Reproducibility and validity of *a posteriori* dietary patterns: a systematic review[a]

Valeria Edefonti[1], Roberta De Vito[2], Michela Dalmartello[1], Linia Patel[1], Andrea Salvatori[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 Computer Science, Princeton University, Princeton, NJ, 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 02-50320853; fax: 0039 02-50320866; email: valeria.edefonti@unimi.it.

Word count: 7561 words.

Number of figures: 2 figures.

[a] Supplemental Tables 1, 2, 3, and 4 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

Number of tables: 4 tables.

Running title: reproducibility and validity of dietary patterns.

List of abbreviations: 24HR/48HR: 24/48 hour recall; ARI: adjusted Rand index; CA: cluster analysis; CFA: confirmatory factor analysis; DP: dietary pattern; EFA: exploratory factor analysis; FFQ: food-frequency questionnaire; FG: food group; m24HR: mean 24 hour recall; mDR: mean dietary record; PCA: principal component analysis; SMC: Swedish Mammography Cohort.

List of financial support: Valeria Edefonti was supported by Università degli Studi di Milano 'Young Investigator Grant Program 2017'. The funder has no role in any phase of this systematic review.

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

Abstract - 298 words

The effective use of dietary patterns (DPs) remains limited. There is a need to assess their consistency over multiple administrations of the same dietary source, different dietary sources or across different studies. Similarly, their generalizability should be based on a previous assessment of DP construct validity. However, to date, no systematic reviews on reproducibility and validity of *a posteriori* DPs have been carried out. In addition, several methodological questions related to their identification are still open and prevent a fair comparison of epidemiological results on DPs and disease.

A systematic review of the literature on the PubMed database was conducted. We identified 218 articles, 64 of which met the inclusion criteria. Of these, the 38 articles dealing with reproducibility, relative and construct validity of DPs were included.

These articles (published in 1999 – 2017, 53% from 2010 onwards) were based on observational studies conducted worldwide. The 14 articles that assessed DP reproducibility across different statistical solutions examined different research questions. Included were: the number of food groups or subjects, input variable format (as well as adjustment for energy intake), algorithms and the number of DPs to retain in cluster analysis, rotation method and score calculation in factor analysis. However, we identified at most 3 articles per research question on DP reproducibility across statistical solutions. From another 15

articles, reproducibility of DPs over shorter (≤ 1 year) time periods was generally good and higher than DP relative validity (as measured across different dietary sources). Confirmatory factor analysis was used in 15 of the included articles. It provided reassuring results in identifying valid dietary constructs characterizing the populations under consideration. Based on the available evidence, only suggestive conclusions can be derived on reproducibility across different statistical solutions. Nevertheless, most identified DPs showed good reproducibility, fair relative validity and good construct validity.

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

a posteriori dietary patterns; cluster analysis; construct validity of dietary patterns; consistency of dietary patterns; factor analysis; generalizability of dietary patterns; reproducibility of dietary patterns; relative validity of dietary patterns; validity of dietary patterns.

Introduction

Since the early 80's, dietary patterns (DPs)[b] have been used to synthesize multiple related dietary components in combined variables representing key dietary habits and/or the overall diet in free-living individuals. Interest in DPs is also motivated by well-known interactive effects of foods that are eaten together and by data dimensionality/multiple testing issues affecting the statistical analysis of many single food groups (FGs) or nutrients (1).

However, the lack of consistent methodology in deriving DPs has severely limited the ability to draw firm conclusions about the health risks or benefits associated with DPs (2). Indeed, only the most recent version of the Dietary Guidelines for Americans (3) has included evidence on DPs.

In 2012, the National Cancer Institute launched the Dietary Patterns Methods Project to support standardized and parallel analyses on selected *a priori* (or index-based) DPs and mortality outcomes in 3 large US cohorts (2). An index-based approach to DPs was chosen because results can be readily translated into dietary recommendations. Based on the application of multivariate statistical analysis to the available data, *the a posteriori* (or data-

[b] ABBREVIATIONS: 24HR/48HR: 24/48 hour recall; ARI: adjusted Rand index; CA: cluster analysis; CFA: confirmatory factor analysis; DP: dietary pattern; EFA: exploratory factor analysis; FFQ: food-frequency questionnaire; FG: food group; m24HR: mean 24 hour recall; mDR: mean dietary record; PCA: principal component analysis; SMC: Swedish Mammography Cohort

78 driven) DPs offer the advantage of representing actual dietary behavior in a population at a
79 certain time-point. If the population variability is well captured, the set of identified *a*
80 *posteriori* DPs provide a realistic representation of eating choices (4). In addition, the *a*
81 *posteriori* approach could capture rare, but well-characterized, dietary behaviors of
82 subpopulations, including ethnic minorities (5).
83 Subjective decisions have been constantly reported as a limitation in studies deriving *a*
84 *posteriori* DPs with principal component analysis (PCA), exploratory factor analysis (EFA),
85 or cluster analysis (CA) (6). These decisions concern input variable format and potential
86 transformation, number of input variables and food grouping schemes, estimation method
87 as well as criteria for model selection, including how to choose the number of DPs to retain
88 (7). Although subjectivity in PCA/EFA and CA is often emphasized, very few papers have
89 provided a formal comparison of different modeling strategies based on objective criteria.
90 The reproducibility of DPs across different statistical solutions has rarely been a concern.
91 Similarly, confirmatory factor analysis (CFA) still has limited use in the validation of EFA-
92 based DPs and in the development of constructs representing correlation structures among
93 FGs and among DPs. Even though this should the first step for the generalization of DPs to
94 other studies, their construct validity has been investigated in a few papers.

More generally, the reproducibility of similar *a posteriori* DPs across time, studies and/or countries have not been extensively assessed so far (5, 8). Although in the literature there is a distinction between consistency of DPs across multiple administrations of the same dietary assessment tool in a short period of time (reproducibility) (i.e. (9)) and consistency over longer time-periods (stability over time) (i.e. (10)), unsolved methodological issues have been reported in both these analyses (11, 12). Similarly, the comparison of *a posteriori* DPs across different dietary assessment tools (relative validity) (i.e. (9)) poses unsolved methodological issues (13).

To our knowledge, no attempts have been carried out so far to collect and summarize the existing evidence on reproducibility and validity of *a posteriori* DPs. This paper provides details on the literature search and selection process and also summarizes the evidence on reproducibility, relative and construct validity of DPs. A companion review will include information on stability of DPs over longer time-periods and reproducibility of DPs across studies.

Methods

Literature search strategy

We carried out a systematic search 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 search was restricted to human studies reported in the English language and published up to January 11, 2019 and followed the guidelines from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) group (14). Two authors (MD and VE) independently selected the articles and retrieved and assessed the potentially relevant ones. The reference lists of the identified articles as well as other systematic reviews focusing on similar topics were also scanned. Discrepancies in article selection were resolved by involving a third researcher (MF).

Inclusion and exclusion criteria

Articles were included or excluded according to the following criteria.

A posteriori dietary patterns

We focused our systematic review on *a posteriori* DPs. However, in the absence of previous known reviews on this topic, we preferred not to add the term *"a posteriori"* to our search string. Therefore, we further excluded papers presenting reproducibility or validity of *a priori* DPs only or applying reduced rank regression. We included in the review papers comparing *a priori* and *a posteriori* DPs as far as they provided information on reproducibility and validity

130 of *a posteriori* DPs. We also considered papers comparing PCA (or EFA) and CA, but we
131 excluded them when concentrating only on the comparison between PCA/EFA- and CA-
132 based DPs (e.g. (15)).

133 *Reproducibility and validity of a posteriori dietary patterns*

134 In recent years, disagreements in terminology across different scientific areas have
135 characterized the concepts of reproducibility, replicability, and validity of scientific findings
136 (16) (17). In **Supplemental Table 1**, we introduce the basic definitions adopted in the current
137 review as well as the statistical tools used for their assessment. We integrate basic
138 terminology within the scientific process of DP identification in nutritional epidemiology.

139 **Figure 1** shows prototypical paths of DP identification processes related to reproducibility
140 and validity of DPs. Dietary patterns are identifiable within any study design and starting
141 from any dietary assessment tool source. If one dietary source is used at one time point, the
142 assessment of DP reproducibility arises from the use of different statistical approaches for
143 DP identification [Panel (A)]. Within the validation study of a new food-frequency
144 questionnaire (FFQ), the same FFQ was administered twice (within 1 year) and compared
145 with a gold standard dietary assessment tool [a dairy record (DR) or (multiple administration
146 of) a 24-hour recall (24HR)] carried out on the same time interval and sample; DP
147 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.

In summary, in the literature review, we distinguished the following definitions of reproducibility of DPs: 1. **across different statistical solutions**: the extent to which similar DPs are consistently seen when a change occurred in: a. input variable format or scale; b. number of input variables; c. estimation method; or d. criteria for model selection (including number of DPs to retain); **2. over time**: the extent to which similar DPs are consistently seen

over short (i.e. ≤1 year) (traditionally called reproducibility in nutritional epidemiology) or longer time periods (i.e. ≥2 years) (stability over time); **3. across centers or studies (potentially representing different populations or countries)**: the extent to which similar DPs are common to diverse subsamples of interest, as opposed to study-specific DPs (cross-study reproducibility).

In the assessment of reproducibility across statistical solutions, we excluded papers that choose the number of clusters to retain with objective criteria (e.g. (18)), within an analysis of the association between DPs and disease. In the assessment of cross-study reproducibility, we excluded papers based on a merged data matrix (generated by combining data from all the studies) approach (e.g. (19)), where it was not possible to identify study-specific DPs and their potential reproducibility. Finally, we included papers using “internal validity” or “internal stability” indexes to choose the optimal number of clusters in the section on reproducibility of DPs across different statistical solutions. Although the terminology looks misleading, the research question was how to choose the number of clusters to retain and this was assessed with validity- or stability-based criteria for optimal solution identification.

The current review included and summarized evidence on reproducibility of DPs over shorter time periods and reproducibility across statistical solutions.

We also distinguished between construct validity and relative (or comparative) validity of DPs (**Supplemental Table 1)**. Construct validity indicates whether a test measures its targeted latent constructs through suitable operationalizations of the constructs; in nutritional epidemiology, it deals with the ability of the empirically derived DP scores to resemble the latent DPs in their composition and correlation with the other DPs. The relative validity of DPs has borrowed its meaning from the relative validity of a FFQ; it indicates the ability of FFQ-based DPs to resemble those derived on the gold-standard tool. We included papers assessing either construct or relative validity of DPs. We excluded papers that only assessed validity of DPs against socio-demographic characteristics, lifestyle habits, nutrient/food profiles from the same dietary source, nutritional biomarkers, markers of disease, or a disease of interest (e.g. (20)).

Finally, we excluded those studies that, while focusing on the association between some identified DPs and a disease, provided assessments of internal reproducibility with the split-half approach and/or reliability measured as internal consistency with Cronbach's alpha (e.g. (20-22)).

Data extraction

Quantitative and qualitative data were extracted from each of the studies selected 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 (type of design, brief description of data collection, number and age of the participants, and years of follow-up); 3. dietary assessment tools used; 4. DP identification method; 5. DP name and composition; 6. statistical methods used for the assessment of reproducibility and/or validity of DPs; and 7. main results on DP reproducibility and validity.

Quality assessment of the included studies

Each article that met the inclusion criteria was independently rated for quality by all researchers, except one (MF), using the “Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies” from the National Institutes of Health, National Heart, Lung, and Blood Institute (23). If the ratings differed, then the remaining author (MF) was considered for quality adjudication. Involved researchers used the available study rating tools on the range of items included in each tool to judge each study to be of "good," "fair," or "poor" quality. The reference tools used depended on the study design and included the “Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies” and the “Quality Assessment Tool for Case-control Studies” (23); for the quality assessment of validation studies, we adopted the “Quality Assessment Tool for Observational Cohort and

220 Cross-Sectional Studies", in accordance with the presence of repeated dietary measures.
221 Since our review was not focused on any specific outcome of interest, the rating system
222 items that dealt with: 1. the presence of an outcome, or 2. the association between exposure
223 and outcome were consistently given a "cannot determine/not reported/not applicable" score
224 across all the studies. Thus, the maximum rating for cohort/cross-sectional studies was
225 equal to 7 (out of the original 14 items) and the one for case-control studies was equal to 9
226 (out of the original 12 items). In addition, we decided that the item asking about reliability,
227 validity, and consistent definition of the exposure (number 9 in the cohort/cross-sectional
228 design tool and 10 in the case-control design tool) was concerned with the dietary
229 assessment tools used to measure dietary information. When the assessment of either
230 reproducibility or validity was performed on a FFQ, we marked "yes" in correspondence to
231 the tool item. When other dietary assessment tools were used instead of a FFQ, we marked
232 "yes" when either multiple administrations of a 24HR or a DR were provided. When a
233 validation study was assessed for quality, we marked this item with a "not applicable" in the
234 absence of any previous publication on FFQ reproducibility and validity. We did not consider
235 applicable to our quality assessment process the part of point 10 asking for reliability of the
236 risk measure in the case-control study design tool.

237 In general terms, a "good" study has the least risk of bias due to flaws in study design or
238 implementation, a "fair" study is susceptible to some bias deemed not sufficient to invalidate
239 its results, whereas a "poor" rating indicates significant risk of bias. We followed the website
240 guidelines (23) and did not base our final evaluation on a cut-off approach on the total score
241 (calculated summing up the 1's corresponding to "yes"), but we carefully evaluated the "no"
242 items to assess the overall risk of bias of the examined study. Finally, we chose not to
243 exclude studies on the basis of their quality, because of the lack of previous evidence on
244 reproducibility and/or validity of DPs.
245
246 **Results**
247 An initial literature search of the PubMed database identified 218 articles, of which 181
248 remained when we limited the search to publications related solely to humans and written
249 in the English language. Their full texts were retrieved for detailed evaluation. After the
250 exclusion of 35 review articles, 124 original research articles were also excluded because
251 they met the exclusion criteria indicated previously. In detail, the most frequent reasons for
252 exclusion were as follows: DPs intended as a synonym of dietary habits; *a posteriori* DPs
253 not identified in the paper [i.e. *a priori* DPs, DPs from reduced rank regression (either
254 exploratory or confirmatory), treelet transform, or latent class models], or just compared with

255 the *a priori* ones; PCA- or EFA-based DPs compared with CA-based DPs, with no separate
256 analyses on either approach; reproducibility and validity of FFQs and not of DPs; split-half
257 or Cronbach's alpha only; DP validity assessed against subjects' characteristics or a disease
258 of interest; conference abstracts not published as a full text article. Forty-two additional
259 articles were identified from manual searches of reference lists of selected original and
260 review articles. Thus, 64 articles were included in our systematic review. Of these, 38 articles
261 were included in the current review and were concentrated on reproducibility, relative and
262 construct validity of DPs; the 34 articles that focused on stability of DPs over time and on
263 their reproducibility across studies were included in an additional review. Eight papers (10,
264 11, 24-29) were common to both reviews (**Figure 2**).
265 General characteristics and study design information from the 38 studies on reproducibility,
266 relative and construct validity of DPs (9-11, 13, 24-57) are presented in **Table 1**. The articles
267 were published between 1999 and 2017, with 53% of them published from 2010 onwards;
268 the studies were carried out in several areas in the world, including Europe and North
269 America, but Asia and Oceania were also well represented with 6 and 2 articles,
270 respectively. A few articles were based on the same studies, including those from the
271 Swedish Mammography Cohort (SMC) (10, 26-28, 33), from the MONItoring of trends and
272 determinants in CArdiovascular Disease (MONICA) study (29, 47), and those from the

European Prospective Investigation into Cancer and Nutrition (EPIC) study (49, 51, 55). All the articles were based on observational studies, including 1 case-control (45), 18 cohort (10, 13, 24, 25, 27, 28, 35, 36, 42, 43, 46, 48, 49, 51, 53-56) and 9 cross-sectional (38, 39, 41, 44, 47, 48, 50, 52, 57) studies, 1 multiple administration of the same survey (29), and 9 validation studies of FFQs (9, 11, 30-34, 37, 40). One study included adult men only (9), 11 studies included adult women only (10, 27, 28, 30, 33, 36, 37, 39, 49, 52, 55), with some of them based on pregnant women (36, 37, 39); one article was based on children (56) and another one on adolescents (13). When available, the (total) follow-up time ranged from 1 month (30) to 14 years (51). Dietary assessment instruments were administered between 1982 – 1983 (29) and 2014 - 2015 (34), with assessments equally carried out in the '80s, '90s and 2000s, and a few ones in 2000 - 2010. With a few exceptions (35, 38, 42, 46, 50), the FFQ was the main dietary assessment tool used; in most studies, the FFQs were self-administered (8 FFQs were interviewer-administered only) and had a reference period of 1 year, with the obvious exception of the FFQs assessing diet during pregnancy (37, 39) and of the SMC FFQ (6 months) (10, 26-28, 33). The number of food items inquired in the FFQs ranged from 26 (29, 47) to 284 (43), with 56% of the FFQs showing ≥100 items. When 2 FFQ administrations were available, the median time interval between them was 12 months. Reproducibility and/or relative validity of the FFQs were directly assessed within the 9

validation studies included in the review (9, 11, 30-34, 37, 40); in addition, 14 articles reported on a previous assessment of FFQ reproducibility and/or relative validity (10, 13, 24, 25, 27, 28, 40, 41, 48, 49, 51-53, 55), whereas 9 articles did not report any information (29, 36, 39, 43-45, 47, 56) or declared that they did not test for them (54).

A different dietary assessment tool was used in 16 articles, including the 9 articles based on validation studies of FFQs (9, 11, 30-34, 37, 40). In 7 articles, information from 1 (35) or multiple administrations of the same 24HR format was collected, with number of collecting occasions ranging from 2 (50) to 18 (6*3 consecutive day 24HRs) (32) and completion of the form in different combinations of time occasions and consecutive/non-consecutive days; a DR was used in 10 articles, with reference time periods varying from 3 (13, 40) to 7 (9, 31, 33, 47) days, weighing system adopted (30, 33, 38, 47) or not, and single (13, 30, 35, 39, 40, 47) or multiple (9, 31, 33, 38) administrations of the same tools provided.

No matter of the dietary assessment tool used, the number of FGs defined from the available food items ranged from 15 (56) to 56 (24, 35), with a median value of 30.5 FGs included in the statistical analysis. When information from more than 1 dietary source was available, the same food grouping scheme was adopted across the different sources in all the articles (9, 11, 13, 30-35, 37-40, 47).

Among the selected papers, 11 (29%) were based on studies of “good” quality, 17 (45%) on studies of “fair” quality, and 10 (26%) on studies of “poor” quality.

Tables 2, 3, and **4** present details on DP identification method, on methods for the assessment of DP reproducibility and/or validity, and main results on their reproducibility and validity. Details on DP composition are presented in **Supplemental Tables 2**, **3**, and **4**.

Among the 38 articles included, 30 performed PCA, EFA, or CFA and 6 performed CA (25, 42-44, 46, 56), whereas 2 articles carried out both EFA/CFA and CA (40, 51). In addition, 7 (22%) of the articles that carried out EFA or PCA assessed matrix factorability before starting the statistical analysis (30, 32, 34, 37, 40, 41, 50) (data not shown).

Table 2 concerned reproducibility of DPs derived from different statistical solutions, with 8 papers considering PCA/EFA (26, 36, 41, 45, 48, 50, 51, 54, 57) and 6 considering CA (42-44, 46, 51, 56). The proposed research questions dealt with: 1. input variable preprocessing [i.e. adjustment by energy intake (26, 36, 42), standardization (46), and dichotomization (26)]; 2. number of input variables (45) and subjects (57) to be included in the analysis; 3. solution method for CA (43, 44, 56); 4. rotation method for PCA/EFA (41, 48) and CFA (50); 5. number of DPs to retain (25, 43, 44, 51); 6. score calculation [natural vs. applied (i.e. calculated using loadings from a separate PCA on subsample 1 and data from subsample

2) scores] in PCA (48). One article (25) proposed the comparison of different statistical solutions within the assessment of DP stability over time.

Concerning input variable preprocessing, 2 articles considered adjustment by energy intake with the residual method (26, 36) in PCA/EFA, whereas the third one (42) considered percent daily energy contribution vs. number of servings in CA; in the comparison between unadjusted and energy-adjusted solutions, 1 article used the correlation coefficient (36) and another one (26) the Procrustes rotation method. Independently of the statistical approach and type of adjustment used, the conclusions on the comparison between energy-adjusted and unadjusted solutions were similar across papers (Supplemental Table 2**)**: 1. With PCA/EFA, the DPs extracted were generally similar (in terms of loadings and percentages of explained variances); 2. With CA, the DPs were similar (in terms of higher/lower mean intakes of the FGs characterizing the clusters) and subgroups with high-energy contribution were consistently clustered across solutions; 3. When available, correlation coefficients between similar DPs under the 2 solutions were >0.8; 4. DPs with high loadings on energy-contributing FGs were lost with energy adjustment (36); and 5. the ability of CA to differentiate FGs with higher-then-mean intakes seemed higher with number of servings variables (42).

342 In addition, 2 articles considered the effect of standardizing or not FG intakes (expressed as
343 percentage of daily energy intake) in CA (46) and of dichotomizing FGs with more than
344 75% of nonusers (26). In the former case (46), both the approaches led to well-separated
345 and interpretable 6-cluster solutions that were stable and equivalent as to discriminant
346 analysis; however, composition and number of subjects per cluster were different. An
347 unstandardized solution was suggested as standardized variables just allowed to isolate
348 one or a few clusters including extreme individuals, whereas the remaining clusters were all
349 very similar one to other. In the latter case (26), the Procrustes rotation method confirmed
350 that dichotomizing variables with a high percentage of nonusers did not affect the FGs with
351 significant factor loadings, the magnitude of the factor loadings or the explained variance,
352 and thus the order of the extracted DPs.
353 Two articles assessed the effect of different numbers of: 1. input variables (from different
354 food grouping schemes) in PCA-derived DPs (45); or 2. subjects to include in PCA and CFA
355 (not based on previous EFA) in a study combining 2 studies from France and Spain (57). In
356 the former case-control study on endometrial cancer (45), the DPs identified according to 3
357 food grouping schemes (168 useable FFQ items, or 56 FGs from nutrient content or use
358 classification, or 36 FGs from the United States Department of Agriculture suggestions) were
359 not materially different except for the total variance explained in food use, which increased

360 as the detail included in the PCA decreased (up to ~17% with 36 FGs). However, for both
361 DPs, exact agreement in tertile classification decreased as the difference in the number of
362 items used for PCA increased and misclassification rates were higher for the Healthy DP. In
363 the latter article (57), PCA and CFA were carried out on 1000 randomly selected samples
364 from 4 different set-ups [100%, 50%, or 25% of the French study (1236 subjects) and 100%
365 of the Spanish study (274 subjects)]. From the bootstrap-based distributions of the factor
366 loadings to each FG for each DP, a more consistent set of CFA-based, rather than PCA-
367 based, DPs was identified across the set-ups. CFA-based DPs outperformed PCA-based
368 ones especially when smaller sample sizes were considered.
369 Three articles (43, 44, 56) were concerned with the choice of the optimal algorithm for
370 performing CA and compared the mostly used k-means and Ward's minimum variance
371 algorithms with flexible beta (43), with k-medians (44), or with Gaussian mixture models
372 (56), in a complex set-up of varying number of clusters. Together with them (43, 44, 56),
373 another 2 articles assessed the simpler issue of the optimal number of clusters to retain
374 when a k-means algorithm was carried out (25, 51). Finally, Fransen et al. (51) considered
375 the same research question for PCA and EFA too. In the comparison of clustering algorithms
376 (43, 44, 56), the k-means provided the highest reproducibility of the cluster solutions with all
377 different numbers of clusters, as compared to the Ward's minimum variance (43, 44), flexible

beta algorithm (43), and k-medians (44). For all possible numbers of solutions, the Gaussian mixture model was more similar to the k-means algorithm than to the Ward's one; however, the best Gaussian mixture model identified from the data implied FG variances to vary within and between clusters and it was therefore more general than the equivalent model subsumed by the k-means algorithm (56). With respect to the choice of the optimal number of clusters, 1 article (43) adopted a split-half cross-validation approach and used the median log-ratio value of between- versus within-cluster variances of the available FGs, after having previously identified the optimal algorithm as the k-means algorithm [with Hubert and Arabie's Adjusted Rand Index (ARI), kappa and Cramer's V statistics]; a similar article (44) identified the optimal combination of clustering method and number of clusters by using the box-plot and average value (over 20 repetitions of each algorithm) of the distribution of Cramer's V statistic and ARI; the paper by Greve et al. (56) assumed that the optimal number of clusters was the one that provided more similar solutions across the different algorithms, based on pairwise comparisons of ARI values.

