

# Reproducibility of A Posteriori Dietary Patterns across Time and Studies: A Scoping Review

Valeria Edefonti,<sup>1</sup> Roberta De Vito,<sup>2,3</sup> Andrea Salvatori,<sup>1</sup> Francesca Bravi,<sup>1</sup> Linia Patel,<sup>1</sup> Michela Dalmartello,<sup>1</sup> and Monica Ferraroni<sup>1</sup>

<sup>1</sup>Branch of Medical Statistics, Biometry and Epidemiology “G. A. Maccacaro,” Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy; <sup>2</sup>Department of Biostatistics, Brown University, Providence, RI, USA; and <sup>3</sup>Data Science Initiative, Brown University, Providence, RI, USA

## ABSTRACT

Few studies have considered if a posteriori dietary patterns (DPs) are generalizable across different centers or studies, or if they are consistently seen over time. To date, no systematic search of the literature on these topics has been carried out. A scoping review was conducted through a systematic search on the PubMed database. In the current review, we included the 34 articles examining the extent to which a posteriori DPs were consistently seen: 1) across centers from the same study or across different studies potentially representing different populations or countries (here indicated as cross-study reproducibility) and 2) over longer time periods (i.e.,  $\geq 2$  y) (here indicated as stability over time). Selected articles (published in 1981–2019, 32% from 2010 onwards) were based on observational studies, mostly from Europe and North America. Five articles were based on children and/or adolescents and 14 articles included adults (2 men; 12 women, of whom 3 were pregnant women). A posteriori DPs were mostly derived (32 articles) with principal components or factor analyses. Among the 9 articles assessing DP reproducibility across studies (number of centers/studies: 2–27; median: 3), 5 provided a formal assessment using statistical methods (4 index-based approaches of different complexity, 1 statistical model). A median of 4 DPs was reproduced across centers/studies (range: 1–7). Among the 25 articles assessing DP stability over time (number of time-occasions: 2–6; median: 3), 19 provided a formal assessment with statistical methods (17 index-based and/or test-based approaches, 1 statistical model, 1 with both strategies). The number and composition of DPs remained mostly stable over time. Based on the limited evidence collected, most identified DPs showed good reproducibility across studies and stability over time. However, when present within the single studies, the criteria for the formal assessment of cross-study reproducibility or stability over time were generally very basic. *Adv Nutr* 2020;11:1255–1281.

**Keywords:** a posteriori dietary patterns, cluster analysis, consistency of dietary patterns, cross-study reproducibility of dietary patterns, factor analysis, generalizability of dietary patterns, reproducibility of dietary patterns, reproducibility of dietary patterns across studies, stability of dietary patterns over time

## Introduction

Over the last 20 y, the analysis of dietary patterns (DPs) has provided a complementary strategy to the traditional single-food or single-nutrient approach. Use of DPs captures the intrinsic complexity of diet, the potential synergistic effects between its different components, as well as the variability in DPs existing within and between populations (1).

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

VE was supported by the Università degli Studi di Milano “Young Investigator Grant Program 2017.”

Author disclosures: The authors report no conflicts of interest.

Supplemental Materials and Methods, Supplemental Tables 1–3, and Supplemental Figure 1 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/advances>.

Address correspondence to VE (email: [valeria.edefonti@unimi.it](mailto:valeria.edefonti@unimi.it)).

Abbreviations used: CA, cluster analysis; CFA, confirmatory factor analysis; DP, dietary pattern; EFA, exploratory factor analysis; FFQ, food-frequency questionnaire; MONICA, MONitoring of trends and determinants in Cardiovascular Disease; NHS, Nurses' Health Study; PCA, principal components analysis; SMC, Swedish Mammography Cohort.

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

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

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

Second, evaluations of generalizability of a posteriori DPs should be based on ad hoc statistical methods tailored to disentangle the true differences in time or populations from time-specific or study-specific effects or simpler artifacts. A few novel methods have been proposed for the assessment of reproducibility of a posteriori DPs across studies (8, 11–14), including the use of the congruence coefficient for factor-loading comparison. Despite the several challenges to confront with, including individual and population-specific

dimensions of stability [e.g., (15, 16)] as well as transitions of target populations to a later stage in life [e.g., (16–18)], fewer research efforts have been focused on methods for the assessment of DP stability over time.

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

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

## Methods

### Literature search strategy

A scoping review was conducted using a systematic search of the literature through MEDLINE via PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) to identify all the articles on reproducibility and validity of a posteriori DPs, based on the following string: “(reproducibility or validity) and dietary pattern\*.” The guidelines from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) group were followed (24). The search was restricted to human studies reported in the English language and published up to 11 January 2019. Two authors (MD and VE) independently screened titles and then abstracts and retrieved the potentially relevant articles. The reference lists of the identified articles and other systematic reviews based on similar topics were also scanned. Discrepancies were resolved by involving a third researcher (MF).

### Inclusion and exclusion criteria

Articles were included or excluded based on the following criteria.

#### *A posteriori DPs.*

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

#### *Reproducibility and validity of a posteriori DPs.*

In the current review, we summarized evidence on cross-study reproducibility of DPs (including both reproducibility

**TABLE 1** Dimensions of stability according to possible levels of analysis<sup>1</sup>

Level	Methods <sup>2</sup>	Forms of stability <sup>2</sup>
Individual-level: Are single subjects stable eaters over time or do they change their DPs?	DPs; CA	DPs <ul style="list-style-type: none"> <li>• Percentages of stable eaters or transitioners</li> <li>• Ranking of clusters with the higher stability</li> </ul>
	Relevant food groups; ANOVA for testing differences in the mean intakes across clusters	Relevant food groups <ul style="list-style-type: none"> <li>• Lower-than- or higher-than-average consumption of food groups within clusters of subjects</li> </ul>
Population-level: <ul style="list-style-type: none"> <li>• Are DPs stable within a target population?</li> <li>• Is there a change in individuals' life course in the period under examination?</li> <li>• If yes, is the entire population experiencing a change in the life course?</li> <li>• Are there parallel subcohorts of different ages who get older, to assess "secular trends"?</li> </ul>	DPs; PCA/EFA with potential CFA on EFA-based results	DPs <ul style="list-style-type: none"> <li>• Number of identified DPs over time: are there DPs gained or lost during the period?</li> <li>• Percentage of explained variance of single DPs: are percentages similar over time for stable DPs?</li> <li>• DP composition: are factor-loading matrices similar over time?</li> <li>• DP scores: do mean scores from similar DPs change over time? Do quantile categories assigned to the same subject change over time?</li> </ul>
	Relevant food groups; MANOVA or ANOVA for testing differences in mean intakes or changes over time for EFA- or CFA-based relevant food groups	Relevant food groups <ul style="list-style-type: none"> <li>• Number of relevant food groups within a DP: is the number of food groups increasing or decreasing consistently over time?</li> <li>• Food-group intakes within a DP: do mean intakes from the same relevant food group change over time?</li> </ul>

<sup>1</sup>CA, cluster analysis; CFA, confirmatory factor analysis; DP, dietary pattern; EFA, exploratory factor analysis; MANOVA, multivariate ANOVA; PCA, principal components analysis.

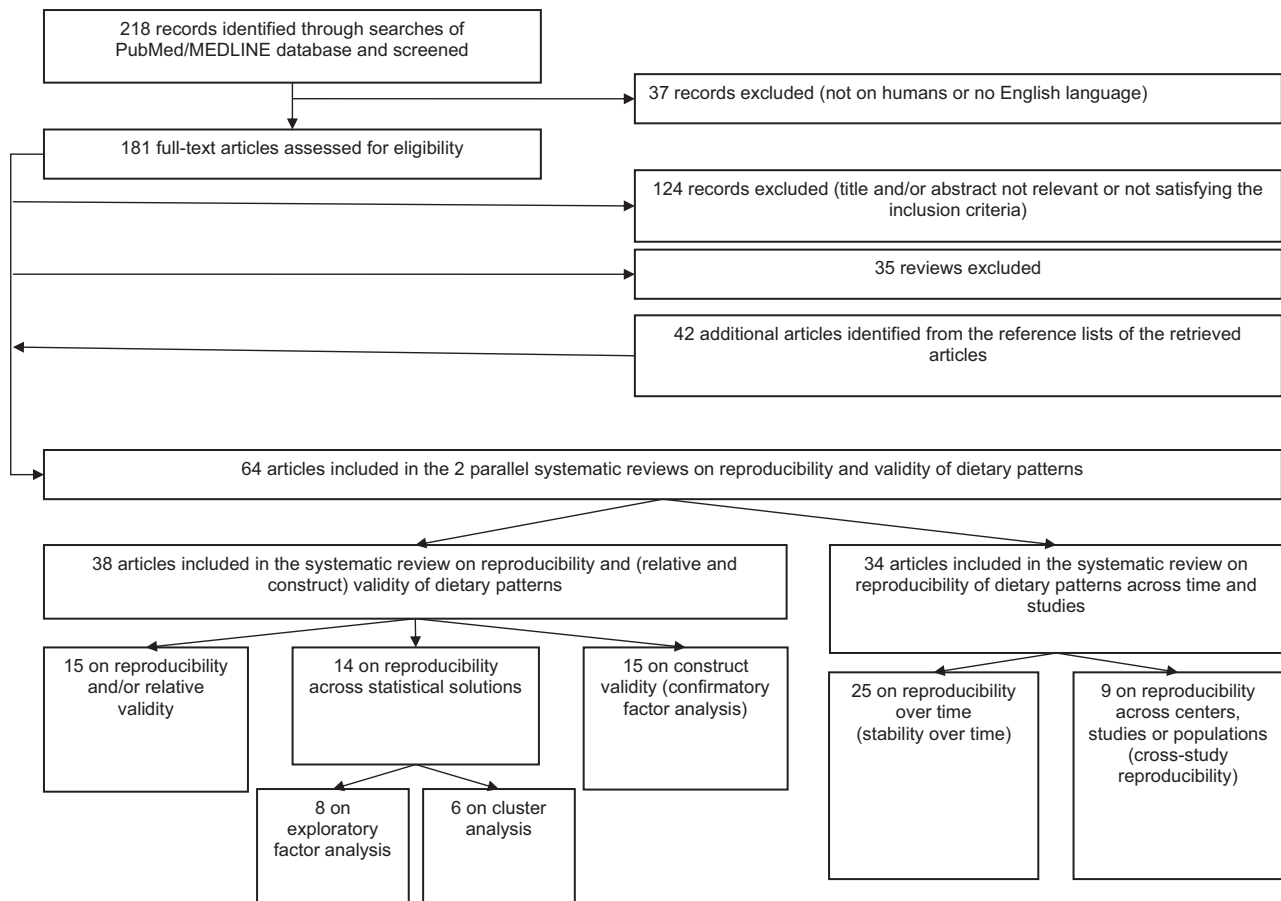
<sup>2</sup>Methods for the assessment of stability over time can target DPs directly as well as the relevant food groups defining these DPs; likewise, stability can be inspected at the DP level or at the relevant food-group level.

across centers from a multicentric study and reproducibility across different studies) and stability of DPs over time. Supplemental Table 1 and Supplemental Figure 1 provide an overview of the general terminology used in this review and of its use in nutritional epidemiology. The definition and use of terms introduced in our earlier review (5) (i.e., reproducibility across different statistical methods, short-term reproducibility, relative validity, and construct validity) were also presented within the **Supplemental Materials and Methods**. We also chose not to exclude studies on the basis of their quality, because of the lack of previous evidence on reproducibility and/or validity of DPs.

#### **Stability of DPs over time: possible forms of assessment.**

**Table 1** provides a detailed description of the different levels of analysis available within an assessment of stability of DPs. In detail, when the primary research question is to target potential transitions of subjects from one DP to

another DP over time (individual-level stability analysis), the most straightforward approach is to apply a CA and to track changes by calculating the percentages of transitioners (or stable eaters) across successive time points. When the primary aim is to describe potential changes over time in the covariance structure among dietary items within a population (population-level stability analysis), the most suitable approach is to apply PCA/EFA; changes can be tracked through the monitoring of the following aspects (in order of importance): 1) number of identified DPs (Are there DPs gained or lost?), 2) percentage of explained variance by each DP (Do stable DPs show similar percentages over time?), 3) DP composition (Are similar DPs characterized by the same relevant food groups or nutrients? Or are factor loadings similar or congruent over time?), 4) DP scores [Do the mean DP scores change (e.g., increase or decrease following some path) over time?]. Additional levels of complexity may arise when important changes



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

in the life course (e.g., from childhood to adolescence, or before and after pregnancy) happen within the period of observation. Within these designs, secular trends can be tracked identifying parallel subcohorts of different ages at baseline and comparing DPs derived on the subcohorts considered at the same age period.

### Data extraction

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

## Results

### Study selection process

**Figure 1** shows the flowchart of the study selection process carried out within the systematic search of the literature supporting this scoping review. From the PubMed database literature search, we identified 218 articles, of which 181 remained for detailed evaluation after the search was limited to human studies and articles written in the English language. Thirty-five review articles were removed, and 124 original research articles were also not included because they met the exclusion criteria. The most frequent reasons for exclusion were previously described in detail in the companion review (5). Forty-two additional articles were identified from manual searches of reference lists of selected original and review articles. Thus, 64 articles were included in our scoping review. Of these, the 34 articles that focused on stability of DPs over time and on their reproducibility across studies were included in this review, whereas the 38 articles on reproducibility and relative and construct validity of DPs were included in the companion paper (5). Eight articles (6, 9, 10, 22, 23, 25–27) were common to both reviews.

