of the 4 groups and both EFA-based and CFA-based scores; 2. t-test of baseline and follow-up differences in mean intakes for the 18 CFA-based FG with at least 1 loading >0.2 for any	Validity: CFA, but no goodness of fit assessment or formal comparison with EFA; Stability over time: 1. Spearman correlation coefficient between EFA-based DP scores equal to 0.59, 0.57, 0.59, and 0.50 for HEALTHY DP, 0.47, 0.48, 0.51, and 0.39 for WESTERN DP, and 0.54, 0.66, 0.58, and 0.46 for ALCOHOL DP after 4, 5, 6, and 7 ys, respectively; Spearman correlation coefficient between CFA-based DPs equal to 0.63, 0.63, 0.62, and 0.54 for HEALTHY DP, 0.60, 0.54, 0.56, and 0.57 for WESTERN DP, and 0.73, 0.76, 0.70, and 0.75 for ALCOHOL DP after 4, 5, 6, and 7 ys, respectively; 2. t-test: no evidence of a difference in the means for 10, 6, 6, and 2 of 25 FG after 4, 5, 6, and 7 ys, respectively, but evidence that 3, 7, 8, and 11 of the 18 FG underwent significant changes after 4, 5, 6, and 7 ys, respectively (P<=0.01); 3. Spearman correlation coefficients between baseline and follow-up intakes of FG consistently decreasing in

				of the 3 DPs in any of the 4 subsamples; 3. Spearman correlation coefficient between baseline and follow-up intakes of 18 CFA-based FG with at least 1 loading >0.2 for any of the 3 DPs in any of the 4 subsamples; Internal stability of DPs: test of significant changes in the covariance matrix for each confirmed DP at baseline and follow-up	size over time (no correlation after 7 ys exceeding the size of the correlations after 4 ys); Internal stability of DPs: no significant instability after 4 and 5 ys of follow-up; significant instabilities for WESTERN DP after 6 ys (P= 0.01) and for WESTERN (P= 0.02) and ALCOHOL DPs (P=0.01) after 7 ys

[1]ABBREVIATIONS: 24HR: 24 hours recall; ALSPAC: Avon Longitudinal Study of Parents and Children; CA: cluster analysis; CC: congruence coefficient; CCS: Case-Control Study, here intended as the full name of one of the included studies and not as the case-control study design; CFA: confirmatory factor analysis; CI: confidence interval; DP: dietary pattern; DR: dietary record; EAT: Eating Among Teens; EFA: exploratory factor analysis; EIG: Eigenvalue; F: female; FFQ: food-frequency questionnaire; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1/2/3; FFQVP: Food-Frequency Questionnaire Validation Project; FG: food groups; HPFS: Health Professionals Follow-up Study; LOA: limits of agreement; M: male; m24HR: mean 24 hours recall; MANOVA: multivariate analysis of variance; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not available; NHS: Nurses' Health Study; PCA: principal component analysis; PCFA: principal component factor analysis; SD: standard deviation; SMC: Swedish Mammography Cohort; SWS: Southampton Women's Survey; TLGS: Teheran Lipid and Glucose Study; y: year

Figure 1. Flowchart of the study selection process performed within the systematic search of the literature supporting the scoping review

Manuscript: "Reproducibility of *a posteriori* dietary patterns across time and studies: a scoping review"

Valeria Edefonti

Online Supplementary Material

Supplemental materials and methods

In **Supplemental Table 1**, we introduce the basic terminology we adopted in the current review, as well as the statistical tools used for their assessment. In **Supplemental Figure 1** we introduce prototypical paths of DP identification processes related to reproducibility and validity of DPs. Dietary patterns are identifiable within any study design and starting from any dietary assessment tool source. If one dietary source is used at one time point, the assessment of DP reproducibility arises from the use of different statistical approaches for DP identification [Panel (A)]. Within the validation study of a new food-frequency questionnaire (FFQ), the same FFQ was administered twice (within 1 year) and compared with a gold standard dietary assessment tool [a dairy record (DR) or (multiple administration of) a 24-hour recall (24HR)] carried out on the same time interval and sample; DP reproducibility is assessed comparing the 2 sets of FFQ-based DPs, whereas relative validity of DPs is assessed comparing FFQ-based and gold-standard-based DPs [Panel (B)]. When either cohort studies or multiple waves of the same survey are available, a dietary assessment tool is administered to the same subjects in multiple occasions over longer time periods and the comparison of sets of DPs at the available measurement occasions allows for the evaluation of stability of DPs over time [Panel (C)]. Finally, to assess cross-study reproducibility of DPs, comparison of different sets of DPs derived from comparable dietary sources (at similar time points) is possible across centers from the same study, or across different studies representing potentially different populations or countries [Panel (D)]. In any of these 4 settings, confirming EFA-based DPs is possible through CFA, which assesses construct validity of DPs; results from the two approaches can be formally compared with suitable statistical tools [Panel (E)]. We re-classified the main findings from the articles included in the systematic review based on these definitions, no matter of the original definitions provided by the authors.

