The American Journal of Clinical Nutrition 123 (2026) 101263

The American Journal of CLINICAL NUTRITION

journal homepage: https://ajcn.nutrition.org/

Original Research Article

Validity and reliability of tools to measure ultraprocessed food intake under the NOVA system: A systematic review

Ivan Cavero-Redondo [1,†], Alicia Saz-Lara [1,**,†], Cristina Bouzas [2,3], Andrea del Saz-Lara [4], Gemma Chiva-Blanch [2,5], Jadwiga Konieczna [2,6], Catalina Picó [2,7], Fernando Fernández-Aranda [2,8], María Jesús Moreno-Aliaga [2,9], Lidia Daimiel [2,10], José Alfredo Martínez [2,11], Jordi Salas-Salvadó [2,12], María P Portillo [2,13], Estefanía Toledo [2,14], Marta Giralt [2,15], Dolores Corella [2,16], Isabel Moreno-Indias [2,17], María-Carmen López de Las Hazas [18], Alberto Dávalos [2,18,*]

[1] *CarVasCare Research Group, Instituto de Biomedicina/Facultad de Enfermería de Cuenca, Universidad de Castilla-La Mancha, Cuenca, Spain;* [2] *Consorcio CIBER de la Fisiopatología de la Obesidad y Nutrición (CIBERObn), Instituto de Salud Carlos III (ISCIII), Madrid, Spain;* [3] *Research Group on Community Nutrition and Oxidative Stress, University of the Balearic Islands-IUNICS, Palma de Mallorca, Spain, Health Research Institute of the Balearic Islands (IdISBa), Palma de Mallorca, Spain;* [4] *Facultad de Ciencias de la Salud, UTAMED (Universidad Tecnológica Atlántico Mediterráneo), Málaga, Spain;* [5] *NUTRALiSS Research Group, Health Sciences Faculty, Universitat Oberta de Catalunya (UOC), Barcelona, Spain;* [6] *Research Group on Nutritional Epidemiology and Cardiovascular Physiopathology (NUTRECOR), Health Research Institute of the Balearic Islands (IdISBa), University Hospital Son Espases, Palma de Mallorca, Spain;* [7] *Nutrigenomics, Biomarkers and Risk Evaluation (NuBE) Group, University of the Balearic Islands (UIB), Palma, Mallorca, Spain, Health Research Institute of the Balearic Islands (IdISBa), Palma, Mallorca, Spain, Artificial Intelligence Research Institute of the Balearic Islands (IAIB), Palma, Mallorca, Spain;* [8] *Psychoneurobiology of Eating and Addictive Behaviours Group, Neurosciences Programme, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain, Department of Clinical Sciences, School of Medicine and Health Sciences, University of Barcelona, L'Hospitalet de Llobregat, Spain;* [9] *Center for Nutrition Research, Department of Nutrition, Food Science and Physiology, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Spain, IdiSNA, Navarra Institute for Health Research, Pamplona, Spain;* [10] *Nutritional Control of the Epigenome Group, Precision Nutrition and Obesity Program, IMDEA Nutrition, Campus of International Excellence (CEI) UAM + CSIC, Madrid, Spain;* [11] *Precision Nutrition and Cardiometabolic Health, IMDEA Nutrition, CEI UAM + CSIC, Madrid, Spain;* [12] *Departament de Bioquímica i Biotecnologia, Alimentació, Nutrició, Desenvolupament i Salut Mental (ANUT-DSM), Institut d'Investigació Sanitària Pere Virgili (IISPV), Universitat Rovira i Virgili, Reus, Spain;* [13] *Department of Nutrition and Food Science, Nutrition and Obesity Group, Faculty of Pharmacy, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain, BIOARABA Health Research Institute, Vitoria-Gasteiz, Spain;* [14] *Department of Preventive Medicine and Public Health, Center for Nutrition Research, School of Medicine, University of Navarra, Pamplona, Spain, IdiSNA, Navarra Institute for Health Research, Pamplona, Spain;* [15] *Departament de Bioquímica i Biomedicina Molecular, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain, Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain;* [16] *Department of Preventive Medicine and Public Health, University of Valencia, Valencia, Spain;* [17] *Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA, BIONAND, Málaga, Spain, Unidad de Gestión Clínica de Endocrinología y Nutrición, Hospital Universitario Virgen de la Victoria, Málaga, Spain;* [18] *Laboratory of Epigenetics of Lipid Metabolism, Instituto Madrileño de Estudios Avanzados (IMDEA) Nutrition, CEI UAM + CSIC, Madrid, Spain*

A B S T R A C T

Background: Ultraprocessed foods (UPFs) contribute substantially to global energy intake and are consistently linked to adverse health outcomes. However, accurately assessing UPF intake remains challenging due to reliance on self-reported data and retrospective processing classification, with performance highly dependent on reference method quality.