## Main characteristics of the included studies

General characteristics and study design information from the 34 articles on stability and cross-study reproducibility of DPs (6–12, 15–18, 22, 23, 25–45) are presented in **Table 2**. The selected articles were published between 1981 and 2019, with 32% of them published from 2010 onwards; the studies were mostly carried out across Europe and North America. Several articles were based on the same studies, including (but not limited to) those from the Swedish Mammography Cohort (SMC) (6, 7, 9, 22, 23), the Avon Longitudinal Study of Parents and Children (ALSPAC) (17, 18, 39, 40), and the Nurses' Health Study (NHS) I and II (35, 36, 38, 42). All the articles were based on observational studies, including 1 case-control (32), 24 cohort (6–10, 15–18, 22, 23, 26, 28, 30, 31, 33, 35–40, 42, 45) and 2 cross-sectional (43, 44) studies; in addition, there were 3 multiple administrations of the same survey (27, 34, 41), 1 validation study of the SMC food-frequency questionnaire (FFQ) (25), and 3 articles including studies with different designs (11, 12, 29). Two articles included men only (33, 45), 12 included women only (9, 11, 12, 15, 22, 23, 28, 30, 35, 36, 38, 40), with 3 studies based on pregnant women (15, 30, 40); 5 articles considered the recruitment of children and/or adolescents (16–18, 31, 39). With a few exceptions (16, 18, 30, 37, 43, 44), dietary information was collected with an FFQ. The FFQs were self-administered [except for the Southampton Women's Survey (SWS) (15, 28)]; the reference period of assessment was generally 1 y, except for diet during pregnancy (15, 28) or the high school period (36, 38). The number of food items inquired in the FFQs ranged from 26 (27, 34) to 276 (6), with a median value of 111.5 items. When >1 FFQ administration was available from cohort studies, the time interval between successive administrations could be fixed or variable [range of the minimum distance between dietary data used for DP identification: 1 mo (during pregnancy) (30) to 7 y (37)]. The reproducibility and/or relative validity of the FFQs was assessed within 1 validation study included in the review (25); in addition, 20 articles reported information on FFQ reproducibility and/or relative validity (6–12, 15, 22, 23, 26, 29, 31–33, 35, 36, 38, 42, 45). DPs were based on data collected through a dietary record and/or a recall of 24 or 48 h in 6 articles (16, 18, 30, 37, 43, 44).

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

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

## Cross-study reproducibility of DPs

**Table 3** concerns the 9 articles on cross-study reproducibility of a posteriori DPs. All the articles applied PCA or EFA, and 1 article (26) added a CFA to validate results from a previous

EFA. The number of involved centers or studies ranged from 2 (12, 43) to 27 (8), with a median of 3 centers/studies included per article.

## Identification of DPs across centers or studies.

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

Another 2 companion articles from Spain formally explored 1) the cross-study reproducibility of PCA-based DPs in 2 different samples extracted from similar Spanish populations (12) and 2) the applicability of 3 "internal" DPs derived from the previous Spanish case-control study (12) to independent ("external") populations with similar characteristics from France, the United States, and Sweden (as identified by a bibliographic search of the literature on the association between DPs and breast cancer) (11). The former article (12) applied a bootstrap-based approach to compare results from separate study-specific PCAs based on the same food-grouping scheme. The latter article (11) proposed to reconstruct the "external" DP scores as linear combinations of the published DP loadings and consumption of the published food groups, as re-calculated on the dietary data from the Spanish study. Similarly, the authors re-calculated the "external" DP loadings as based on the reference set of Spanish food groups to allow for direct comparison between loadings (11).

Finally, when individual-level data were available from studies of the same collaborative project, multistudy factor analysis was proposed in 1 article (32) to extend standard maximum-likelihood EFA and allowed for a partial sharing of EFA-based DPs across studies. Some DPs were derived to be common across all the studies; in addition to them, each study may express extra study-specific DPs. The number of shared and study-specific DPs was identified using a combination of standard criteria for EFA and information criteria for model selection (32).

The number of described DPs ranged from 2 (7) to 8 (44), with a median of 4 DPs per article; 2 articles (6, 7) reported the presence of additional population-specific DPs not described in detail (Supplemental Table 2).

## Assessment of cross-study reproducibility of DPs.

Four articles (6, 7, 43, 44) did not formally assess cross-study reproducibility and concluded that the study-specific sets of PCA/EFA-based DPs were qualitatively similar based on loadings and percentages of explained variances. A formal



**TABLE 2** Basic characteristics of observational studies on cross-study reproducibility and stability over time of a posteriori DPs<sup>1</sup>

Reference	Location; study	Study design	Participants, n	Age, y	Follow-up, y	Questionnaire
Asghari, 2012 (25)	Iran; TLGS	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 five 10-y age intervals and sexes plus extra wave of the cohort study with FFQ administration	132 (89 completed FFQ3)	35.6 ± 16.8 (20–70)	8, until 2011 (baseline: 1999–2001)	FFQ (based on a Willett format): 1 y; SA; reproducibility and validity to be assessed in this study, but validity granted for the analysis of stability over time; 168 FI; 12 24HRs: collected monthly on 2 formal weekend days and 10 weekdays; FFQ1: completed 1 mo before collection of the first 24HRs; FFQ2: completed 1 mo after the last 24HR, 14 mo between FFQ1 and FFQ2; FFQ3: completed at the end of the follow-up; 19 FGs common to all dietary sources
Balder, 2003 (6)	Netherlands, Sweden, Finland, and Italy; DIETSCAN Project (NLSC, SMC, ATBC, ORDET)	Parallel analysis of 4 prospective cohort studies according to the same strategy (no pooled analysis); NLSC (random subcohort of); population-based cohort of Ms and Fs from Dutch municipalities that began in 1986; SMC: population-based cohort of Fs based on a mammography screening in 2 countries in central Sweden from 1987 to 1990; ATBC: randomized placebo-controlled intervention study conducted among M smokers who lived in southwestern Finland (1985–1988); ORDET: cohort study of Italian healthy volunteer Fs from the province of Varese, northern Italy (1987–1992)	NLSC: 3123 Ms and Fs (1598 Fs and 1525 Ms); SMC: 61,469 Fs; ATBC: 27,111 Ms; all numbers referred to subjects with complete dietary data; ORDET: 9208 Fs	NLSC: at baseline 61.4 ± 4.2 for Ms and ± 4.3 for Fs (55–69); SMC: at baseline 53.7 ± 9.7 (40–74); ATBC: at baseline 57.7 ± 5.1 (50–69); ORDET: at baseline 48 ± 8.5 (35–69)	7 for NLSC (baseline: 1986), 13 for SMC (baseline: 1987–1990), and NA for ATBC (baseline: 1985–1988, intervention ended in 1993 after 5–8 y, follow-up later on), 9 for ORDET (baseline: 1987–1992)	4 different but validated FFQs: NLSC-FFQ: 1 y; SA; NA reproducibility but valid; 150 FI (51 FG, but final number equal to 49); SMC-FFQ: 6 mo; 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 FGs, but smaller final number of FGs); ORDET-FFQ: 1 y; SA; reproducible and valid; 107 FI (51 FGs, but final number equal to 32)
Borland, 2008 (28)	UK; SWS	SWS: prospective study including Fs from the general population living in the western part of Southampton; subset of Fs interviewed 2 y later at the same time of the year as the first interview (1998, November 13–December 22) from the cohort of 6129 SWS nonpregnant Fs; a subset of 29 diet changers out of all included in a separate analysis	94 nonpregnant Fs	At baseline (20–34)	2 y (baseline: 1998)	FFQ: 3 mo; IA; 100 FI (49 FGs); NA reproducibility and validity; FFQ administered 2 times, at baseline and after 2 y
Castello, 2016 (12)	Spain; EpiGEICAM, DDM-Spain	EpiGEICAM: case-control study on F breast cancer based on 14 Spanish provinces (2006–2011); DDM-Spain: cross-sectional study based on a random sample of Fs from 7 screening centers (minimum 500 from each center) (2007–2008)	EpiGEICAM: 973 healthy Fs; DDM-Spain: 3550 Fs	EpiGEICAM: 50.63 ± 9.47 (22–71); DDM-Spain: 56.20 ± 5.46 (45–69)	Not applicable	EpiGEICAM: FFQ: 5 y; NA SA; based on a validated FFQ; 117 FI (26 FGs); DDM-Spain: FFQ: 1 y; IA; based on a validated FFQ; 99 FI (all in common with EpiGEICAM FFQ) (26 FGs)
Castello, 2016 (11)	Spain; EpiGEICAM	EpiGEICAM: case-control study on F breast cancer patients based in 14 Spanish provinces (2006–2011); selection of 3 studies (Bessaud et al, Adebamowo et al, and Terry et al) from a systematic review of the literature on DPs and breast cancer	EpiGEICAM: 973 case-control pairs of Fs (1946 Fs in total)	EpiGEICAM: 50.63 ± 9.47 (22–71); other studies: NA	Not applicable	EpiGEICAM: FFQ: 5 y; NA SA; based on a validated FFQ; 117 FI (26 FGs); other studies: FFQs described in the paper

(Continued)

**TABLE 2** (Continued)

Reference	Location; study	Study design	Participants, n	Age, y	Follow-up, y	Questionnaire
Chen, 2015 (29)	Canada CCS, FFQVP	Two time-separated studies (over a decade) in Newfoundland and Labrador, including noninstitutionalized adult residents; CCS: case-control study with a frequency matching on age (5 y) and sex (2001–2005)—controls only from the CCS study; FFQVP: validation study conducted with a stratified random-digit dialing with proportional allocation (2011–2012)	CCS: 554 controls; FFQVP: 192	CCS: 58.7 ± 7.7 (35–70) (20–4 in all CCS cases and controls); FFQVP: 56.2 ± 8.7 (35–70)	Not applicable in either studies	Modified FFQ based on a Hawaii FFQ; 2 y; SA; 169 FI (39 FGs); valid; same FFQ administered in both studies
Crozier, 2009 (15)	UK; SWS	SWS: prospective cohort study including Fs from the general population living in the western part of Southampton (1998–2002)	2270 (early pregnancy) and 2649 (late pregnancy) from a cohort of 12,572 nonpregnant Fs; 2057 Fs with complete information at the 3 time points of interest used for the stability analysis	At baseline (20–34)	Before pregnancy (median time to conception: 1.8 y from initial interview)—late pregnancy (34 wk of gestation) (baseline: 1998–2002)	FFQ: 3 mo; IA; 98 FI (48 FGs); valid; FFQ administered at 3 time points, before pregnancy, in early pregnancy, and late pregnancy
Cuco, 2006 (30)	Spain; NA	Longitudinal cohort study based in the city of Reus, Spain, including healthy F volunteers who had planned and completed a pregnancy and had complete dietary information at all assessment occasions (1992–1996)	80 Fs	Mean: 29 (at baseline: 18–35; final range: 24–35)	Last pre-conception visit (1–3 menstrual cycles) to weeks 6, 10, 26, and 38 of pregnancy, and 6 mo postpartum (baseline: 1992–1996)	1 7-consecutive-day DR at each time point; check with trained interviewers; 22 FGs common to all time points
Cutler, 2009 (31)	USA (Minnesota); EAT	EAT: cohort study of ethnically diverse youth from Minnesota schools during early and middle adolescence; EAT-I (Time 1) and EAT-II (Time 2)	Time 1: 4746; time 2: 2516	Time 1: at baseline [12–13: early adolescence or middle school (younger cohort)] and 15–16: middle adolescence or high school (older cohort)]; time 2: same students 5 y later	5 y (Time 2: 2003–2004) (baseline: time 1: 1998–1999)	YAO-FFQ, based on the NHS FFQ; NA reference period; NA SA; reproducible and valid in children and adolescents 9–18 y old; 152 FI (152 FGs); pretested in a low-income, ethnically diverse middle school population with good results for comprehension
De Vito, 2019 (32)	USA, Italy, and Switzerland; INHANCE	INHANCE: consortium of case-control studies on head and neck cancer; subsample of 7 case-control studies providing information on a common set of 23 nutrients derived from study-specific FFQs: North Carolina (2002–2006); Milan (2006–2009); New York MSKCC (1992–1994); Los Angeles (1999–2004); Switzerland (1991–1997); Italy, multicenter (1990–1999)	10,668 (3844 cases; 6824 controls)	NA, but adults	Not applicable	5 study-specific FFQs; as the European studies (Italy Multicenter, Switzerland, and Milan (2006–2009)) shared the same FFQ; 1 y for the 4 US studies and 2 y for the 3 European studies; IA for 3 studies and SA for 4 studies; either reproducible and valid or based on previously validated FFQs; number of FI varying from 72 to 138 (23 common nutrients)

(Continued)