Supplemental Table 1. Definition of terms used in the current review and brief description of the statistical approaches used to assess these concepts in the current review[1]

Term	*General definition*	*Additional details within dietary pattern analysis*[2]	*Statistical method*[3]
Agreement	How close two measurements made on the same subject are? It is measured on the same scale as the measurements themselves. Agreement between measurements is a characteristic of the measurement methods involved (51)		Bland-Altman method with 95% LOA (limits are defined such that we expect that, in the long run, 95% of future differences between measurements made on the same subject will lie within the LOA) (15); Proportions of subjects classified into the same, adjacent, or opposite quantile category of score, or proportions of misclassified subjects (16); Kappa coefficient on score quantile categories (37); Sequence index plot (18)
Reliability	How inherent variability in the 'true' level of the quantity between subjects relates to the global variability of a phenomenon (variability in true levels plus variability in measurement error in observed measurements)? If reliability is high, measurement errors are small in comparison to the true differences between subjects, so that subjects can be relatively well distinguished (in terms of the quantity being measured) on the basis of the error-prone measurements. Conversely, if measurement errors tend to be large compared with the true differences between subjects, reliability will be low		Intraclass correlation coefficient between scores (52); Test-retest reliability on scores or on dominant food groups defining the identified dietary patterns (52) (see (5) for details)

	(51)		
Repeatability	How much is the variation in repeat measurements made on the same subject under identical conditions? Measurements are made by the same instrument or method, the same observer and they are made over a short period of time (over which underlying value considered constant). Variability in measurements made on the same subject in a repeatability study can then be ascribed only to errors due to the measurement process itself (51)		Pearson or Spearman or Kendall tau correlation coefficient between scores
Reproducibility	How much is the variation in measurements made on a subject under changing conditions? The changing conditions may be due to different measurement methods or instruments, measurements being made by different observers, or measurements being made over a longer period of time (within which the 'error-free' level of the variable could undergo non-negligible change) (51)	*Reproducibility across different statistical solutions:* Do different choices in the method used for the identification of DPs lead to similar sets of DPs? *Short-term reproducibility or reproducibility*: Are the sets of DPs derived at two administrations of the same dietary assessment tool to the same subjects within 1 year similar? Reproducibility of DPs is typically assessed following a previous assessment of reproducibility of a food-frequency questionnaire within a validation study *Long-term reproducibility or stability over time*: Are the sets of DPs derived at two or more administrations of the same dietary assessment tool to the same subjects over	Pearson (33) or Spearman (28) or Kendall tau correlation coefficient between scores; Intra-class correlation coefficient between scores (25); Congruence coefficient between loadings (11, 30)

		longer time periods (i.e., 2 years or more) similar? *Cross-study reproducibility:* Are the sets of DPs derived across centers (within the same study) or across different studies (potentially representing different populations or countries) similar?	
Validity	Does a test accurately measure what it claims to be measuring?		
Relative validity	Does a test compare well with a gold standard test? (53)	Are the sets of DPs derived on data from two different dietary sources similar? Relative validity of DPs is typically assessed following a previous assessment of relative validity of a food-frequency questionnaire against a gold standard tool within a validation study	Pearson (54) or Spearman (55) or Kendall tau (56) correlation coefficient between scores [crude or corrected (de-attenuated) for accounting for variation in time (57)]; Congruence coefficient between loadings (see (5) for details)
Construct validity	Does a test well measure the latent constructs that it is supposed to measure through operationalizations of the construct? (58)	Do the empirically derived DP scores resemble the latent DPs they should represent (in their composition and correlation with the other DPs)?	CFA (9, 27)

[1]ABBREVIATIONS: CFA: confirmatory factor analysis; DP: dietary pattern; LOA: limits of agreement

[2]See Supplemental Figure 1 for additional details

[3]For each statistical method mentioned, we provided an example study reference from the current review or from the companion one (5) to facilitate the association between research question and statistical method used to accomplish the objective

Supplemental Table 2. Cross-study reproducibility of *a posteriori* dietary patterns: details on dietary pattern composition[1]