Finally, when no algorithm choice was allowed and the k-means algorithm was carried out (25, 51), the optimal number of clusters to retain was identified with internal cluster validity (e.g. Calinski-Harabasz index, Davies-Bouldin index, and prediction-strength method) and

395 stability (e.g. Jaccard) indexes (25, 51); for PCA/EFA the usual criteria for identifying the
396 optimal number of factors to retain were adopted (51).
397 Three articles were concerned with the choice of the optimal rotation method in EFA (41) or
398 PCA (48) and of a combination of rotation method and cut-off for FG inclusion in EFA and
399 CFA (50). Based on 2 close administrations (at 15 days apart) of the same FFQ, the first
400 article (41) assessed the effect on DP repeatability of 2 orthogonal (varimax and quartimax)
401 and 2 non-orthogonal (promax and oblimin) rotations, as compared to an unrotated solution.
402 The main conclusions were the following ones: 1. In the unrotated solutions, the identified
403 DPs were similar over the 2 FFQ administrations, although the limits of agreement were
404 wide; 2. For either orthogonal or non-orthogonal rotation, the agreement was poorer
405 between corresponding DPs at the 2 time-points, as compared to the unrotated solution; 3.
406 Between the orthogonal rotations, a better agreement was found for the quartimax rotation;
407 4. Between the non-orthogonal rotations, a better agreement was found for the oblimin
408 rotation (41). Based on the baseline data from a population survey, the second article (48)
409 concluded that DPs derived from varimax and promax rotations were qualitatively similar
410 and opted for the promax solution which allows correlations between DPs. Based on another
411 population-based survey, the third article (50) assessed the effect on DP reproducibility of
412 different cut-offs (i.e. |0.20| or |0.25|) for FG inclusion and rotation method (i.e. varimax,

promax, and oblimin), with the following conclusions: 1. A |0.25| cut-off for FG inclusion in EFA provided reproducible results for any rotation methods; 2. A |0.25| cut-off for FG inclusion in CFA defined a valid CFA model; 3. A better model fit was observed for CFA with promax and then varimax, and last oblimin rotation solution, with small but significant correlations between factors.

Finally, 1 article (48) assessed the difference between using natural and applied ~~(i.e. calculated using loadings from a separate PCA on subsample 1 and data from subsample 2)~~ PCA-based scores. It concluded that: 1. Correlation coefficients between natural and applied scores for the same DP were high (≥0.89) and significant; 2. No systematic bias was found in the Bland-Altman plot comparing natural and applied scores; 3. For both DPs, the agreement was relatively weak in men and only acceptable in women, as indicated by the relative variation measure (48).

Table 3 concerned reproducibility and/or relative validity of DPs, with 7 articles assessing DP reproducibility and relative validity together (9, 11, 30-34), 7 articles assessing relative validity of DPs only (13, 35, 37-40, 47), and 1 article assessing DP reliability (54). All the articles derived DPs from PCA or EFA and 1 article additionally derived DPs with CA (40). Dietary patterns were separately identified on FFQ data at time 1 and 2 (9, 11, 30-34, 54), and/or on mean intakes from multiple administrations of the gold standard dietary

assessment tool [mean 24HR (m24HR) or mean DR (mDR)] (9, 11, 31-34, 37, 38). The DP identification process was similar in all the articles and generally included a combination of eigenvalue>1, scree test, and interpretability to choose the number of DPs to retain, a varimax rotation to improve DP interpretation and descriptive labeling for naming the identified DPs. Three articles proposed standardization [with (47) or without Kaiser normalization (39)] or log-transformation of input variables (31, 38, 54) and adjustment by energy intake with the residual method for either input variables (38) or DP scores (31).

The number of described DPs ranged from 2 to 5, with 47% of the articles naming and describing 2 DPs; however, 7 articles (9, 13, 32-34, 37, 39) reported the existence of additional DPs not common to all dietary sources (Supplemental Table 3). The described DPs were generally similar across different dietary sources (in terms of factor loadings and percentages of explained variance) and their names reflected these similarities; some variation in DP composition was reported across available dietary sources or different time-points in 1 article (35), whereas, in another article (40), additional DPs were identified for FFQ data only (Supplemental Table 3). The described DPs generally included a Healthy/Health-aware/Fruits and vegetables/Prudent/Mediterranean profile and a Less Healthy/Western/Processed Food(s) pattern, but we also identified variants of a Traditional (11, 31, 34, 35, 38, 47, 54), Sweet-based (34, 40, 47), Sandwich-based (30, 35), or Alcohol-

449 based DPs (33, 34, 40) (Supplemental Table 3). Reproducibility of DPs was assessed with
450 1 (9, 11, 31, 33) or more than 1 statistical approaches (30, 32, 34); similarly, relative validity
451 was assessed with 1 (9, 31, 33, 35, 47) or more (11, 13, 30, 32, 34, 37-40) approaches, and
452 reliability was assessed with more than 1 statistical method (54). The intra-class correlation
453 coefficient (11, 32, 34, 54), the (Pearson, Spearman, or Kendall) correlation coefficient (9,
454 11, 13, 30-35, 37-40, 47), the Bland-Altman method (11, 13, 30, 32, 34, 37-40), the
455 proportions of subjects classified into the same, adjacent, opposite quantiles, and the
456 weighted kappa coefficient (30, 32, 34, 37) were used alone or in combination for the
457 assessment of reproducibility and/or relative validity. Partial, de-attenuated or corrected
458 correlation coefficients were also introduced in some articles to account for the effect of
459 energy intake, and/or of repeated administration of the gold standard dietary assessment
460 tool (9, 11, 32, 33).
461 Among the 7 articles assessing simultaneously reproducibility and relative validity of DPs
462 (9, 11, 30-34), the main results were the following ones: 1. The different statistical
463 approaches used led to concordant results, except for 1 article (30) where only the Bland-
464 Altman approach consistently highlighted increasing differences in DP scores with
465 increasing scores; 2. Under the same statistical approach, the assessment of DP
466 reproducibility provided generally stronger results than relative validity (9, 11, 30, 31, 33,

34); 3. Well-characterized DPs based on a few identifiable FGs were more likely to be reproducible and valid than DPs including different aspects of the diet simultaneously (Supplemental Table 3); for example, the Sandwich and drinks DP (30), the Animal and Plant Protein DP (34), and the Drinker DP (33) had higher reproducibility and relative validity than others from the same papers.

Among the 7 articles assessing relative validity of DPs only (13, 35, 37-40, 47), we distinguished between those comparing FFQs and DR (13, 39, 40, 47), the one comparing the FFQ with a 24HR (37), and those studies not based on FFQ data (35, 38). In the first group (13, 39, 40, 47), the relative validity of all DPs was questionable with any approach in 1 article (40) and it was poor for the Western DP in another article (39); however, the Healthy/Prudent/Green DPs showed a higher degree of relative validity, as compared to the corresponding Western/Western/Traditional DPs in (13, 39, 47). On the contrary, when comparing FFQ-based DPs with those on m24HR (37), the Less-Healthy DP was found to be more valid than the Healthy DP in pregnant women, although results for both DPs were stronger than in previous articles. When 24HR or 48HR were compared with DR data (35), relative validity was moderate-to-good with 48HR-based DPs, but less strong with 24HR-based DPs; the Health-aware DP showed the highest validity on the 48HR-based comparison. Finally, when a Diet History Questionnaire was compared with a DR (38), the

485 Healthy DP was found to be valid, but the same was not true for the other 2 DPs, which
486 showed wider limits of agreement in women, based on DR data.
487 When the reliability of CFA-based DPs was evaluated by Ryman et al. (54), composite
488 reliability of DPs was good and similar across DPs, but test-retest reliability of DPs was
489 moderate. In addition, indicator and test-retest reliabilities of CFA-based FGs were similar
490 and poor-to-fair. The Processed foods and the Fruits and Vegetables DPs showed better
491 reliability overall.
492 Table 4 provides details on the 15 articles assessing construct validity of DPs through the
493 application of CFA (10, 24, 27-29, 47, 48, 50-55) to validate previous EFA-based DPs or as
494 an alternative one-step approach to be compared with PCA/EFA (49, 57). Some of them
495 used CFA-based DPs for assessing more general research questions on relative validity of
496 DPs (47), DP reproducibility (50, 57) or reliability (54), DP stability over time (10, 27-29) or
497 cross-study reproducibility (24); other studies simply used CFA to represent DPs of a
498 population of interest in a more ideal way (48, 49, 51-53, 55).
499 When CFA was used after a previous EFA, the cut-offs for FG inclusion in the CFA models
500 ranged from |0.20| (10, 24, 27, 28) to |0.60| (52, 53) and the CFA model was estimated on
501 a different (validation) sample in 5 articles (24, 51-53, 55).

Among the 15 CFA-based papers, 4 (49-51, 57) provided a formal model selection procedure, where different numbers of DPs, cut-offs for FGs (and rotation methods), and/or correlation structures between DPs were considered. In addition, in 10 articles, the goodness of fit of the selected CFA model was formally tested according to 1 (47, 48) or more (24, 49-55) indexes, whereas 1 article (57) used descriptive statistics from the bootstrap-based distributions of the factor loadings of each FG to each DP. None of the 4 articles that assessed stability of CFA-based DPs over time (10, 27-29) gave details on model fitting. Finally, some articles provided results on values and statistical significance of standardized factor loadings (50, 52-55) and a few compared EFA- and CFA-based DPs with correlation coefficients between factor scores of similar DPs (47, 49).

Among the 10 articles using goodness of fit indexes (24, 47-55), the final CFA model was considered a good model in 8 articles and a slightly inappropriate model in 1 article (52), whereas, in another article (50), a cut-off of |0.25| for FG inclusion provided a good model fitting, as compared to a CFA with |0.20| cut-off. In general, FG standardized loadings were high and reached statistical significance (50, 52-55) and correlation coefficients between EFA- and CFA-based DP scores were very high (47, 49). In another paper (57), CFA outperformed PCA in terms of DP interpretability on a bootstrap-based comparison. Overall, the different statistical criteria pointed to reassuring results: most CFA models confirmed

their utility in identifying the minimal constructs characterizing the overall diet in the populations under consideration.

Concerning the quality assessment of the included studies, those of “good” quality consistently identified highly reproducible and/or valid DPs; studies of “poor” quality still tended to identify DPs with a fair-to-good reproducibility and/or validity. However, for some papers (10, 27-29) it was not possible to formally evaluate DP validity, in the absence of CFA goodness of fit statistics.

Conclusions

The concept of healthy eating patterns has been adopted by the Dietary Guidelines for Americans over time and there is an emerging body of evidence on the beneficial or detrimental effects of DPs on health. Nevertheless, the key issues of reproducibility and validity of DPs have been assessed by a limited number of articles (mostly based on *a priori* DPs) and using very different approaches. This review included 38 articles on *a posteriori* DPs, with ~15 articles dealing with each research question. To our knowledge, this is the first attempt to collect the overall evidence on these issues and it is therefore valuable, yet it is still limited in its ability to draw strong conclusions.

537 The identification of DP with PCA/EFA or CA has traditionally used standard statistical
538 approaches and software. However, since 2011, 7 of our articles have assessed matrix
539 factorability before starting PCA/EFA (30, 32, 34, 37, 40, 41, 50) and 3 recent articles (43,
540 44, 56) have proposed some innovation in CA procedures, with sound conclusions. Some
541 novelties have been therefore introduced in the identification of *a posteriori* DPs over the
542 last decade. However, there are essentially no specific investigations on fundamental
543 questions that researchers should consider when using EFA or CA. For example, this
544 happened for input variable format (e.g. nutrients or FGs, and, in the latter case, number of
545 servings or percentage daily energy intake), transformation (e.g. log-transformation or not)
546 and/or potential adjustment by energy intake (on input data or on DP scores, with the
547 residual method or with other solutions), with only 4 articles (26, 36, 42, 46) included in the
548 current review. Similarly, many other relevant topics were investigated in at most 3 or 4
549 articles, so evidence is too weak to draw any conclusions on reproducibility of DPs across
550 different statistical solutions.
551 We found more convincing results from the assessment of reproducibility of DPs over short
552 time periods and of relative validity of DPs. Before reporting the key findings, some general
553 concerns have to be introduced. First of all, during this review, it has often happened that
554 the Results sections described those DPs that were similar across the available dietary

555 datasets, whereas the Discussion sections were left with a short note on the presence of
556 additional DPs which were not common to all dietary datasets (9, 13, 32-34, 37, 39). Second,
557 dietary pattern similarities were defined qualitatively, looking at factor loading matrices and
558 percentages of explained variances or at FGs that contributed higher-than-mean intakes for
559 each cluster. Third, when present, the quantification of similarities relied mostly on
560 elementary statistics, with no statistical models assumed. Forth, the optimal number of DPs
561 to retain was chosen separately for each dietary dataset. Any assessment of reproducibility
562 or relative validity of DPs is based on these critical points.
563 An opposite solution to independent sets of DPs (to be later analyzed for reproducibility and
564 validity) is to work on a merged data matrix and force the dietary data to express the same
565 set of DPs across dietary datasets. We recently introduced multi-study factor analysis (58)
566 to allow for the simultaneous identification of common and study-specific DPs across
567 different studies, within a statistical model that includes a formal assessment of the number
568 of shared and study-specific DPs. A similar idea of partial sharing of DP could be applied in
569 the assessment of DP reproducibility and relative validity, after multiple measures from each
570 subject are taken into account. Use of a statistical model would solve most of the inherent
571 limitations of correlation coefficients, cross-classification and weighted kappa coefficients.

572 In the validation studies of FFQs that we analyzed, the assessment of reproducibility of DPs
573 provided systematically better results than the corresponding assessment of relative validity,
574 independently of the statistical approach used. This suggests that multiple administrations
575 of the same dietary tool improve consistency of the corresponding DPs, as compared to
576 having 2 different dietary sources. In the latter case, reference periods, number of collected
577 food items and the administration process are deeply different. An effort is generally made
578 to create a common set of FGs that fits both the instruments, however other differences
579 cannot be eliminated and are reflected in the weaker agreement between corresponding
580 DPs.
581 It is reassuring that results on DP relative validity were similar no matter if reproducibility
582 was assessed in the same study design or not (13, 35, 37-40, 47). However, in papers
583 assessing DP relative validity only, the presence of different study designs, dietary
584 assessment tools (24HR or DR), reference period of collection and timing of administration
585 made the comparison of results even more difficult.
586 Reproducibility of DPs across multiple administrations of the same FFQ was good and the
587 differences between corresponding factor scores were not systematically biased. However,
588 we detected some variability in factor scores that was reflected in wider-than-expected limits

589 of agreement. A 1-year (median) time-interval between FFQ administrations across studies
590 could be at the origin of this extra variability.
591 Confirmatory factor analysis should have a wider use in nutritional epidemiology, either for
592 describing dietary habits of a population in a more ideal way or for assessing more general
593 questions on reproducibility of DPs over time (10, 27-29), across populations (24) or dietary
594 sources (47). The current review showed that, when used to identify synthetic dietary profiles
595 from a previous EFA or as a one-step approach, CFA provided models with good fit and
596 interpretable DPs. Publication bias is likely to be present in this case, especially with those
597 articles that simply confirm a previous EFA. Some caution is therefore needed before
598 concluding on the effective power of CFA. On the other hand, we lacked information on
599 model goodness-of-fit for most of the articles assessing more general research questions
600 through CFA-based DPs (10, 27-29). Researchers should have in mind that using CFA to
601 assess reproducibility of DPs in time or across studies requires giving details on CFA
602 performance too.
603 We have speculated on the possibility that some DPs would have been more likely to be
604 reproducible and valid than others across the articles included in the review. Unfortunately,
605 CFA does not allow to evaluate the validity of single DPs. The goodness of fit measures
606 represent global model fitting, whereas the significance tests on standardized CFA loadings

607 are not informative, when based on highly selected FGs and a reasonable sample size. In
608 regard with reproducibility and relative validity, there is some evidence that DPs built on a
609 few characteristic FGs were more likely to be reproducible and valid; for example, Sandwich-
610 based (30) or Alcohol-based (33) DPs gave better results on reproducibility and relative
611 validity than other DPs presented in the same articles. Similarly, well-characterized
612 traditional DPs (e.g. (31)) could be more likely to be reproducible and valid, although this
613 was not always true (e.g. (34)). Western-like or Prudent-like DPs were generally based on
614 a higher number of dominating FGs and those FGs represent different aspects of Western
615 (e.g. processed food, red meat, sausages, butter, French fries, eggs, high-fat dairy products)
616 or Prudent (e.g. fruits, vegetables, fish, poultry, low-fat dairy products, nuts and seeds) diets.
617 These aspects may explain why these DPs reached only fair-to-moderate levels of
618 agreement. A similar argument was already presented in a previous review on empirically
619 derived DP (6).
620 It is crucial to evaluate the quality of the original studies included in a systematic review
621 using standardized and validated quality assessment tools, like the one (23) we referred to
622 in the current analysis. However, our topic did not fit well within the typical research question
623 of a possible association between exposure and disease. In addition, any evaluation of
624 reproducibility and/or validity of DPs depends strongly on how well DPs were originally

identified in the sample under consideration. Finally, the way the assessment of reproducibility and validity of *a posteriori* DPs is carried out (e.g., how many criteria were considered and which criteria were used) should deserve additional attention. A standard quality assessment tool is not able to capture all these aspects, which are fundamental in a systematic review on reproducibility and validity of *a posteriori* DPs. Nevertheless, we showed that better-designed studies were more likely to provide highly reproducible and/or valid DPs. This conclusion reflects the general idea that good results are more likely to come from well-designed and carefully implemented studies, based on a sound statistical analysis. In conclusion, although some caution is worthy, this preliminary attempt to collect evidence on reproducibility, relative and construct validity of *a posteriori* DPs provides several reasonable conclusions on a topic that has not been fully considered so far. In addition, we provide those new to factor or cluster analyses with a small guide that summarizes evidence on several subjective decisions involved in the DP identification process.

Acknowledgements

VE and MF designed research; VE and MD collected the relevant papers and selected those to be included in the systematic review; all the authors performed the quality assessment of the original studies included in the systematic review; MD and LP prepared the first draft of Table 1 and of part of Tables 2, 3 and 4; VE, RDV, and MF completed and refined Tables 2, 3, and 4; AS revised all the tables and checked their consistency with the text; AS prepared Figure 1; RDV prepared Figure 2; VE wrote the paper and had primary 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. Liese AD, Krebs-Smith SM, Subar AF, George SM, Harmon BE, Neuhouser ML, Boushey CJ, Schap TE, Reedy J. The Dietary Patterns Methods Project: synthesis of findings across cohorts and relevance to dietary guidance. J Nutr 2015;145(3):393-402. doi: 10.3945/jn.114.205336.

3. Health UDo, Services H. Dietary guidelines for Americans 2015-2020: Skyhorse Publishing Inc., 2017.

4. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM. Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. J Am Diet Assoc 2007;107(7):1233-9. doi: 10.1016/j.jada.2007.03.014.

5. Tucker KL. Dietary patterns, approaches, and multicultural perspective. Appl Physiol Nutr Metab 2010;35(2):211-8. doi: 10.1139/H10-010.

6. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev 2004;62(5):177-203.

7. Edefonti V, Randi G, La Vecchia C, Ferraroni M, Decarli A. Dietary patterns and breast cancer: a review with focus on methodological issues. Nutr Rev 2009;67(6):297-314. doi: 10.1111/j.1753-4887.2009.00203.x.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. Bamia C, Orfanos P, Ferrari P, Overvad K, Hundborg HH, Tjonneland A, Olsen A, Kesse E, Boutron-Ruault MC, Clavel-Chapelon F, et al. Dietary patterns among older Europeans: the EPIC-Elderly study. Br J Nutr 2005;94(1):100-13.

16. Peng RD. Reproducible research and Biostatistics. Biostatistics 2009;10(3):405-8. doi: 10.1093/biostatistics/kxp014.

17. Broman K, Cetinkaya-Rundel M, Nussbaum A, Paciorek C, Peng R, Turek Dl, Wickham H. Recommendations to Funding Agencies for Supporting Reproducible Research, American Statistical Association, Available online at: http://www.amstat.org/asa/files/pdfs/pol-reproducibleresearchrecommendations.pdf (5 March 2019, date last accessed).

18. Wirfalt AK, Jeffery RW. Using cluster analysis to examine dietary patterns: nutrient

intakes, gender, and weight status differ across food pattern clusters. J Am Diet Assoc 1997;97(3):272-9.

19. Edefonti V, Hashibe M, Ambrogi F, Parpinel M, Bravi F, Talamini R, Levi F, Yu G, Morgenstern H, Kelsey K, et al. Nutrient-based dietary patterns and the risk of head and neck cancer: a pooled analysis in the International Head and Neck Cancer Epidemiology consortium. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23(7):1869-80. doi: 10.1093/annonc/mdr548.

20. Tseng M, Breslow RA, DeVellis RF, Ziegler RG. Dietary patterns and prostate cancer risk in the National Health and Nutrition Examination Survey Epidemiological Follow-up Study cohort. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2004;13(1):71-7.

21. Edefonti V, Bravi F, Garavello W, La Vecchia C, Parpinel M, Franceschi S, Dal Maso L, Bosetti C, Boffetta P, Ferraroni M, et al. Nutrient-based dietary patterns and laryngeal cancer: evidence from an exploratory factor analysis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2010;19(1):18-27. doi: 10.1158/1055-9965.EPI-09-0900.