**TABLE 2** (Continued)

Reference	Location; study	Study design	Participants, n	Age, y	Follow-up, y	Questionnaire
Dekker, 2013 (10)	Netherlands; Doetinchem Cohort Study	3 successive surveys (surveys 2, 3, and 4, after the first one) within the same population-based cohort study including at baseline an age- and sex-stratified random sample of residents from Doetinchem town (1987–1991; follow-up available for 2/3 of the original random sample by design)	4007 subjects with information available for the 3 rounds. In detail: 6113 (survey 2); 4916 (survey 3); 4520 (survey 4)	~47–66	6 y (survey 2; 1993–1997), 11 y (survey 3; 1998–2002), 16 y (survey 4; 2003–2007) after the first survey, so 10-y follow-up from survey 2 to survey 4 (baseline: 1987–1991) 1990 and 1994 waves (baseline: 1986)	FFQ: 1 y; NA SA; reproducible and valid; 178 FI (32 FGs)
Fung, 2001 (33)	USA; HPFS	HPFS; prospective cohort study of US M health professionals started in 1986; random sample from the 18,255 subjects of the HPFS recruited between 1993 and 1994 who volunteered to provide blood samples	466 Ms	At baseline (40–75)	1990 and 1994 waves (baseline: 1986)	FFQ: 1 y; SA; reproducible and valid; 131 FI (42 FGs)
Gerdes, 2002 (34)	Denmark; MONICA	Three consecutive surveys from MONICA project, including at baseline (DAN-MONICA I, 1982–1984) a random sample of Danish citizens who lived in the western part of the Copenhagen County and were 30, 40, 50, and 60 y at baseline and further re-examined in 2 successive surveys (DAN-MONICA II and DAN-MONICA III)	3317 Fs (1822 + 737 + 778) and 3378 Ms (1876 + 725 + 777)	At baseline: 30, or 40, or 50, or 60	1982–1984 (baseline, DAN-MONICA I) – 1986–1987 (DAN-MONICA II) and 1991–1992 (DAN-MONICA III)	FFQ: 1 y; NA SA; NA reproducibility and validity; 26 FI (23 FI, with 3 excluded, no FGs built)
Judd, 2014 (26)	USA; REGARDS	Population-based cohort study including a random sample of black and white individuals and designed to oversample black participants and people residing in the stroke belt, a US region at particularly high risk for stroke (8 US states) (2003–2007)	21,636	>45	No follow-up	FFQ: 1 y; SA; NA reproducibility, but valid; 107 FI (58 FGs, but final analysis on 56 FGs due to low communalities and zero consumption)
Lopez-Garcia, 2004 (35)	USA; NHS	NHS; prospective cohort study of US F registered nurses started in 1976; sample of Fs who were selected as control subjects for a nested case-control study on diabetes and who did not have cardiovascular disease, cancer, or diabetes mellitus at the time of blood drawing	732 Fs	At blood drawing, mean: 56 (43–69) (1989–1990)	1986 and 1990 waves (baseline: 1976)	FFQ: 1 y; SA; reproducible and valid; administered 2 times in 1986 and 1990; 116 FI (37 FGs)
Malik, 2012 (36)	USA NHS II	NHS II; prospective cohort study of US F registered nurses started in 1989; sample of Fs who returned an FFQ on high school diet in 1998 and did not have confirmed diabetes/history of diabetes/gestational diabetes, cancer, or cardiovascular disease	37,038 Fs	At baseline in 1989 (24–44), in 1997 at high school FFQ completion (34–53)	1997–2005 (baseline: 1989)	HS-FFQ: high school period; SA; reproducible and valid; 124 FI (37 FGs); NHS II FFQ: 1 y; SA; reproducible and valid; 131 FI (40 FGs); NHS II administered 4 times to assess adult diet (in 1991, 1995, 1999, and 2003)
Mannisto, 2005 (7)	Netherlands, Sweden, and Italy; DIETSCAN Project (NLSC, SMC, ATBC, ORDET)	Parallel analysis of 3 prospective cohort studies according to the same strategy (no pooled analysis); NLSC (random subcohort of): population-based cohort of Ms and Fs from Dutch municipalities that began in 1986; SMC: population-based cohort of Fs based on a mammography screening in 2 countries in central Sweden from 1987 to 1990; ORDET: cohort study of Italian healthy volunteer Fs from the province of Varese, northern Italy (1987–1992)	NLSC: 1598 Fs; SMC: 61,463 Fs; ORDET: 10,788 Fs	NLSC: 61.4 ± 4.3 at baseline (55–69); SMC: 53.7 ± 9.7 at baseline (40–74); ORDET: 48 ± 8.5 at baseline (34–70)	7 y for NLSC (baseline: 1986), and 13 y for SMC (baseline: 1987–1990), 9 y for ORDET (baseline: 1987–1992)	3 different but validated FFQs: NLSC-FFQ: 1 y; SA; NA reproducibility but valid; 150 FI (51 FGs, but final number equal to 49); SMC-FFQ: 6 mo; SA; NA reproducibility but valid; 67 FI (51 FGs, but final number equal to 42); ORDET-FFQ: 1 y; SA; reproducible and valid; 107 FI (51 FGs, but final number equal to 31)

(Continued)



**TABLE 2** (Continued)

Reference	Location; study	Study design	Participants, n	Age, y	Follow-up, y	Questionnaire
Mikkilä, 2005 (16)	Finland; Cardiovascular Risk in Young Finns Study	Cardiovascular Risk in Young Finns Study: multicenter prospective cohort study of children, adolescents, and young adults started in 1980 in Finland; random sample of 50% of the participants who had dietary information and was followed at 2 time points	1768 subjects in 1980, 1200 in 1986, and 1037 in 2001, giving a total of 1037 subjects with complete information at the 3 time points	At baseline (3–18), in 2001 (24–39)	1980 (baseline) –2001, with a first wave of follow-up in 1986	1 48HR for each time point (in 1980, 1986, and 2001); different number of recorded FI for each time point (23 FGs)
Mishra, 2006 (37)	UK; Medical Research Council National Survey of Health and Development (1946 British Birth Cohort)	1946 British Birth Cohort: longitudinal cohort study based on a social class stratified, random sample of 5362 singleton births in England, Scotland, or Wales during the first week of March 1946, with 21 occasions for collecting information throughout the life course until published paper; data from interviews at 3 time points in 1982, 1989, and 1999	1265 subjects with dietary information at the 3 time points	36 in 1982, 43 in 1989, 53 in 1999	1946 (baseline) –1999	1 5-d DR completed between spring and autumn for each time point in 1982, 1989, and 1999; different number of recorded FI for each DR (126 FGs)
Moskal, 2014 (8)	Europe; EPIC	EPIC: cohort study on healthy Ms and Fs from 23 centers representing 10 European countries, including a calibration study based on a random sample of 5–12% subjects from each EPIC center	477,312 (including 34,436 from the calibration study with 24HR)	At baseline (35–70)	1992–1998 (for FFQ); 1995–2000 (for 24HR)	Country-specific dietary questionnaires, mostly FFQs; NA reference period; SA; valid; NA FI (23 nutrients); 1 24HR recall via face-to-face interview to describe the identified DPs
Newby, 2006 (23)	Sweden; SMC	SMC: population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; subsample of SMC including healthy Fs at baseline with complete information on FFQ1 and FFQ2	33,840 Fs	Mean: 52 at baseline (all Fs born between 1914 and 1948)	From 1987–1990 (baseline) to 1997–onwards	FFQ1 (1987–1990): 6 mo; SA; reproducible and valid; 67 FI (29 FGs); FFQ2 (1997): 1 y; SA; based on the 1987 reproducible and valid FFQ; 97 FI (32 FGs); mean time interval between FFQs: 8.8 y
Newby, 2006 (22)	Sweden; SMC	SMC: population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; subsample of SMC including healthy Fs at baseline with complete information on FFQ1 and FFQ2	33,840 Fs	Mean: 52 at baseline (all Fs born between 1914 and 1948)	From 1987–1990 (baseline) to 1997, 9 y of follow-up	FFQ1 (1987–1990): 6 mo; SA; reproducible and valid; 67 FI (29 FGs); FFQ2 (1997): 1 y; SA; based on the 1987 reproducible and valid FFQ; 97 FI (32 FGs)
Nimptsch, 2014 (38)	USA; NHS II	NHS II: prospective cohort study of US F registered nurses started in 1989; sample of Fs who returned an FFQ on high school diet in 1998, underwent at least 1 lower bowel endoscopy between 1998 and 2007, and had no history of cancer, colorectal adenomas, hyperplastic polyps	17,221 Fs	At baseline in 1989 (24–42), in 1997 at high school FFQ completion (34–51)	1997–2007 (baseline: 1989)	HS-FFQ: high school period (1960–1980); SA; reproducible and valid; 124 FI (37 FGs); NHS II FFQ: 1 y; SA; reproducible and valid; 131 FI (40 FGs); NHS II administered 5 times to assess adult diet (in 1991, 1995, 1999, 2003, and 2007)
Northstone, 2005 (39)	UK; ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant F residents in the former Avon Health Authority with expected delivery date between 1 April 1991 and 31 December 1992; subset of ALSPAC study including 4- and 7-y-old children (2 waves)	9550 and 8286 children at 4 and 7 y, respectively	4 and 7	2 waves for the children (4 and 7 y of age) (baseline: 1991–1992)	FFQ adapted from the one used to assess maternal diet at 32 wk of pregnancy; NA reference period; SA; completed by the mother/main carer; NA reproducibility and validity; 90 FI (57 FGs)

(Continued)

**TABLE 2 (Continued)**

Reference	Location; study	Study design	Participants, n	Age, y	Follow-up, y	Questionnaire
Northstone, 2013 (18)	UK; ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant F residents in the former Avon Health Authority with expected delivery date between 1 April 1991 and 31 December 1992; subset of ALSPAC study including 7-, 10-, and 13-y-old children (3 waves)	7285, 7473, and 6105 children, at 7, 10, and 13 y, respectively	~7, 10, and 13	3 waves for the children (7, 10, and 13 y of age) (baseline: 1991–1992)	1 3-d DR for each time point, including 2 weekdays and 1 weekend; at 7 y caregiver completion, at 10 and 13 y, child completion; 62 FGs at each time point
Northstone, 2008 (17)	UK; ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant F residents in the former Avon Health Authority with expected delivery date between 1 April 1991 and 31 December 1992; subset of ALSPAC study including 3-, 4-, 7-, and 9-y-old children (4 waves)	10,139, 9550, 8286, and 8010 children, at 3, 4, 7, and 9 y, respectively; 6177 children with information at 4 time points for stability analysis	~3, 4, 7, and 9	4 waves for the children (3, 4, 7, and 9 y of age) (baseline: 1991–1992)	Slightly different FFQs adapted from the one used to assess maternal diet at 32 wk of pregnancy; NA reference period; SA, completed by the mother/main carer; NA reproducibility and validity; NA FI, increasing number for increasing study wave number; 34, 35, 41, and 41 FGs at 3-, 4-, 7-, and 9-y data
Northstone, 2008 (40)	UK; ALSPAC	ALSPAC: longitudinal cohort study including a sample of pregnant F residents in the former Avon Health Authority with expected delivery date between 1 April 1991 and 31 December 1992; subset of ALSPAC study including F s during pregnancy and 4 y after delivery (2 waves)	12,053 and 9504 F s pregnant at baseline and at 4 y of the child, respectively; 8953 F s with complete information at both time points	NA, but pregnant F s	4 y (47 mo after birth) (baseline: 1991–1992)	Slightly different FFQs with extra information added in the second FFQ; NA reference period; SA; NA reproducibility and validity; NA FI (44 FGs at pregnancy assessment and 52 FGs at 4-y wave)
Prevost, 1997 (41)	UK; HALS	Two consecutive surveys (1984–1985: HALS1, 1991–1992: HALS2); HALS1: random stratified sample of adult residents in England, Scotland, and Wales	HALS1: 9003; HALS2: 5352 from HALS1, still alive and able to participate	(18–74)	1991–1992 (HALS2) (baseline: 1984–1985, HALS1)	FFQ: NA reference period; NA SA; 39 FI (39 FGs); NA reproducibility and validity; FFQ administered 2 times, at baseline and at follow-up
Schulze, 2006 (42)	USA; NHS II	NHS II: prospective cohort study of US F registered nurses started in 1989; sample of F s who returned 3 plausible FFQs and did not have history of diabetes, cancer, cardiovascular disease, or were pregnant at FFQ administration time	51,67	At baseline (24–44), in 1991 (26–46)	1991–1999 (baseline: 1989)	NHS II FFQ: 1 y; SA; reproducible and valid; 133 FI (39 FGs); NHS II administered 3 times to assess adult diet (in 1991, 1995, and 1999)
Schwerin, 1981 (43)	USA; Ten-State Nutrition Survey (Ten-State); HANES I	Merging of 2 cross-sectional studies; Ten-State (1968–1970): sample disproportionately poor, with few young adults, and a disproportionate number of blacks and Spanish Americans from geographically scattered states; subjects are provided with detailed information from special clinics; HANES I (1971–1974): broad-based national sample including all age groups between 1 and 74 y	Ten-State: 11,337; HANES I: 20,749	1–74	No follow-up	1 24HR (15 FGs) for both surveys

(Continued)

**TABLE 2 (Continued)**

Reference	Location; study	Study design	Participants, n	Age, <sup>2</sup> y	Follow-up, y	Questionnaire
Schwerin, 1982 (44)	USA; Ten-State Nutrition Survey (Ten-State), HANES I, NCFCS	Merging of 3 cross-sectional studies; Ten-State (1968–1970); sample disproportionately poor, with few young adults, and a disproportionate number of blacks and Spanish Americans from geographically scattered states; subjects are provided with detailed information from special clinics; HANES I (1971–1974); broad-based national sample including all age groups between 1 and 74 y; NCFCS (1977–1978); representative sample of US population	Ten-State: 11,337; HANES I: 20,749; NCFCS: 28,030	1–74	No follow-up	1 24HR (15 FGs) for all 3 surveys, plus for NCFCS 2-d DR; for NCFCS, combination of information from 24HR and 2-d DR into a 3-d food consumption in grams
Togo, 2004 (27)	Denmark; MONICA	Three consecutive surveys from MONICA project, including at baseline (M-82) a random sample of Danish citizens who lived in the western part of the Copenhagen County and were 30, 40, 50, and 60 y at baseline (1982–1984) and further re-examined in 2 successive surveys (M-87, M-93)	2436 subjects participating in all 3 surveys, including 1806 subjects in M-82	30, or 40, or 50, or 60 at baseline	At 5 y (1987–1988, M-87) and 11 y (1993–1994, M-93)	FFQ: 1 y; NA SA; NA reproducibility and validity; 26 FI (21 FGs)
van Dam, 2002 (45)	USA; HPFS	HPFS; prospective cohort study of US M health professionals started in 1986; all Ms without diagnosed diabetes, cardiovascular disease, or cancer at baseline	42,504 Ms	At baseline in 1986 (40–75)	1986–1998	FFQ: 1 y; SA; reproducible and valid; 131 FI (37 FGs)
Weismayer, 2006 (9)	Sweden; SMC	SMC; population-based cohort based on a mammography screening in 2 counties in central Sweden from 1987 to 1990; subsample of SMC including 4 randomly selected subsamples of 1000 Fs each (giving a total of 4000 Fs), who completed 2 identical FFQs, to avoid survey learning effects	3606 Fs (871, 864, 887, and 967, at 4, 5, 6, 7 y after baseline, respectively)	49–70	4, 5, 6, 7 y after baseline (1987–1990) depending on the subsample	FFQ (1987–1990); 6 mo; SA; reproducible and valid; 67 FI (25 FGs); FFQ completed at baseline and after 4, 5, 6 or 7 y depending on the subsample

<sup>1</sup>ALSPAC, Avon Longitudinal Study of Parents and Children; ATBC, Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; CCS, case-control study (here intended as the full name of one of the included studies and not as the case-control study design); DDM-Spain, Determinantes de la Densidad Mammográfica en España; DIETSCAN, Dietary patterns and CANcer in 4 European countries project; DR, dietary record; EAT, Eating Among Teens; EPIC, European Prospective Investigation into Cancer and Nutrition; EpiGEICAM, Grupo Español de Investigación en Cáncer de Mama; F, female; FFQ, food-frequency questionnaire; FFQ1/FFQ2/FFQ3, food-frequency questionnaire at time 1, 2, and 3; FFQVP, Food-Frequency Questionnaire Validation Project; FG, food group; FI, food items; HALS, Health and Lifestyle Survey; HANES, Health and Nutrition Examination Survey; HS, high school; HPFS, Health Professionals Follow-Up Study; IA, interviewer-administered; INHANCE, International Head and Neck Cancer Epidemiology Consortium; M, male; MONICA, Monitoring of trends and determinants in Cardiovascular Disease; MSKCC, Memorial Sloan Kettering Cancer Center; NA, not available; NCFCS, Nationwide Food Consumption Survey; NHS, Nurses' Health Study; NLS, Netherlands Cohort Study on Diet and Cancer; ORDET, Ormoni e Dieta nella Etiologia dei Tumori in Italy; REGARDS, Reasons for Geographic and Racial Differences in Stroke; SA, self-administered; SMC, Swedish Mammography Cohort; SWS, Southampton Women's Survey; TUGS, Teheran Lipid and Glucose Study; YAO, Youth Adolescent Questionnaire; 24HR, 24-h recall; 48HR, 48-h recall.