Reference	*Location and Study*	*Dietary pattern composition*
Balder, 2003 (6)	Netherlands, Sweden, Finland, and Italy DIETSCAN (NLCS, SMC, ATBC, ORDET)	*From PCFAs based on unadjusted variables for energy intake*: (SALAD) VEGETABLE (common to all studies and different genders): high in raw leaf vegetables, tomatoes, carrots, cabbages and sometimes oil, poultry, rice, pasta and fish; PORK, PROCESSED MEAT, POTATOES (common to all studies and different genders): high in pork, processed meat, and potatoes; COOKED VEGETABLES (common to NCLS Ms and ORDET): high in cooked leaf vegetables, cabbages, legumes, and carrots; ALCOHOL (common to ATBC, SMC and ORDET): high in wine, beer, and spirits; SWEET AND/OR SAVORY SNACKS (common to NCLS Ms and Fs): high in savory snacks, nuts, sweets/candies, cakes/cookies; BROWN/WHITE BREAD SUBSTITUTION (common to NCLS Ms and Fs): high in bread substituters; *plus other 2 population-specific DPs not described in detail*
Castello, 2016 (12)	Spain EpiGEICAM, DDM-Spain	*From PCA on EpiGEICAM study data*: WESTERN: high in high-fat dairy products, processed meat, refined grains, sweets, energetic drinks and other convenience foods and sauces and low in low-fat dairy products and whole grains; PRUDENT: high in low-fat dairy products, vegetables, fruits, whole grains, and juices; MEDITERRANEAN: high in fish, vegetables, legumes, boiled potatoes, fruits, olives, and vegetable oil and low in juices *From PCA on DDM-Spain study data*: WESTERN: in addition with previous foods, low in white fish; PRUDENT: high in whole grains and juices but not on low-fat dairy products, vegetables and fruits; MEDITERRANEAN: high in some vegetables, legumes, potatoes, nuts, low-fat dairy products, sweets, and sugary and convenience foods, but not in fish, olive oil and fruits
Castello, 2016 (11)	Spain EpiGEICAM	*Dietary patterns based on original PCA from Castello and on reconstruction of loadings from Bessaoud, Adebamowo, and Terry*: WESTERN (Castello, Bessaoud, Adebamowo, Terry): high in high-fat dairy products (only cheese in Bessaoud), red and processed meat, refined grains, sweets, caloric drinks (not present in Bessaoud), and convenience food and sauces; Castello-PRUDENT, Bessaoud-MEDITERRANEAN, Adebamowo-PRUDENT, and Terry-HEALTHY: high in fish, fruits, and vegetables, also high in low-fat products ("dairy products" in Bessaoud); Castello-MEDITERRANEAN, Bessaoud-MEDITERRANEAN, Adebamowo-PRUDENT, and Terry-HEALTHY: high in fish, fruits, and vegetables, also high legumes (not present in Terry), nuts (not present in Bessaoud and Terry), and olive oil (not present in Adebamowo and Terry)

De Vito, 2019 (32)	USA, Italy, and Switzerland INHANCE	*From multi-study factor analysis on all the 7 available studies*: ANIMAL PRODUCTS AND CEREALS: high in total protein, zinc, phosphorus, riboflavin, sodium, niacin, thiamin, cholesterol, calcium, vitamin B6, iron, potassium, and total carbohydrates; ANTIOXIDANT VITAMINS AND FIBER: high in vitamin C, total fiber, total folate, potassium, total carotene, and vitamin B6; FATS: high in monounsaturated and polyunsaturated fatty acids, vitamin E, and saturated fatty acids *Study-specific DPs for the US studies only: 4 DPs with some variation but basically summarized as*: DAIRY PRODUCTS AND BREAKFAST CEREALS: high in calcium and low in niacin (or viceversa)
Judd, 2014 (26)	USA REGARDS	*From final PCA solution on the whole sample*: CONVENIENCE: high in mixed dishes with meat, pasta dishes, Mexican dishes, pizza, red meat, soup, fried potatoes, and Chinese dishes; PLANT-BASED: high in cruciferous, green leafy, dark yellow, and other vegetables, fruits, beans, and fish; SWEETS/FATS: miscellaneous sugar, desserts, bread, sweet breakfast foods, chocolate, candy, solid fats, and oils; SOUTHERN: high in added fats, eggs, fried food, organ meats, processed meats, and sugar-sweetened beverages; ALCOHOL/SALADS: high in salad dressing, green leafy vegetables, tomatoes, wine, butter, and liquor
Mannisto, 2005 (7)	Netherlands, Sweden, and Italy DIETSCAN (NLCS, SMC, ATBC, ORDET)	*From PCFAs on each study: common DPs*: VEGETABLES - VEG: high in vegetables, legumes, fruit, pasta, fish and oil; PORK, PROCESSED MEAT, POTATOES - PPP: high in pork, beaf, processed meats, potatoes, rice, poultry, liver, butter/low-fat margarine, pasta, and coffee; *plus other population-specific DPs not described in detail*
Moskal, 2014 (8)	Europe EPIC	*From overall PCA*: PC1: high in dietary fibre, vitamin C, beta-carotene and folate, low in saturated fatty acids, cholesterol, vitamin B12, retinol, and vitamin D; PC2: high in riboflavin, B6, folate, vitamin B12, vitamin C, beta-carotene, retinol, phosphorus, potassium and magnesium, low in starch; PC3: high in vitamin D, polyunsaturated fatty acids, thiamin, vitamin B6, and fibre, low in saturated fatty acids and retinol; PC4: high on calcium, total proteins, riboflavin, and phosphorus, low in polyunsaturated fatty acids and vitamin E
Schwerin, 1981 (43)	USA Ten-State Nutrition Survey (Ten-State), HANES I	*From PCA on Ten-State*: I: high in dairy products, and soups, and low in foods primarily sugar; II: high in nonsugary beverages and condiments and low in dairy products; III: high in cereals and grains, legumes and nuts, and eggs; IV: high in fruits, vegetables and juices, desserts and meats; V: high in poultry; VI: high in mixed dishes - protein, and shellfish; VII: high in fish and fats