Objectives: To evaluate the validity, reliability, agreement, and risk of bias of instruments used to quantify UPF intake according to the NOVA classification and to examine the agreement between food processing classification systems.

Methods: We conducted a 2020 PRISMA–compliant systematic review of PubMed/MEDLINE, Scopus, and Web of Science from inception to 12 January, 2026, without language restrictions. Two reviewers (IC-R and AS-L) independently screened studies, extracted data, and assessed

Abbreviations: FFQ, food frequency questionnaire; ICC, intraclass correlation coefficient; PABAK, prevalence-adjusted bias-adjusted kappa; UPF, ultraprocessed food; κ, Cohen's kappa.

* Corresponding author.

** Corresponding author.

E-mail addresses: alberto.davalos@imdea.org (A. Dávalos), ivan.cavero@uclm.es (I. Cavero-Redondo), alicia.delsaz@uclm.es (A. Saz-Lara).

† IC-R and AS-L contributed equally to this work.

https://doi.org/10.1016/j.ajcnut.2026.101263

Received 11 September 2025; Received in revised form 18 January 2026; Accepted 3 March 2026; Available online 6 March 2026

methodological quality via the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) risk of bias checklist. The evidence was synthesized narratively and through structured visual summaries of test–retest reliability (intraclass correlation coefficients), relative (criterion/convergent) validity against dietary reference methods, classification agreement (Cohen's κ and prevalence-adjusted bias-adjusted κ), and internal consistency (Cronbach's α or McDonald's ω).

Results: Thirty studies were included: 19 evaluated basic dietary intake instruments (NOVA-oriented food frequency questionnaires, short screening tools, or UPF scores), 4 examined biomarker-based or indirect tools, and 7 assessed contextual or auxiliary instruments. The test–retest reliability of the dietary intake instruments was generally moderate-to-high (intraclass correlation coefficients ≈ 0.46–0.94). Relative validity against dietary reference methods was modest to moderate (correlations typically r ≈ 0.47–0.72), and agreement between processing classification systems ranged from fair-to-substantial (Cohen's κ and prevalence-adjusted bias-adjusted κ ≈ 0.36–0.84). The contextual and auxiliary instruments showed high internal consistency (Cronbach's α/McDonald's ω ≥0.74). Fifteen studies were rated as having a low risk of bias, and 15 were rated as unclear.

Conclusions: Instruments assessing NOVA-defined UPF intake demonstrate acceptable reproducibility but variable, reference-dependent validity and system-specific classification uncertainty. Future research should prioritize higher-quality reference methods, transparent coding rules, calibration strategies, and routine evaluation of measurement error and responsiveness.

This trial was registered at PROSPERO CRD420251135087.

Keywords: intraclass correlation coefficient, NOVA, psychometric properties, reliability, validity, ultraprocessed foods

Introduction

Ultraprocessed foods (UPFs), defined as industrial formulations that combine refined ingredients with additives to increase palatability, shelf life, and convenience, constitute a substantial proportion of dietary energy intake in many countries worldwide. A growing body of epidemiologic evidence links increased UPF consumption to adverse cardiometabolic outcomes, certain cancers, neurocognitive disorders, and increased all-cause mortality [1]. Consequently, accurate and reproducible quantification of UPF intake is essential for population surveillance, etiologic research, and policy evaluation [2].

Operationalizing food processing within nutritional epidemiology remains challenging. Detailed information on industrial processing, recipes, and ingredient composition is often proprietary and unavailable, limiting classification precision [3]. As a result, most studies rely on self-reported dietary instruments, such as food frequency questionnaires (FFQs), UPF screening tools, and 24-h dietary recalls, which were developed before food processing became a central research construct and are retrospectively mapped to processing categories [4].