22. Sun J, Buys NJ, Hills AP. Dietary pattern and its association with the prevalence of obesity, hypertension and other cardiovascular risk factors among Chinese older adults. Int J Environ Res Public Health 2014;11(4):3956-71. doi: 10.3390/ijerph110403956.

23. National Heart L, Institute B. Quality assessment tool for observational cohort and cross-sectional studies. Bethesda: National Institutes of Health, Department of Health and Human Services 2014.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. Nanri A, Shimazu T, Ishihara J, Takachi R, Mizoue T, Inoue M, Tsugane S, Group JFVS. Reproducibility and validity of dietary patterns assessed by a food frequency questionnaire used in the 5-year follow-up survey of the Japan Public Health Center-Based Prospective Study. J Epidemiol 2012;22(3):205-15.

32. Liu X, Wang X, Lin S, Song Q, Lao X, Yu IT. Reproducibility and validity of a Food Frequency Questionnaire for assessing dietary consumption via the dietary pattern method in a chinese rural population. PLoS One 2015;10(7):e0134627. doi: 10.1371/journal.pone.0134627.

33. Khani BR, Ye W, Terry P, Wolk A. Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. J Nutr 2004;134(6):1541-5.

34. Hong X, Ye Q, Wang Z, Yang H, Chen X, Zhou H, Wang C, Chu W, Lai Y, Sun L, et al. Reproducibility and validity of dietary patterns identified using factor analysis among Chinese populations. Br J Nutr 2016;116(5):842-52. doi: 10.1017/S000711451600249X.

35. McNaughton SA, Mishra GD, Bramwell G, Paul AA, Wadsworth ME. Comparability of dietary patterns assessed by multiple dietary assessment methods: results from the 1946 British Birth Cohort. Eur J Clin Nutr 2005;59(3):341-52. doi: 10.1038/sj.ejcn.1602079.

36. 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.

37. Loy SL, Jan Mohamed HJ. Relative validity of dietary patterns during pregnancy assessed with a food frequency questionnaire. Int J Food Sci Nutr 2013;64(6):668-73. doi: 10.3109/09637486.2013.787398.

38. Okubo H, Murakami K, Sasaki S, Kim MK, Hirota N, Notsu A, Fukui M, Date C. Relative validity of dietary patterns derived from a self-administered diet history questionnaire using factor analysis among Japanese adults. Public Health Nutr

2010;13(7):1080-9. doi: 10.1017/S1368980009993211.

39. Crozier SR, Inskip HM, Godfrey KM, Robinson SM. Dietary patterns in pregnant women: a comparison of food-frequency questionnaires and 4 d prospective diaries. Br J Nutr 2008;99(4):869-75. doi: 10.1017/S0007114507831746.

40. 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.

41. Bountziouka V, Panagiotakos DB. The role of rotation type used to extract dietary patterns through principal component analysis, on their short-term repeatability. J Data Sci 2012;10:19-36.

42. Bailey RL, Gutschall MD, Mitchell DC, Miller CK, Lawrence FR, Smiciklas-Wright H. Comparative strategies for using cluster analysis to assess dietary patterns. J Am Diet Assoc 2006;106(8):1194-200. doi: 10.1016/j.jada.2006.05.012.

43. Lo Siou G, Yasui Y, Csizmadi I, McGregor SE, Robson PJ. Exploring statistical approaches to diminish subjectivity of cluster analysis to derive dietary patterns: The Tomorrow Project. Am J Epidemiol 2011;173(8):956-67. doi: 10.1093/aje/kwq458.

44. Sauvageot N, Schritz A, Leite S, Alkerwi A, Stranges S, Zannad F, Streel S, Hoge A,

Donneau AF, Albert A, et al. Stability-based validation of dietary patterns obtained by cluster analysis. Nutr J 2017;16(1):4. doi: 10.1186/s12937-017-0226-9.

45. McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr 2001;4(5):989-97.

46. Wirfalt E, Mattisson I, Gullberg B, Berglund G. Food patterns defined by cluster analysis and their utility as dietary exposure variables: a report from the Malmo Diet and Cancer Study. Public Health Nutr 2000;3(2):159-73.

47. Togo P, Heitmann BL, Sorensen TI, Osler M. Consistency of food intake factors by different dietary assessment methods and population groups. Br J Nutr 2003;90(3):667-78.

48. Lau C, Glumer C, Toft U, Tetens I, Carstensen B, Jorgensen T, Borch-Johnsen K. Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study. Br J Nutr 2008;99(5):1089-98. doi: 10.1017/S0007114507837494.

49. Bedard A, Garcia-Aymerich J, Sanchez M, Le Moual N, Clavel-Chapelon F, Boutron-Ruault MC, Maccario J, Varraso R. Confirmatory Factor Analysis Compared with Principal Component Analysis to Derive Dietary Patterns: A Longitudinal Study in Adult Women. J Nutr 2015;145(7):1559-68. doi: 10.3945/jn.114.204479.

50. 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.

51. Fransen HP, May AM, Stricker MD, Boer JM, Hennig C, Rosseel Y, Ocke MC, Peeters PH, Beulens JW. A posteriori dietary patterns: how many patterns to retain? J Nutr 2014;144(8):1274-82. doi: 10.3945/jn.113.188680.

52. Maskarinec G, Novotny R, Tasaki K. Dietary patterns are associated with body mass index in multiethnic women. J Nutr 2000;130(12):3068-72.

53. Park SY, Murphy SP, Wilkens LR, Yamamoto JF, Sharma S, Hankin JH, Henderson BE, Kolonel LN. Dietary patterns using the Food Guide Pyramid groups are associated with sociodemographic and lifestyle factors: the multiethnic cohort study. J Nutr 2005;135(4):843-9. doi: 10.1093/jn/135.4.843.

54. 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.

55. Schulze MB, Hoffmann K, Kroke A, Boeing H. Risk of hypertension among women in the EPIC-Potsdam Study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. Am J Epidemiol 2003;158(4):365-73.

56. Greve B, Pigeot I, Huybrechts I, Pala V, Bornhorst C. A comparison of heuristic and model-based clustering methods for dietary pattern analysis. Public Health Nutr 2016;19(2):255-64. doi: 10.1017/S1368980014003243.

57. 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.

58. 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.

Table 1. Basic characteristics of observational studies on reproducibility, relative and construct validity of *a posteriori* dietary patterns[a]

Reference	Study Design	Subjects [number and age (ys)] and follow-up	Questionnaire
Ambrosini, 2011 Australia Western Australian Pregnancy Cohort (Raine) Study Fair quality (13)	14-year follow-up of the Raine cohort study, including adolescents from 2900 pregnant Fs originally recruited at 16 - 20 weeks of gestation between 1989 and 1991	1613 adolescents who completed the FFQ, 822 adolescents who completed the DR, 783 adolescents who completed both FFQ and DR 14 (mean: 14, SD: 0.2) Follow-up: Not applicable	FFQ: 1 y; SA; validity assessed but no comments on the results; 212 FI; FFQ completed by primary caregiver and adolescent; 3-day DR completed by adolescents, and verified by a dietician; interest on representative DR; 38 FG common to all dietary sources
Asghari, 2012 Iran TLGS Fair quality (11)	TLGS: cohort study on urban residents in Tehran in 1999 - 2001; Validation study of the TLGS FFQ based on a random sample of participants who were proportionately distributed across 5 10-year age intervals and 2 sexes plus extra wave of the cohort study with FFQ administration	132 (89 completed FFQ3) 20 - 70 (mean: 35.6, SD: 16.8) Follow-up: 8 ys, until 2011	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
Bailey, 2006 USA (Pennsylvania) Geisinger Rural Aging Study Fair quality (42)	Geisinger Rural Aging Study: longitudinal cohort study of rural older adults in Pennsylvania enrolled within a Medicare-managed health maintenance organization; random sample of participants to an intensive cross-sectional research study, not depressed or with functional limitations	179 66 - 87 (mean: 73, SD: 5) Follow-up: No follow-up	5 24HRs collected on random and nonconsecutive days over 10 months using a multi-pass technique; m24HRs used for the analysis; 24 FG for all time-points
Balder, 2003 Netherlands, Sweden, Finland, and Italy DIETSCAN (NLCS, SMC, ATBC, ORDET) Good quality (26)	Parallel analysis of 4 studies (no pooled analysis); NLCS (random subcohort of): population-based cohort of Ms and Fs from Dutch municipalies; 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; ORDET: cohort study of Italian healthy volunteer Fs from the province of Varese, northern Italy	NLCS: 3123 (1598 Fs and 1525 Ms); SMC: 61,469 Fs; ATBC: 27,111 Ms; ORDET: 9208 Fs NLCS: 55 - 69 at baseline in 1986 (mean: 61.4, SD: 4.2 for Ms and 4.3 for Fs); SMC: 40 - 74 when invited to mammography screening in 1987 to 1990 (mean: 53.7, SD: 9.7); ATBC: 50 - 69 at baseline between 1985 and 1988 (mean: 57.7, SD: 5.1); ORDET: 35 - 69 between 1987 and 1992 (mean: 48, SD: 8.5) Follow-up: 7 for NLCS (baseline: 1986); 13 for SMC	4 different but validated FFQs: NLCS-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)

		(baseline: 1987-1990); NA for ATBC (baseline: 1985 - 1988, intervention ended in 1993 after 5-8 ys, follow-up later on); 9 for ORDET (baseline: 1987-1992)	
Beck, 2012 New Zealand NA Poor quality (30)	Validation study of a new FFQ; convenient sample of Fs living in Auckland in 2009 free of chronic disease, recruited with a magazine advertisement or invitation to potential volunteers	115 Fs 18 - 44 (median: 33) Follow-up: 1 month	FFQ: 1 month; SA; reproducibility and validity to be assessed in this study; FFQ1: completed at baseline; FFQ2: completed 1 month later; 4-day weighted DR: completed between FFQ1 and FFQ2; 144 FI for FFQ and DR (30 FG - most frequently consumed on FFQ1)
Bedard, 2015 France E3N (EPIC-France) Fair quality (49)	1993 wave of the prospective cohort Study E3N, after exclusion of current or former smokers, and of Fs with prevalent asthma at baseline	30,589 Fs 40 - 65 at baseline (mean: 53) Follow-up: 1993 - 2005	FFQ: NA reference period; SA; reproducible and valid; 208 FI (27 FG)
Bountziouka, 2011 Greece NA Poor quality (40)	Validation study based on a convenience sample, representative of the general population of Athens residents (stratified sample by age group and gender according to 2001 Census)	500 mean: 46, SD: 16 Follow-up: No follow-up	FFQ: 1 month; IA; reproducible and valid; 76 FI; 3-day DR: based on 2 weekdays and 1 weekend day, over the same time span of the FFQ; DR FI matched with FFQ FI;

			24 FG common to all dietary sources
Bountziouka, 2012 Greece NA Fair quality (41)	Nutrition survey	500 mean: 37, SD: 15 Follow-up: No follow-up	FFQ: 1 month; IA; reproducible and valid; 76 FI (24 FG); FFQ completed twice, within a 15 day interval
Castro, 2015 Brazil Healthy Survey of the City of Sao Paulo Poor quality (50)	Cross-sectional population-based survey (using a complex multistage sampling design to have a representative sample of Sao Paulo residents)	1102 (424 Ms; 678 Fs) >= 20, 46% with 60 ys or more Follow-up: No follow-up	2 non-consecutive 24HRs, former collected face to face (USDA 5 Step Multiple Pass Method) and latter with telephone interview; 1169 FI (38 FG, but final analysis on 34 FG)
Crozier, 2008 UK NA Fair quality (39)	Cross-sectional study including Fs in early pregnancy (median gestation: 15.3 weeks) booked for delivery under 2 consultants in Southampton	585 Fs in early pregnancy with complete information on FFQ and DR 16 or more (mean: 26.4, SD: 4.9) Follow-up: Not applicable	FFQ: 3 months (first trimester of pregnancy); IA; NA reproducible and valid; 100 FI (49 FG); 4-day DR: filled in immediately after completion of the FFQ, at the end of the first trimester of pregnancy; DR FI mapped into the 100 FFQ FI and then grouped in the 49 FG used for the FFQ data

Dekker, 2013 Netherlands Doetinchem Cohort Study Good quality (25)	3 successive surveys (surveys 2, 3, and 4, at 3, 11, and 16 ys 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; follow-up available for 2/3 of the original random sample by design	4007 subjects with information available for the 3 rounds. In detail: 1993 - 1997: 6113 (survey 2); 1998 - 2002: 4916 (survey 3); 2003 - 2007: 4520 (survey 4) 47 - 66 Follow-up: 6, 11, 16 ys after the first survey, so 10-y follow-up from survey 2 to survey 4	FFQ: 1 y; NA SA; reproducible and valid; 178 FI (32 FG)
Fransen, 2014 Netherlands EPIC-NL Fair quality (51)	Cohort study consisting of Prospect-EPIC and the Monitoring Project on Risk Factors for Chronic Diseases (MORGEN)-EPIC cohorts	39,678 (Prospect-EPIC Fs, MORGEN-EPIC Ms and Fs), of which 19,837 in the derivation sample and 19,841 in the replication sample Prospect-EPIC: 50 - 69; MORGEN-EPIC: 20 - 64 Follow-up: 1993 - 2007	FFQ: 1 y; SA; reproducible and valid; 178 FI (31 FG)
Greve, 2016 Germany IDEFICS Fair quality (56)	Baseline survey of the GerM subsample of the IDEFICS study (a European longitudinal multicentre study in children and infants from 8 European countries)	1791 children 2-9 Follow-up: No follow-up	FFQ: NA reference period; SA (caregiver); NA reproducibility and validity; 45 FI (15 FG)
Hong, 2016 China NA Good quality	Validation study of FFQ; subsample of 250 participants from the community-based, cross-sectional, nutrition and health	203 31 - 80 (mean: 50.4, SD: 12) Follow-up: 1 y	FFQ: 1 y; IA; reproducibility and validity to be assessed in this study; 87 FI; FFQ

(34)	survey in Nanjing, presenting a multi-stage random sampling design based on 6 communities of residents		completed twice (FFQ1 and FFQ2), at the beginning (June 2014) and end (May 2015) of the study; 4 3-consecutive day (including 2 weekdays and 1 weekend day in a usual week) 24HRs collected at intervals of 3 months during the 1-year period by trained interviewers; 28 FG common to all dietary sources
Hu, 1999 USA (Massachusetts) HPFS Good quality (9)	HPFS: prospective cohort study of US M health professionals started in 1986; Validation study of the FFQ used in the 1986 wave of the HPFS cohort study; random sample of cohort members (men) from the Boston area	127 Ms 40 -75 ys at baseline in 1986 Follow-up: 6-7 months for validity analysis, 1 y for reproducibility analysis	FFQ: 1 y; SA; reproducibility and validity to be assessed in this study; 131 FI; FFQ1: completed during the following ys; FFQ2: completed 1 y after FFQ1; 2 7-day DRs 6-7 months apart; DR1: completed ~3 months after FFQ1; DR2: completed 2-3 months before FFQ2; 1217 DR food codes used for creating FG; 40 FG common to all dietary sources

Judd, 2014 USA REGARDS Fair quality (24)	Population-based random sample of black and white individuals designed to oversample black participants and people residing in the stroke belt (8 US states)	21,636 > 45 Follow-up: 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)
Khani, 2004 Sweden SMC Fair quality (33)	SMC: population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990 with 57,881 Fs who have completed the baseline SMC FFQ; Validation study of the SMC FFQ; 2 random samples, one for FFQ reproducibility assessment and the other for FFQ validity assessement, reference FFQ completed at baseline for both samples	197 Fs included in the FFQ reproduciblity sample; 111 Fs included in the FFQ validity sample 40 - 74 at baseline Follow-up: 1 y	FFQ: 6 months; SA; reproducibility and validity to be assessed in this study; 60 FI; FFQ1: completed at baseline within the reproducibility sample; FFQ2: completed 1 y after FFQ1 within the reproducibility sample; FFQ: completed at baseline within the validity sample; 4 7-day open ended weighted DR 3 months apart to cover a ys; 543 DR food codes matched to the FFQ items; 26 FG common to all dietary sources
Lau, 2008 Denmark Inter99 Study Fair quality (48)	Age- and sex- stratified random sample of participants to a health survey derived from baseline data of the population-based intervention study Inter99 (1999 - 2001), that included residents from the south-western part of the Copenhagen County	6563 (3372 Fs; 3191 Ms) 30 - 60 (mean: 46.3, SD: 7.9) Follow-up: No follow-up	FFQ: 1 month; SA; NA reproducibility, valid; 198 FI (34 FG)

Liu, 2015 China NA Poor quality (32)	Validation study of a new FFQ developed from a NCI FFQ to capture DPs of rural chinese population; random sample of subjects from an underdeveloped rural area of southwest China, free of chronic malignant diseases	179 40 - 70 at baseline in 2012 (mean: 55, SD: 8.2) Follow-up: 1 y	FFQ: 1 y; IA; reproducibility and validity to be assessed in this study; 131 FI; FFQ1: completed at baseline; FFQ2: completed 1 y after FFQ1; 6 3-day 24HRs completed in between the 2 FFQs (18 24HRs in 1 y, 3 24HRs every 2 months, on consecutive days, given by 2 weekdays and 1 weekend day); 18 FG common to all dietary sources
Lo Siou, 2011 Canada Tomorrow Project Fair quality (43)	Tomorrow Project: longitudinal cohort study with 2-stage random sampling design including Albertans Ms and Fs with no personal history of cancer recruited between 2001 and 2007; subset of participants with complete data by November 2007	16,674 (6445 Ms; 10,229 Fs) 35 - 69 (mean: 50.5, SD: 9.1 for Ms and 9.2 for Fs) Follow-up: No follow-up	FFQ: 1 y; SA; NA reproducibility and validity; 284 FI (55 FG)
Loy, 2013 Malaysia USM Birth Cohort Study Good quality (37)	Validation study of the FFQ from USM Birth Cohort study, based on a convenience sample of pregnant healthy Fs from the north-east of Peninsular Malaysia	162 pregnant Fs 19 - 40 (mean: 28.67) Follow-up: mid pregnancy - late pregnancy	FFQ; 6 months of pregnancy; IA; validity to be assessed in this study; 82 FI; FFQ conducted immediately after completing the 24HRs in late pregnancy; 6 24HRs, 3 24HRs in mid (mean gestation: 15.6 weeks) and late (mean gestation: 34.3 weeks) pregnancy (2 weekdays

			and 1 weekend dietary intake); 23 FG common to all dietary sources
Maskarinec, 2000 USA (Hawaii) NA Fair quality (52)	Cross-sectional study based on an ethnically diverse population, with recruitment at different mammography facilities on Oahu	514 Fs 35 - 85 (mean: 53.9, SD: 10.1) Follow-up: Not applicable	FFQ: NA reference period; SA; valid; ~209 FI (39 FG, but final analysis on 23 FG due to skewness in FG distributions)
McCann, 2001 USA (New York) Western New York Diet Study Fair quality (45)	Case-control study on endometrial cancer with population-based controls frequency-matched to cases on age and county of residence, conducted between October 1986 and March 1991 in the Buffalo area	1095 (232 cases; 863 controls) 40 - 85 for cases Follow-up: Not applicable	FFQ: 2 ys; IA; NA reproducibility and validity; 190 FI (different numbers of FG in the analysis corresponding to 3 different food grouping schemes: 168 FG, as to useable information from FFQ, 56 FG, as to nutrient content and use, and 36 FG, as to USDA suggestions)
McNaughton, 2005 UK Medical Research Council National Survey of Health and Development (1946 British Birth Cohort) Good quality (35)	1946 British Birth Cohort: longitudinal 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 current paper; data from 1989 interview	2265 subjects who completed the 48HR recall and the DR in 1989 43 in 1989 Follow-up: No follow-up	1 48HR at interview; 1 5-day DR completed in the 5 days following the 48HR collection; 1 24HR recall relative to the 24-hour period preceding the interview; 56 FG common to all dietary sources

Nanri, 2012 Japan JPHC Poor quality (31)	Validation study of JPHC study FFQ; sub-sample of married couples from 5-year follow-up survey of the JPHC study (cohort 1: baseline: 1990, and cohort 2: baseline: 1993) who provided complete information on 2 FFQs and DRs	498 (244 Ms and 254 Fs, 290 in cohort 1 and 289 in cohort 2) Cohort 1: 40 - 59 at baseline; cohort 2: 40 - 69 at baseline Follow-up: 1 y	FFQ: 1 y; SA; reproducibility and validity to be assessed in this study; 147 FI, but 134 FI used for the final analysis; FFQ_R: completed 1 y after or before FFQ_V; FFQ_V: completed after DRs, and compared with DR; 28 - 14 DRs: completed in 1 y (i.e. 7-day DRs collected 4 (or 2) times at 3 month (or 6 month) intervals during the ys); 558 DR FI matched to 134 FFQ FI; 48 FG common to all dietary sources
Newby, 2006 Sweden SMC Good quality (10)	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) Follow-up: from 1987 - 1990 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 Sweden SMC Good quality (27)	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) Follow-up: from 1987 - 1990 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)

Northstone, 2008 UK ALSPAC Fair quality (36)	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 (1 wave)	12,053 pregnant Fs Age: NA ys Follow-up: NA	FFQ: NA reference period; SA; NA reproducibility and validity; NA FI (44 FG)
Okubo, 2010 Japan NA Good quality (38)	Cross-sectional study including apparently healthy volunteer Fs and their husbands from 3 areas of Japan [rural and urban Osaka (urban), Nagano (rural inland) and Tottori (rural coastal)]; Fs of 30 - 69 ys, such that 8 Fs were equally distributed in each 10 ys age stratum, but no age requirement for Ms	184 (92 Fs; 92 Ms) 31 - 69 for Fs (mean: 49.6, SD: 11.4); 32 - 76 for Ms (mean: 52.8, SD: 12.1) Follow-up: Not applicable	DHQ; 1 month; SA, valid; 150 FI (145 effective FI); DHQ administered 4 times (1 for each season over 1 y), 2 days before the start of the DRs; 4 4-day weighed DRs (1 in each season over 1 y); 3 weekdays and 1 weekend day; 1299 FI (1259 FI used); 30 FG common to all dietary sources
Park, 2005 USA (Hawaii and Los Angeles) Hawaii - Los Angeles Multiethnic Cohort Study Poor quality (53)	Baseline wave of the Multiethnic Cohort Study including the 5 principal ethnic groups (African Americans, Hawaiians, Japanese Americans, Latinos, and Whites) who lived in Hawaii and Los Angeles	195,298 45 - 75 Follow-up: No follow-up	FFQ: NA reference period; SA; valid; NA FI (30 FG, but final analysis on 20 FG due to null values and non-normality in FG distributions)