<sup>2</sup>Values are means ± SDs (ranges).

**TABLE 3** Cross-study reproducibility of a posteriori DPs<sup>1</sup>

Reference	Location; study	DP identification methods	Explained variance % (number of factors) or vs CFA/CA model	Assessment of reproducibility/validity	Main results
Balder, 2003 (6)	Netherlands, Sweden, Finland, and Italy; DIETSCAN (NLCS, SMC, ATBC, ORDET)	Separate EFAs on each of the 4 studies: standardization and separate analysis by sex; within each study, sensitivity analyses assessing the effect of: 1) untransformed vs dichotomized variables (for FGs with >75 % of nonusers); 2) unadjusted vs energy-adjusted variables using residual method; 3) solutions with 2–6 factors; 4) split-half analysis using the procrustes rotation to compare different solutions; Scree test to assess the final number of factors to retain in a range from 2 to 6 factors; Varimax rotation; Loading $\geq 0.35$ cutoff	NLCS: 23 (5) for Ms, 23.2 (5) for Fs; SMC: 21.8 (4); ATBC: 20.3 (3); ORDET: 28.5 (4); final results based on unadjusted variables for energy	Internal reproducibility: see (5) for details; cross-study reproducibility: no formal assessment	Internal reproducibility: see (5) for details; cross-study reproducibility: 2 of the identified DPs were qualitatively similar across studies and between Ms and Fs
Castello, 2016 (12)	Spain; EpiGEICAM, DDM-Spain	Separate PCAs on EpiGEICAM and DDM studies; PCA on EpiGEICAM data: PCA on controls only; EIG > 1; No rotation; Loading $\geq 0.30$ cutoff; PCA on DDM data: separate PCAs on 5000 replicates of the DDM-Spain study within bootstrap estimation with selection of the 3 DPs that were more similar to those from EpiGEICAM study; PCA on controls only; EIG > 1; No rotation; Loading $\geq 0.30$ cutoff	37 (3) with PCA on EpiGEICAM data	Cross-study reproducibility: CC (95% CI) between factor loadings (with values of 0.85–0.94 indicate fair similarity and values $\geq 0.95$ indicate 2 DPs were equivalent); Spearman correlation coefficient (Corr) (95% CI) between factor scores (considering any significant correlation as being indicative of DP similarity)	Cross-study reproducibility: satisfactory reproducibility of WESTERN DP; but not of PRUDENT and MEDITERRANEAN DPs [WESTERN DPs: CC = 0.90 (95% CI: 0.58–0.95), Corr = 0.92 (95% CI 0.55–0.98); PRUDENT: CC = 0.76 (95% CI 0.40–0.84), Corr = 0.83 (95% CI 0.47–0.91); MEDITERRANEAN: CC = 0.77 (95% CI 0.65–0.83), Corr = 0.74 (95% CI 0.63–0.79)]; had we considered any significant correlation as being indicative of similarity, all DPs from the EpiGEICAM data were reproducible in the DDM-Spain study
Castello, 2016 (11)	Spain; EpiGEICAM	PCA on EpiGEICAM study: PCA on controls only; EIG > 1; No rotation; Loading $\geq 0.30$ cutoff; food consumption information from EpiGEICAM study grouped into FG proposed in 3 other papers (Bessaoud et al., Adebamowo et al., and Terry et al.) and factor scores calculated with loadings from the original papers and FGs defined as in the original papers but recalculated on EpiGEICAM data; factor loadings recalculated using the definition of FG from (10)	37 (3) with PCA on EpiGEICAM data	Cross-study reproducibility: CC (95% CI) between factor loadings (with values of 0.85–0.94 indicate fair similarity and values $\geq 0.95$ indicate 2 DPs were equivalent); Spearman correlation coefficient (Corr) (95% CI) between factor scores (considering any significant correlation as being indicative of DP similarity)	Cross-study reproducibility: 5 of the 6 reconstructed DPs showed high CC (>0.9) to their corresponding DP derived on the EpiGEICAM study data [CC (Castello-WESTERN, Bessaoud-WESTERN) = 0.82, Corr (Castello-WESTERN, Bessaoud-WESTERN) = 0.57; CC (Castello-WESTERN, Adebamowo-WESTERN) = 0.92, Corr (Castello-WESTERN, Adebamowo-WESTERN) = 0.83; CC (Castello-WESTERN, Terry-WESTERN) = 0.94, Corr (Castello-WESTERN, Terry-WESTERN) = 0.85; CC (Castello-PRUDENT, Bessaoud-MEDITERRANEAN) = 0.86, Corr (Castello-PRUDENT, Bessaoud-MEDITERRANEAN) = 0.67; CC (Castello-MEDITERRANEAN, Bessaoud-MEDITERRANEAN) = 0.95, Corr (Castello-MEDITERRANEAN, Bessaoud-MEDITERRANEAN) = 0.85; CC (Castello-PRUDENT, Adebamowo-PRUDENT) = 0.95, Corr (Castello-PRUDENT, Adebamowo-PRUDENT) = 0.85; CC (Castello-MEDITERRANEAN, Adebamowo-PRUDENT) = 0.88, Corr (Castello-MEDITERRANEAN, Adebamowo-PRUDENT) = 0.73; CC (Castello-PRUDENT, Terry-HEALTHY) = 0.95, Corr (Castello-PRUDENT, Terry-HEALTHY) = 0.89; CC (Castello-MEDITERRANEAN, Terry-HEALTHY) = 0.77, Corr (Castello-MEDITERRANEAN, Terry-HEALTHY) = 0.52]; some smaller CC between comparable DPs depended on lack of FGs in the original studies

(Continued)

**TABLE 3** (Continued)

Reference	Location; study	DP identification methods	Explained variance % (number of factors) or vs CFA/CA model	Assessment of reproducibility/validity	Main results
De Vito, 2019 (32)	USA, Italy, and Switzerland; INHANCE	Multistudy factor analysis on the merged dataset including the 7 studies; within-study log-transformation (base e) and standardization; controls-only analysis; identification of shared (among all studies) and (potential) study-specific dietary patterns within an integrated statistical model based on the maximum likelihood approach; number of factors to retain chosen according to a combination of standard techniques for FA, including Horn's parallel analysis, Cattell's scree plot, and the Steiger's RMSEA index, for the best number of total factors allowed, and to Akaike Information Criterion, for the number of shared factors; Varimax rotation on the shared factor loading matrix; Loading $\geq 0.60$ cutoff for the shared (rotated) factors and loading $\geq 0.25$ cutoff for the study-specific (unrotated) factors; robustness analyses and stratified multistudy factor analysis by sex	75–81 (3 common DPs shared among all the studies plus 1 additional study-specific DP for each of the 4 US studies)	Cross-study reproducibility: multistudy factor analysis	Cross-study reproducibility: Study populations from Italy, Switzerland, and the United States shared 3 reproducible DPs characterized by consumption of animal products and cereals, vitamin-rich foods, and fats, respectively; each of the American studies was characterized by a somewhat similar additional DP, which opposed calcium and niacin as dominant nutrients
Judd, 2014 (26)	USA; REGARDS	EFA on the first split-sample, CFA on the second split-sample, and final PCA on the whole sample as far as the model is correctly identified; EFA: 3 separate PCAs by population subgroups [region (southeastern US stroke belt/non-belt), sex (male/female), and race (black/white)] to identify the optimal number of factors in a range from 3 to 6 factors; EIG > 1.5; Scree test, interpretability of results from stratified PCAs; Varimax rotation; Descriptive labeling; CFA: Loading >0.20 cutoff on EFA results; No different correlation structures specified; RMSEA and CFI	NA (5)	Cross-study reproducibility: CC determined for each stratification pair for each of the factor number solutions ("excellent" when the smallest coefficient was >0.8, "good" between 0.65 and 0.8, "acceptable" between 0.5 and 0.65, and "poor" <0.5); Validity: CFA	Cross-study reproducibility: PCA stratified by region of residence on the first half-sample: excellent CC for the 4- and 5-factor solutions, and acceptable CC for the 3- and 6-factor solutions; PCA stratified by gender: good CC for the 5- and 6-factor solutions and poor CC for the 3- and 4-factor solutions; PCA stratified by race: acceptable CC in the 5-factor solution, but poor CC for the other 3; the 5-factor solution had an acceptable CC in all stratified analyses and it was interpretable, so this was the final model selected for CFA; CFA on the second half-sample using the 5-factor solution: very good results, even when removing FG with low factor loadings (RMSEA values <0.05)

(Continued)



**TABLE 3** (Continued)

Reference	Location; study	DP identification methods	Explained variance % (number of factors) or xs CFA/CA model	Assessment of reproducibility/validity	Main results
Mannisto, 2005 (7)	Netherlands, Sweden, and Italy; DIETSCAN (NLCS, SMC, ATBC, ORDET)	Separate PCFAs on each of the 3 studies: Scree test; Varimax rotation; Loading $\geq 0.35$ cutoff	NLCS: 23.2 (5); ORDET: 29 (4); SMC: 21.8 (4)	Cross-study reproducibility: no formal assessment	Cross-study reproducibility: both the identified DPs remained quite consistent across cohort studies
Moskal, 2014 (8)	Europe; EPIC	Overall PCA on combined but country-specific questionnaire intakes and separate PCAs by center; log-transformation (base e) and energy adjustment with energy density method (based on alcohol-free energy) but no adjustment for center; separate analysis by sex; PCA on covariance matrix; Scree-plot, interpretability; Varimax rotation; Loading $> 0.45$ cutoff	Overall PCA: 67 (4)	Cross-study reproducibility: Kuznowski's index, $B_k$ , which measures the proportion of variance captured by k center-specific PCs, which is also captured by overall PCA	Cross-study reproducibility: $> 75\%$ of the variance that would be captured by center-specific PCs was captured by the PCs from the overall PCA ( $B_k > 0.76$ for all $j \geq 2$ , $B_2 > 0.85$ for 23 of 27 centers); retaining $\geq 4$ PCs was sufficient to capture at least 80% of variance in any center ( $B_k > 0.80$ for all $j \geq 4$ ); differences between sexes in each center were small when $k > 2$
Schwerin, 1981 (43)	USA; Ten-State Nutrition Survey (Ten-State), HANES I	Separate PCAs on the 2 surveys: standardization; $ELG > 1$ ; Varimax rotation; Alpha-numeric labeling; assignment algorithm of subjects based on the highest factor score; (probably) applied scores on HANES I data based on Ten-State DP loadings in the final solution	55.3 (7)	Cross-study reproducibility: no formal assessment	Cross-study reproducibility: the identified DPs were similar in the 2 surveys in terms of FGs consumed
Schwerin, 1982 (44)	USA; Ten-State Nutrition Survey (Ten-State), HANES I, NFCS	Separate PCAs on the 3 surveys: standardization; $ELG > 1$ ; Varimax rotation with Kaiser normalization; Alpha-numeric labeling	NA (6, or 7, or 8)	Cross-study reproducibility: no formal assessment	Cross-study reproducibility: 4 of the identified DPs remained quite consistent across studies that covered a decade

<sup>1</sup> ATBC, Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; CA, cluster analysis; CC, congruence coefficient; CFA, confirmatory factor analysis; CFI, comparative fit index; DDM-Spain, Determinantes de la Densidad Matrográfica en España; DIETSCAN, DIETary patterns and CANCER in 4 European countries project; DP, dietary pattern; EFA, exploratory factor analysis; ELG, eigenvalue; EPIC, European Prospective Investigation into Cancer and Nutrition; EPiGEICAM, Grupo Español de Investigación en Cáncer de Mama; F, female; FA, factor analysis; FG, food group; HANES, Health and Nutrition Examination Survey; INHANCE, International Head and Neck Cancer Epidemiology Consortium; M, male; NA, not available; NFCS, Nationwide Food Consumption Survey; NLCS, Netherlands Cohort Study on Diet and Cancer; ORDET, Ormoni e Dieta nella Etiologia dei Tumori in Italy; PC, principal component; PCA, principal components analysis; PCFA, principal components factor analysis; REGARDS, Reasons for Geographic and Racial Differences in Stroke; RMSEA, root mean square error of approximation; SMC, Swedish Mammography Cohort.

assessment was carried out in the remaining 5 articles (8, 11, 12, 26, 32). Congruence coefficients between factor loadings and correlation coefficients between factor scores were used in 3 articles (11, 12, 26), whereas the other 2 articles used the Krzanowski's index (8) and multistudy factor analysis (13), respectively. The aim of the analyses was also different across the 5 articles. In 2 articles (8, 26) the statistical analysis was meant to support an overall PCA/EFA model where the single centers/studies were merged in 1 database. Another 2 studies (11, 12) were aimed at testing the extent to which a posteriori DPs are generalizable within and between countries. One article (32) was in between the 2 approaches as it was focused on an assessment of cross-study reproducibility in an international context as in reference 11; however, the availability of consortia data allowed to fit a statistical model that accounted simultaneously for common and study-specific DPs.