Multiple food processing classification systems coexist, including NOVA, the University of North Carolina system, the International Food Information Council framework, and the International Agency for Research on Cancer processing scheme. Applying different taxonomies to identical dietary data can yield markedly different UPF exposure estimates [5]. Moreover, UPF-related instruments have frequently prioritized feasibility over measurement rigor: item selection, scoring rules, reference methods, and reported psychometric metrics vary widely. These features complicate synthesis and raise concerns regarding misclassification, responsiveness, and cross-cultural validity [6].

Recent years have seen the development of NOVA-focused FFQs, short UPF screening tools, and cohort-level UPF scores derived from existing dietary data. Complementary approaches, including metabolomic proxies, behavior-based measures, and food environment audits, have also emerged [7–12]. However, the available evidence indicates substantial variability in reproducibility and in criterion and construct (convergent) validity across instruments, populations, and settings [7,8]. Agreement between processing classification systems is often only moderate [5,13], and performance may decline when tools are adapted across cultural contexts or applied to children or older adults [14–17]. Together, these findings suggest that existing instruments may yield imperfect estimates of UPF exposure [8].

Beyond these measurement challenges, an ongoing conceptual debate surrounds the definition of "ultraprocessed" foods, particularly within the NOVA framework. Critics argue that some definitional elements lack fully objective, operationalizable criteria, introduce subjectivity into food classification, and contribute to system-dependent exposure estimates (indicating that associations may be influenced by systematic classification and measurement error specific to a given system rather than reflecting true causal effects) [18].

Therefore, the present systematic review focuses on instruments used to quantify NOVA-defined UPF intake, not because NOVA represents the gold standard but rather because it is the most widely applied processing framework in epidemiologic research. This pragmatic focus enables structured synthesis while acknowledging the contested nature of the construct and the existence of alternative taxonomies. The review pursued 2 objectives: (objective 1) to assess the measurement properties of instruments quantifying NOVA-defined UPF intake and (objective 2) to examine the agreement between processing classification systems applied to the same dietary data. Instruments were organized into dietary intake instruments, broader UPF-related measures, and contextual or auxiliary tools informing determinants of UPF exposure.

Methods

Design and registration

We conducted a systematic review in accordance with PRISMA 2020 [19] and the Cochrane Handbook for Systematic Reviews [20]. A completed PRISMA 2020 checklist is provided in Supplementary Table 1, indicating the page numbers where each reporting item is addressed. The review protocol was registered in PROSPERO (CRD420251135087) after completion of the literature search but prior to data extraction and synthesis.

Eligibility criteria

Study eligibility was defined via a Population, Exposure, Comparator, and Outcome (PECO)-based strategy adapted for measurement research, in which the comparator component was not needed, and the study design was explicitly specified as part of the framework.

The inclusion criteria were as follows: *1*) Population (P): children, adolescents, or adults from community-based or clinical settings. *2*) Exposure (E): assessment of UPF intake or UPF-related constructs using instruments explicitly linked to food processing classifications (e.g., NOVA or comparable processing-based systems). *3*) Outcome (O): reporting ≥1 measurement property of the instrument (e.g.,

validity, reliability, agreement, responsiveness), agreement between processing classification systems applied to the same dietary data, and/or associations with processing categories or percentage of energy (%E) from UPF. *4*) Study design (S): cross-sectional studies or cross-sectional analyses (including validation studies, reproducibility or test–retest studies, method comparison studies, and instrument development studies) conducted within cohorts or randomized controlled trials, as well as published study protocols reporting relevant methodological information.

The exclusion criteria were as follows: *1*) studies not reporting any measurement property of a UPF-related instrument and not addressing agreement between processing classification systems; *2*) studies assessing dietary intake without explicit linkage to food processing classifications; *3*) studies focused exclusively on health outcomes, dietary patterns, or nutrient-based exposures without evaluation of UPF measurement; *4*) animal studies, in vitro studies, narrative reviews, editorials, commentaries, and methodological papers without original empirical data; and *5*) gray literature sources (e.g., conference abstracts, dissertations, technical reports, or registry-only records) due to insufficient methodological detail for appraisal of measurement properties.

Eligible studies were restricted to peer-reviewed publications. Gray literature sources (e.g., conference abstracts, dissertations, technical reports, or registry-only records) were not searched or included, as the review aimed to synthesize instruments with sufficiently detailed methodological reporting to allow appraisal of measurement properties.