Ryman, 2015 USA (Southwest Alaska) CANHR Fair quality (54)	Cohort study based on a convenience sampling of Alaska native (Yup'ik or Cup'ik) adults participating in CANHR Study and completing at least 1 FFQ between September 2009 and May 2013	358 for EFA (1st FFQ, September 2009 - August 2011), 272 for CFA (1st FFQ, September 2011 - May 2013), 113 for test-retest (2nd FFQ, September 2009 - May 2013) >18 (median: 37, IQR: 23 - 54, in September 2009) Follow-up: September 2009 - May 2013	CANHR FFQ: 1 y; IA; 163 FI (22 FG, but final CFA on 18 FG); not tested for reproducibility and validity; FFQ1 in September 2009 - August 2011 for EFA (358 subjects); FFQ1 in September 2011 - May 2013 for CFA (272 subjects); FFQ2 in September 2009 - May 2013 for test-retest (113 subjects)
Sauvageot, 2017 Luxembourg, Belgium, and France NESCaV Good quality (44)	NESCaV: cross-border cardiovascular health population-based cross-sectional study, based on a stratified random sample of 3133 subjects recruited from 3 neighboring regions (Grand-Duchy of Luxembourg, Wallonia in Belgium, and Lorraine in France) from the Greater Region	2298 18 – 69 Follow-up: Not applicable	FFQ: 2 ys; NA SA; NA reproducibility and validity; 134 FI (45 FG)
Schulze, 2003 Germany EPIC-Potsdam Good quality (55)	Cohort study participating into the EPIC project and including 27,548 Ms and Fs; Fs without a previous diagnosis of hypertension or intake of antihypertensive medication within a 4-week period prior to the baseline examination were included at baseline, between August 1994 and September 1998	10,489 Fs, divided into learning (1937 Fs with normal blood pressure) and study (8552 Fs followed for 2-4 ys for incident hypertension, and including 123 incident verified cases) samples 35 - 64 at baseline	FFQ: 1 y; SA; reproducible and valid; 148 FI (44 FG)

		Follow-up: 2 - 4 ys (until May, 15, 2002)	
Togo, 2004 Denmark MONICA Poor quality (29)	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 and further rexamined in 1987-88 (M-87) and 1993-1994 (M-93)	2436 subjects participating in all 3 surveys, including 1806 subjects in M-82 30, or 40, or 50, or 60 at baseline in 1982 - 1984 Follow-up: at 5 ys (1987 - 1988) and 11 ys (1993 - 1994)	FFQ: 1 y; NA SA; NA reproducibility and validity; 26 FI (21 FG)
Togo, 2003 Denmark MONICA Poor quality (47)	Danish part of MONICA 1 (1982 - 1984) survey, including 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	3785 (879 Ms and 927 Fs) 30, or 40, or 50, or 60 at baseline in 1982-1984 Follow-up: No follow-up	FFQ: 1 y; NA SA; NA reproducibility and validity; 26 FI; 7-day weighted DR completed in a normal week within 3 weeks following the baseline investigation; 111 FI; 21 FG common to all dietary sources
Varraso, 2012 France and Spain EGEA2-France, Spanish PAC-COPD Poor quality (57)	EGEA2-France: cross-sectional study, 2003-2007 (12-year follow-up of EGEA study which is a case-control and family asthma study); Spanish PAC-COPD, 2004-2007: cross-sectional study of patients hopitalized for the first time for a COPD exacerbation between 2004 and 2006	EGEA2-France: 1236; Spanish PAC-COPD: 274 EGEA2-France: mean: 43, SD: 16; Spanish PAC-COPD: mean: 68, SD: 8 Follow-up: Not applicable	EGEA2-France: FFQ: 1 y; SA; based on a validated FFQ; 118 FI (46 FG); Spanish PAC-COPD: FFQ: 2 ys; IA; NA reproducible and valid; 122 FI (43 FG all shared with EGEA2-France FG)

Weismayer, 2006 Sweden SMC Poor quality (28)	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) 49 - 70 Follow-up: 4, 5, 6, 7 ys after baseline 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
Wirfalt, 2000 Sweden MDC Fair quality (46)	MDC: population-based prospective cohort study in Malmo, with baseline examinations conducted from March 1991 to October 1996; subset of participants with complete dietary data belonging to a substudy of the MDC Study	5357 50 - 73 for Ms and 45 - 73 for Fs Follow-up: No follow-up	Modified DHQ combining a 7-day menu book with a 168 item FFQ: NA reference period; IA; reproducibility and validity assessed; 48 original FG, but 43 FG used in the final analysis due to negligible energy contribution and non-consumption

[a]ABBREVIATIONS: 24HRs/48HRs: 24/48 hours recall; ALSPAC: Avon Longitudinal Study of Parents and Children; ATBC: Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; CANHR: Center for Alaska Native Health Research study; CFA: confirmatory factor analysis; DIETSCAN: DIETary patternS and CANcer in four European countries project; DHQ: diet history questionnaire; DR: dietary record; E3N: Mutuelle Generale de l'Education Nationale (EPIC - France); EFA: exploratory factor analysis; EGEA2-France: Epidemiological Study on the Genetics and Environment of Asthma 2–France; EPIC-NL: European Prospective Investigation into Cancer and Nutrition-The Netherlands; EPIC-Potsdam: European Prospective Investigation into Cancer and Nutrition-Potsdam; F: female; FFQ: food-frequency questionnaire; FFQ_R: food-frequency questionnaire from the reproducibility study; FFQ_V: food-frequency questionnaire from the relative validity study; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1, 2, or 3; FG: food groups; FI: food items; HPFS: Health Professionals Follow-up Study; JPHC: Japan Public Health Center-based Prospective study; IA: interviewer-administered; IDEFICS: Identification and Prevention of Dietary and Lifestyle-induced Health Effects in Children and Infants; IQR: interquartile range; M: male; m24: mean 24 hour recall; MDC: Malmo Diet and Cancer study; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not

available; NCI: National Cancer Institute; NESCaV: Nutrition, Environment and Cardiovascular Health; NLCS: Netherlands Cohort Study on diet and cancer; ORDET: Ormoni e Dieta nella Eziologia dei Tumori in Italy; PAC-COPD: Phenotype and Course of Chronic Obstructive Pulmonary Disease study–Spain; REGARDS: Reasons for Geographic and Racial Differences in Stroke; SA: self-administered; SD: standard deviation; SMC: Swedish Mammography Cohort; TLGS: Teheran Lipid and Glucose Study; USDA: US Department of Agriculture; USM: Universiti Sains Malaysia; y: year

Table 2. Reproducibility of *a posteriori* dietary patterns across statistical solutions[a]

Reference	DP identification methods	Percent Explained Variance (# factors) or CFA/CA model	Assessment of reproducibility/validity	Main Results
Bailey, 2006 USA (Pennsylvania) Geisinger Rural Aging Study (42)	**Separate CAs using either number of servings or percent daily energy contribution from the same FG and according to the same CA approach**; NA algorithm (PROC FASTCLUS); Euclidean distance; varying number of cluster from 2 to 6; screeplot of eigenvalues and within-cluster sum of squares plot to choose the optimal number of clusters	Not applicable, 2-cluster solution chosen examining screeplot of eigenvalues and within-cluster sum of squares plot	**Reproducibility**: No formal assessment	**Reproducibility**: Both methods consistently clustered subgroups with high energy contribution (e.g. fats and oils and dairy desserts); clusters resulting from the percent energy method were less likely to discern differences between FG and in particular to differentiate fruit and vegetable subgroups, as compared to number of servings method
Balder, 2003 Netherlands, Sweden, Finland, and Italy DIETSCAN (NLCS, SMC, ATBC, ORDET) (26)	**Separate PCFAs 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	NLCS: 23 (5) with Ms, 23.2 (5) with Fs; ORDET: 28.5 (4); SMC: 21.8 (4); ATBC: 20.3 (3); final results based on unadjusted variables for energy	**Reproducibility:** comparison of different scenarios within each study with Procrustes rotation; Cross-study reproducibility: no formal assessment	**Reproducibility:** 1. Dichotomization: no effect (correlations of 0.98 - 1.00 on the diagonal of the Procrustes rotation matrix and low mutual correlations between factors); 2. Energy-adjustment: when using the energy-adjusted FG, the factor solutions were mostly comparable with the unadjusted factor solutions; mainly the DPs with high loadings on energy-contributing FG changed; by using energy-adjusted food variables, substitution of foods such as brown vs. white bread and low fat vs. medium and full-

	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			fat dairy products became more important, but other DPs unaffected by adjustment for energy (high correlations on the diagonal of the Procrustes rotation matrix); 3. Solutions with 2-6 factors: use of the Procrustes rotation matrix to track similar DPs across solutions with different number of factors: study-specific numbers of factors described with percentages of explained variance; 4. Split-half analysis: very similar results on the 2 subsamples **Cross-study reproducibility:** Two of the identified DPs were qualitatively similar across studies and between Ms and Fs
Bountziouka, 2012 Greece NA (41)	**Separate PCAs conducted on the 2 administrations of the FFQ with different rotation methods**; EIG>1; Varimax and quartimax rotation among the orthogonal rotations and promax and oblimin rotation among the non-ortoghonal rotations; Loading >|0.30|cut-off	Unrotated: 38 (4) with FFQ1 data and 40 (4) with FFQ2 data; Varimax rotation: 32.5 (4) with FFQ1 data and 35.6 (4) with FFQ2 data; Quartimax rotation: 32.8 (4) with FFQ1 data and 38.7 (4) with FFQ2 data; Promax rotation: NA (3); Oblimin rotation: NA (3)	**Reproducibility:** Kendall tau-b correlation coefficient between corresponding scores derived from solutions at different time-points with no rotation and with different rotation methods; Bland-Altman method (with 95% LOA) between scores from solutions at different time-points with no rotation	**Reproducibility:** 1. Unrotated solutions: All the 4 identified DPs were qualitately similar and the following measures witnessed a good agreement between scores at the 2 time-points; Kendall tau-b correlation coefficient between FFQ1 and FFQ2 scores ranged from 0.50 to 0.63 (all P<0.0001); Bland-Altman method: mean differences were equal to 0 but wide LOA especially for the LOW-FAT DP; 2. Orthogonal rotation solutions: 3 DPs were qualitately similar across the 2 orthogonal solutions, but the agreement was low-to-moderate between scores at the 2 time-points; Kendall tau-b correlation coefficient between FFQ1 and FFQ2 scores ranged from 0.15 to 0.44 for the varimax (all P<0.0001) and from 0.28 to 0.46 for

			and with different rotation methods	the quartimax rotation method (all P<0.0001); Bland-Altman method: mean differences were equal to 0, but wider LOA than with unrotated solutions; from both approaches, better agreement with quartimax (than varimax) rotation; 3. Non-orthogonal rotation solutions: 3 DPs were qualitatively similar, but the agreement was low-to-moderate between scores at the 2 time-points; Kendall tau-b correlation coefficient between FFQ1 and FFQ2 scores ranged from 0.21 to 0.41 for the promax (all P<0.0001) and from 0.31 to 0.46 for the oblimin rotation method (all P<0.0001); Bland-Altman method: mean differences were equal to 0 but wider LOA than with unrotated solution; from both approaches, better agreement with oblimin (than promax) rotation
Castro, 2015 Brazil Healthy Survey of the City of Sao Paulo (50)	**EFA:** adjustment for within-person variation via Multiple Source Method; robust maximum likelihood estimation; EIG>1, Scree test, interpretability; Varimax among the orthogonal rotations and promax (power=4) and oblimin rotation among the non-orthogonal rotations; Alphanumeric labelling; **CFA:** Loading >=|0.20| or	EFA: ~10 with any rotation method used (2); CFA:2-factor model with |0.25| cut-off and promax rotation method	**Reproducibility and Validity**: CFA; different cut-off for FG inclusion; within CFA with and without different cut-offs for FG inclusion, comparison of rotation methods	**Validity:** 1. CFA with |0.20| cut-off: regardless of rotation method, factor loadings were statistically significant for all DPs (P< 0.05) and similar to those from EFA; (Reproducibility: promax and oblimin produced DPs with small but significant correlations (r = 0.17, P< 0.01); irrespective of rotation method, unacceptable model fits except for SRMR (SRMR < 0.08)); 2. CFA with |0.25| cut-off: regardless of rotation method, factor loadings were statistically significant for all DPs (P< 0.05) and similar to those from EFA; (Reproducibility: better model fit with promax (best

	|0.25|cut-offs on EFA results based on different rotation methods; robust maximum likelihood estimation; adjusted chi-squared test, CFI, NNFI, RMSEA (90% CI), and SRMR			values of CFI, NNFI, RMSEA, and SRMR) and then varimax, and last oblimin rotation solution (CFI and NNFI < 0.90); small but significant correlations between factors, with both promax (r = 0.19, P< 0.01) and oblimin rotations (r = 0.18, P< 0.01))
Dekker, 2013 Netherlands Doetinchem Cohort Study (25)	**CA:** 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 indexes and internal cluster validity indexes	**Reproducibility:** internal cluster validity and stability (Jaccard indexes with 0.85 cut-off) indexes; **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;	**Reproducibility:** 1. internal cluster stability: highly stable clusters, with Jaccard indexes >0.85 for most cluster numbers from 2 to 6, but highest stability for the 2-cluster solution; 2. internal cluster validity: indexes pointing to 2-cluster solution, although with some exceptions; **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 were >5%, but the third was not; 2. Transitions of individuals

			2. Transitions of individuals between DPs over time: proportion of stable eaters (those assigned to the same cluster) and transitioners (those assigned to different clusters) in all 3 surveys and in survey 2 and 4 (over the higher 10-year 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	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 energy contributed by important FG was found during the 10-year 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)

Fransen, 2014 Netherlands EPIC-NL (51)	**PCA:** percentage energy contributed variables from both subsamples and the whole study population based on varying number of factors retained from 2 to 6; EIG>1, Scree test, Scree test optimal coordinate, interpretability; Varimax rotation; Alphanumeric labelling; **EFA:** percentage energy contributed variables from both subsamples and the whole study population based on varying number of factors retained from 2 to 6; EIG>1, Scree test, Scree test optimal coordinate, interpretability; Varimax rotation; Alphanumeric labelling; **CA:** top-coding of percentage energy contributed variables from both subsamples and the whole study population; k-means algorithm; Calinski-Harabasz and Davies-Bouldin indexes to assess the number of clusters to retain; **CFA:** Loading >=|0.25| cut-offs on	PCA/EFA: NA (2); CA: 2-cluster solution according to Calinski-Harabasz and Davies-Bouldin indexes; CFA: 3-factor model chosen according to confirmation success measure	**Reproducibility**: 1. comparison of results from either PCA/EFA or CA on derivation and replication samples; 2. comparison of results from either PCA/EFA or CA on derivation and whole samples; 3. cluster stability with Jaccard similaries; 4. internal validity indexes for PCA/EFA (EIG>1, Scree test, Scree test optimal coordinate, interpretability) and CA (Calinski-Harabasz and Davies-Bouldin) to identify the number of DPs to retain; **Validity**: CFA on replication sample starting from PCA/EFA on derivation sample with indexes of confirmation success (ratio of FG not	**Reproducibility:** 1. comparison between derivation and replication samples: PCA/EFA: good reproducibility; CA: good reproducibility (small deviations between the 2 subsamples, although increasing with increasing number of clusters); 2. comparison between derivation and whole samples: PCA/EFA: almost identical DPs on the subsamples and whole population study; CA: almost identical clusters on the subsamples and whole population study; 3. cluster stability: highly stable cluster solutions (Jaccard similaries for all solutions >0.85), with the best solution given by 2 clusters; 4. internal validity indexes: PCA/EFA: no optimal number of DPs to retain common to all indexes (EIG>1: 11 DPs, Scree test: 3 DPs, Scree test optimal coordinate: 8 DPs); CA: 2-cluster solution was optimal according to the Calinski-Harabasz and Davies-Bouldin indexes; **Validity:** CFA on replication sample starting from PCA/EFA on derivation sample: high concordance between confirmation success measures; different confirmation success indexes between DPs within the same solution; all solutions contained 1 or more poorly confirmed DP (deviation >30%); 3-component solution was better confirmed that the others

	PCA results (with a different number of DPs) for variables in the replication sample; Loading >=|0.20| cut-offs to name DPs		confirmed to the total number of FG and deviations in factor loadings between PCA/EFA and CFA)	
Greve, 2016 Germany IDEFICS (56)	**CA**: rescaled relative frequencies (variances equal to 1); k-means (10 000 starting values), Ward's method and Gaussian mixture models (with automatic model selection via the Bayesian Information Criterion) in comparison; varying number of clusters to retain between 2 and 6; Labelling based on the difference between the cluster-specific mean consumption frequency and the overall mean consumption	Not applicable, 3-cluster solution chosen because of the highest similarities of the cluster solutions derived with each method	**Reproducibility**: ARI to assess pairwise agreement between clustering solutions	**Reproducibility**: Very little agreement between the 3 clustering methods; for all possible numbers of solutions, the Gaussian mixture model solution was constantly more similar to the k-means solution than to the Ward's solution; the best fitting Gaussian mixture model was the one that allowed the variances of the food consumption frequencies to vary within and between clusters; comparing the 3 clustering methods, the solutions with 3 clusters were most similar to each other (ARI equal to 0.47 comparing Gaussian mixture model vs. k-means, 0.23 for Gaussian mixture model vs. Ward's method and 0.20 for k-means vs. Ward's method)

	frequency measured in units of overall SDs for the FG			
Lau, 2008 Denmark Inter99 Study (48)	**Subsample 1: PCA 1:** overall analysis and separate analyses by sex; PCFA; Scree test, interpretability; Varimax and Promax rotations compared; Loading >=|0.40| cut-off; **Subsample 1: PCA 2:** as PCA 1 but including only FI whose loading was >= |0.40| cut-off; **Subsample 2: PCA 3**: overall analysis and separate analyses by sex; same criteria of PCA 1; natural scores; **Subsample 2: PCA 4**: overall analysis and separate analyses by sex; same criteria of PCA 1; applied scores with PCA 1-based loadings; **Subsample 1: CFA**: Loading	PCA 1: 17.1 (2) for entire subsample 1, 17.0 (2) for Ms, and 15.4 (2) for Fs; PCA 2, 3, and 4: NA (2); CFA: No model selection	**Reproducibility:** Pearson correlation coefficient between scores based on PCA 1 and PCA 2 in subsample 1; Pearson correlation coefficient between scores based on PCA 3 and PCA 4 in subsample 2; Bland-Altman plot between scores based on PCA 1 and PCA 2 in subsample 1, RV (95% CI of the difference of factor scores/95% CI of the average of factor scores) measure; Bland-Altman plot between scores based on PCA 3 and PCA 4 in subsample	**Reproducibility:** Rotation method on PCA 1: no significant differences in the final DPs derived from varimax vs. promax transformation, so promax rotation used for the PCA 1 analysis; Pearson correlation coefficient between scores based on PCA 1 and PCA 2 in subsample 1 was equal to 0.93 (P<0.0001) for TRADITIONAL and MODERN DPs; Pearson correlation coefficient between scores based on PCA 3 (natural scores) and PCA 4 (applied scores) in subsample 2 was equal to 0.89, 0.98, and 0.90 (P<0.0001) for the TRADITIONAL DP in all, Fs and Ms, respectively, and 0.89, 0.99, and 0.93 (P<0.0001) for MODERN DP in all, Fs and Ms, respectively; Bland-Altman method: no systematic bias between scores based on PCA 1 and PCA 2 in subsample 1; relatively poor agreement (RV=39.9% for TRADITIONAL DP and 37.6% for MODERN DP and PCA 1 and PCA 2 scores); no systematic bias between scores based on PCA 3 and PCA 4 in subsample 2; relatively poor agreement (RV=47.5%

	>=|0.40| cut-off on PCA 1 results; RMSEA		2, with RV; Validity: CFA	for TRADITIONAL DP and 47.7% for MODERN DP and PCA 3 and PCA 4 scores); for Fs acceptable RV, whereas for Ms larger variations than for Fs; Validity: CFA: good fit (RMSEA equal to 0.008 < 0.10)
Lo Siou, 2011 Canada Tomorrow Project (43)	**Separate CAs using 3 different algorithms (k-means, Ward's minimum variance, and flexible beta with beta equal to -0.25 and -0.50)**: standardization ((value - minimum) divided by range) of the percentage of daily total energy intake; varying number of clusters from 2 to 7; between- versus within-cluster variance criterion to choose the optimal number of clusters; checking of potential outliers but no need to remove them	Not applicable, 4-cluster solution chosen for Ms according to median (natural) log-transformed ratios of between- versus within-cluster variances for the 55 FG (best cluster had many FG with large ratios) and with number of clusters varying from 2 to 7 obtained from applying the k-means method, and 3-cluster solution chosen for Fs according to interpretability of the results	**Reproducibility:** 1. Optimal clustering method: separately for Ms and Fs, average values over 20 repetitions for Hubert and Arabie's ARI and kappa and Cramer's V statistics to identify the optimal clustering method based on a split-half cross validation approach considering the different numbers of clusters; 2. Optimal number of clusters: median log-ratio value of between- versus within-cluster variances for the 55 FG (best cluster had many FG with large ratios) and with	**Reproducibility:** 1. Optimal clustering method: for Ms, as the number of clusters increased, the agreement and association between cluster assignments decreased when the k-means and Ward's methods were applied; a similar pattern was observed for Fs with the k-means method; agreement and association between cluster assignments remained low when applying the flexible-beta method; compared with the other 2 clustering methods, the k-means method had the highest reproducibility of the cluster solutions for Ms and Fs and with all different numbers of clusters; 2. Optimal number of clusters: in Ms, the median log-ratio value jumped from -3.45 to -3.03 between the 3-cluster and 4-cluster solutions, suggesting the optimal number of clusters for Ms was 4; in Fs, the median log-ratio values varied little across different numbers of clusters, suggesting no clear choice for the number of clusters

			number of clusters varying from 2 to 7 obtained from applying the k-means method	
McCann, 2001 USA (New York) Western New York Diet Study (45)	**Separate PCAs for each of the 3 food classification methods**: controls-only PCA; Percentage of variance explained by each factor, interpretability; Varimax rotation; Descriptive labelling; Loading >=0.30 or <= -0.20 cut-offs used for the calculation of factor scores	7.7 (2) with 168 FG data, 13.4 (2) with 56 FG data, and 16.9 (2) with 36 FG data	**Reproducibility**: percentage exact agreement of classification along the diagonal for tertiles of DP scores by the 3 food classification schemes	**Reproducibility**: Food classification method affected neither the number nor character of the DPs identified, although total variance explained in food use increased as the detail included in the PCA decreased (~8%, with 168 FG to ~17%, with 36 FG); Percentage exact agreement: for both DPs, exact agreement in tertile classification decreased as the difference in the number of items used for PCA increased; for the HEALTHY DPs, almost half the subjects were misclassified on DP score by the broader food-use classification method including 36 FG, as compared to 168 FG; for the HIGH FAT DPs, the effect was similar but less dramatic, with percentage exact agreement decreasing from 81% (168 FG vs. 56 FG) to 76% (168 FG vs. 36 FG)
Northstone, 2008 UK ALSPAC (36)	**Separate PCAs on unadjusted (weekly frequency of consumption) and adjusted (residual method) dietary variables**: standardization; Scree test, interpretability; Varimax rotation; Loading >|0.3| cut-off	32.4 (5) with unadjusted data and 26.9 (4) with energy-adjusted data	**Reproducibility**: Pearson correlation coefficient between scores from similar DPs on the unadjusted and energy-adjusted data	**Reproducibility:** Slight differences seen in terms of components extracted and factor loadings obtained; strong correlations (all > 0.8) between scores from analogous unadjusted and energy-adjusted DPs; PROCESSED DP obtained from the unadjusted data was negatively correlated with both HEALTH-CONSCIOUS and CONFECTIONERY DPs obtained

				using the energy-adjusted data (-0.538 and -0.492, respectively)
Sauvageot, 2017 Luxembourg, Belgium, and France NESCaV (44)	**Separate CAs using 3 different algorithms (k-means, k-medians, and Ward's minimum variance):** standardization ((value - minimum) divided by range) of the residuals calculated according to Willett method; varying number of clusters from 2 to 6; 20 repetitions of the algorithm; for k-means and k-medians, 1000 runs with different random starting seeds, and solution that had the minimum total within-cluster sum of squares distances was selected; stability measure representing empirical misclassification rate across solutions on training and test samples (with k-nearest-means classifier for k-means and Ward's	Not applicable, 3-cluster solution with k-means chosen according to Cramer's V and ARI	**Reproducibility**: Optimal clustering method and number of clusters: box-plots and average values over 20 repetitions of each algorithm for each index	**Reproducibility**: Regardless of stability indices and number of clusters, more stable solutions were obtained with k-means; the most stable solution was obtained with 3 clusters

	methods and k-nearest-medians classifier for the k-medians algorithm), Cramer's V and ARI to choose the optimal combination of clusteriong method and number of clusters; truncation of >6 SDs values			