Schwerin, 1982 (44)	USA Ten-State Nutrition Survey (Ten-State), HANES I, NFCS	*From PCA on Ten-State*: 7 DPs I: high in dairy products, and soups; II: high in nonsugary beverages and condiments; III: high in cereals and grains, legumes and nuts, and eggs; IV: high in vegetables and fruit, meats, and desserts; V: high in poultry; VI: high in mixed protein dishes and shellfish; VII: high in fish and fats and oils *From PCA on HANES I*: 8 DPs, of which 7 DPs similar to the Ten-State ones and 1 extra DP with greater consumption of sugary food and beverages *From PCA on NFCS*: 6 DPs, of which 5 were either identical over the decade or combinations of previous DPs

[1]ABBREVIATIONS: ATBC: Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; DDM-Spain: Determinantes de la Densidad Mamográfica en España; DIETSCAN: DIETary patternS and CANcer in four European countries project; DP: dietary pattern; EPIC: European Prospective Investigation into Cancer and Nutrition; EpiGEICAM: Grupo Español de investigación en Cáncer de Mama; F: female; HANES: Health and Nutrition Examination Survey; INHANCE: International Head and Neck Cancer Epidemiology consortium; M: male; NFCS: Nationwide Food Consumption Survey; NLCS: Netherlands Cohort Study on diet and cancer; ORDET: Ormoni e Dieta nella Eziologia dei Tumori in Italy; PC: principal component; PCA: principal component analysis; PCFA: principal component factor analysis; REGARDS: Reasons for Geographic and Racial Differences in Stroke; SMC: Swedish Mammography Cohort

Supplemental Table 3. Stability over time of a *posteriori* dietary patterns: details on dietary pattern composition[1]

Reference	Location and Study	Dietary pattern composition
Asghari, 2012 (25)	Iran TLGS	*From PCAs on different dietary sources and time-points*: IRANIAN TRADITIONAL (common to all 4 dietary data): high in vegetables, fruits, potatoes, dairy products, legumes and nuts, whole grains, tea and coffee, olives, eggs, red meat, and organ meat; WESTERN (common to all 4 dietary data): high in carbonated drinks, salty snacks and salty vegetables, sugars, sweets, desserts, vegetable oil, animal fats, fast foods, poultry, fish and other seafood and refined grains; COMBINED (FFQ3 data only): high in potatoes, tea and coffee, vegetable oils, eggs, legumes and nuts, sugar, whole grains and salty snacks
Borland, 2008 (28)	UK SWS	*From PCAs at baseline and at follow-up*: PRUDENT DIET: high in vegetables, fruit, wholemeal bread, rice/pasta, yogurt, breakfast cereals, low in 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; HIGH-ENERGY DIET: high in puddings, cakes/biscuits, potatoes/chips, vegetables, fruit, red/processed meat, fish, eggs, oils and full-fat spreads
Chen, 2015 (29)	Canada CCS, FFQVP	*From EFA on CCS study*: MEAT: high in red meat, cured/processed red meat, cured/processed meat, and mixed dishes; PLANT-BASED DIET: high in fruit, cruciferous vegetables, other green vegetables, beans, peas, other vegetables, tomato sauce, total cereals and grains, and whole grains; FISH: high in fish, processed fish, berries, and other local fruits, and low in cheese *From EFA on FFQVP study*: as above for the MEAT and FISH DPs, but the PLANT-BASED DP becomes: VEGETABLES/FRUITS: high in greens, tomato sauce, berries, and other vegetables; *plus an additional DP*: GRAINS: high in whole grains, cereal, grains and low in beer, white wine, and coffee