### *Summary of the evidence on cross-study reproducibility of DPs.*

No matter the statistical approach used, the number of DPs reproduced across the studies ranged from 1 (12) to 7 (43), with a median value of 4 common DPs identified. In addition, 2 articles (6, 32) described 1 (32) and 4 (6) DPs that were reproducible among subsets of the included studies. Among the reproducible DPs, most studies identified variants of a Western-like DP (6–8, 11, 12, 26, 32) and/or a Prudent-like DP (6–8, 11, 26, 32, 43, 44); in addition, some articles identified a variant of a Fat- or Condiment-based DP (8, 11, 26, 32, 43, 44), whereas another article added to its reproducible set of DPs a Traditional (Southern) and Alcohol/Salads DP across 8 US regions (26).

### **Stability of DPs over time**

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

### *Identification of DPs over multiple time occasions.*

With the exception of a single article (27), DPs were separately identified at each time point following the same standardized approach across time occasions. While most of the articles simply proposed separate time-specific statistical analyses (9, 16, 17, 22, 23, 29–31, 33–36, 38, 39, 42, 45), a few others proposed either applied (15, 25, 28, 40, 41) or simplified (37) scores to harmonize PCA- or EFA-based DPs derived at different time points. As opposed to standard or “natural” scores, applied scores were calculated at a later time point combining loadings from a PCA/EFA at a previous (analysis at 2 time points) or reference time point (analysis at  $\geq 3$  time points) with dietary information at the current time point (40); at a fixed time point, simplified scores (46) were

calculated as an unweighted sum of dominant food groups, where only the sign (and not the value) of the loading is used.

To further improve comparability of DPs at different time points, the article by Togo et al. (27) used a mean-structure CFA model that allowed the jointly modeling of dietary data at the 2 time points within a formal statistical approach that explored trends in (potentially correlated) DP scores across time.

The number of described DPs ranged from 2 to 6, with 11 of the articles naming and describing 2 DPs; however, in 5 articles (9, 15, 16, 22, 23), the authors reported additional DPs not common to all time points and/or not relevant/interpretable (Supplemental Table 3). The described DPs were generally similar across time points in terms of factor loadings and percentages of explained variance; their names reflected these similarities. Some variation in DP composition was reported, either leading to a change in the DP name across time points (29) or not (16, 22, 30, 31, 36, 40). Additional DPs were identified at earlier (17) and/or later time points (17, 25, 29, 31); some other DPs were lost at later time points (17, 30, 40) (Supplemental Table 3).

### *Assessment of stability over time: DPs and their relevant food groups.*

Six articles (22, 29, 31, 35, 39, 42) did not formally assess stability of DPs over time; except for 1 DP in 2 studies (22, 29), the main conclusion from these articles was that the time-specific sets of PCA/EFA-based DPs were qualitatively similar based on loadings and percentages of explained variances.

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

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

**TABLE 4** Stability over time of a posteriori DPs<sup>1</sup>

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Asghari, 2012 (25)	Iran; TLGS	Separate PCFAs on FFQ1, FFQ2, FFQ3, and m24HRs: Scree test and interpretability; Varimax rotation; Descriptive labeling; Applied scores from previous EFAs to data from FFQ3 were reported but their use was not clear	27.4 (2) with FFQ1 data, 31.6 (2) with FFQ2 data, 39.0 (3) with FFQ3 data, and 32.0 (2) with m24HR data	Reproducibility: see (5) for details; Relative validity: see (5) for details; Stability over time: intraclass correlation coefficient between continuous scores from FFQ2 and FFQ3 data, weighted $\kappa$ -coefficient and proportions of subjects at the same quintile, adjacent quintile and opposite quintile when comparing quintiles classification of factor scores between baseline and follow-up data	Reproducibility: see (5) for details; Relative validity: see (5) for details; Stability over time: intraclass coefficients between FFQ2- and FFQ3-based scores equal to $-0.09$ ( $P = 0.653$ ) for the IRANIAN TRADITIONAL and $0.49$ ( $P < 0.001$ ) for the WESTERN DPs; percentage of subjects at the same quintile higher for the WESTERN DP vs the IRANIAN TRADITIONAL DP (27.1% vs 20.2%); proportion of individuals at the opposite quintile reversed (35.8% vs 41.5%); weighted $\kappa$ -coefficient: $0.09$ (95% CI: $-0.05, 0.23$ ) for the IRANIAN TRADITIONAL and $0.20$ (95% CI: $0.05, 0.34$ ) for the WESTERN DP
Borland, 2008 (28)	UK; SWS	Separate PCAs at baseline and at follow-up: Interpretability: NA varimax rotation; Descriptive labeling: Applied scores calculated with loadings from the PCA on the whole cohort with complete FFQ (6125 subjects); Scores expressed in units of SD at initial visit (scores at both time points divided by the SD of the scores at initial visit)	NA (2)	Stability over time: Spearman correlation coefficient between DP scores at 2 time points; Bland-Altman method	Stability over time: Reasonable Spearman correlation coefficients (on the overall sample of 94 Fs: $0.81$ for PRUDENT DP and $0.64$ for the HIGH-ENERGY DP; higher correlations among the no-major-change group than in the diet-changers group for both DPs); Bland-Altman method: average change (repeat – initial visit) equal to $0.13$ SD for the PRUDENT DP score and equal to $-0.01$ SD for the HIGH-ENERGY DP; wider LOA for the HIGH-ENERGY than for the PRUDENT DP; narrower LOA in the no-major-change group than in the diet-changers group for both DPs
Chen, 2015 (29)	Canada; CCS, FFQVP	Separate EFAs in the 2 studies: $EIG > 1.5$ , Scree test, $> 50$ % variance explained by a factor, interpretability: Varimax rotation; Loading $> 0.35$ cutoff for CCS and $> 0.5$ cut-off for FFQVP study	54 (3) for the CCS study and 63 (4) for the FFQVP study	Stability over time: no formal assessment	Stability over time: The DPs of the Newfoundland and Labrador adult population have remained reasonably stable over almost a decade, although the PLANT-BASED DP derived from CCS study was a combination of the VEGETABLES/FRUITS DP and the GRAINS DP in the FFQVP study
Crozier, 2009 (15)	UK; SWS	Separate PCAs at 3 time points: standardization; NA criteria for choosing the number of factors; NA rotation; Descriptive labeling; Natural scores calculated with the factor loadings derived at each time point; Applied scores calculated at a follow-up time with loadings obtained from PCA at the baseline time point	14.5 (2) before pregnancy, 14.2 (2) in early pregnancy, and 14.5 (2) in later pregnancy	Stability over time: Spearman correlation coefficient between pairs of DP scores across the 3 time points; Bland-Altman method; formal comparison between natural and applied scores	Stability over time: The identified DPs were strikingly similar at all 3 time points in terms of factor loadings and explained variances; high Spearman correlation coefficients for both natural and applied DP scores before pregnancy and during early pregnancy and late pregnancy (natural scores with range: $0.51-0.81$ ; applied scores with range: $0.52-0.80$ ); Bland-Altman method: minimal change in PRUDENT DP score in early ( $-0.01$ SD; $P = 0.35$ ) and late ( $-0.03$ SD; $P = 0.11$ ) pregnancy compared with before pregnancy; no overall change in HIGH-ENERGY DP score in early pregnancy compared with before pregnancy ( $0.01$ SD; $P = 0.49$ ), but a small significant increase in late pregnancy compared with before pregnancy ( $0.07$ SD; $P = 0.0002$ ); narrower LOA for the PRUDENT score than the HIGH-ENERGY DP score

(Continued)

**TABLE 4 (Continued)**

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Cuco, 2006 (30)	Spain; NA	Separate PCFAs at each of the 6 time points: EIG > 1, Scree test, interpretability; No rotation; Descriptive labeling starting from a 0.20 cutoff	21.48 (2) at preconception, 20.91 (2) at 6th week, 21.64 (2) at 10th week, 24.23 (2) at 26th week, 24.21 (2) at 38th week, and 12.79 (1) at 6th month of the child	Stability over time: CC between loadings from similar DPs across different available time-points; MANOVA for the analysis of consumption trend of dominant FG for each DP using standardized consumptions	Stability over time: coefficients of congruence: for the SWEETENED BEVERAGES AND SUGARS DP, quite high coefficients, ranging between 0.39 and 0.88 in absolute values, with high coefficients also between pregnancy and postpartum periods; for the VEGETABLES AND MEAT DP high coefficients of congruence, ranging between 0.30 and 0.79 in absolute values; analysis of trend in dominant FGs; no significant differences in the standardized mean consumption of dominant FGs for both DPs Stability over time: The same set of 4 DPs found in boys and girls in early and middle adolescence was relatively stable over a 5-y time period; when examining age-matched secular trends in middle adolescents at Time 1 and Time 2, almost identical DPs 5 y apart were identified, except for the FAST-FOOD DP that emerged in the middle adolescent boys at Time 2
Cutler, 2009 (31)	USA (Minnesota); EAT	Separate PCFAs by cohort (older/younger) and sex (boys/girls) based on responses at Time 1 and responses at Time 2; standardization and energy-density transformation; EIG > 1, Scree test, interpretability; Varimax rotation; Descriptive labeling	NA (4 at Time 1, 4 or 5 at Time 2, depending on subgroup)	Stability over time: stability between DPs at Time 1 and Time 2 not formally assessed; secular trends [examined comparing DPs of middle adolescents at Time 1 (older cohort) with DPs in middle adolescents at Time 2 (younger cohort)] not formally assessed	Reproducibility: see (5) for details; Stability over time: 7) stability of DPs over time in terms of contribution of an 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% of total energy contributed in survey 2, 3, and 4, respectively, so 2 changes on 3 were > 5%; 2) Transitions of individuals between DPs over time: 30.7% of the 4007 subjects with complete FFQ information were stable eaters assigned to HIGH-FIBER BREAD DP in all 3 surveys; and 11.1% were stable eaters assigned to LOW-FIBER BREAD DP in all 3 surveys, giving a total of 41.8%; when comparing survey 2 and 4 on the longest time frame (10 y), 57.8% of participants assigned to HIGH-FIBER BREAD DP in both surveys, 15.2% assigned to LOW-FIBER BREAD DP at both surveys, 18.7% went from the HIGH- to LOW-FIBER BREAD DP, and 9.6% went from the LOW- to HIGH-FIBER BREAD DP; among stable eaters over time, no significant differences in percentage of energy intake contributed by important FGs was found during the 10-y period; transitioners had higher relative differences in percentage of energy intake for important FGs than stable eaters (0.27-3.01 as compared with 0.86-1.88)
Dekker, 2013 (10)	Netherlands; Doetinchem Cohort Study	Separate CAs at each of the 3 surveys: percentage energy contributed variables (nutrient density); k-means algorithm; Bootstrap and internal cluster validity indexes (Calinski-Harabasz index, Davies-Bouldin index, and prediction-strength method) to assess the optimal number of clusters to retain between 2 and 6 clusters; Labeling based on FGs that contributed the highest percentage of total energy compared with other DPs within the same survey (>40% higher energy indicated an important FG); Robustness analysis with partitioning around medoids method	Not applicable, 2-cluster solution chosen according to Jaccard similarity and internal cluster validity indexes	Reproducibility: see (5) for details; Stability over time: 1) stability of DPs over time in terms of contribution of an 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% cutoff; 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-y period); relative change in mean percentage of total energy a specific FG contributed from survey 2 to survey 4 between individuals with stable and unstable behavior	Reproducibility: see (5) for details; Stability over time: 1) stability of DPs over time in terms of contribution of an 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% of total energy contributed in survey 2, 3, and 4, respectively, so 2 changes on 3 were > 5%; 2) Transitions of individuals between DPs over time: 30.7% of the 4007 subjects with complete FFQ information were stable eaters assigned to HIGH-FIBER BREAD DP in all 3 surveys; and 11.1% were stable eaters assigned to LOW-FIBER BREAD DP in all 3 surveys, giving a total of 41.8%; when comparing survey 2 and 4 on the longest time frame (10 y), 57.8% of participants assigned to HIGH-FIBER BREAD DP in both surveys, 15.2% assigned to LOW-FIBER BREAD DP at both surveys, 18.7% went from the HIGH- to LOW-FIBER BREAD DP, and 9.6% went from the LOW- to HIGH-FIBER BREAD DP; among stable eaters over time, no significant differences in percentage of energy intake contributed by important FGs was found during the 10-y period; transitioners had higher relative differences in percentage of energy intake for important FGs than stable eaters (0.27-3.01 as compared with 0.86-1.88)

(Continued)