Varraso, 2012 France and Spain EGEA2-France, Spanish PAC-COPD (57)	**PCA and CFA used as equivalent approaches on 1000 randomly selected samples from 4 different set-ups**: **1. 100% of EGEA2-France study;** **2. 50% of EGEA2-France study;** **3. 25% of EGEA2-France study;** **4. 100% of Spanish PAC-COPD study;** **PCA**: Scree-plot, interpretability; Varimax rotation; Distribution of the factor loading of FG to each DP represented via box-plot and median loading > |0.30| as cut-off; **CFA:** not based on previous EFA; 4 different models tested (3-factor and 2-factor models with correlated latent variables, 3-factor and 2-factor models with independent latent variables); chi-squared test, GFI, and RMSEA; Distribution of the factor loading of FG to each DP represented via box-plot and median loading > |0.30| as cut-off	PCA: NA (3); CFA: 3-factor model with no correlation among latent variables (highest GFI and lowest RMSEA)	**Reproducibility and Validity**: statistical properties (min, quartile 1, median, quartile 3, max) of the distribution of the factor loading of each FG to each DP in each of the 4 subsamples considered	**Reproducibility and validity**: Two consistent DPs were identified by CFA in each of the subsamples, whereas PCA led to less interpretable (smaller median of factor loadings and higher dispersion) DPs, especially for the smallest sample

Wirfalt, 2000 Sweden MDC (46)	**Separate CAs using 2 different input variable formats:** standardization or not of the percentage of daily total energy intake; k-means algorithm; varying number of clusters from 2 to 10; interpretability (cluster size and ability to differentiate FG intakes) to choose the optimal number of clusters	Not applicable, 6-cluster solution chosen according to interpretability of results	**Reproducibility**: 1. Optimal number of clusters: no formal assessment; 2. Choice of the optimal input variable format: for each set of input variables, discriminant analysis after assuming the optimal 6-cluster solution (discriminant function chosen with all 43 FG and with stepwise regression to identify FG contributing significantly to the formation of clusters)	**Reproducibility**: 1. Optimal number of clusters: the 6-cluster solution produced reasonably sized and well-separated clusters for both input variable formats considered; 2. Choice of the optimal input variable format: the 6-cluster solution identified for each set of input variables was reproducible: with standard discriminant analysis, the agreement between actual and predicted cluster allocation ranged between 91.0% and 95.2% for the unstandardized variables, and between 91.1% and 100% for the z-scored variables; when using the stepwise function of the discriminant analysis, 18 unstandardized variables and 31 z-scored variables contributed significantly to the predicted cluster allocations

[a]ABBREVIATIONS: ALSPAC: Avon Longitudinal Study of Parents and Children; ARI: adjusted Rand index; ATBC: Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; CA: cluster analysis; CFA: confirmatory factor analysis; CFI: comparative fit index; CI: confidence interval; DIETSCAN: DIETary patternS and CANcer in four European countries project; DP: dietary pattern; EFA: exploratory factor analysis; EGEA2-France: Epidemiological Study on the Genetics and Environment of Asthma 2–France; EIG: Eigenvalue; EPIC-NL: European Prospective Investigation into Cancer and Nutrition-The Netherlands; FFQ: food-frequency questionnaire; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1, 2, or 3; F: female; FG: food groups; FI: food items; GFI: goodness of fit index; IDEFICS: Identification and Prevention of Dietary and Lifestyle-induced Health Effects in Children and Infants; LOA: limits of agreement; M: male; MDC: Malmo Diet and Cancer study; NA: not available; NESCaV: Nutrition, Environment and Cardiovascular Health; NLCS: Netherlands Cohort Study on diet and cancer; NNFI: non-normed fit index or Tucker-Lewis index; ORDET: Ormoni e Dieta nella Eziologia dei Tumori in Italy; PAC-COPD: Phenotype and Course of Chronic Obstructive

Pulmonary Disease study–Spain; PCA: principal component analysis; PCFA: principal component factor analysis; RMSEA: root mean square error of approximation; RV: relative variation; SD: standard deviation; SMC: Swedish Mammography Cohort; SRMR: standardized root mean square residual.

Table 3. Reproducibility and/or relative validity of *a posteriori* dietary patterns[a]

Reference	DP identification methods	Percent Explained Variance (# factors) or CFA/CA model	Assessment of reproducibility/validity	Main Results
Ambrosini, 2011 Australia Western Australian Pregnancy Cohort (Raine) Study (13)	**Separate EFAs (maximum likelihood method) conducted on the FFQ and DR data with all available information used (1613 subjects for FFQ and 822 subjects for DR data):** EIG>1 on FFQ data only, Scree test; Varimax rotation; Loading >|0.20|cut-off; 4 FG removed from the final analysis due to small loadings on all factors	84 (2) with FFQ data and 53 (2) with DR data	**Relative validity**: Sperman correlation coefficient (crude and partial, with adjustment by total energy intake) and Bland-Altman method (with 95% LOA) between scores from FFQ and DR data	**Relative validity**: The identified DPs were similar although not identical in terms of loadings; modest sperman correlation coefficient between DP scores from FFQ and DR given by 0.43 (crude) and 0.45 (partial and corrected) (P<0.001) for HEALTHY DP and 0.27 (crude) and 0.36 (partial and corrected) (P<0.001) for WESTERN DP; correlations improved after ajustment for energy intake; Bland-Altman method: acceptable (not significantly different from 0) mean agreement for both DP scores; 95% LOA given by (-1.69, 1.65) for HEALTHY DP and (-1.89, 1.82) for WESTERN DP, so slightly narrower for HEALTHY DP; minor differences between girls and boys in all previous analyses

Asghari, 2012 Iran TLGS (11)	**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:** intra-class correlation coefficient between scores from FFQ1 and FFQ2 data; Relative validity: sperman correlation coefficient, and Bland-Altman method (with 95% LOA) between scores from FFQ2 and scores from m24HR data, deattenuated correlation coefficient (Rosner and Willett formula) between each DP score to reduce the random within-person month-to month variability in 24HR-based DPs; **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	**Reproducibility:** intra-class correlation coefficients between FFQ1- and FFQ2-based scores equal to 0.72 (P<0.001) for the IRANIAN TRADITIONAL DP and 0.80 (P<0.001) for the WESTERN DP; Relative validity: crude and corrected sperman correlation coefficients between FFQ2 and m24HRs similar and equal to 0.48 for the IRANIAN TRADITIONAL and 0.75 for the WESTERN DPs; Bland-Altman plot: 95% LOA for the difference between factor scores from FFQ2 and m24HR lay between -1.58 and 1.58 for the IRANIAN TRADITIONAL and between -1.33 and 1.33 for the WESTERN DP; **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

			between baseline and follow-up data	

Beck, 2012 New Zealand NA (30)	**Separate PCFAs on FFQ1, FFQ2, and DR**: EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling	~20 (2) with each of the 3 dietary sources	**Reproducibility:** Pearson correlation coefficient and Bland-Altman method (with 95% LOA) between scores from FFQ1 and FFQ2 data; weighted kappa statistics and proportions of subjects at the same third, or the opposite third when comparing tertiles classification of factor scores between FFQ1 and FFQ2 data; **Relative validity**: Pearson correlation coefficient and Bland-Altman method (with 95% LOA) between scores from FFQ1 and DR data; weighted kappa coefficient and proportions of subjects at the same third, or the opposite third when comparing tertiles classification of factor scores between FFQ1 and DR data	**Reproducibility:** good Pearson correlation coefficients between FFQ1 and FFQ2 DP scores (0.76 for the HEALTHY DP and 0.76 for the SANDWICH AND DRINKS DP (P<0.001)); Bland-Altman method: the difference between DP scores from FFQ1 and FFQ2 increased with increasing scores for both DPs; Cross-classification of DP scores: >50% of participants classified in the same third and <10% misclassified into the opposite third for both the DPs between FFQ1 and FFQ2; Weighted kappa coefficient between FFQ1 and FFQ2 moderate (HEALTHY) and good (SANDWICH AND DRINKS DP); **Relative validity**: reasonable Pearson correlation coefficients between FFQ1 and DR DP scores (0.34 for the HEALTHY DP and 0.62 for the SANDWICH AND DRINKS DP) (P<0.001)); Bland-Altman method: the difference between DP scores from FFQ1 and DR increased with increasing scores for both DPs; Cross-classification of DP scores: >50% of participants classified in the same third and <10% misclassified into the opposite third for both the DPs between FFQ1 and DR; Weighted kappa coefficient between FFQ1 and DR DP scores fair (HEALTHY) and moderate (SANDWICH AND DRINKS);

Bountziouka, 2011 Greece NA (40)	**Separate PCAs conducted on FFQ and DR data:** EIG>1.4, Scree test; Varimax rotation; Loading >|0.30|cut-off; **Separate CAs conducted on FFQ and DR data**: k-means method; Euclidean and Mahalanobis distances; maximum achieved distances between cluster's centers; 2-, 3-, and 5- cluster solutions considered	PCA: 35 (4) with FFQ data and 29 (4) with DR data; CA: not applicable, 2-cluster solution chosen according to maximum achieved distances between cluster's centers	**Relative validity**: Kendall tau-b correlation coefficient between scores from FFQ and DR; Bland-Altman method (with 95% LOA) between scores from FFQ and DR; Kendall tau-b correlation coefficient and exact classification rate for CA	**Relative validity:** PCA: Kendall tau-b correlation coefficient: significant but low correlation coefficient equal to 0.22 for the WESTERN and 0.23 for the MEDITERRANEAN DPs (P<0.001 for both DPs); Bland-Altman method: 95% LOA given by -2.35, 2.30 for WESTERN and -2.23, 2.26 for MEDITERRANEAN DP; CA: Kendall tau-b correlation coefficient: very good agreement between clusters derived from FFQ and DR (0.81, P<0.001); exact classification rate: 48% and 59% depending on the distance used
Crozier, 2008 UK NA (39)	**Separate PCFAs conducted on FFQ and DR data**: standardization; NA criteria for choosing the number of factors; NA rotation; Descriptive labelling; Fisher-Yates transformation of scores to improve adherence to normality	15.9 (2) with FFQ data and 14.3 (2) with DR data	**Relative validity**: Pearson correlation coefficient between scores from FFQ and DR; Bland-Altman method (with 95% LOA) between scores from FFQ and DR	**Relative validity**: The corresponding DPs from FFQ and DR data were strikingly similar, especially the PRUDENT DP; Pearson correlation coefficient between FFQ and DR scores were 0.67 (P < 0.001) for PRUDENT DP and 0.35 (P < 0.001) for WESTERN DP; Bland-Altman method: good agreement between scores from FFQ and DR for PRUDENT DP (95% of the differences laying within −1.58 and +1.58 SDs), but less good for WESTERN DP (95% of the differences lying within −2.22 and +2.22 SDs); consistently wider limits for the WESTERN DP with generally similar variations across characteristics

Hong, 2016 China NA (34)	**Separate EFAs on FFQ1, FFQ2, and m24HRs**: EIG, Scree test, interpretability; Varimax rotation; Loading >0.30 cut-off	40.0 (4) for FFQ1 data, 44.9 (4) for FFQ2 data, and 32.4 (4) for m24HR data	**Reproducibility:** intra-class correlation coefficient between DP scores from FFQ1 and FFQ2 data; cross-classification: range of agreement rates for the same or adjacent quartile classifications and misclassification into opposite quartiles; kappa coefficient; **Relative validity**: Pearson correlation coefficient between DP scores from FFQ1 and FFQ2, respectively, and m24HR data, after adjusting for energy intake using the residual method; cross-classification: range of agreement rates for the same or adjacent quartile classifications and misclassification into opposite quartiles; kappa coefficient; Bland-Altman method and 95% LOA considering mFFQ, in comparison with m24HR scores	**Reproducibility:** The 4 derived DPs were qualitatively similar across 3 sources of dietary data, although loadings were partly different; good intra-class correlation coefficient between DP scores from FFQ1 and FFQ2 data (>0.6 for all DPs, all P<0.001); cross-classification: range of agreement rates for the same or adjacent quartile classifications equal to 29.2-66.3% (both for ANIMAL AND PLANT PROTEIN DP, with adjacent and same quartile, respectively) and misclassification into opposite quartiles was <5% for all DPs; kappa coefficient: fair-to-moderate (range: 34-68% with minimum for NUTS AND SWEETS and maximum for ANIMAL AND PLANT PROTEIN DPs, respectively); **Relative validity:** reasonable adjusted Pearson correlation coefficient between DP scores from FFQ and m24HR data (range of adjusted values: 0.387 - 0.838 with minimum for CHINESE TRADITIONAL DP and maximum for ANIMAL AND PLANT PROTEIN DP); cross-classification: range of agreement rates for the same or adjacent quartile classifications equal to 32.4 (for CHINESE TRADITIONAL DP, same quartile, FFQ1) - 47.0% (for ANIMAL AND PLANT PROTEIN DP, same quartile, FFQ1) and misclassification into opposite quartiles was <5% for all DPs; kappa coefficient: fair-to-moderate (range: 25.9-48.1% for BEVERAGE AND ALCOHOL DP with FFQ1 and ANIMAL AND PLANT PROTEIN with FFQ1, respectively); Bland-Altman method: mean agreement between DP scores derived from mFFQ and m24HR were not significantly different from 0 in all comparisons; mean differences were 0.0 (95% LOA: -1.03 - 1.04) for ANIMAL AND PLANT PROTEIN DP, 0.0 (95% LOA:

				−reasonable adjusted Pearson correlation coefficient between DP scores from FFQ and m24HR data (range of adjusted values: 0.387 - 0.838 with minimum for CHINESE TRADITIONAL DP and maximum for ANIMAL ANOTEIN DP had better performance that the other DPs
Hu, 1999 USA (Massachusetts) HPFS (9)	**Separate PCFAs on FFQ1, FFQ2, and mDRs**: EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling	20 (2)	**Reproducibility**: crude Pearson correlation coefficient between DP scores from FFQ1 and FFQ2; **Relative validity**: crude and corrected (for week-to-week variation in DRs) Pearson correlation coefficient between DP scores from either FFQ1 or FFQ2 and DR	**Reproducibility**: good crude Pearson correlation coefficient between DP scores from FFQ1 and FFQ2 (0.70 for the PRUDENT and 0.67 for the WESTERN DPs); **Relative validity**: (crude and corrected) Pearson correlation coefficients between DP scores from either FFQ1 or FFQ2 and DR ranged from 0.34 to 0.74
Khani, 2004 Sweden SMC (33)	**Separate PCFAs on FFQ1 and FFQ2 within the reproducibility sample, and on FFQ and mDRs within the validity sample**: EIG>1.8; Varimax rotation; Descriptive labelling	Within the reproducibility sample: 29 (3) for FFQ1 data and 30 (3) for FFQ2 data; within the validity sample: 30 (3) for DR data and 34 (3) for FFQ data	**Reproducibility:** crude sperman correlation coefficient between DP scores from FFQ1 and FFQ2 data; **Relative validity**: crude and corrected (for unreproducibility of the FFQ) sperman correlation coefficient between DP scores from FFQ and DR	**Reproducibility:** good crude sperman correlation coefficient between DP scores from FFQ1 and FFQ2 data (range: 0.63 - 0.73 across DPs), with highest results for the DRINKER DP; Relative validity: reasonable (crude and corrected) sperman correlation coefficient between DP scores from FFQ and DR (range of crude values: 0.41 - 0.73; range of corrected values: 0.50 - 0.85), with highest results for the DRINKER DP
Liu, 2015 China	**Separate PCFAs on FFQ1, FFQ2, and**	30 (2)	**Reproducibility:** Pearson correlation coefficient (crude and	**Reproducibility:** between FFQ1 and FFQ2, crude Pearson correlation coefficients equal to 0.58 for the PRUDENT DP and 0.60

NA (32)	**m24HRs**: EIG>1.5, Scree test, interpretability; Varimax rotation; Loading >|0.4| cut-off		partial, with adjustment for log10-transformation of total energy intake), intra-class correlation coefficient (to adjust for the effect of different scales of measures), and Bland-Altman method (with 95% LOA) between scores from FFQ1 and FFQ2 data; weighted kappa coefficient and proportions of subjects at the same third, or the opposite third when comparing tertiles classification of factor scores between FFQ1 and FFQ2 data; **Relative validity**: Pearson correlation coefficient [crude, partial (with adjustment for log10-transformation of total energy intake), and de-attenuated, to correct monthly and seasonal variation] and Bland-Altman method (with 95% LOA) between scores from either FFQ1 or FFQ2 and 24HR data; weighted kappa coefficient and proportions of	for the PROCESSED FOOD DP, partial Pearson correlation coefficient equal to 0.51 for PRUDENT DP and 0.56 for PROCESSED FOOD DP, intra-class correlation coefficient equal to 0.57 for PRUDENT DP and 0.55 for PROCESSED FOOD DP; Bland-Altman method: divergence not obvious between DP scores on FFQ1 and FFQ2; Cross-classification analysis: >54% of the participants correctly classified into the same tertile and <9% misclassified into an opposite tertile for both DPs when 2 FFQs compared; moderate weighted kappa coefficient (0.45 for PRUDENT and 0.56 for PROCESSED FOOD) between the 2 FFQs; **Relative validity**: between FFQs and 24HRs, crude Pearson correlation coefficients ranged from 0.45 to 0.64 for PRUDENT DP and from 0.46 to 0.50 for PROCESSED FOOD DP, de-attenuated correlation coefficients ranged from 0.54 to 0.78 for the PRUDENT DP and from 0.55 to 0.61 for the PROCESSED FOOD DP; partial Pearson correlation coefficients ranged from 0.41 to 0.56 for the PRUDENT DP and from 0.42 to 0.44 for the PROCESSED FOOD DP; Bland-Altman method: divergence not obvious between DP scores on FFQ1 or FFQ2 and 24HR data; cross-classification analysis: >54% of the participants correctly classified into the same tertile and <9% misclassified into an opposite tertile for both DPs when FFQs and 24HRs compared; moderate weighted kappa coefficient (range: 0.42-0.60 across the 2 DPs and FFQs)

			subjects at the same third, or the opposite third when comparing tertiles classification of factor scores between either FFQ1 or FFQ2 and 24HRs data	
Loy, 2013 Malaysia USM Birth Cohort Study (37)	**Separate PCAs on FFQ and m24HRs**: EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling	22.4 (2) with FFQ data and 20.7 (2) with m24HRs data	**Relative validity**: Pearson correlation coefficient and Bland-Altman method (with 95% LOA) between scores from FFQ and 24HR data; weighted kappa coefficient and proportions of subjects at the same third, or the opposite third when comparing tertiles classification of factor scores between FFQ and m24HR data	**Relative validity**: relatively high sperman correlation coefficient between DP scores from FFQ and m24HR data given by 0.59 (HEALTHY) and 0.63 (LESS-HEALTHY) (P<0.001); Bland-Altman method: good agreement for both DPs, with 95% of the differences within +-1.87 SD (HEALTHY) and 1.69 SD (LESS-HEALTHY), no association between the difference and the average for both DPs; cross-classification: acceptable (<10%) degrees of misclassification and lower than recommended percentage of classified in the same third (~50% or more) for both DPs: moderate (0.56) and good (0.72) agreement from weighted kappa coefficient for the HEALTHY and LESS-HEALTHY DPs, respectively; from all criteria, LESS-HEALTHY DP more valid than HEALTHY DP
McNaughton, 2005 UK Medical Research Council National Survey of Health and	**Separate PFCAs on 24HR recall, 48HR recall, and DR data**: Separate analyses by sex; EIG>1, Scree test; Varimax rotation; Loading > 0.3 cut-off	Range: 19 (5) with 24HR data - 22 (5) with DR data	**Relative validity**: correlation coefficient between scores from similar DPs across dietary assessment tools	**Relative validity**: Five distinct DPs were identified using the DR and 48HR, but were less consistent on the 24HR data; Moderate-to-good correlations between factor scores on 48HR and DR data (0.13–0.67, all P<0.001), with the highest values for the HEALTH-AWARE DP in both Ms and Fs; correlations with 48HR data were higher than those between the 24HR and DR data (-0.01 – 0.59, with most P-values<0.001)

Development (1946 British Birth Cohort) (35)				
Nanri, 2012 Japan JPHC (31)	**Separate PCAs on logtransformed data from FFQ_R, FFQ_V, and mDR data**: Separate analyses by sex; EIG>1, Scree test, interpretability; Varimax rotation; Descriptive labelling; Energy-adjusted scores using the residual method	In Ms: 23.9 for mDR data, 29.4 for FFQ_R data, and 26.5 for FFQ_V data (3); in Fs: 23.0 for mDR data, 24.9 for FFQ_R data, and 32.9 for FFQ_V data (3)	**Reproducibility:** sperman correlation coefficient between DP scores from the FFQ_R and FFQ_V data in both Ms and Fs; **Relative validity**: sperman correlation coefficient between DP scores from mDR and FFQ_V data	**Reproducibility:** acceptable sperman correlation coefficient between DP scores from the FFQ_R and FFQ_V data in both Ms and Fs for the 3 DPs (TRADITIONAL JAPANESE DP in Ms and WESTERNIZED JAPANESE DP in Fs given by 0.77 and 0.71, respectively, range of correlation coefficients: 0.55-0.77 across DPs); **Relative validity**: acceptable sperman correlation coefficient between DP scores from mDR and FFQ_V (TRADITIONAL JAPANESE DP in Ms and in Fs given by 0.49 and 0.63, respectively, range of correlation coefficients: 0.32-0.63 across DPs)
Okubo, 2010 Japan NA (38)	**Separate PCFAs conducted on DHQ1, mDHQs, and mDRs data**: log-transformation and adjustment by energy intake with residual method; Separate analyses by sex; Scree test, interpretability; Varimax rotation; Descriptive labelling	In Fs, 30.1 (3) with DHQ1 data, 31.2 (3) with mDHQ data, and 30.8 (3) with mDR data; in Ms, 21.5 (2) with DHQ1 data, 24.4 (2) with mDHQ data, and 25.8 (2) with mDR data	**Relative validity**: Pearson correlation coefficient between DHQ1 and mDR data and between mDHQ and mDR data; Bland-Altman method (with 95% LOA) between scores from DHQ1 and mDRs	**Relative validity**: The identified factor loadings were similar in magnitude and direction across DHQ1, mDHQ, and mDR data; Pearson correlation coefficient for the HEALTHY, WESTERN, and JAPANESE TRADITIONAL DPs in Fs were equal to 0.57, 0.36, and 0.44, and for the HEALTHY and WESTERN in Ms were 0.62 and 0.56; when mDHQ was examined, correlation coefficients improved for Fs (0.45 – 0.69); Bland-Altman method: for both Ms and Fs, mean differences between scores derived from DHQ1 and DR were 0; 95% LOA for the difference between factor scores derived from DHQ1 and DR lay within -1.81 and 1.81 for HEALTHY, within -2.22 and 2.22 for WESTERN and within -2.08 and 2.08 for JAPANESE TRADITIONAL