Crozier, 2009 (15)	UK SWS	*From PCAs at 3 time-points*: PRUDENT: high in fruit and vegetables, whole-meal bread, rice and pasta, yogurt, and low in chips and roast potatoes, sugar, white bread, processed meat, full-fat dairy products, crisps, Yorkshire puddings and savory pancakes, confectionery, and tea and coffee; HIGH-ENERGY DIET: high in fruit and vegetables, puddings, meat and fish, eggs and egg dishes, cakes and biscuits, full-fat spread, potatoes, crisps, and confectionery; *plus extra DPs not shared across time-points and not described in detail (low total variance explained and less interpretable)*
Cucò, 2006 (30)	Spain NA	*From PCFAs at the 5 time-points with some variation*: SWEETENED BEVERAGES AND SUGARS: high in sweetened beverages and sugars, and low in fresh fruit, vegetables, roots and tubers (signs inverted in some of the time-points); VEGETABLES AND MEAT (not present in the postpartum period): high in vegetables, roots and tubers, red meat, cured cold meats, olive oil, and eggs
Cutler, 2009 (31)	USA (Minnesota) EAT	*From PCFA at Time 1*: across cohort and gender, same set of 4 DPs identified: VEGETABLE: high in zucchini, squash, eggplant, kale and greens, spinach, peas and lima beans; FRUIT: high in oranges and grapefruit, apples and apple sauce, pears, grapes, bananas, strawberries, cantaloupe and melons, peaches, and plums and apricots; SWEET/SALTY SNACK FOOD: high in chocolate bars, other candy bars, candy with chocolate, brownies, cake, potatoes chips, and nachos; STARCHY FOOD: high in English muffins/bagels, grilled cheese, pancakes, and crackers for 3 subgroups, and high in mashed potatoes, lasagna, pretzels, macaroni and cheese, and spaghetti with sauce for 2 subgroups; *From PCFA at Time 2*: previous 4 DPs, not identical anymore across cohort and gender (except for young girls), but fairly similar: VEGETABLE AND FRUIT DPs: combined in older boys and girls and separate in younger boys (and girls); SWEET/SALTY SNACK FOOD: identical across cohort and gender; STARCHY FOOD: in younger and older girls only; FAST FOOD: high in hamburgers, French fries, fried food, nondiet soda; identified in all age/sex groups except young girls
Dekker, 2013 (10)	Netherlands Doetinchem Cohort Study	*From CA on each of the 3 surveys*: HIGH-FIBER BREAD: high percentage of total energy from high-fibre bread, cakes and cookies, and cheese; LOW-FIBER BREAD: high percentage of total energy from low-fibre bread, sugar-sweetened beverages, other alcoholic drinks and fries

Fung, 2001 (33)	USA HPFS	*From PCFAs at the 3 time-points*: PRUDENT: high in fruit, vegetables, poultry, fish, whole grains, and legumes; WESTERN: high in red and processed meat, French fries, eggs, high-fat dairy products, sweets, and refined grains
Gerdes, 2002 (34)	Denmark MONICA	*From PCFAs at each of the 3 surveys*: COARSE BREAD: high in coarse bread; BAKED GOODS AND SWEETS: high in cakes and biscuits, jam, honey, candy, ice cream, and soda; FRUIT AND VEGETABLES: high in fruit, juice, vegetables, and cheese; MEAT, POTATOES AND FAT: high in meat, sausages, potatoes, butter, fat, and margarine; PASTA AND RICE: high in Mediterranean and Asian cooking; BREAKFAST: high in porridge, oatmeal, milk, yogurt, jam, and honey
Lopez-Garcia, 2004 (35)	USA NHS	*From PCFAs at the 2 time-points and on the average consumption from the 2 time-points*: PRUDENT: high in vegetables, fruit, legumes, whole grains, fish, and poultry; WESTERN: high in red meat, processed meat, refined grains, sweets, desserts, French fries, and high-fat dairy products
Malik, 2012 (36)	USA NHS II	*From EFAs at the 5 time-points with some variation in the 2003 DPs*: PRUDENT: high in vegetables, fruit, legumes, fish, and better-quality grains, low in snacks and soda; WESTERN: high in desserts, snacks, processed meat, French fries, and refined grains, and low in vegetables, fruit, and fish
Mikkila, 2005 (16)	Finland Cardiovascular Risk in Young Finns Study	*From PCFAs at the 3 time-points with some variation described*: PATTERN 1: high in rye, potatoes, milk, butter, sausages, coffee (at all time-points), low in fruit and berry, and other dairy products (in 1980 and 2001); PATTERN 2: high in rye, vegetables, legumes and nuts, tea, rye, cheese and other dairy products (at all time-points), and alcoholic beverages (in 2001); *plus one extra DP not described in detail but not easily interpretable*