**TABLE 4 (Continued)**

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Fung, 2001 (33)	USA; HPFS	Separate PCFAs at the 3 time points (in 1986, 1990, and 1994); NA criteria for choosing the number of factors; Varimax rotation; Descriptive labeling	NA (2)	Stability over time: Pearson correlation coefficient between scores from similar DPs across time points	Stability over time: The 2 identified DPs were qualitatively similar across time; Pearson correlation coefficient between 1986 and 1990 equal to 0.65 for PRUDENT and 0.70 for WESTERN DP; Pearson correlation coefficient between 1990 and 1994 equal to 0.67 for PRUDENT and 0.69 for WESTERN DP; Pearson correlation coefficient between 1986 and 1994 equal to 0.58 for both PRUDENT and WESTERN DPs
Gerdes, 2002 (34)	Denmark; MONICA	Separate PCFAs at each of the 3 surveys; separate analyses by sex and age group; Scree test, interpretability; Varimax rotation; Descriptive labeling	45 (6) with single survey data	Stability over time: trends in mean DP scores with pooled and age-specific data from linear regression models including time per age interaction term	Stability over time: Profound changes happened in the period, with coarse bread, rice, and pasta much more frequently chosen at the expense of traditional Danish main meals; DP scores showed both variance heterogeneity and heterogeneity in trends across age groups; for Ms, COARSE BREAD and PASTA AND RICE DPs both increased $7 (95\% \text{ CI: } 6-8) \times 10^{-2}$ points/y, i.e., $\sim 0.7$ SDs per 10 y; BAKED GOODS AND SWEETS score increased $4 (95\% \text{ CI: } 3-5) \times 10^{-2}$ points/y; FRUIT AND VEGETABLES DP score did not change; MEAT, POTATOES, AND FAT score declined $4 (95\% \text{ CI: } 3-5) \times 10^{-2}$ points/y; and BREAKFAST declined $2 (95\% \text{ CI: } 1-3) \times 10^{-2}$ points/y; for Fs, survey-specific levels differed from the findings in Ms, notably for COARSE BREAD, FRUIT AND VEGETABLES, and MEAT, POTATOES, AND FAT, but showed the same trends: COARSE BREAD and PASTA AND RICE DP scores increased $6 (95\% \text{ CI: } 5-7) \times 10^{-2}$ and $8 (95\% \text{ CI: } 7-9) \times 10^{-2}$ points/y, respectively; BAKED GOODS AND SWEETS score increased $3 (95\% \text{ CI: } 2-4) \times 10^{-2}$ points/y; FRUIT AND VEGETABLES score remained constant; MEAT, POTATOES, AND FAT score declined $6 (95\% \text{ CI: } 5-7) \times 10^{-2}$ points/y; and BREAKFAST score declined $3 (95\% \text{ CI: } 2-4) \times 10^{-2}$ points/y
Lopez-Garcia, 2004 (35)	USA; NHS	Separate PCFAs on FFO in 1986 and 1990 and average consumption across FFQ data; EIG > 1, Scree test, interpretability; Varimax rotation; Descriptive labeling	NA (2)	Stability over time: no formal assessment	Stability over time: The 2 major DPs were qualitatively similar across time
Malik, 2012 (36)	USA; NHS II	Separate EFAs at the 5 time points (during high school and in adulthood in 1991, 1995, 1999, and 2003); EIG > 1, Scree test, interpretability; Varimax rotation; Loading $\geq 0.30$ cutoff; Adjustment of DP scores by total energy with residual method	NA (2)	Stability over time: Spearman correlation coefficient between scores from similar DPs obtained during high school and in adulthood (cumulative updated average)	Stability over time: The 2 identified DPs were qualitatively similar across time; Spearman correlation between high school and adult DP scores equal to 0.49 for PRUDENT and 0.40 for WESTERN DP
Mikkila, 2005 (16)	Finland; Cardiovascular Risk in Young Finns Study	Separate PCFAs at each of the 3 time points (in 1980, 1986, and 2001); EIG > 1, Scree test, interpretability; Varimax rotation; Alphanumeric labeling; Adjustment of DP scores by total energy with residual method	18 (2) with 1980 data, 21 (2) with 1986 data, and 17 (2) with 2001 data	Stability over time: Spearman correlation coefficient between scores from similar DPs in 1980 and 2001; Tracking analysis (cross-classification); proportion of subjects originally in the lowest or highest quintile of factor scores who remained in the same category over 6 (from 1980 to 1986) or 21 (from 1980 to 2001) y, separately for those who were children (3 to 12 y old) and adolescents (15 to 18 y old) at the beginning of the study	Stability over time: The 2 identified DPs were qualitatively similar across time, over a 21-y period; Spearman correlation coefficient between factor scores in 1980 and 2001 were equal to 0.32 for PATTERN 1 and 0.38 for PATTERN 2; Tracking analysis: the proportion of subjects in the lowest or highest quintile of pattern scores remaining in the same quintile after 6 and 21 y was 1.5 to 2 times the expected in both DPs; if no stability is assumed; tracking was stronger among 15-18-y-old subjects at baseline with 30-42% and 27-41% of subjects originally belonging to the extreme quintile of the energy-adjusted DP scores persisted in the same quintile 6 and 21 y later, respectively; highest stability found in the uppermost quintile in both DPs

(Continued)



**TABLE 4** (Continued)

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Mishra, 2006 (37)	UK; Medical Research Council National Survey of Health and Development (1946 British Birth Cohort)	Separate EFAs at the 3 time points (in adulthood in 1982, 1989, and 1999) on binary data (nonconsumption/consumption); separate analyses by sex; EIG > 1, Scree test, interpretability, root mean square residual; Varimax rotation; Loading $\geq 0.25$ cutoff; Simplified DP scores to calculate individual DP scores in 1982 (36 y) and 1989 (43 y) based on EFA performed in 1999 (53 y)	In 1999, 18.9 (3) among Fs and 17.4 (2) among Ms; in 1982 and 1989: NA (3 for Fs and 2 for Ms)	Stability over time: Number of FGs consumed over time for each DP; Weighted $\kappa$ -coefficient (95% CI) between thirds of DP scores between 1982 and 1989, between 1982 and 1999, and between 1989 and 1999	Stability over time: The identified DPs were similar over time among Ms and Fs; Number of FGs consumed over time: for Fs, increased number of FG consumed in the ETHNIC FOOD AND ALCOHOL and FRUIT, VEGETABLES AND DAIRY DPs, and a decrease in MEAT, POTATOES, AND SWEET FOODS DP; for Ms, number of FGs consumed from both DPs increased significantly over time; fair-to-moderate values of $\kappa$ -coefficients, except for MEAT, POTATOES, AND SWEET FOODS DP, which showed poor agreement in Fs across time
Newby, 2006 (23)	Sweden; SMC	Separate PCFAs at each of the 2 time points: Scree test, interpretability; Varimax rotation; Descriptive labeling; Separate CFAs at each time point: Loading $\geq 0.15$ cutoff based on loadings $\geq 0.20$ cutoff 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 FGs; 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 FGs 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 coefficients: between DP scores from PCFA and CFA at fixed time point were $\geq 0.90$ (all $P < 0.0001$ )
Newby, 2006 (22)	Sweden; SMC	Separate PCFAs at each of the 2 time points: Scree test, interpretability; Varimax rotation; Descriptive labeling; Separate CFAs at each time point: Loading $\geq 0.15$ cutoff based on loadings $\geq 0.20$ cutoff 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 FGs and factor loadings for each DP were seen in 1987 and 1997; some variation was observed for HEALTHY DP
Nimptsch, 2014 (38)	USA; NHS II	Separate EFAs at the 5 time points (during high school and in adulthood in 1991, 1995, 1999 and 2003); EIG > 1, Scree test, interpretability; Varimax rotation; Descriptive labeling; Adjustment of DP scores by total energy with residual method	NA (2)	Stability over time: Pearson correlation coefficient between scores from similar DPs obtained during high school and in adulthood (cumulative updated average)	Stability over time: The 2 identified DPs were qualitatively similar across time; Spearman correlation between high school and adult DP scores equal to 0.48 for PRUDENT and 0.39 for WESTERN DP
Northstone, 2005 (39)	UK; ALSPAC	Separate PCAs on 4- and 7-y data: standardization; Scree test, interpretability; Varimax rotation; Loading $> 0.3$ cutoff	17.7 (3) with 4-y-old children data and 18.3 (3) with 7-y-old children data	Stability over time: no formal assessment scores from PCFA and CFA at fixed time point	Stability over time: The 3 DPs were similar at both time points in terms of loadings and explained variances

(Continued)

**TABLE 4 (Continued)**

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Northstone, 2013 (18)	UK; ALSPAC	Separate CAs at each of the 3 time points; standardization with division by the range; k-means algorithm run 100 times with different starting positions to find the solution with the smallest sum of squares differences; internal stability testing of the final solution; number of clusters ranging from 2 to 6; outliers removed from the analysis at each time point; 62 FGs based on average consumption at each time occasion	Not applicable, 4-cluster solution chosen at each-time point according to internal stability measures based on split-half technique performed 5 times (number of children allocated to a different cluster) and interpretability of the results	Stability over time: changes in mean scores of relevant FGs characterizing the cluster; cross-tabulation of cluster solutions at different ages and proportion of subjects who remained in the same cluster between each pair of ages; sequence index plot to illustrate changes in cluster membership over time	Stability over time: 1) Internal stability based on 5 sets of split-sample testing; 4-cluster solution is the most stable, with <10% misclassified children at each time-point; 2) Changes in mean consumption for relevant FGs: mean amount of FGs consumed within each cluster differed between ages, generally increasing as the children got older, although the patterns of foods consumed and the foods in each cluster with higher- and lower-than-average consumptions were similar at each age; 3) Cross-tabulation of subjects at different ages: reasonably high number of children remaining in the same cluster at different ages (50% and 43% of children in the HEALTHY and PROCESSED clusters, respectively, at age 7 y were in the same clusters at age 13 y; proportion of children who stayed in the same cluster at all 3 ages equal to 20%; for individual clusters, the greatest stability was seen for the HEALTHY cluster at 33%, with the PROCESSED cluster second at 22%; less stable results for TRADITIONAL and PACKED LUNCH clusters, with 25–34% remaining in those clusters over time); 4) Sequence index plot: the most consistent cluster membership over time was for the HEALTHY cluster, followed by the PROCESSED cluster
Northstone, 2008 (17)	UK; ALSPAC	Separate PCAs on 3-, 4-, 7-, and 9-y data on subjects available at each time point and on subjects with information at 4 time points: standardization; Scree test; interpretability; Varimax rotation; Loading >0.3 cutoff	23.4 (4) with 3-y-old children data, 17.7 (3) with 4-y-old children data, 18.1 (3) with 7-y-old children data, and 19.2 (3) with 9-y-old children data	Stability over time: Stability assessed for the PROCESSED, TRADITIONAL, and HEALTH CONSCIOUS DPs; Spearman correlation coefficient between DP scores at each time point, paired t test for the change in mean DP scores between periods of questioning; Bland-Altman method and LOA (95% CI) across time points using z scores of each DP score with mean and SD depending on the comparison under consideration; Cross-classification using quintiles; weighted $\kappa$ -coefficient to compare scores between each pair of time points	Stability over time: High Spearman correlation coefficients between the same DP score at each pair of time points (range: 0.35–0.69, all $P < 0.001$ ), but low Spearman correlation coefficients between different DPs across time, except for the HEALTH CONSCIOUS/VEGETARIAN (9 y only) that negatively correlated with the TRADITIONAL DP at previous time points (but no significant $P$ values); paired t test on mean differences in DP scores across time: consistent increase in the mean PROCESSED DP scores at the later ages compared with 3 y old (all $P < 0.001$ ), but no differences for the other DPs; 95% LOA for the adjusted scores: widest LOA for all pairings between 3- and 9-y-old data, narrowest LOA between 4- and 7-y-old data, narrowest LOA for the HEALTH CONSCIOUS between both 3 and 4 y of age and 7 y of age; weighted $\kappa$ -coefficient: reasonable level of agreement between categorized scores from each time point (range: 0.25–0.47), with higher levels between 4 and 7 y of age and 7 and 9 y of age

(Continued)

**TABLE 4** (Continued)

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Northstone, 2008 (40)	UK; ALSPAC	Separate PCAs on pregnancy data and on 4-y data: standardization; Scree test, interpretability; Varimax rotation; Loading >0.3 cutoff; Natural and applied scores at 4 y with applied scores calculated with loadings obtained from pregnancy data PCA	31.3 (5) with pregnancy data and 25.1 (4) with 4-y follow-up data	Stability over time: Stability assessed for the HEALTH CONSCIOUS, PROCESSED, CONFECTIONARY, and VEGETARIAN DPs; Pearson correlation coefficient between scores from similar DPs; obtained at pregnancy and at 4-y follow-up using both natural and applied scores; paired <i>t</i> test to assess the change in mean scores over the 4-y period between questioning; Bland-Altman method and LOA (95% CI) between scores at the 2 time points; cross-tabulation between pregnancy score quintiles and the 2 (natural and applied) sets of 4-y score quintiles; weighted <i>k</i> -coefficient (95% CI) on quintile of factor scores across time (pregnancy vs 4 y; pregnancy vs applied 4 y; 4 y vs applied 4 y)	Stability over time: Similar Pearson correlation coefficients across DPs for the natural and applied scores, although slightly larger using the applied method; paired <i>t</i> test: considerably lower 4-y applied scores on average as compared with corresponding natural mean scores from the separate PCA at 4-y follow-up, but SDs were much larger with applied scores; Fs decreased their scores on the HEALTH-CONSCIOUS DP over time (mean difference: -0.075 and -0.284; <i>P</i> < 0.0001, with natural and applied scores, respectively), but results for natural and applied scores were inconsistent in sign and/or statistical significance for the other DPs; Bland-Altman method: LOA were wider for applied scores; weighted <i>k</i> -coefficient: reasonable level of agreement (0.267 < <i>k</i> < 0.306) between categorized scores from pregnancy and 4-y natural scores; weighted <i>k</i> -coefficient generally higher when comparing pregnancy and 4-y applied scores; cross-classification: agreement was slightly better for the applied score of the HEALTH-CONSCIOUS DP compared to the 4-y natural score, but this was not true for the PROCESSED DP where the applied score was much less stable than the natural score
Prevost, 1997 (41)	UK; HALS	Separate PCAs on HALS1 (previous publication), HALS2 (current publication) and PCA on the merged dataset including subjects from HALS1 and HALS2; Scree test, Chi-square test of isotropic variation; No rotation; Loading >0.3 cut-off; In the final analysis, HALS2 scores calculated with loadings from PCA on HALS1 data, as factor loadings from HALS2 were identical to those originally derived from the full sample and to the HALS2 subset at HALS1	NA (4)	Stability over time: graphical representation of unadjusted mean DP scores for HALS1 and HALS2 by 7-y age groups at HALS1 (separately for Ms and Fs); unadjusted changes (HALS2 score - HALS1 score) in mean DP scores, with corresponding <i>F</i> test	Stability over time: Marked stability of DPs, in terms of variety of foods consumed, from the 1984-1985 survey to the 1991-1992 survey; graphical representation: COMPONENT 1 (high in fruit and vegetables; low in fat): the scores had risen by HALS2, in each age group, considerably more than would be expected for the 7-y advance in age, with the greatest increase in scores occurring in the youngest subjects ( <i>P</i> -interaction between survey indicator and age at survey < 0.05); COMPONENT 2 (high in energy-dense foods); HALS2 scores were all less than would have been expected for the 7-y advance in age and the score decreases were not uniform across the age groups, but were smaller in the older subjects ( <i>P</i> -interaction < 0.05); COMPONENT 3 (high in convenience foods): in Ms (except those aged 67-73 y) and Fs aged ≥ 39 y at HALS1, the scores had decreased in each age group, but the changes were small and less than expected just for the 7-y advance in age; in the younger Fs there was an increase in score by HALS2, contrary to the expected age trend; COMPONENT 4 (high in sugary foods, low in vegetables): same behavior of DP scores for both Ms and Fs at HALS1 and HALS2 (high in youth and older age, and low in middle age), but HALS2 scores were all higher, and higher than would have been expected for the 7-y advance in age; unadjusted mean scores increased significantly for Ms and Fs on COMPONENT 1 and 4 and fell significantly on COMPONENT 2 (Ms and Fs) and 3 (Ms only) ( <i>P</i> < 0.001)
Schulze, 2006 (42)	USA; NHS II	Separate PCFAs at each of the 3 time points (in adulthood in 1991, 1995, and 1999); EIG > 1, Scree test, interpretability; Varimax rotation; Loading ≥ 0.30 cutoff; Adjustment of DP scores by total energy with residual method	NA (2)	Stability over time: no formal assessment	Stability over time: The 2 identified DPs were qualitatively similar across time