				DP in Fs; and within -1.83 and 1.83 for the HEALTHY and within -1.71 and 1.71 for the WESTERN DP in Ms; agreements generally improved between mDHQ and DR data
Ryman, 2015 USA (Southwest Alaska) CANHR (54)	**EFA**: logtransformation (base e) on 358 subjects; NA criteria for choosing the number of factors; NA rotation; Loading >=|0.60|cut-off; **CFA**: Loading >=|0.35|cut-offs (and a priori knowledge of Alaska native diet) on EFA results on 272 subjects; 3-factor model with correlated factors; GFI, AGFI, RMSEA, CFI, and NNFI	EFA: NA (3); CFA: 3-factor model with correlated factors	**Validity:** CFA; **Reliability:** composite reliability of DPs with CFA (squared standardized loadings and sum of the error variances), test-retest reliability of DPs with intra-class correlation coefficient (FFQ1 and FFQ2), indicator reliability of individual FG included in the CFA-based DPs with CFA (square of the standardized factor loadings for each FG), and test-retest reliability of individual FG included in the CFA-based DPs with intra-class correlation coefficient (FFQ1 and FFQ2)	**Validity:** CFA: significant and high (>0.40) standardized coefficients of FG on the given factor, except for 1 FG; satisfactory goodness of fit indexes (GFI, AGFI, CFI, and NNFI values were 0.93, 0.91, 0.92, and 0.91, respectively, all >0.90, and RMSEA was equal to 0.004 < 0.005); **Reliability:** composite reliability of DPs: ranged from 0.56 to 0.73; test-retest reliability of DPs: ranged from 0.34 to 0.66; indicator reliability of individual FG included in CFA-based DPs: ranged from 0.07 to 0.46; test-retest reliability of individual FG included in CFA-based DPs: ranged from 0.11 to 0.50, with better reliability for market-based FG
Togo, 2003 Denmark MONICA (47)	**Separate PCFAs on FFQ and DR data (in octiles):** Separate analyses by sex; standardization with Kaiser normalization; Scree test, interpretability;	In Ms: 30.5 (3) with FFQ data and 26.2 (3) with DR data; in Fs: 23.8 (2) with FFQ	**Relative validity**: Pearson correlation coefficient between scores based on FFQ and DR data;	**Relative validity**: EFA on FFQ and DR data: The identified DPs were very similar, although the percentages of explained variance were lower on DR data; Pearson correlation coefficient between FFQ-based and DR-based scores ranged between 0.34 (TRADITIONAL DP among Ms) and 0.61 (both GREEN DPs, among Ms and Fs); CFA on FFQ and DR data: Pearson correlation coefficient between FFQ-

	Promax rotation; Loading >|0.30| cut-off; **Separate CFAs on FFQ and DR data**: Loading >=|0.30| cut-off on EFA results; polychoric correlation matrix; RMSEA; weighted least square variable estimates with robust standard errors and mean- and variance-adjusted chi-squared test statistic	data and 19.8 (2) with DR data	**Validity:** CFA; Pearson correlation coefficient between scores based on EFA and CFA	based and DR-based scores ranged between 0.37 (TRADITIONAL DP among Ms) and 0.64 (GREEN DP among Fs); higher correlations with CFA than with EFA; **Validity** (EFA vs. CFA with the same dietary source): CFA-based DPs were similar across dietary sources and came from models with reassuring model fit (RMSEA < 0.10 no matter of the dietary source and among Ms and Fs); FFQ data: Pearson correlation coefficient between EFA-based and CFA-based scores ranged between 0.91 (TRADITIONAL DP among Ms) and 0.96 (SWEET-TRADITIONAL DP among Fs); DR data: Pearson correlation coefficient between EFA-based and CFA-based scores ranged between 0.82 (GREEN DP among Ms) and 0.94 (both GREEN and SWEET-TRADITIONAL DPs among Fs); higher correlations were found when using the same dietary data

[a]ABBREVIATIONS: 24HRs/48HRs: 24/48 hours recall; AGFI: adjusted goodness of fit index; CA: cluster analysis; CANHR: Center for Alaska Native Health Research study; CFA: confirmatory factor analysis; CFI: comparative fit index; CI: confidence interval; DHQ1/DHQ2/DHQ3: diet history questionnaire at time 1, 2, or 3; DP: dietary pattern; DR: dietary record; EFA: exploratory factor analysis; EIG: Eigenvalue; F: female; FFQ: food-frequency questionnaire; FFQ_R: food-frequency questionnaire from the reproducibility study; FFQ_V: food-frequency questionnaire from the relative validity study; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1/2/3; FG: food groups; GFI: goodness of fit index; HPFS: Health Professionals Follow-up Study; JPHC: Japan Public Health Center-Based Prospective study; LOA: limits of agreement; M: male; m24HRs/m48HRs: mean 24/48 hours recall; mDHQ: mean diet history questionnaire; mDR: mean dietary record; mFFQ: mean food frequency questionnaire; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not available; NNFI: non-normed fit index or Tucker-Lewis index; PCA: principal component analysis; PCFA: principal component factor analysis; RMSEA: root mean square error of approximation; SD: standard deviation; SMC: Swedish Mammography Cohort; TLGS: Teheran Lipid and Glucose Study; USM: Universiti Sains Malaysia

Table 4. Construct validity of *a posteriori* dietary patterns[a]

Reference	DP identification methods	Percent Explained Variance (# factors) or CFA/CA model	Assessment of reproducibility/validity	Main Results
Bedard, 2015 France E3N (EPIC-France) (49)	**PCA**: Scree test, interpretability; Varimax rotation; Descriptive labelling; **CFA**: not based on previous EFA; 4 different models tested (3-factor and 2-factor models with correlated latent variables, 3-factor and 2-factor models with independent latent variables); overall chi-squared test of fit, GFI, and RMSEA with 90% CI	PCA: 24 (3); CFA: 3-factor model with no correlation among latent variables (highest GFI and lowest RMSEA)	**Validity**: CFA; Pearson correlation coefficient between corresponding scores from PCA and CFA	**Validity**: CFA: good fitting of selected model; Pearson correlation coefficients between corresponding scores from EFA and CFA ranged from 0.83 to 0.87
Castro, 2015 Brazil Healthy Survey of the City of Sao Paulo (50)	**EFA:** adjustment for within-person variation via Multiple Source Method; robust maximum likelihood estimation; EIG>1, Scree test, interpretability; Varimax among the orthogonal rotations and promax (power=4) and oblimin rotation among the non-orthogonal rotations; Alphanumeric labelling; **CFA:** Loading >=|0.20|or |0.25|cut-offs	EFA: ~10 with any rotation method used (2); CFA:2-factor model with |0.25| cut-off and promax rotation method	**Reproducibility and Validity**: CFA; different cut-off for FG inclusion; within CFA with and without different cut-offs for FG inclusion, comparison of rotation methods	**Validity:** 1. CFA with |0.20| cut-off: regardless of rotation method, factor loadings were statistically significant for all DPs (P< 0.05) and similar to those from EFA; (Reproducibility: promax and oblimin produced DPs with small but significant correlations (r = 0.17, P< 0.01); irrespective of rotation method, unacceptable model fits except for SRMR (SRMR < 0.08)); 2. CFA with |0.25| cut-off: regardless of rotation method, factor loadings were statistically significant for all DPs (P< 0.05) and similar to those

	on EFA results based on different rotation methods; robust maximum likelihood estimation; adjusted chi-squared test, CFI, NNFI, RMSEA (90% CI), and SRMR			from EFA; (Reproducibility: better model fit with promax (best values of CFI, NNFI, RMSEA, and SRMR) and then varimax, and last oblimin rotation solution (CFI and NNFI < 0.90); small but significant correlations between factors, with both promax (r = 0.19, P< 0.01) and oblimin rotations (r = 0.18, P< 0.01))
Fransen, 2014 Netherlands EPIC-NL (51)	**PCA:** percentage energy contributed variables from both subsamples and the whole study population based on varying number of factors retained from 2 to 6; EIG>1, Scree test, Scree test optimal coordinate, interpretability; Varimax rotation; Alphanumeric labelling; **EFA:** percentage energy contributed variables from both subsamples and the whole study population based on varying number of factors retained from 2 to 6; EIG>1, Scree test, Scree test optimal coordinate, interpretability; Varimax rotation; Alphanumeric labelling; **CA:** top-coding of percentage energy contributed variables from both	PCA/EFA: NA (2); CA: 2-cluster solution according to Calinski-Harabasz and Davies-Bouldin indexes; CFA: 3-factor model chosen according to confirmation success measure	**Reproducibility**: 1. comparison of results from either PCA/EFA or CA on derivation and replication samples; 2. comparison of results from either PCA/EFA or CA on derivation and whole samples; 3. cluster stability with Jaccard similaries; 4. internal validity indexes for PCA/EFA (EIG>1, Scree test, Scree test optimal coordinate, interpretability) and CA (Calinski-Harabasz and Davies-Bouldin) to identify the number of DPs to retain; **Validity**: CFA on replication sample starting from	**Reproducibility:** 1. comparison between derivation and replication samples: PCA/EFA: good reproducibility; CA: good reproducibility (small deviations between the 2 subsamples, although increasing with increasing number of clusters); 2. comparison between derivation and whole samples: PCA/EFA: almost identical DPs on the subsamples and whole population study; CA: almost identical clusters on the subsamples and whole population study; 3. cluster stability: highly stable cluster solutions (Jaccard similaries for all solutions >0.85), with the best solution given by 2 clusters; 3. internal validity indexes: PCA/EFA: no optimal number of DPs to retain common to all indexes (EIG>1: 11 DPs, Scree test: 3 DPs, Scree test optimal coordinate: 8 DPs); CA: 2-cluster solution was optimal according to the Calinski-Harabasz and Davies-Bouldin indexes; **Validity:** CFA on replication sample starting from

	subsamples and the whole study population; k-means algorithm; Calinski-Harabasz and Davies-Bouldin indexes to assess the number of clusters to retain; **CFA:** Loading >=|0.25|cut-offs on PCA results (with a different number of DPs) for variables in the replication sample; Loading >=|0.20|cut-offs to name DPs		PCA/EFA on derivation sample with indexes of confirmation success (ratio of FG not confirmed to the total number of FG and deviations in factor loadings between PCA/EFA and CFA)	PCA/EFA on derivation sample: high concordance between confirmation success measures; different confirmation success indexes between DPs within the same solution; all solutions contained 1 or more poorly confirmed DP (deviation >30%); 3-component solution was better confirmed that the others
Judd, 2014 USA REGARDS (24)	**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	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 congruence 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)

	results; No different correlation structures specified; RMSEA and CFI			
Lau, 2008 Denmark Inter99 Study (48)	**Subsample 1: PCA 1:** overall analysis and separate analyses by sex; PCFA; Scree test, interpretability; Varimax and Promax rotations compared; Loading >=|0.40| cut-off; **Subsample 1: PCA 2:** as PCA 1 but including only FI whose loading was >= |0.40| cut-off; **Subsample 2: PCA 3**: overall analysis and separate analyses by sex; same criteria of PCA 1; natural scores; **Subsample 2: PCA 4**: overall analysis and separate analyses by sex; same criteria of PCA 1; applied scores with PCA 1-based loadings; **Subsample 1: CFA**: Loading >=|0.40| cut-off on PCA 1 results; RMSEA	PCA 1: 17.1 (2) for entire subsample 1, 17.0 (2) for Ms, and 15.4 (2) for Fs; PCA 2, 3, and 4: NA (2); CFA: No model selection	**Reproducibility:** Pearson correlation coefficient between scores based on PCA 1 and PCA 2 in subsample 1; Pearson correlation coefficient between scores based on PCA 3 and PCA 4 in subsample 2; Bland-Altman plot between scores based on PCA 1 and PCA 2 in subsample 1, RV (95% CI of the difference of factor scores/95% CI of the average of factor scores) measure; Bland-Altman plot between scores based on PCA 3 and PCA 4 in subsample 2, with RV; **Validity:** CFA	**Reproducibility:** Rotation method on PCA 1: no significant differences in the final DPs derived from varimax vs. promax transformation, so promax rotation used for the PCA 1 analysis; Pearson correlation coefficient between scores based on PCA 1 and PCA 2 in subsample 1 was equal to 0.93 (P<0.0001) for TRADITIONAL and MODERN DPs; Pearson correlation coefficient between scores based on PCA 3 (natural scores) and PCA 4 (applied scores) in subsample 2 was equal to 0.89, 0.98, and 0.90 (P<0.0001) for the TRADITIONAL DP in all, Fs and Ms, respectively, and 0.89, 0.99, and 0.93 (P<0.0001) for MODERN DP in all, Fs and Ms, respectively; Bland-Altman method: no systematic bias between scores based on PCA 1 and PCA 2 in subsample 1; relatively poor agreement (RV=39.9% for TRADITIONAL DP and 37.6% for MODERN DP and PCA 1 and PCA 2 scores); no systematic bias between scores based on PCA 3 and PCA 4 in subsample 2; relatively poor agreement (RV=47.5% for TRADITIONAL DP and 47.7% for MODERN DP and PCA 3 and PCA 4 scores); for Fs acceptable RV, whereas for Ms larger variations than for Fs;

				Validity: CFA: good fit (RMSEA equal to 0.008 < 0.10)
Maskarinec, 2000 USA (Hawaii) NA (52)	**EFA**: logtransformation (base e) on the first half of the population; Scree test, interpretability; Varimax rotation; Loading >=|0.60|cut-off; **CFA**: Loading >=|0.60|cut-offs on EFA results for variables in the second half of the population; chi-squared test, RMSEA, CFI, NNFI, Parsimonious NFI; t-test on factor loadings; final CFA results applied on the whole sample	EFA: 93 (4); CFA: No model selection	**Validity**: CFA	**Validity**: CFA: significant standardized coefficients of FG on the given factor, but goodness of fit indexes slightly inappropriate (significant chi-squared test P<0.0001; RMSEA equal to 0.14 >0.10; CFI equal to 0.82 < 0.90; NNFI equal to 0.83 < 0.90; parsimonious NFI equal to 0.68 > 0.60)
Newby, 2006 Sweden SMC (10)	**Separate PCFAs at each time point**: 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 Sweden SMC (27)	**Separate PCFAs at each time point**: 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**: 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 (seafood, poultry, and eggs also contributed to HEALTHY DP in 1987, whereas legumes and soy products contributed to HEALTHY DP in 1997)
Park, 2005 USA (Hawaii and Los Angeles) Hawaii - Los Angeles Multiethnic Cohort Study (53)	**PCFA**: Box-Cox transformation on the first half of the population and in the 10 separate ethnic-gender groups defined on this first half of the sample; EIG>1.25, Scree test, interpretability; Varimax rotation; Loading >=|0.60| cut-off to exclude other 7 FG from the analysis; **CFA**: Loading >=|0.60| cut-off on PCFA results for variables in the second half of the population and in the 10 separate ethnic-gender groups defined on this second half of the sample; RMSEA, CFI and NNFI; t-test	PCFA: 63.5 (3); CFA: No model selection	**Validity**: CFA	**Validity**: CFA: significant and high (>0.6) standardized loadings (all P<0.001); acceptable goodness of fit indexes (RMSEA equal to 0.095 <0.10; CFI equal to 0.90 = 0.90; NNFI equal to 0.88 < 0.90)

	on factor loadings; final PCFA results applied on the whole sample			
Ryman, 2015 USA (Southwest Alaska) CANHR (54)	**EFA**: logtransformation (base e) on 358 subjects; NA criteria for choosing the number of factors; NA rotation; Loading >=|0.60|cut-off; **CFA**: Loading >=|0.35|cut-offs (and a priori knowledge of Alaska native diet) on EFA results on 272 subjects; 3-factor model with correlated factors; GFI, AGFI, RMSEA, CFI, and NNFI	EFA: NA (3); CFA: 3-factor model with correlated factors	**Validity:** CFA; **Reliability:** composite reliability of DPs with CFA (squared standardized loadings and sum of the error variances), test-retest reliability of DPs with intra-class correlation coefficient (FFQ1 and FFQ2), indicator reliability of individual FG included in the CFA-based DPs with CFA (square of the standardized factor loadings for each FG), and test-retest reliability of individual FG	**Validity:** CFA: significant and high (>0.40) standardized coefficients of FG on the given factor, except for 1 FG; satisfactory goodness of fit indexes (GFI, AGFI, CFI, and NNFI values were 0.93, 0.91, 0.92, and 0.91, respectively, all >0.90, and RMSEA was equal to 0.004 < 0.005); **Reliability:** composite reliability of DPs: ranged from 0.56 to 0.73; test-retest reliability of DPs: ranged from 0.34 to 0.66; indicator reliability of individual FG included in CFA-based DPs: ranged from 0.07 to 0.46; test-retest reliability of individual FG included in CFA-based DPs: ranged from 0.11 to 0.50, with better reliability for market-based FG

			included in the CFA-based DPs with intra-class correlation coefficient (FFQ1 and FFQ2)	
Schulze, 2003 Germany EPIC-Potsdam (55)	**EFA**: on the learning sample with following re-analyses limiting the number of included FG until 8 FG; EIG>1, Scree test; No rotation; Descriptive labelling; **CFA**: Loading >=0.40 cut-off on EFA results using the sample study; CFA: 2-factor model with uncorrelated factors; GFI, RMSEA, CFI, NNFI; simplified scores	EFA: NA (2); CFA: No model selection	**Validity**: CFA	**Validity**: CFA: significant standardized loadings; acceptable goodness of fit indexes, except for borderline significance of NNFI (Goodness of Fit equal to 0.98 > 0.90; RMSEA equal to 0.07 <0.10; CFI equal to 0.93 > 0.90; NNFI equal to 0.90 = 0.90)

Togo, 2004 Denmark MONICA (29)	**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-year 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 technique to calculate factor scores	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-strcuture 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
Togo, 2003 Denmark MONICA (47)	**Separate PCFAs on FFQ and DR data (in octiles):** Separate analyses by sex; standardization with Kaiser normalization; Scree test, interpretability; Promax rotation; Loading >|0.30|cut-off; **Separate CFAs on FFQ and DR data**:	In Ms: 30.5 (3) with FFQ data and 26.2 (3) with DR data; in Fs: 23.8 (2) with FFQ data and 19.8 (2) with DR data	**Relative validity**: Pearson correlation coefficient between scores based on FFQ and DR data; **Validity:** CFA; Pearson correlation coefficient between	**Relative validity**: EFA on FFQ and DR data: The identified DPs were very similar, although the percentages of explained variance were lower on DR data; Pearson correlation coefficient between FFQ-based and DR-based scores ranged between 0.34 (TRADITIONAL DP among Ms) and 0.61 (both GREEN DPs, among Ms and Fs); CFA on FFQ and

	Loading >=|0.30|cut-off on EFA results; polychoric correlation matrix; RMSEA; weighted least square variable estimates with robust standard errors and mean- and variance-adjusted chi-squared test statistic		scores based on EFA and CFA	DR data: Pearson correlation coefficient between FFQ-based and DR-based scores ranged between 0.37 (TRADITIONAL DP among Ms) and 0.64 (GREEN DP among Fs); higher correlations with CFA than with EFA; **Validity** (EFA vs. CFA with the same dietary source): CFA-based DPs were similar across dietary sources and came from models with reassuring model fit (RMSEA < 0.10 no matter of the dietary source and among Ms and Fs); FFQ data: Pearson correlation coefficient between EFA-based and CFA-based scores ranged between 0.91 (TRADITIONAL DP among Ms) and 0.96 (SWEET-TRADITIONAL DP among Fs); DR data: Pearson correlation coefficient between EFA-based and CFA-based scores ranged between 0.82 (GREEN DP among Ms) and 0.94 (both GREEN and SWEET-TRADITIONAL DPs among Fs); higher correlations were found when using the same dietary data

Varraso, 2012 France and Spain EGEA2-France, Spanish PAC-COPD (57)	**PCA and CFA used as equivalent approaches on 1000 randomly selected samples from each of 4 different set-ups**: **1. 100% of EGEA2-France study;** **2. 50% of EGEA2-France study;** **3. 25% of EGEA2-France study;** **4. 100% of Spanish PAC-COPD study;** **PCA**: Scree-plot, interpretability; Varimax rotation; Distribution of the factor loading of FG to each DP represented via box-plot and median loading > |0.30| as cut-off; **CFA:** not based on previous EFA; 4 different models tested (3-factor and 2-factor models with correlated latent variables, 3-factor and 2-factor models with independent latent variables); chi-squared test, GFI, and RMSEA; Distribution of the factor loading of FG to each DP represented via box-plot and median loading > |0.30|as cut-off	PCA: NA (3); CFA: 3-factor model with no correlation among latent variables (highest GFI and lowest RMSEA)	**Reproducibility and Validity**: statistical properties (min, quartile 1, median, quartile 3, max) of the distribution of the factor loading of each FG to each DP in each of the 4 subsamples considered	**Reproducibility and validity**: Two consistent DPs were identified by CFA in each of the subsamples, whereas PCA led to less interpretable (smaller median of factor loadings and higher dispersion) DPs, especially for the smallest sample

Weismayer, 2006 Sweden SMC (28)	**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 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	**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 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

[a]ABBREVIATIONS: AGFI: adjusted goodness of fit index; CA: cluster analysis; CANHR: Center for Alaska Native Health Research study; CC: congruence coefficient; CFA: confirmatory factor analysis; CFI: comparative fit index; CI: confidence interval; DP: dietary pattern; DR: dietary record; E3N: Mutuelle Generale de l'Education Nationale (EPIC - France); EFA: exploratory factor analysis; EGEA2-France: Epidemiological Study on the Genetics and Environment of Asthma 2–France; EIG: Eigenvalue; EPIC-NL: European Prospective Investigation into Cancer and Nutrition-The Netherlands; EPIC-Potsdam: European Prospective Investigation into Cancer and Nutrition–Potsdam; F: female; FFQ: food frequency questionnaire; FFQ1/FFQ2/FFQ3: food frequency questionnaire at time 1, 2, or 3; FG: food group; GFI: goodness of fit index; M: male; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not available; NFI: normed fit index; NNFI: non-normed fit index or Tucker-Lewis index; PAC-COPD: Phenotype and Course of Chronic Obstructive Pulmonary Disease study–Spain; 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; RV: relative variation; SD: standard deviation; SMC: Swedish Mammography Cohort; SRMR: standardized root mean square residual