Mishra, 2006 (37)	UK Medical Research Council National Survey of Health and Development (1946 British Birth Cohort)	*From EFAs on 1999 data, but similar to 1982 and 1989 data*: Among Fs; ETHNIC FOOD AND ALCOHOL: high in Indian and Chinese meals, rice, pasta, oily fish and shellfish, olive oil, some vegetables, and alcoholic beverages; MEAT, POTATOES AND SWEET FOODS: high in red meat, bacon, ham, potatoes, sweet pies, cakes, puddings and desserts, and low in pasta, and skimmed milk; FRUIT, VEGETABLES AND DAIRY: low-fat and reduced-fat dairy products, fruit, some vegetables and whole-meal bread, and low in meat, meat products, and white bread Among Ms; ETHNIC FOOD AND ALCOHOL: high in Indian and Chinese meals, rice, pasta, shellfish, olives, some vegetables and legumes, and alcoholic beverages, and low in meat pies, fried chips, and animal fats; MIXED: high in many fruits and vegetables, low-fat/low-calorie yogurt and soya milk and cakes, sweet biscuits, sweet pies, puddings, desserts, confectionery, and ice cream
Newby, 2006 (23)	Sweden SMC	*From PCFA at both time-points (1987 and 1997) and confirmed with CFA at both time-points (1987 and 1997)*: HEALTHY: high in vegetables, fruit, whole grains, fruit juice, and cereal; WESTERN/SWEDISH: high in meat, processed meat, liver, refined grains, and potatoes; ALCOHOL: high in wine, spirits, snacks beer, and chocolate; SWEETS: high in sweet baked goods, chocolate, sugary foods, dairy desserts, soda, fruit soup, and refined grains; *plus 2 extra DPs not shared among the 2 time-points*
Newby, 2006 (22)	Sweden SMC	*From PCFA at both time-points (1987 and 1997) and confirmed with CFA at both time-points (1987 and 1997)*: with some variation in the Healthy DP (seafood, poultry, and eggs also contributed to HEALTHY DP in 1987, whereas legumes and soy products contributed to HEALTHY DP in 1997) HEALTHY: high in vegetables, fruit, whole grains, fruit juice, and cereal; WESTERN/SWEDISH: high in meat, processed meat, liver, refined grains, and potatoes; ALCOHOL: high in wine, spirits, snacks beer, and chocolate; SWEETS: high in sweet baked goods, chocolate, sugary foods, dairy desserts, soda, fruit soup, and refined grains; *plus 2 extra DPs not shared among the 2 time-points*
Nimptsch, 2014 (38)	USA NHS II	*From EFAs at the 5 time-points*: PRUDENT: high in vegetables, fruit, better-quality grains, fish, and poultry; WESTERN: high in desserts and sweets, snack foods, red and processed meat, French fries, and refined grains