(Continued)

**TABLE 4 (Continued)**

Reference	Location; study	DP identification methods	Explained variance% (number of factors) or CFA/CA model	Assessment of reproducibility/validity	Main results
Togo, 2004	Denmark; MONICA	EFA; on a subsample of the M-82 data (who filled a DR too); Separate analyses by sex; Scree test, interpretable; Varimax rotation; Descriptive labeling; CFA; Loading $\geq 0.30$ cutoff on EFA results; CFA: 3-factor model with correlated factors; CFA performed on M-82 data (all M-82 participants) and on the subgroup including M-82-87 data; to include diet information at 5-y follow-up, CFA performed as a mean-structure factor analysis with group mean factor scores at baseline equal to 0 (but free to be estimated at M-87) and fixed loadings and factor-factor correlations over time; minimization 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-structure factor analysis on the subgroup with data at both time points (M-82-87)	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 (Ms) and the SWEET-TRADITIONAL (Fs) DPs decreased to -0.27 and -0.18, and the SWEET DP (Ms) was virtually unchanged
van Dam, 2002	USA; HPFS	Separate PCFAs at each of the 3 time points (1986, 1990, and 1994); EIG $> 1$ , Scree test, interpretable; Varimax rotation; Descriptive labeling; Robustness analyses to assess the effect of number of factors retained, estimation method, and type of rotation	NA (2)	Stability over time: Pearson correlation coefficient between scores from similar DPs across time points	Stability over time: The 2 major DPs were qualitatively similar across time; Pearson correlation between the PRUDENT DP score was 0.59 between 1986 and 1990, 0.60 between 1990 and 1994, and 0.55 between 1986 and 1994; for the WESTERN DP scores, the Pearson correlation was 0.69 between 1986 and 1990, 0.72 between 1990 and 1994, and 0.64 between 1986 and 1994
Weismayer, 2006 (9)	Sweden; SMC	Separate EFAs at baseline and at follow-up for each of the 4 subgroups: Scree test, interpretability; Varimax rotation; Descriptive labeling; Separate CFAs at baseline and at follow-up for each of the 4 subgroups: Loading $\geq 0.20$ cutoff 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 FGs 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 FGs 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 y, 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 y, respectively; 2) t test: no evidence of a difference in the means for 10, 6, 6, and 2 of 25 FGs after 4, 5, 6, and 7 y, respectively, but evidence that 3, 7, 8, and 11 of the 18 FGs underwent significant changes after 4, 5, 6, and 7 y, respectively ( $P \leq 0.01$ ); 3) Spearman correlation coefficients between baseline and follow-up intakes of FGs consistently decreasing in size over time (no correlation after 7 y exceeding the size of the correlations after 4 y); Internal stability of DPs: no significant instability after 4 and 5 y of follow-up; significant instabilities for WESTERN DP after 6 y ( $P = 0.01$ ) and for WESTERN ( $P = 0.02$ ) and ALCOHOL DPs ( $P = 0.01$ ) after 7 y

<sup>1</sup>ALSPAC, Avon Longitudinal Study of Parents and Children; CA, cluster analysis; CC, congruence coefficient; CCS, case-control study (here intended as the full name of one of the included studies and not as the case-control study design); CFA, confirmatory factor analysis; DP, dietary pattern; DR, dietary record; EAT, Eating Among Teens; EFA, exploratory factor analysis; EIG, eigenvalue; F, female; FFQ, food-frequency questionnaire; FFQ1/FFQ2/FFQ3, food-frequency questionnaire at time 1, 2, or 3; FFQ/P, Food-Frequency Questionnaire Validation Project; FG, food group; HALS, Health and Lifestyle Survey; HPFS, Health Professionals Follow-Up Study; LOA, limits of agreement; M, male; m24HR, mean 24-h recall; MANOVA, multivariate ANOVA; MONICA, Monitoring of trends and determinants in Cardiovascular Disease; NA, not available; NHS, Nurses' Health Study; PCA, principal components analysis; PCFA, principal components factor analysis; SMC, Swedish Mammography Cohort; SWS, Southampton Women's Survey, T.L.G.S., Teheran Lipid and Glucose Study; 24HR, 24-h recall.

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

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

#### ***Summary of the evidence on stability over time: DPs.***

Besides the weak evidence from the 6 articles (22, 29, 31, 35, 39, 42) based on a qualitative assessment, a summary of the evidence from articles formally evaluating DP stability is provided below. In addition (31, 39), the stability of DPs from childhood onwards was formally evaluated in 3 articles (16–18), with 2 of them exploring the issue in subjects who moved from childhood to adolescence (18) or from childhood/adolescence to adulthood (16). The main conclusions were as follows: 1) during childhood, the identified DPs were very stable, with the highest agreement found between successive waves (4 and 7 y; 7 and 9 y) and for the Health-conscious DP (17); 2) from childhood to adolescence, the number of children remaining in the same cluster across time-occasions was still reasonably high, with the greatest stability found for the Healthy cluster (33% of subjects in the same cluster at all 3 ages) (18); 3) from childhood/adolescence to adulthood (~20-y period), both the correlation coefficients between time-specific scores and the proportion of subjects remaining in the extreme quintiles over time pointed to DP stability, with the highest stability found for the uppermost quintile category of subjects and for the subjects aged 15–18 y old at baseline (16).

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

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

Finally, 11 articles (9, 10, 23, 25, 27, 28, 33, 34, 37, 41, 45) assessed the stability of DPs identified in successive waves on men and/or women. Three of them (25, 27, 34) showed instability over time for most or all the identified DPs. In detail, at 12 y from the validation study of the Teheran Lipid and Glucose Study, the Iranian traditional DP was found to be unreproducible according to all criteria,

whereas quintile categories of the Western DP showed poor agreement over time (25). Going from the 1982–1984 to the 1987–1988 survey of the Danish MONItoring of trends and determinants in Cardiovascular Disease (MONICA) Study (27), increasing mean scores were found for the Green DP, but the Traditional (in men) and the Sweet-Traditional (in women) DPs showed decreased mean scores, within an overall mean-structure CFA model. However, while going from the 1982–1984 to the 1991–1992 survey of the Danish MONICA study, both men and women showed the same trend of increasing consumption of Coarse Bread, Pasta, and Rice and Baked Goods and Sweets DPs at the expense of a decrease in mean intakes of the Meat, Potatoes, and Fats DP and the Breakfast DP (34). In 1 pioneering article that compared 2 consecutive US surveys (41), 2 [component 1 (high in fruit and vegetables) and component 4 (high in sugary foods)] out of the 4 identified DPs increased over time more than would have been expected for the 7-y advance in age.

A weaker form of instability over time concerned single DPs within a set of substantially stable DPs. This issue was evident across the 1980s and 1990s for the Meat, Potatoes, and Sweet Foods DP in 36-y-old females from the United Kingdom over 17 y of follow-up (37) and for the Western/Swedish DP in 52-y-old females from Sweden over 9 y of follow-up (23). Finally, several studies (9, 10, 28, 33, 45) showed good stability of all DPs found during adulthood.

When identified (SMC study) (9, 23), the Alcohol DP showed the best reproducibility; however, the more refined analysis of changes in the time-specific covariances matrices revealed instability after 7 y in one of the articles (9). With 2 exceptions (23, 37), the Western-like (e.g., Western; High-energy; Low-fiber Bread; Meat, Potatoes, and Sweet Foods; and Western/Swedish) and the Prudent-like (e.g., Prudent; High-fiber Bread; Healthy; and Fruit, Vegetables, and Dairy) DPs generally showed a similar and moderate stability over time. Traditional-like DPs (e.g., Iranian Traditional, Sweet-Traditional, and Traditional) were less likely to be stable over time (18, 25, 27, 40).

In addition, most of the articles with  $\geq 3$  measurement occasions [i.e., (16, 17, 33, 37, 45)] showed that the agreement was higher when the DPs were identified on data from consecutive, as compared with nonconsecutive, waves.

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

#### ***Summary of the evidence on stability over time: relevant food groups within DPs.***

The analysis of trends of consumption of relevant food groups within each DP (9, 10, 18, 23, 30, 37) supported or further strengthened results on DP stability over time. When the DPs were stable (9, 10, 30), no material differences in mean consumption of relevant food groups were found in 1 article (30) or less than a half of them underwent significant



changes (9, 10). When 1 DP was not stable over time (23, 37), the mean intakes (23) [or the number (37)] of relevant food groups changed over time, and this also had an impact on relevant food groups for other DPs over time; a change might also occur in the number of relevant food groups that characterized stable DPs over time, reflecting an increasing variety in consumption over time within the same DP (37). When moving from childhood to adolescence, the mean amount of food groups consumed generally increased over time, but the foods in each cluster with higher- and lower-than-average consumptions were similar at each age (18).

## Conclusions

The present scoping review provides a preliminary summary of the current results on reproducibility of a posteriori DPs across studies and over long time periods. The evidence collected is still limited, with only 9 articles identified on cross-study reproducibility. In addition, only 55% (cross-study reproducibility) and 76% (stability over time) of the articles adopted a formal statistical approach, which, however, relied on elementary statistics (i.e., correlation coefficients) in most of the cases and on a statistical model in 3 articles only. Based on the evidence collected, most identified DPs (in particular, Alcohol, Prudent, and Western DPs) showed good reproducibility across studies and stability over time.

The assessment of cross-study reproducibility has gained recent attention in the literature (8, 11, 12, 26, 32), after some sparse pioneering attempts in the 1980s (43, 44) and 2000s (6, 7). Recent articles (8, 11, 12, 26, 32) have definitely confirmed the merits of the assessment of cross-study reproducibility of PCA/EFA-based DPs. Besides having found a high congruence between apparently similar pairs of DPs in terms of food composition and association with cancer risk, some novelties in methods (11, 12, 32) have been introduced. These include multistudy factor analysis (13) [when individual-level dietary data are available, see the corresponding R package “MSFA” (13) from GitHub] and the approach by Castello and colleagues (11, 12) [when published factor-loading matrices and food-grouping schemes are available, see the Supplementary Material of reference (11)]. Moreover, following 2 articles identified in the current review (26, 30), Castello and colleagues (11, 12) popularized the use of the congruence coefficient between factor loadings to assess DP similarity. In addition to set up specific cutoffs to identify DP similarity or equivalence, they showed that the congruence coefficient outperforms the correlation coefficient between factor scores and overcomes the misuse of its statistical significance.

Although the assessment of cross-study reproducibility has undergone a major improvement in statistical methods, researchers have still to face with the interpretation of similarities and differences across centers/studies: which latent factors (e.g., climate, influence of media or society, or food supply) are responsible for the identification of DPs in a country, but not in another one, or for the different variants of similar DPs across countries? For example, given

the same climate and food supply, groups with different age, religion, ethnicity, or socioeconomic background may show different versions of a similar DP (4). Similarly, sources of beneficial or detrimental nutrients differ across populations or subpopulations with varied age, ethnicity, or socioeconomic background. For example, in 10 case-control studies from the International Head and Neck Cancer Epidemiology Consortium (47), we have shown that the primary sources of vitamin C were different across countries: within the European studies subjects mainly derived natural vitamin C from citrus fruits, kiwi, tomatoes, green salad, and apples/pears, whereas, in the US studies, fruit juices and potatoes were relevant contributors, too. Within countries, sources were different in (otherwise comparable) populations from urban or rural areas (e.g., miso soup in the rural, vegetables and green tea in a more industrialized area from Japan), among young people or blacks from the United States, where fortified drinks and Southern greens were the major contributors of vitamin C, respectively. Besides the complexity of DP analysis, these considerations suggest the importance of working at a subpopulation level and the need for statistical criteria assessing similarity of subpopulation-specific DPs to allow for the merging of data from different subpopulations.

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

The analysis of DP stability can be very complicated. For example, research can target the individual- and/or the population-specific levels of stability and can assess stability of the identified DPs and/or the relevant food groups. Also, the stability of DPs identified across different life-course periods can be the focus of the research [e.g., (15, 16, 31)]. Even when considering adults only, differences in the study designs arose from subjects' age at baseline, the time intervals between successive waves, and the maximum time interval between the first and the last wave considered. In addition, the statistical methods used for the assessment of DP stability differ markedly across articles: 25% of them did not use any statistical procedure (but simply inspected the factor-loading matrices over time), whereas 50% considered 2 criteria.

Within this complicated scenario, we can only comment on some preliminary results. First, the closer the examined waves of dietary information collection are the better is the stability of the identified DPs. This conclusion is very well supported, without any restriction on the statistical approach used for the analysis. When the dietary assessment tool, subject's life, and the DP identification process are stable over successive administrations, DP instabilities are either unexpected or due to essential and timely modifications of diet-related policies (e.g., the ban on *trans* fats), which lead to changes in behavior and food-product development and marketing (4). Second, in 75% of the articles, the number

of identified DPs and the percentage of explained variance were substantially stable over time. We can conclude that, to date, overall dietary habits have been generally expressed in a stable number of constructs over time, with a few new or lost DPs over 10 or 20 y. Also, the ability of the identified DPs to capture the overall variance did not change over time, although the relative importance of the single DPs (in terms of percentage of the total variance explained) may vary. Third, within an identified DP, the correlation structure among food groups is still stable over time, although changes in relevant food groups have been reported in more refined statistical analyses. DPs are more likely to evolve, rather than disappear or emerge as brand-new ones. This conclusion may reflect the combination of several aspects. Among the most relevant ones, we mention early-life experiences with various tastes and flavors and parental feeding practices, which tend to persist over the lifespan (48); however, later food choices could be influenced by media/society or aging. At a population-level, several other factors may influence the potential evolution of DPs over time, including changes in food supply (e.g., preferences for ethnic foods) as well as in nutrition-related policies. For example, we might hypothesize that the ban on *trans* fats will favor a change in the DP structure of those putative DPs named Snacks, or Sweets, or Desserts (based on bakery products, baked goods, commercially fried foods, and spreads, which are likely to contain *trans* fats) in favor of similar processed foods made with nonhydrogenated oils.