Figure legends

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 or studies. All of these settings may include confirmation of the identified dietary patterns with confirmatory factor analysis [Panel (E)][a]

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

Figure 2. Sankey diagram showing the selection process used in the systematic review on reproducibility and validity of dietary patterns[a]

[a]In the current review, we provided details on the 38 papers that assessed reproducibility, relative and construct validity of *a posteriori* dietary patterns

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[a]

Term	Definition	Statistical method
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 method(s) involved, which does not depend on the population in which measurements are made, unless bias or measurement precision varies with the true value being measured	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); Proportions of subjects classified into the same, adjacent, or opposite quantile category of score, or proportions of misclassified subjects; Kappa coefficient on score quantile categories or clusters
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, because differences between measurements of two subjects could be due purely to error rather than to a genuine difference in their true values	Intraclass correlation coefficient between scores; Test-retest reliability on scores or on dominant food groups defining the identified dietary patterns

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 (or rater), if human input is required, and they are made over a short period of time, over which the underlying value can be considered to be 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	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 being used, measurements being made by different observers or raters, or measurements being made over a period of time, within which the 'error-free' level of the variable could undergo non-negligible change	Pearson or Spearman or Kendall tau correlation coefficient between scores; Intra-class correlation coefficient between scores; Congruence coefficient between loadings
Validity	Does a test accurately measure what it claims to be measuring?	
Relative validity	Does a test compare well with a gold standard test?	Pearson or Spearman or Kendall tau correlation coefficient between scores [crude or corrected (de-attenuated) for accounting for variation in time]; Congruence coefficient between loadings
Construct validity	Does a test well measure the latent constructs that it is supposed to measure through operationalizations of the construct?	CFA

[a]ABBREVIATIONS: CFA: confirmatory factor analysis; LOA: limits of agreement

Supplemental Table 2. Reproducibility of *a posteriori* dietary patterns across statistical solutions: details on dietary pattern composition[a]

Reference	Dietary pattern composition
Bailey, 2006 USA (Pennsylvania) Geisinger Rural Aging Study (42)	**From CA on the number of servings:** CLUSTER 1: higher mean amounts of bread, sweet breads, dairy desserts, processed meats, eggs, and fats/oils; CLUSTER 2: higher mean amounts of most fruit/vegetable subgroups, fish, milk, and poultry **From CA on the percent daily energy contribution:** CLUSTER 1: higher mean amounts of pasta/noodles/rice, starchy vegetables, vegetable soups/sauces/juices, dairy desserts, cheese, most meat subgroups, and fats/oils; CLUSTER 2: higher mean amounts of sweet breads (e.g. cookies, muffins, and doughnuts), snacks, other fruit, fish, and sweets
Balder, 2003 Netherlands, Sweden, Finland, and Italy DIETSCAN (NLCS, SMC, ATBC, ORDET) (26)	**Dietary patterns 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, and 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**

Bountziouka, 2012 Greece NA (41)	**From unrotated PCA solution at both time-points**: WESTERN: high in white starchy products, eggs, potato, red meat, poultry, full-fat delicatessen, bakery, sweets and sodas; MEDITERRANEAN: high in low-fat dairy products, whole meal products, fish, legumes, fruit and vegetables; DRINKING: high in wine, beer, spirits, and stimulants; LIGHT PRODUCTS: high in low-fat dairy products, low-fat delicatessens, and light sodas **From orthogonal rotation (varimax and quartimax) solutions**: 3 DPs similar at the 2 time-points, except for the percentage of explained variances (WESTERN, HIGH-PROTEIN, and DRINKING DPs), but a LOW-CALORIE DP was found for FFQ2 and quartimax rotation. **From non-orthogonal rotation (promax and oblimin) solutions**: 3 DPs similar at the 2 time-points, except for the percentage of explained variances (UNFAVOURABLE, HEALTHY, and DRINKING DPs)
Castro, 2015 Brazil Healthy Survey of the City of Sao Paulo (50)	**From EFA and CFA, with different cut-offs for FG inclusion and with different rotation methods**: major differences in the first factor for EFA and |0.20| cut-off, but minimal with EFA (or CFA) and |0.25| cut-off: TRADITIONAL: high in typically consumed Brazilian foods like rice, beans, sugar, white breads, plus some additional FG in EFA with |0.20| cut-off (high in butter, margarine, beef and low in low fat milk); VEGETABLE-BASED DIET: high in salad dressings, leafy vegetables, non leafy vegetables, and spices, plus whole breads in CFA with oblimin rotation, or plus whole breads and white cheese, fruits and fruit juices in EFA with |0.20| cut-off
Dekker, 2013 Netherlands Doetinchem Cohort Study (25)	**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

Fransen, 2014 Netherlands EPIC-NL (51)	**From PCA/EFA**: 2-6 DPs possibly retained and to be confirmed with CFA 2-component solution: WESTERN: high in French fries, fast food, and soft drinks; PRUDENT: high in fish, vegetable, and high-fiber products; 3-component solution: PRUDENT DP was subdivided into 2 DPs; 4-component solution: WESTERN DP was subdivided into 2 DPs **From CA:** 2-6 DPs possibly retained: first 5 solutions which had 1 PRUDENT DP that included fish, high-fiber products, vegetables, and fruit (DP 2A, 3C, 4B, 5B, and 6E); WESTERN DP obtained for the 2-cluster solution (DP 2B) was subdivided into different clusters when more DPs retained
Greve, 2016 Germany IDEFICS (56)	**From application of all 3 CA methods, with some variation**: NON-PROCESSED: higher-than-average consumption of fruits, vegetables and wholemeal bread and lower-than-average consumption of refined cereals, sweet drinks and fast food; BALANCED: slightly higher-than-average consumption of sauces and butter, sweet drinks, meat and refined cereals, and slightly lower-than-average consumption of breakfast cereals, dairy products and fruits; JUNK FOOD: higher-than-average consumption of fast food, breakfast cereals, meat alternatives and dairy products, as well as sweet snacks for Gaussian mixture and k-means models only, and lower-than-average consumption of wholemeal bread, fruits and vegetables

Lau, 2008 Denmark Inter99 Study (48)	Subsample 1: PCA 1: for both Ms and Fs, with small differences: TRADITIONAL: high in paté or high-fat meat for sandwiches, mayonnaise salads, red meat, potatoes, butter and lard, low-fat fish, low-fat meat for sandwiches and sauces; MODERN: high in vegetables, fruit, mixed vegetables dishes, vegetable oil and vinegar dressing, poultry, and pasta, rice, and wheat kernels; Subsample 1: PCA 2: same DPs as PCA 1 (differences in factor loadings < 0.007); Subsample 2: PCA 3: same DPs as PCA 1, except for low-fat fish and margarine (differences in factor loadings < 0.15 except for low-fat fish and margarine); Subsample 2: PCA 4: same DPs as PCA 3; Subsample 1: CFA: same as PCA 1 (differences in factor loadings < 0.15)
Lo Siou, 2011 Canada Tomorrow Project (43)	**From CA among Ms**: DAIRY AND SWEETS: higher mean energy contributions from pasta/pizza, soda (regular), or chips, as well as from low-fat dairy and sweets; WESTERN: no comments; HEALTHY: higher mean energy contributions from fruits, poultry (no skin), vegetables (cooked, raw, tomatoes, cabbage, or legumes), and fish, and lower mean energy contributions from meat (processed or not), sweets, soda (regular), other bread, French fries, butter, margarine, or mayonnaise; WHOLEMEAL BREAD AND JAM: higher mean energy contributions from wholemeal bread, jam, cooked potatoes, margarine, or mayonnaise **From CA among Fs**: WESTERN AND SWEETS: no comments; HEALTHY: higher mean energy contributions from fruits, poultry (no skin), vegetables (cooked, raw, tomatoes, cabbage, or legumes), and fish, and lower mean energy contributions from meat (processed or not), sweets, soda (regular), other bread, French fries, butter, margarine, or mayonnaise; LOW FAT DAIRY AND BREAKFAST CEREAL: no comments

McCann, 2001 USA (New York) Western New York Diet Study (45)	**From PCAs on each of the 3 food classification methods:** HEALTHY: high in fruits and vegetables, poultry, fish, and whole grains; HIGH FAT: high in refined grains, fast foods, high-fat mixed dishes, and meats
Northstone, 2008 UK ALSPAC (36)	**From PCA on unadjusted data:** HEALTH-CONSCIOUS: high in salad, fresh fruit, rice, pasta, fish, pulses, and non-white bread; TRADITIONAL: high in all types of vegetables, some red meat and poultry; PROCESSED: high in meat pies, sausage and burgers, fried foods, pizza, and chips; CONFECTIONERY: high in chocolate, sweets, biscuits, cakes, and other puddings; VEGETARIAN: high in meat substitutes, pulses, nuts, and herbal tea **From PCA on energy-adjusted data:** PROCESSED DP lost, but the other ones were present and in the same order, with slight differences in factor loadings for FG that used to load on the PROCESSED DP and now loaded negatively on the HEALTH-CONSCIOUS DP
Sauvageot, 2017 Luxembourg, Belgium, and France NESCaV (44)	**From final CA solution:** PRUDENT: higher mean residual intake of brown bread, fruits, oleaginous fruits, dried fruits, soups, vegetables, pulses, preserved vegetables, offal, fish, smoked and canned fish, shellfish and mussels, dairy products, soya products, olive oil, oil-rich in omega 3 or 6, water and tea and lower mean residual intake of white bread, pastries, rice and pasta, fried foods, lean and fatty meat, processed smoked meat, processed meat, ready meals, minarine and margarine, fresh cream and dressing, sugar and sweets, salty biscuits, soft drinks, diet soft drinks, beer and aperitifs, and spirits; NON-PRUDENT: higher mean residual intake of white bread, potatoes, fried foods, lean and fatty meat, offal, processed meat, shellfish and mussels, minarine and margarine, fresh cream and dressings, coffee, diet soft drinks, beer and wine, and lower mean residual intake of cereals, rice/pasta, fruits, oleaginous fruits, dried fruits,

	vegetables, pulses, preserved vegetables, fish, smoked and canned fish, dairy products, soya products, olive oil and oil-rich in omega 3 or 6, light fresh cream and dressings, sugar and sweets, water, fruit or vegetable juice and tea; CONVENIENT: higher mean residual intake of cereals, pastries, rice and pasta, preserved vegetables, smoked and canned fish, ready meals, high-fat dairy products, soya products, fresh cream and dressings, sugar and sweets, salty biscuits, fruit or vegetable juice, soft drinks and aperitifs, and spirits, and lower mean residual intake of brown bread, potatoes, oleaginous fruits, soups, vegetables, pulses, offal, fish, shellfish and mussels, oil-rich in omega 3, coffee and wine
Varraso, 2012 France and Spain EGEA2-France, Spanish PAC-COPD (57)	**100% of EGEA2-France study**: PCA: PRUDENT: high in vegetables, fruit, oil, legumes, and fish; WESTERN: high in prepared meals, French fries, processed meats, sandwiches, snack, soda, pods and peas, cakes, condiments, high-fat dairy products, and potatoes; ALCOHOL AND WINE: high in alcoholic beverages, and low in low-fat dairy products; CFA: PRUDENT: high in vegetables, fruit, oils, whole-grain cereals, and fish; WESTERN: high in prepared meals, French fries, processed meats, condiments, alcohol, beer/cider, sandwiches, potatoes, pods and peas, snack, soda, cakes, red meats, high-fat dairy products, nuts and seeds, offal, shellfish, sorbet, high-fat dairy products, coffee, fruit juice, refined cereals, butter, chocolate, and red wine **50% of EGEA2-France study:** PCA: VEGETABLES, OIL, AND FISH: high in vegetables, oil, and fish; WESTERN: high in prepared meals, French fries, processed meats, sandwiches, snack, soda, cakes, pods and peas, beer, condiments, high-fat dairy products, and fruit juice; ALCOHOL: high in alcoholic beverages, shellfish, and coffee; FRUIT: high in fruit; CFA: PRUDENT: high in vegetables, fruit, oils, whole-grain cereals, and fish; WESTERN: high in prepared meals, French fries, processed meats, condiments, alcohol, sandwiches, potatoes,

	pods and peas, snack, soda, cakes, beer/cider, high-fat dairy products, red meats, sorbet, nuts and seeds, offal, shellfish, coffee, fruit juice, refined cereals, butter, chocolate, and red wine **25% of EGEA2-France study:** PCA: VEGETABLES, OIL, AND FRUIT: high in vegetables, oil, and fruit; WESTERN: high in prepared meals, French fries, processed meats, sandwiches, soda, snack, cakes, beer/cider, pods and peas, and condiments; ALCOHOL: high in alcoholic beverages; CFA: PRUDENT: high in vegetables, fruit, oils, whole-grain cereals, and fish; WESTERN: high in prepared meals, French fries, processed meats, condiments, alcohol, sandwiches, potatoes, legumes, poultry, pods and peas, snack, soda, cakes, beer/cider, high-fat dairy products, red meats, sorbet, nuts and seeds, offal, shellfish, coffee, fruit juice, egg, refined cereals, butter, chocolate, and red wine **100% of Spanish PAC-COPD study:** PCA: VEGETABLES AND MEATS: high in other oils, fruity vegetables, red meats, offal, cured meats, and potatoes; LEAFY VEGETABLES AND LOW-FAT DAIRY: high in leafy vegetables and low-fat dairy products. CFA: PRUDENT: high in fruity vegetables, other vegetables, blue fish, leafy vegetables, white fish, other oil, red meats, pods and peas, and dark-yellow vegetables; WESTERN: high in high-fat dairy products, chocolate, potatoes, soda, snack, nuts and seeds, butter, and refined cereal and low in low-fat dairy products and citrus

Wirfalt, 2000 Sweden MDC (46)	**From CA on unstandardized variables:** MANY FOODS AND DRINKS: highest mean consumption of cheese, and high-fat meats; FIBRE BREAD: highest mean consumption of fibre bread, and high-fat meats; LOW FAT AND HIGH FIBRE: highest mean consumption of fruits, and low-fat milk; WHITE BREAD: highest mean consumption of white bread, high-fat meats, sweets, low-fat spread, and low-fat meats; MILK FAT: highest mean consumption of Bregott spread, sweets, white bread, and high-fat meats; SWEETS AND CAKES: highest mean consumption of sweets, and high-fat meat **From CA on z-scored variables:** DRINKS AND FRIES: highest mean consumption of low-fat dressing, liquor, fried potatoes, and wine; ICE-CREAM AND CAKE: highest mean consumption of ice-cream, chocolates, and sherbet; DIETERS: highest mean consumption of sherbet, cottage cheese, fruit, high-fat fish, coffee, low-fat milk, miscellaneous, vegetables, fibre crisp-bread, and low-fat spread; HEALTHY: highest mean consumption of cottage cheese, low-fat milk, low-fat spread, crackers, fibre bread, fruit, fibre crisp-bread, miscellaneous, low-fat cake, and boiled potatoes; TRADITIONAL: highest mean consumption of white bread, sweets, Bregott spread, and whole milk; MEDITERRANEAN: highest mean consumption of wine, oil, vegetables, rice/pasta, low-fat fish, fruit, low-fat meats, egg, dressing, fibre crisp-bread, high-fat fish, nuts, tea, and cheese

[a]ABBREVIATIONS: ALSPAC: Avon Longitudinal Study of Parents and Children; ATBC: Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; CA: cluster analysis; CFA: confirmatory factor analysis; DIETSCAN: DIETary patternS and CANcer in four European countries project; DP: dietary pattern; EFA: exploratory factor analysis; EGEA2-France: Epidemiological Study on the Genetics and Environment of Asthma 2–France; EPIC-NL: European Prospective Investigation into Cancer and Nutrition-The Netherlands; F: female; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1, 2, or 3; FG: food groups; IDEFICS: Identification and Prevention of Dietary and Lifestyle-induced Health Effects in Children and Infants; M: male; MDC: Malmo Diet and Cancer study; NA: not available; NESCaV: Nutrition, Environment and Cardiovascular Health; NLCS: Netherlands Cohort Study on diet and cancer; ORDET: Ormoni e Dieta nella Eziologia dei Tumori in Italy; PAC-

COPD: Phenotype and Course of Chronic Obstructive Pulmonary Disease study–Spain; PCA: principal component analysis; SMC: Swedish Mammography Cohort

Supplemental Table 3. Reproducibility and/or relative validity of *a posteriori* dietary patterns: details on dietary pattern composition[a]

Reference	Dietary pattern composition
Ambrosini, 2011 Australia Western Australian Pregnancy Cohort (Raine) Study (13)	**From EFAs on FFQ and DR data**: HEALTHY: high in several vegetable types, fresh fruit, fish (steamed, grilled, or canned), whole grains, low-fat dairy products, and mineral water; WESTERN: high in takeaway foods, confectionery, soft drinks, crisps, fried potato chips, soft drinks; **plus extra DPs not shared among the 2 dietary sources data and not described in detail (small percentages of explained variance, few foods loading highly on them)**
Asghari, 2012 Iran TLGS (11)	IRANIAN TRADITIONAL (**common to all 4 dietary sources**): 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 sources**): 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
Beck, 2012 New Zealand NA (30)	**From PCFAs on FFQ1, FFQ2, and DR data:** HEALTHY: high in tomatoes, lettuce, capsicum, broccoli, carrots, onions, apples, almonds, yogurt, brown bread, crackers, porridge, herbal tea, and water; SANDWICH AND DRINKS: high in brown bread, butter, cheese, beef, coffee, black tea, and milk added to drinks

Bountziouka, 2011 Greece NA (40)	**From PCA: on both FFQ and DR:** WESTERN: high in full-fat dairy products, refined grains, potatoes, red meat, full-fat delicatessen, and bakery products; MEDITERRANEAN: high in low-fat dairy, whole-wheat grains, fish and seafood, vegetables, and fruit **From PCA: on FFQ data only:** LOW-FAT PRODUCTS: high in low-fat delicatessen, bakery, light sodas, full-fat delicatessen, whole-grains, and red meat; DRINKING: high in wine, beer, spirits, refined grains, and stimulants **From PCA: on DR data only:** SWEETS: high in wholegrains, sweets, and low-fat dairy products, and low in poultry, wine, fish and seafood, and potatoes; STIMULANTS: high in legumes and stimulants, and low in low-fat delicatessen and eggs **From CA:** UNHEALTHY: high in full-fat dairy products, refined grains, potatoes, and red meat; HEALTHY: high in low-fat dairy products, whole-wheat grains, fish and seafood, vegetables, and fruit
Crozier, 2008 UK NA (39)	**From PCFA on FFQ data:** PRUDENT: high in fruit and vegetables, wholemeal bread, rice and pasta, yogurt, cheese, fish and reduced-fat milk, and low in white bread, added sugar, tinned vegetables, full-fat milk and crisps; WESTERN: high in red and processed meat, cakes and biscuits, puddings, Yorkshire puddings and savory pancakes, chips, roast and boiled potatoes, sugar, sweets and chocolate, and low in reduced-fat milk **From PCFA on DR data:** PRUDENT: high in wholemeal bread, fruit and vegetables, cheese, yogurt and reduced-fat milk, and low in chips and roast potatoes, white bread and tinned vegetables; WESTERN: high in full-fat spread, cooking fats and salad oils, full-fat milk, sweets and chocolate, white bread, crisps, tea and coffee, chips and roast potatoes, Yorkshire puddings and savory pancakes, and low in

	reduced-fat spread, reduced-fat milk, wholemeal bread, decaffeinated tea and coffee; **plus extra DPs not shared among the 2 dietary sources data and not described in detail (few foods loading highly on them)**
Hong, 2016 China NA (34)	**From EFAs on FFQ1, FFQ2, and m24HR data:** ANIMAL AND PLANT PROTEIN: high in poultry meats, fish and shrimp, bean curd, livestock meats, dry bean and other soy bean products; NUTS AND SWEETS: high in nuts, sweets and desserts, and snacks; CHINESE TRADITIONAL: high in other grains and products, potatoes, fresh vegetables, fried food, high-fat dairy products, wheat and products, rice and products, and pickled vegetables; BEVERAGES AND ALCOHOL DP: high in sodas, juice, beer, wine, processed meats and liquor; **plus extra DPs less interpretable and highly variable and not described in detail**
Hu, 1999 USA (Massachusetts) HPFS (9)	**From PCFAs on FFQ1, FFQ2, and mDR data:** PRUDENT: high in vegetables, legumes, wholegrains, fruit, oil and vinegar salad dressing, and fish and other seafood; WESTERN: high in processed meat, red meat, butter, high-fat dairy products, refined grains, eggs, and French fries; **plus extra DPs not shared among all available dietary sources data and not described in detail (small amount of total variance explained)**
Khani, 2004 Sweden SMC (33)	**From PCFAs on FFQ1 and FFQ2 within the reproducibility sample, and on FFQ and mDRs within the validity sample:** HEALTHY: high in vegetables, fruit, fish, poultry, tomato, whole grain, cereal and low-fat dairy products; WESTERN: high in processed meat, meat, refined grains, sweets, margarine, high-fat dairy products, potatoes, and soda; DRINKER: high in beer, wine, liquor, and snacks;

	plus extra DPs not shared among the compared dietary sources data and not described in detail (<7% total variance explained for each of them)
Liu, 2015 China NA (32)	**From PCFAs on FFQ1, FFQ2, and m24HR data:** PRUDENT: high in rice, wheat, total fruits, fresh vegetables, bean products, white meat, red meat, nuts and fresh eggs; PROCESSED FOOD: high in pickled vegetables, preserved vegetables, salted meat, and salted eggs; **plus extra DPs not shared among all available dietary sources data and not described in detail (less interpretable and highly variable)**
Loy, 2013 Malaysia USM Birth Cohort Study (37)	**From PCAs on FFQ and m24HR data**: HEALTHY: high in fish and other seafood, fruit, dairy products, vegetables, nuts and legumes; LESS HEALTHY: high in confectioneries, condiments, oils and fats, tea and coffee, cereals, meat and offal; **plus extra DPs not shared among the 2 dietary sources data and not described in detail (small percentages of explained variance, few foods loading highly on them)**
McNaughton, 2005 UK Medical Research Council National Survey of Health and Development (1946 British Birth Cohort) (35)	**From PCFAs across the 3 dietary data in both Ms and Fs**: with some variation on HEALTH-AWARE and SANDWICH: HEALTH-AWARE: high in high-fibre breakfast cereals, wholemeal breads, apples, and bananas; DINNER PARTY: high in coffee, white wine, and cream; TRADITIONAL: high in potatoes, green vegetables, carrots, red meat, and peas; REFINED GRAINS: high in sugar, butter, white bread (for Fs only), and whole milk; SANDWICH: high in tomatoes, lettuce, and onions
Nanri, 2012 Japan JPHC (31)	**From PCAs on FFQ_R, FFQ_V, and mDR data**: PRUDENT JAPANESE: high in vegetables, fruit, potatoes, soy products, mushrooms, seaweed, oily fish, and green tea; WESTERNIZED JAPANESE: high in bread, meat, processed meat, fruit juice, coffee, black tea, soft drinks,

	sauces, mayonnaise and dressing; TRADITIONAL JAPANESE: high in rice, miso soup, pickles, salmon, salty fish, seafood other than fish, fruit and sake (Ms only)
Okubo, 2010 Japan NA (38)	**From PCFAs on DHQ1, mDHQ, and mDR data, among Fs**: HEALTHY: high in green and yellow vegetables, fish, fruits, mushrooms, white vegetables, sea products, seaweeds, pickled vegetables, shellfish, and pulses, and low in beef and pork; WESTERN: high in vegetable oil, processed meat, butter, and eggs; JAPANESE TRADITIONAL: high in miso soup, rice, and low in shellfish and bread **From PCFAs on DHQ1, mDHQ, and mDR data, among Ms:** HEALTHY: high in green and yellow vegetables, fruits, mushrooms, white vegetables, seaweeds, daily products, sugar, miso soup, and pulses; WESTERN: high in chicken, vegetable oil, processed meat, and beef and pork, and low in rice
Ryman, 2015 USA (Southwest Alaska) CANHR (54)	**From final CFA solution**: PROCESSED FOODS: high in salty snacks, sweetened cereals, pizza, sweetened drinks, hot dogs and lunch meat, fried chicken, and canned tuna; FRUITS AND VEGETABLES: high in fresh citrus, potato salad, citrus juice, corn, green beans, green salad, and market berries in akutaq; SUBSISTENCE FOODS: high in seal or walrus soup, non-oil fish, wild greens, and bird soup
Togo, 2003 Denmark MONICA (47)	**From PCFA on FFQ data among Ms, but similar with PCFA on DR data and CFA on both datasets**: GREEN: high in wheat bread, and rye bread with wholegrain and/or bran, raw and boiled vegetables, and fruit; SWEET: high in cake, biscuits and baked goods, candy or chocolate, soft drink or ice-cream, jam, and marmalade or honey; TRADITIONAL: high in meat, paté, meat for bread, potatoes, butter, lard and hard margarine; **From PCFA on FFQ data among Fs, but similar with PCFA on DR data and CFA on both datasets:** GREEN: Same as for Ms