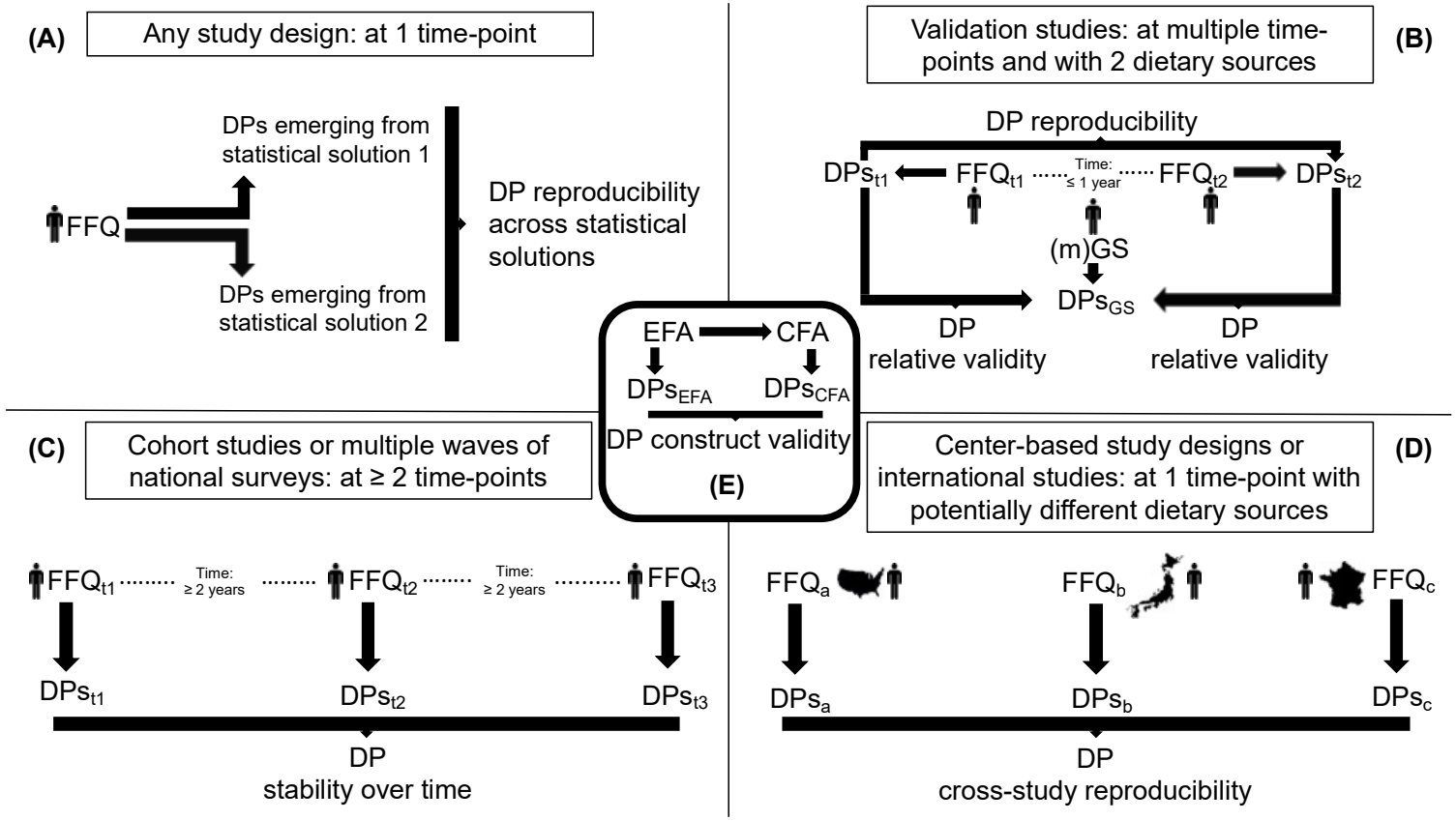
Northstone, 2005 (39)	UK ALSPAC	*From PCAs at 2 time points*: JUNK: high in high-fat and sugar content, processed and convenience foods; TRADITIONAL: high in meat, potatoes and vegetables; HEALTH-CONSCIOUS: high in vegetarian style foods, rice, pasta, cheese, salad, fish, and fruit
Northstone, 2013 (18)	UK ALSPAC	*From CAs at all time-points, in order of size*: PROCESSED: higher mean consumption of processed meat, pies and pasties, coated and fried chicken and white fish, pizza, chips, baked beans and tinned pasta, chocolate, sweets, sugar and diet and regular fizzy drinks; HEALTHY: higher mean consumption of non-white bread, reduced fat milk, cheese, yogurt and fromage frais, butter, breakfast cereal, rice, pasta, eggs, fish, vegetable and vegetarian dishes, soup, salad, legumes, fruit, crackers and crispbreads, high-energy-density sauces (e.g. mayonnaise), fruit juice, and water; TRADITIONAL: higher mean consumption of red meat, poultry, potatoes, vegetables, starch-based products (e.g. Yorkshire pudding), low-energy-density sauces (e.g. gravy), puddings, tea and coffee; PACKED LUNCH: higher mean consumption of white bread, margarine, ham and bacon, sweet spreads (e.g. honey), salty flavourings (e.g. yeast extract), crisps, biscuits, diet squash, tea and coffee
Northstone, 2008 (17)	UK ALSPAC	*From PCA on 4 time-points*: PROCESSED (at all time-points): high in high-fat and sugar content foods and processed and convenience foods; TRADITIONAL (at all time-points): high in meat, poultry, potatoes and vegetables; HEALTH CONSCIOUS (at 3, 4, and 7 ys only): high in salads, fruit, vegetables, fish, pasta and rice; HEALTH CONSCIOUS/VEGETARIAN (at 9 ys only): high in salads, fruit, vegetables, fish, pasta and rice, but also high in meat substitutes, pulses, nuts and vegetarian pies; SNACK (at 3 ys only): high in cheese, fruit, puddings, cakes, biscuits, and crisps
Northstone, 2008 (40)	UK ALSPAC	*From PCA on pregnancy data*: HEALTH-CONSCIOUS: high in salad, fresh fruit, rice, pasta, fish, pulses, and non-white bread; TRADITIONAL (British): high in all types of vegetables, some red meat, and poultry; PROCESSED: high in meat pies, sausage and burgers, fried foods, pizza, and chips; CONFECTIONARY: high in chocolate, sweets, biscuits, cakes and other puddings; VEGETARIAN: high in meat substitutes, pulses, nuts, and herbal tea; *From PCA on 4-y follow-up data*: TRADITIONAL DP lost, HEALTH-CONSCIOUS similar, the other 3 DP virtually identical in the dominant FG across time

Prevost, 1997 (41)	UK HALS	*From PCAs at HALS1 and HALS2 and also similar for Ms and Fs*: COMPONENT 1 (HIGH IN FRUIT AND VEGETABLES, LOW IN FAT): high in fresh fruit, salads, brown bread, fruit juice, green vegetables, spread (low-fat), milk (semi-skimmed), other vegetables, and root vegetables, low in chips, fried foods, and processed meats; COMPONENT 2 (HIGH IN ENERGY-DENSE FOODS): high in puddings/pies, cake, potatoes, biscuits, preserves, pulses, carcass meat, root vegetables, cream, cooked fruit, confectionery, green vegetables, milk, eggs, light desserts; COMPONENT 3 (HIGH IN CONVENIENCE FOODS): high in crisps, soft drinks, chips, fried food, coffee, pasta/rice, processed meat, low in tea and preserves; COMPONENT 4 (HIGH IN SUGARY FOODS, LOW IN VEGETABLES): high in confectionery, biscuits, cake, and low in green vegetables, root vegetables, pulses, other vegetables, and potatoes
Schulze, 2006 (42)	USA NHS II	*From EFAs at all time-points*: PRUDENT: high in fruits, vegetables, whole grains, fish, poultry, and salad dressings; WESTERN: high in red and processed meat, refined grains, sweets and desserts, and potatoes
Togo, 2004 (27)	Denmark MONICA	*From CFA among Ms, at both baseline and follow-up*: GREEN: high in wheat bread and rye bread with whole grains and/or bran; raw and boiled vegetables, fruit, rice, cheese, fish, milk products and low in white (wheat) bread; SWEET: high in cake, biscuits, or other baked goods, candy or chocolate, soft drink or ice-cream, and jam/marmalade or honey; TRADITIONAL: high in meat, paté and meat for bread, potatoes, white (wheat) bread, sausage, butter, lard and hard margarine, and eggs; *From CFA among Fs, at both baseline and follow-up*: GREEN: same as for Ms; SWEET-TRADITIONAL: high in candy or chocolate, cake, biscuits, or other baked goods, paté and meat for bread, white (wheat) bread, butter, lard and hard margarine, soft drink or ice-cream, jam/marmalade or honey, potatoes, meat, and sausage
van Dam, 2002 (45)	USA HPFS	*From PCFAs at the 2 time-points*: PRUDENT: high in vegetables, legumes, fruit, whole grains, fish, and poultry; WESTERN: high in red meat, processed meat, refined grains, French fries, high-fat dairy products, sweets and desserts, high-sugar drinks, and eggs