Evidence from the current review is still too limited to provide a firm conclusion on the most suitable timeframe to administer successive dietary assessment tools within longitudinal studies or repeated surveys. In the absence of major life changes in the target population, DPs still show a good stability within 6–7 y after the previous dietary assessment; however, within a more refined statistical model, marked signs of instability were found after the same number of years for 1 (at 6 y) or 2 (at 7 y) DPs, but not for the last DP identified on the same dataset (9). Thus, scheduling successive administrations of the dietary assessment tool every 4 y, like in the NHS II, and updating the Dietary Guidelines for Americans every 5 y are recommended strategies to monitor DPs at their maximum potential stability over time.

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

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

Future efforts should be directed on defining the generalizability of a posteriori DPs within a statistical model where time or study variables are explicitly modeled and the selection of the type and number of DPs to retain at each measurement occasion is carried out borrowing information across any levels of the analysis. The use of multistudy factor analysis (13) in nutritional epidemiology (32) has provided an example of a fruitful application of a novel statistical modeling strategy to tackle cross-study reproducibility of a posteriori DPs. Similarly, multilevel latent class analysis (49) may offer insights in cross-study reproducibility, and latent class transition models (i.e., latent Markov models) (50) can offer a natural framework to track changes in DPs over time. These possibilities rely not only on statistical skills but also on an effort of integration of study protocols and data. As far as studies are conceived as isolated attempts at knowledge, any assessment of reproducibility will likely end up into a unified but distorted combination of results from separate studies with their own decisions and limitations. In the short term, as researchers, we can at least contribute to expanding a general culture of reproducibility by assessing the reproducibility of DPs according to a series of different criteria, although based on elementary statistics.

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

## Acknowledgments

The authors' responsibilities were as follows—VE and MF: designed the research; VE and MD: collected the relevant articles and selected those to be included in the systematic review; MD and LP: prepared the first draft of Table 2 and of part of Tables 3 and 4; VE, RDV, and MF: completed and refined Tables 3 and 4; AS and FB: revised all the tables and checked their consistency with the text; AS: prepared Supplemental Figure 1 and prepared Table 1; MD: prepared Figure 1; VE: wrote the manuscript and had primary responsibility for final content; and all authors: read and approved the final manuscript.

## References

1. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002;13(1):3–9.
2. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* 2004;62(5):177–203.
3. Weikert C, Schulze MB. Evaluating dietary patterns: the role of reduced rank regression. *Curr Opin Clin Nutr Metab Care* 2016;19(5):341–6.
4. Tucker KL. Dietary patterns, approaches, and multicultural perspective. *Appl Physiol Nutr Metab* 2010;35(2):211–8.
5. Edefonti V, De Vito R, Dalmartello M, Patel L, Salvatori A, Ferraroni M. Reproducibility and validity of a posteriori dietary patterns: a systematic review. *Adv Nutr* 2020;11(2):293–326.
6. Balder HF, Virtanen M, Brants HA, Krogh V, Dixon LB, Tan F, Mannisto S, Bellocco R, Pietinen P, Wolk A, et al. Common and country-specific dietary patterns in four European cohort studies. *J Nutr* 2003;133(12):4246–51.
7. Mannisto S, Dixon LB, Balder HF, Virtanen MJ, Krogh V, Khani BR, Berrino F, van den Brandt PA, Hartman AM, Pietinen P, et al. Dietary patterns and breast cancer risk: results from three cohort studies in the DIETSCAN project. *Cancer Causes Control* 2005;16(6):725–33.
8. Moskal A, Pisa PT, Ferrari P, Byrnes G, Freisling H, Boutron-Ruault MC, Cadeau C, Nailler L, Wendt A, Kuhn T, et al. Nutrient patterns and their food sources in an international study setting: report from the EPIC study. *PLoS One* 2014;9(6):e98647.
9. Weismayer C, Anderson JG, Wolk A. Changes in the stability of dietary patterns in a study of middle-aged Swedish women. *J Nutr* 2006;136(6):1582–7.
10. Dekker LH, Boer JM, Stricker MD, Busschers WB, Snijder MB, Nicolaou M, Verschuren WM. Dietary patterns within a population are more reproducible than those of individuals. *J Nutr* 2013;143(11):1728–35.
11. Castello A, Buijse B, Martin M, Ruiz A, Casas AM, Baena-Canada JM, Pastor-Barrisio R, Antolin S, Ramos M, Munoz M, et al. Evaluating the applicability of data-driven dietary patterns to independent samples with a focus on measurement tools for pattern similarity. *J Acad Nutr Diet* 2016;116(12):1914–24 e6.
12. Castello A, Lope V, Vioque J, Santamarina C, Pedraz-Pingarron C, Abad S, Ederria M, Salas-Trejo D, Vidal C, Sanchez-Contador C, et al. Reproducibility of data-driven dietary patterns in two groups of adult Spanish women from different studies. *Br J Nutr* 2016;116(4):734–42.
13. De Vito R, Bellio R, Trippa L, Parmigiani G. Multi-study factor analysis. *Biometrics* 2019;75(1):337–46.
14. Murakami K, Shinozaki N, Fujiwara A, Yuan X, Hashimoto A, Fujihashi H, Wang HC, Livingstone MBE, Sasaki S. A systematic review of principal component analysis-derived dietary patterns in Japanese adults: are major dietary patterns reproducible within a country? *Adv Nutr* 2019;10(2):237–49.
15. Crozier SR, Robinson SM, Godfrey KM, Cooper C, Inskip HM. Women's dietary patterns change little from before to during pregnancy. *J Nutr* 2009;139(10):1956–63.
16. Mikkila V, Rasanen L, Raitakari OT, Pietinen P, Viikari J. Consistent dietary patterns identified from childhood to adulthood: the Cardiovascular Risk in Young Finns Study. *Br J Nutr* 2005;93(6):923–31.
17. Northstone K, Emmett PM. Are dietary patterns stable throughout early and mid-childhood? A birth cohort study. *Br J Nutr* 2008;100(5):1069–76.
18. Northstone K, Smith AD, Newby PK, Emmett PM. Longitudinal comparisons of dietary patterns derived by cluster analysis in 7- to 13-year-old children. *Br J Nutr* 2013;109(11):2050–8.
19. Northstone K, Ness AR, Emmett PM, Rogers IS. Adjusting for energy intake in dietary pattern investigations using principal components analysis. *Eur J Clin Nutr* 2008;62(7):931–8.
20. Castro MA, Baltar VT, Selem SS, Marchioni DM, Fisberg RM. Empirically derived dietary patterns: interpretability and construct validity according to different factor rotation methods. *Cad Saude Publica* 2015;31(2):298–310.
21. Varraso R, Garcia-Aymerich J, Monier F, Le Moual N, De Batlle J, Miranda G, Pison C, Romieu I, Kauffmann F, Maccario J. Assessment of dietary patterns in nutritional epidemiology: principal component analysis compared with confirmatory factor analysis. *Am J Clin Nutr* 2012;96(5):1079–92.
22. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. *J Nutr* 2006;136(10):2580–7.
23. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Long-term stability of food patterns identified by use of factor analysis among Swedish women. *J Nutr* 2006;136(3):626–33.
24. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P. Preferred Reporting Items for Systematic Review and Meta-Analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
25. Asghari G, Rezazadeh A, Hosseini-Esfahani F, Mehrabi Y, Mirmiran P, Azizi F. Reliability, comparative validity and stability of dietary patterns derived from an FFQ in the Tehran Lipid and Glucose Study. *Br J Nutr* 2012;108(6):1109–17.
26. Judd SE, Letter AJ, Shikany JM, Roth DL, Newby PK. Dietary patterns derived using exploratory and confirmatory factor analysis are stable and generalizable across race, region, and gender subgroups in the REGARDS study. *Front Nutr* 2014;1:29.
27. Togo P, Osler M, Sorensen TI, Heitmann BL. A longitudinal study of food intake patterns and obesity in adult Danish men and women. *Int J Obes* 2004;28(4):583–93.
28. Borland SE, Robinson SM, Crozier SR, Inskip HM, Group SWSS. Stability of dietary patterns in young women over a 2-year period. *Eur J Clin Nutr* 2008;62(1):119–26.
29. Chen Z, Wang PP, Shi L, Zhu Y, Liu L, Gao Z, Woodrow J, Roebotham B. Comparison in dietary patterns derived for the Canadian Newfoundland and Labrador population through two time-separated studies. *Nutr J* 2015;14:75.
30. Cuco G, Fernandez-Ballart J, Sala J, Viladrich C, Iranzo R, Vila J, Arijia V. Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. *Eur J Clin Nutr* 2006;60(3):364–71.
31. Cutler GJ, Flood A, Hannan P, Neumark-Sztainer D. Major patterns of dietary intake in adolescents and their stability over time. *J Nutr* 2009;139(2):323–8.
32. De Vito R, Lee YCA, Parpinel M, Serraino D, Olshan AF, Zevallos JP, Levi F, Zhang ZF, Morgenstern H, Garavello W, et al. Shared and study-specific dietary patterns and head and neck cancer risk in an international consortium. *Epidemiology* 2019;30(1):93–102.
33. Fung TT, Rimm EB, Spiegelman D, Rifai N, Tofler GH, Willett WC, Hu FB. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *Am J Clin Nutr* 2001;73(1):61–7.
34. Gerdes LU, Bronnum-Hansen H, Osler M, Madsen M, Jorgensen T, Schroll M. Trends in lifestyle coronary risk factors in the Danish MONICA population 1982–1992. *Public Health* 2002;116(2):81–8.
35. Lopez-Garcia E, Schulze MB, Fung TT, Meigs JB, Rifai N, Manson JE, Hu FB. Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. *Am J Clin Nutr* 2004;80(4):1029–35.
36. Malik VS, Fung TT, van Dam RM, Rimm EB, Rosner B, Hu FB. Dietary patterns during adolescence and risk of type 2 diabetes in middle-aged women. *Diabetes Care* 2012;35(1):12–8.
37. Mishra GD, McNaughton SA, Bramwell GD, Wadsworth ME. Longitudinal changes in dietary patterns during adult life. *Br J Nutr* 2006;96(4):735–44.
38. Nimptsch K, Malik VS, Fung TT, Pischon T, Hu FB, Willett WC, Fuchs CS, Ogino S, Chan AT, Giovannucci E, et al. Dietary patterns during high school and risk of colorectal adenoma in a cohort of middle-aged women. *Int J Cancer* 2014;134(10):2458–67.
39. Northstone K, Emmett P. Multivariate analysis of diet in children at four and seven years of age and associations with socio-demographic characteristics. *Eur J Clin Nutr* 2005;59(6):751–60.

40. Northstone K, Emmett PM. A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis. *Br J Nutr* 2008;99(5):1099–106.
41. Prevost AT, Whichelow MJ, Cox BD. Longitudinal dietary changes between 1984–5 and 1991–2 in British adults: association with socio-demographic, lifestyle and health factors. *Br J Nutr* 1997;78(6):873–88.
42. Schulze MB, Fung TT, Manson JE, Willett WC, Hu FB. Dietary patterns and changes in body weight in women. *Obesity (Silver Spring)* 2006;14(8):1444–53.
43. Schwerin HS, Stanton JL, Riley AM, Jr, Schaefer AE, Leveille GA, Elliott JG, Warwick KM, Brett BE. Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. *Am J Clin Nutr* 1981;34(4):568–80.
44. Schwerin HS, Stanton JL, Smith JL, Riley AM, Jr, Brett BE. Food, eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. *Am J Clin Nutr* 1982;35(5 Suppl):1319–25.
45. van Dam RM, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Dietary patterns and risk for type 2 diabetes mellitus in U.S. men. *Ann Intern Med* 2002;136(3):201–9.
46. Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *Br J Nutr* 2003;89(3):409–19.
47. Edefonti V, Hashibe M, Parpinel M, Turati F, Serraino D, Matsuo K, Olshan AF, Zevallos JP, Winn DM, Moysich K, et al. Natural vitamin C intake and the risk of head and neck cancer: a pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Int J Cancer* 2015;137(2):448–62.
48. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing children's eating behaviours. *Nutrients* 2018;10(6). doi: 10.3390/nu10060706.
49. Vermunt JK. Multilevel latent class models. *Soc Method* 2003;33(1):213–39.
50. Sotres-Alvarez D, Herring AH, Siega-Riz AM. Latent transition models to study women's changing of dietary patterns from pregnancy to 1 year postpartum. *Am J Epidemiol* 2013;177(8):852–61.
51. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 2008;31(4):466–75.
52. Ryman TK, Boyer BB, Hopkins S, Philip J, O'Brien D, Thummel K, Austin MA. Characterising the reproducibility and reliability of dietary patterns among Yup'ik Alaska Native people. *Br J Nutr* 2015;113(4):634–43.
53. Willett W. *Nutritional Epidemiology*. Oxford University Press; 2012.
54. Beck KL, Kruger R, Conlon CA, Heath AL, Coad J, Matthys C, Jones B, Stonehouse W. The relative validity and reproducibility of an iron food frequency questionnaire for identifying iron-related dietary patterns in young women. *J Acad Nutr Diet* 2012;112(8):1177–87.
55. Ambrosini GL, O'Sullivan TA, de Klerk NH, Mori TA, Beilin LJ, Oddy WH. Relative validity of adolescent dietary patterns: a comparison of a FFQ and 3 d food record. *Br J Nutr* 2011;105(4):625–33.
56. Bountziouka V, Tzavelas G, Polychronopoulos E, Constantinidis TC, Panagiotakos DB. Validity of dietary patterns derived in nutrition surveys using a priori and a posteriori multivariate statistical methods. *Int J Food Sci Nutr* 2011;62(6):617–27.
57. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* 1999;69(2):243–9.
58. Strauss ME. Introduction to the special section on construct validity of psychological tests: 50 years after Cronbach and Meehl (1955). *Psychol Assess* 2005;17(4):395.