	SWEET TRADITIONAL: high in cake, biscuits and baked foods, candy or chocolate, paté and meat for bread, white and wheat, butter, lard and hard margarine

[a]ABBREVIATIONS: CA: cluster analysis; CANHR: Center for Alaska Native Health Research study; CFA: confirmatory factor analysis; DHQ1/DHQ2/DHQ3: diet history questionnaire at time 1, 2, or 3; DP: dietary pattern; DR: dietary record; EFA: exploratory factor analysis; F: female; FFQ: food-frequency questionnaire; FFQ_R: food-frequency questionnaire from the reproducibility study; FFQ_V: food-frequency questionnaire from the relative validity study; FFQ1/FFQ2/FFQ3: food-frequency questionnaire at time 1, 2, or 3; HPFS: Health Professionals Follow-up Study; JPHC: Japan Public Health Center-Based Prospective study; M: male; m24HRs: mean 24 hours recall; mDHQ: mean diet history questionnaire; mDR: mean dietary record; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not available; PCA: principal component analysis; PCFA: principal component factor analysis; SMC: Swedish Mammography Cohort; TLGS: Teheran Lipid and Glucose Study; USM: Universiti Sains Malaysia

Supplemental Table 4. Construct validity of *a posteriori* dietary patterns: details on dietary pattern composition[a]

Reference	Dietary pattern composition
Bedard, 2015 France E3N (EPIC-France) (49)	**From PCA and CFA:** PRUDENT: high in vegetables, condiments, sauces, fish, fresh diary products, fruit, olive oil; WESTERN: high in rice/pasta/grain, potatoes, processed meat, red meat and offal, bread, fats except olive oil and sunflower oils, dough and pastry; APERITIF: high in crackers, nuts and seeds, alcoholic beverages, canned fish, seaweed, eggs
Castro, 2015 Brazil Healthy Survey of the City of Sao Paulo (50)	**From EFA and CFA**, with different cut-offs for FG inclusion and with different rotation methods: major differences in the first factor for EFA and |0.20| cut-off, but minimal with EFA (or CFA) and |0.25| cut-off: TRADITIONAL: high in typically consumed Brazilian foods like rice, beans, sugar, white breads, plus some additional FG in EFA with |0.20| cut-off (high in butter, margarine, beef and low in low fat milk); VEGETABLE-BASED DIET: high in salad dressings, leafy vegetables, non leafy vegetables, and spices, plus whole breads in CFA with oblimin rotation, or plus whole breads and white cheese, fruits and fruit juices in EFA with |0.20| cut-off
Fransen, 2014 Netherlands EPIC-NL (51)	**From PCA/EFA**: 2-6 DPs possibly retained and to be confirmed with CFA 2-component solution: WESTERN: high in French fries, fast food, and soft drinks; PRUDENT: high in fish, vegetable, and high-fiber products; 3-component solution: PRUDENT DP was subdivided into 2 DPs; 4-component solution: WESTERN DP was subdivided into 2 DPs; **From CA**: 2-6 DPs possibly retained: first 5 solutions which had 1 PRUDENT DP that included fish, high-fiber products, vegetables, and fruit (DP 2A, 3C, 4B, 5B, and 6E);

	WESTERN DP obtained for the 2-cluster solution (DP 2B) was subdivided into different clusters when more DPs retained
Judd, 2014 USA REGARDS (24)	**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
Lau, 2008 Denmark Inter99 Study (48)	Subsample 1: PCA 1: for both Ms and Fs, with small differences: TRADITIONAL: high in paté or high-fat meat for sandwiches, mayonnaise salads, red meat, potatoes, butter and lard, low-fat fish, low-fat meat for sandwiches and sauces; MODERN: high in vegetables, fruit, mixed vegetables dishes, vegetable oil and vinegar dressing, poultry, and pasta, rice, and wheat kernels; Subsample 1: PCA 2: same DPs as PCA 1 (differences in factor loadings < 0.007); Subsample 2: PCA 3: same DPs as PCA 1, except for low-fat fish and margarine (differences in factor loadings < 0.15 except for low-fat fish and margarine); Subsample 2: PCA 4: same DPs as PCA 3; Subsample 1: CFA: same as PCA 1 (differences in factor loadings < 0.15)
Maskarinec, 2000 USA (Hawaii) NA (52)	**From final CFA solution**: MEAT: high in processed and red meats, fish, poultry, eggs, fats and oils, and condiment; VEGETABLES: high in different vegetables (dark yellow, green leaf and other vegetables); BEAN: high in legumes, tofu and soy products; COLD FOODS: high in fruit, fruit juice and cold breakfast cereals

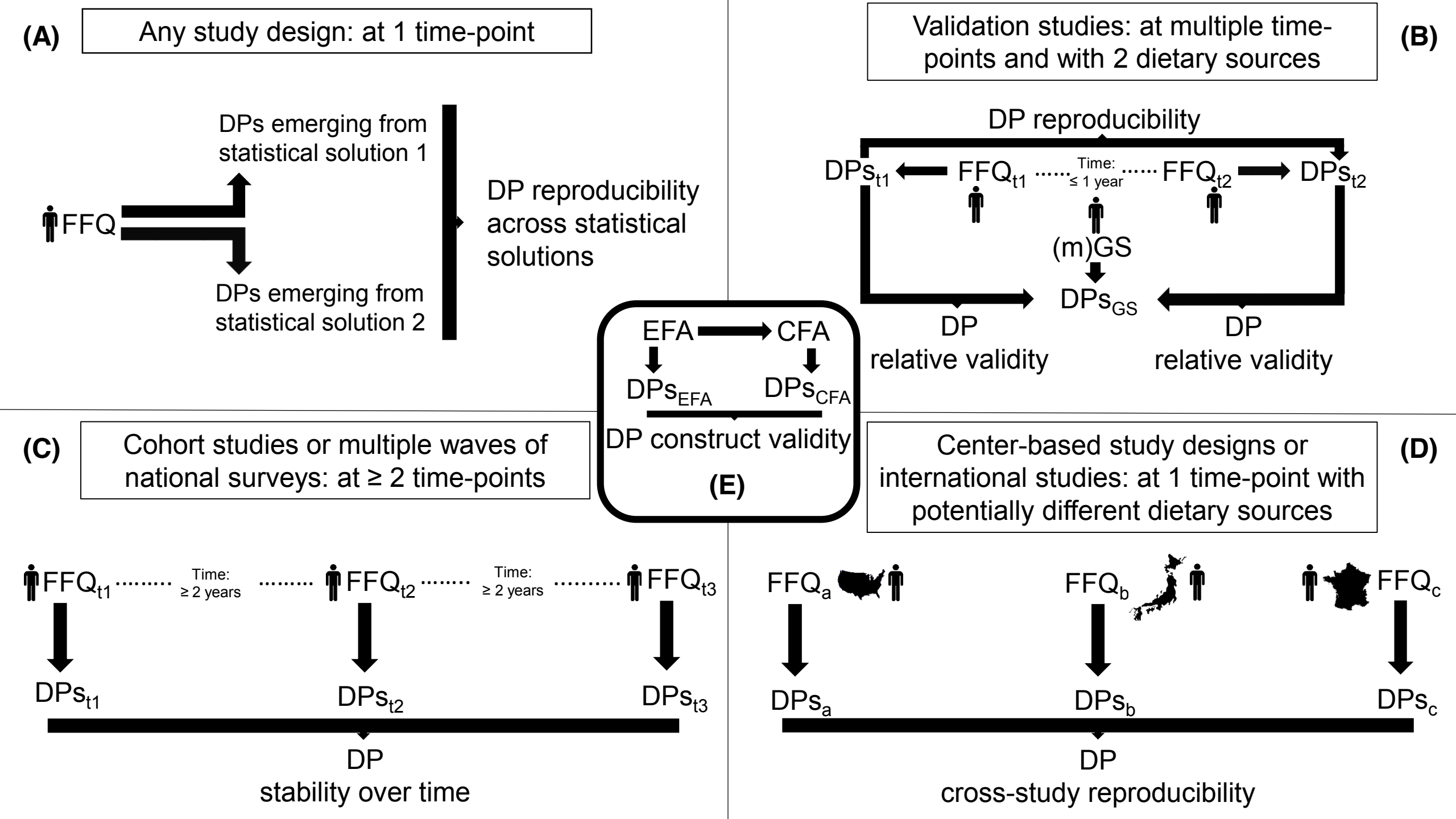
Newby, 2006 Sweden SMC (10)	**From PCFA at both time-points (1987 and 1997) and confirmed with CFA at both time-points (1987 and 1997)**: HEALTHY: high in vegetables, fruit, whole grains, fruit juice, and cereal; WESTERN/SWEDISH: high in meat, processed meat, liver, refined grains, and potatoes; ALCOHOL: 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 FFQ data**
Newby, 2006 Sweden SMC (27)	**From PCFA at both time-points (1987 and 1997) and confirmed with CFA at both time-points (1987 and 1997):** with some variation 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 FFQ data**
Park, 2005 USA (Hawaii and Los Angeles) Hawaii - Los Angeles Multiethnic Cohort Study (53)	**From final PCFA solution on the overall sample:** FAT AND MEAT: high in discretionary fat, meat, eggs, and cheese; VEGETABLES: high in dark-green, deep yellow and other vegetables; FRUIT AND MILK: high in milk and yogurt, and fruit groups
Ryman, 2015 USA (Southwest	**From final CFA solution**: PROCESSED FOODS: high in salty snacks, sweetened cereals, pizza, sweetened drinks, hot dogs and lunch meat, fried chicken, and canned tuna;

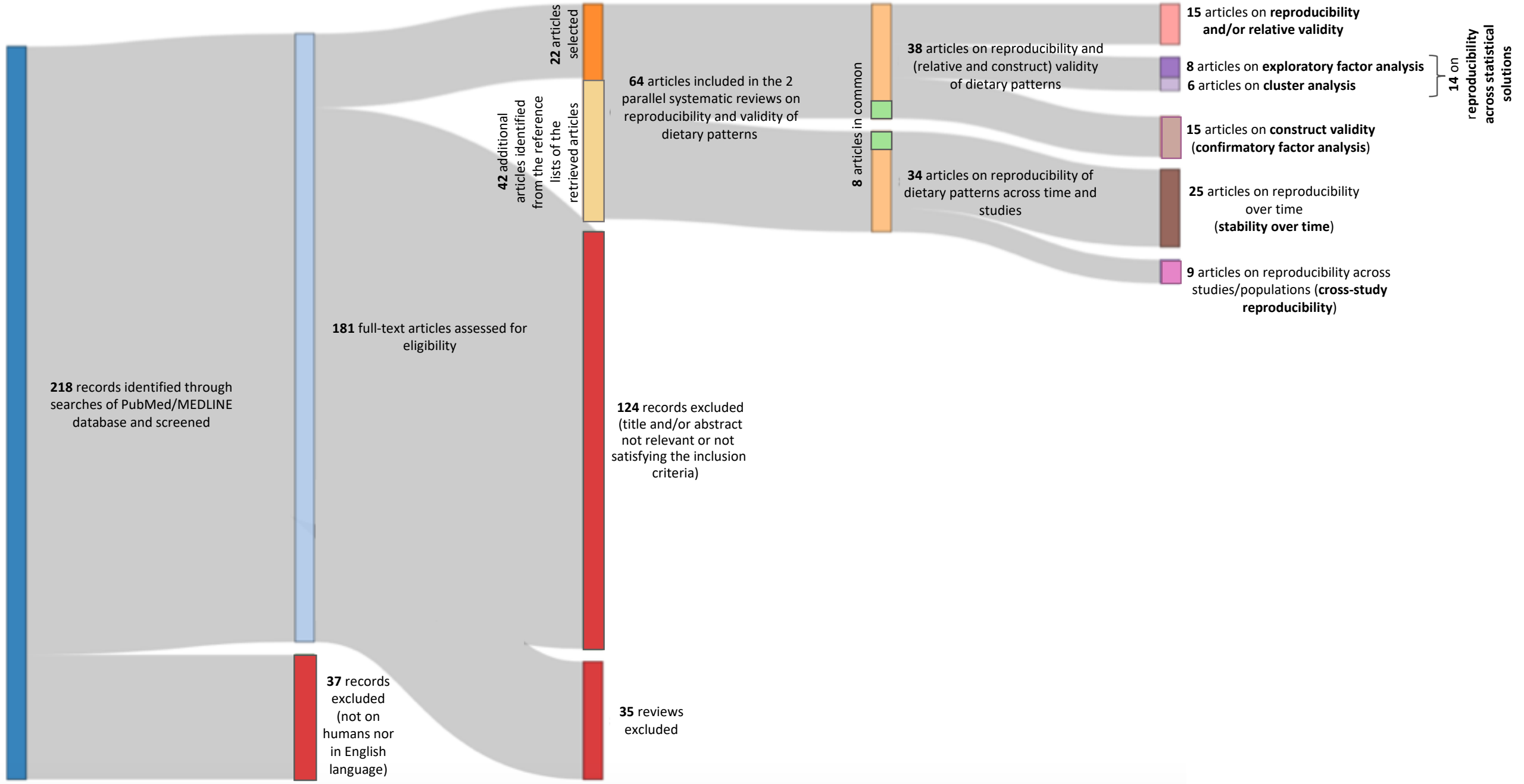
Alaska) CANHR (54)	FRUITS AND VEGETABLES: high in fresh citrus, potato salad, citrus juice, corn, green beans, green salad, and market berries in akutaq; SUBSISTENCE FOODS: high in seal or walrus soup, non-oil fish, wild greens, and bird soup
Schulze, 2003 Germany EPIC-Potsdam (55)	**From EFA on the learning sample:** TRADITIONAL COOKING: high in meat, sauce, poultry, potatoes, and cooked vegetables; FRUITS AND VEGETABLES: high in fruits, raw vegetables, and vegetable oils
Togo, 2004 Denmark MONICA (29)	**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
Togo, 2003 Denmark MONICA (47)	**From PCFA on FFQ data among Ms, but similar with PCFA on DR data and CFA on both datasets**: GREEN: high in wheat bread, and rye bread with wholegrain and/or bran, raw and boiled vegetables, and fruit; SWEET: high in cake, biscuits and baked goods, candy or chocolate, soft drink or ice-cream, jam, and marmalade or honey; TRADITIONAL: high in meat, paté, meat for bread, potatoes, butter, lard and hard margarine; **From PCFA on FFQ data among Fs, but similar with PCFA on DR data and CFA on both datasets:**

	GREEN: Same as for Ms SWEET TRADITIONAL: high in cake, biscuits and baked foods, candy or chocolate, paté and meat for bread, white and wheat, butter, lard and hard margarine
Varraso, 2012 France and Spain EGEA2-France, Spanish PAC-COPD (57)	**100% of EGEA2-France study:** PCA: PRUDENT: high in vegetables, fruit, oil, legumes, and fish; WESTERN: high in prepared meals, French fries, processed meats, sandwiches, snack, soda, pods and peas, cakes, condiments, high-fat dairy products, and potatoes; ALCOHOL AND WINE: high in alcoholic beverages, and low in low-fat dairy products; CFA: PRUDENT: high in vegetables, fruit, oils, whole-grain cereals, and fish; WESTERN: high in prepared meals, French fries, processed meats, condiments, alcohol, beer/cider, sandwiches, potatoes, pods and peas, snack, soda, cakes, red meats, high-fat dairy products, nuts and seeds, offal, shellfish, sorbet, high-fat dairy products, coffee, fruit juice, refined cereals, butter, chocolate, and red wine **50% of EGEA2-France study:** PCA: VEGETABLES, OIL, AND FISH: high in vegetables, oil, and fish; WESTERN: high in prepared meals, French fries, processed meats, sandwiches, snack, soda, cakes, pods and peas, beer, condiments, high-fat dairy products, and fruit juice; ALCOHOL: high in alcoholic beverages, shellfish, and coffee; FRUIT: high in fruit; CFA: PRUDENT: high in vegetables, fruit, oils, whole-grain cereals, and fish;

	WESTERN: high in prepared meals, French fries, processed meats, condiments, alcohol, sandwiches, potatoes, pods and peas, snack, soda, cakes, beer/cider, high-fat dairy products, red meats, sorbet, nuts and seeds, offal, shellfish, coffee, fruit juice, refined cereals, butter, chocolate, and red wine **25% of EGEA2-France study:** PCA: VEGETABLES, OIL, AND FRUIT: high in vegetables, oil, and fruit; WESTERN: high in prepared meals, French fries, processed meats, sandwiches, soda, snack, cakes, beer/cider, pods and peas, and condiments; ALCOHOL: high in alcoholic beverages; CFA: PRUDENT: high in vegetables, fruit, oils, whole-grain cereals, and fish; WESTERN: high in prepared meals, French fries, processed meats, condiments, alcohol, sandwiches, potatoes, legumes, poultry, pods and peas, snack, soda, cakes, beer/cider, high-fat dairy products, red meats, sorbet, nuts and seeds, offal, shellfish, coffee, fruit juice, egg, refined cereals, butter, chocolate, and red wine **100% of Spanish PAC-COPD study:** PCA: VEGETABLES AND MEATS: high in other oils, fruity vegetables, red meats, offal, cured meats, and potatoes; LEAFY VEGETABLES AND LOW-FAT DAIRY: high in leafy vegetables and low-fat dairy products. CFA: PRUDENT: high in fruity vegetables, other vegetables, blue fish, leafy vegetables, white fish, other oil, red meats, pods and peas, and dark-yellow vegetables; WESTERN: high in high-fat dairy products, chocolate, potatoes, soda, snack, nuts and seeds, butter, and refined cereal and low in low-fat dairy products and citrus
Weismayer, 2006 Sweden SMC (28)	**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**

[a]ABBREVIATIONS: CA: cluster analysis; CANHR: Center for Alaska Native Health Research study; CFA: confirmatory factor analysis; DP: dietary pattern; DR: dietary record; E3N: Mutuelle Generale de l'Education Nationale (EPIC - France); EFA: exploratory factor analysis; EGEA2-France: Epidemiological Study on the Genetics and Environment of Asthma 2–France; EPIC-NL: European Prospective Investigation into Cancer and Nutrition-The Netherlands; EPIC-Potsdam: European Prospective Investigation into Cancer and Nutrition-Potsdam; F: female; FFQ: food frequency questionnaire; FG: food group; M: male; MONICA: MONItoring of trends and determinants in CArdiovascular Disease; NA: not available; PAC-COPD: Phenotype and Course of Chronic Obstructive Pulmonary Disease study–Spain; PCA: principal component analysis; PCFA: principal component factor analysis; REGARDS: Reasons for Geographic and Racial Differences in Stroke; SMC: Swedish Mammograpy Cohort


(A)
Any study design: at 1 time-point
DPs emerging from statistical solution 1
FFQ
DPs emerging from statistical solution 2
DP reproducibility across statistical solutions
Validation studies: at multiple time-points and with 2 dietary sources
(B)
DP reproducibility
DPs$_{t1}$
FFQ$_{t1}$
Time: ≤ 1 year
FFQ$_{t2}$
DPs$_{t2}$
(m)GS
DPs$_{GS}$
DP relative validity
DP relative validity
EFA
CFA
DPs$_{EFA}$
DPs$_{CFA}$
DP construct validity
(E)
(C)
Cohort studies or multiple waves of national surveys: at ≥ 2 time-points
FFQ$_{t1}$
Time: ≥ 2 years
FFQ$_{t2}$
Time: ≥ 2 years
FFQ$_{t3}$
DPs$_{t1}$
DPs$_{t2}$
DPs$_{t3}$
DP stability over time
Center-based study designs or international studies: at 1 time-point with potentially different dietary sources
(D)
FFQ$_a$
FFQ$_b$
FFQ$_c$
DPs$_a$
DPs$_b$
DPs$_c$
DP cross-study reproducibility


218 records identified through searches of PubMed/MEDLINE database and screened
181 full-text articles assessed for eligibility
37 records excluded (not on humans nor in English language)
22 articles selected
42 additional articles identified from the reference lists of the retrieved articles
64 articles included in the 2 parallel systematic reviews on reproducibility and validity of dietary patterns
124 records excluded (title and/or abstract not relevant or not satisfying the inclusion criteria)
35 reviews excluded
8 articles in common
38 articles on reproducibility and (relative and construct) validity of dietary patterns
34 articles on reproducibility of dietary patterns across time and studies
15 articles on **reproducibility and/or relative validity**
8 articles on **exploratory factor analysis**
6 articles on **cluster analysis**
14 on **reproducibility across statistical solutions**
15 articles on **construct validity** (**confirmatory factor analysis**)
25 articles on reproducibility over time (**stability over time**)
9 articles on reproducibility across studies/populations (**cross-study reproducibility**)