Weismayer, 2006 (9)	Sweden SMC	*From EFAs at baseline and follow-up and confirmed by CFAs at baseline and follow-up*: HEALTHY: high in fruits, tomatoes, vegetables, cereal, and fish; WESTERN: high in meat, processed meat, fried potatoes, soft drinks, and sweets; ALCOHOL: high in beer, wine, and liquor consumption as well as snack consumption; *plus extra DPs difficult to interpret or dominated by only 1 high loading*

[1]ABBREVIATIONS: ALSPAC: Avon Longitudinal Study of Parents and Children; CA: cluster analysis; CCS: Case-Control Study, here intended as the full name of one of the included studies and not as the case-control study design; CFA: confirmatory factor analysis; DP: dietary pattern; EAT: Eating Among Teens; EFA: exploratory factor analysis; F: female; FFQ: food-frequency questionnaire; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1/2/3; FFQVP: Food-Frequency Questionnaire Validation Project; FG: food groups; HALS: Health and Lifestyle Survey; HPFS: Health Professionals Follow-up Study; M: male; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not available; NHS: Nurses' Health Study; PCA: principal component analysis; PCFA: principal component factor analysis; SMC: Swedish Mammography Cohort; SWS: Southampton Women's Survey; TLGS: Teheran Lipid and Glucose Study; y: year

Supplemental Figure 1. Schemes of dietary pattern identification processes related to the assessment of their reproducibility and validity. Specifically, reproducibility and/or validity of dietary patterns can be assessed in the following set-ups: Panel (A): at one time point and with one dietary source; Panel (B): at multiple time points and with two dietary source, Panel (C): at multiple time points; Panel (D): across centers from the study or across different studies. All of these settings may include confirmation of the identified dietary patterns with confirmatory factor analysis [Panel (E)][1]



[1]ABBREVIATIONS: CFA: confirmatory factor analysis; DPs: dietary patterns; EFA: exploratory factor analysis: FFQ: food-frequency questionnaire; GS: gold standard dietary assessment tool; mGS: mean of intakes from multiple administrations of the same gold standard tool

Supplemental References
[In addition to references cited in the main text]

51. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 2008;31(4):466–75.
52. Ryman TK, Boyer BB, Hopkins S, Philip J, O'Brien D, Thummel K, Austin MA. Characterising the reproducibility and reliability of dietary patterns among Yup'ik Alaska Native people. Br J Nutr 2015;113(4):634-43. doi: 10.1017/S0007114514003596.
53. Willett W. Nutritional Epidemiology. Oxford University Press; 2012, Oxford, UK.
54. Beck KL, Kruger R, Conlon CA, Heath AL, Coad J, Matthys C, Jones B, Stonehouse W. The relative validity and reproducibility of an iron food frequency questionnaire for identifying iron-related dietary patterns in young women. J Acad Nutr Diet 2012;112(8):1177-87. doi: 10.1016/j.jand.2012.05.012.
55. Ambrosini GL, O'Sullivan TA, de Klerk NH, Mori TA, Beilin LJ, Oddy WH. Relative validity of adolescent dietary patterns: a comparison of a FFQ and 3 d food record. Br J Nutr 2011;105(4):625-33. doi: 10.1017/S0007114510004137.
56. Bountziouka V, Tzavelas G, Polychronopoulos E, Constantinidis TC, Panagiotakos DB. Validity of dietary patterns derived in nutrition surveys using a priori and a posteriori multivariate statistical methods. Int J Food Sci Nutr 2011;62(6):617-27. doi: 10.3109/09637486.2011.561783.
57. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr 1999;69(2):243-9.
58. Strauss ME. Introduction to the special section on construct validity of psychological tests: 50 years after Cronbach and Meehl (1955). Psychol Assess 2005;17(4):395. doi: 10.1037/1040-3590.17.4.395.