Published study protocols were included only when identified through bibliographic database searches and when they provided relevant methodological information on the development or planned validation of UPF assessment instruments. Dedicated protocol registries or trial registries were not searched, as unpublished or registry-only records typically do not report the measurement properties required to address the objectives of this review.

Eligibility criteria were defined to ensure that the included studies directly addressed objective 1 (evaluation of measurement properties of instruments used to assess NOVA-defined UPF intake) and/or objective 2 (assessment of agreement between processing classification systems applied to the same dietary data). The absence of a required comparator reflects the nature of measurement-focused research, where reference methods vary in quality and purpose.

Analytical grouping and classification strategy

To facilitate synthesis and interpretation, the included studies were grouped a priori into 3 nonexclusive analytical blocks, which were used solely for analytical purposes and did not constitute eligibility criteria:

- Basic dietary intake instruments with validity or reliability data, including NOVA-focused FFQs, brief UPF screening questionnaires, and cohort-level UPF scores reporting $\geq$1 measurement property [e.g., test–retest intraclass correlation coefficients (ICCs)], validity against 24-h recalls, weighed records or food diaries, agreement indices such as Cohen's κ (κ) and prevalence-adjusted bias-adjusted κ (PABAK), or Bland–Altman analyses].
- Broader UPF-related tools, including biomarkers (e.g., polymetabolite signatures), diet quality or guideline-adherence indices, and behavioral instruments demonstrating convergent relationships with UPF exposure.
- Contextual or auxiliary instruments, such as food environment audits, attitudes, acceptance or knowledge scales, parental feeding practices, and additive-specific questionnaires (e.g., Allura Red AC). These instruments provide information on exposure contexts or determinants but do not directly quantify UPF intake.

This analytical framework was applied after study inclusion to structure the Results and Discussion sections and does not imply equivalence in measurement purpose or performance across blocks.

Search strategy

The full electronic search strategy was developed in consultation with an information specialist and implemented across PubMed/MEDLINE, Scopus, and Web of Science from the inception of the databases to 12 January, 2026. Database-specific search strings, including Boolean operators, filters, and limits, are reported in Supplementary Table 2. In addition, the reference lists of all included studies were manually screened to identify additional eligible publications. No language or publication date limits were applied in the search phase.

Study selection

This systematic review was conducted in accordance with the PRISMA 2020 guidelines. After removal of duplicates, titles and abstracts were independently screened by 2 reviewers (IC-R and AS-L) to identify potentially eligible studies. Full-text articles were subsequently assessed for inclusion on the basis of predefined eligibility criteria. Discrepancies at any stage were identified by direct comparison of the reviewers' decisions and were first addressed through consensus discussion between the 2 reviewers. When consensus could not be reached, a third reviewer (AD) acted as an adjudicator and made the final decision. The complete study selection process is summarized in the PRISMA flow diagram.

Data extraction

Data extraction was performed independently by 2 reviewers (IC-R and AS-L) via a standardized, pilot-tested extraction form. For each included study, data were extracted on study objectives, design, population characteristics, the dietary construct or instrument under evaluation, the reference or comparator method used (when applicable), the time interval between repeated assessments in reliability or replication studies, UPF exposure metrics and units, and the main psychometric or validity outcomes. Information on study funding sources was extracted when reported.

The extracted data were cross-checked for consistency, and discrepancies were resolved through discussion and consensus between reviewers. When disagreements persisted, they were resolved by consultation with a third reviewer (AD). These elements were selected to harmonize reporting across studies and to clarify what aspect of UPF measurement was evaluated and under which reference conditions.

The extracted variables were prespecified to enable objective synthesis, allowing separate evaluations of measurement performance (objective 1) and cross-taxonomy agreement (objective 2).

Handling of missing or unclear information

The authors of the included studies were not contacted to request missing data or additional methodological clarification. When

TABLE 1
Characteristics of studies including basic dietary intake instruments with validity/reliability.

Reference	Country	Objective	Design	*n*	Population	Instrument/ construct	UPF unit/metric	Reference method	Time interval between assessments	Psychometrics/validity	Main UPF-related finding
Abutbul Vered et al., [22] 2022	Israel	Evaluate the validity and reproducibility ofa web-based FFQ	Cross-sectional	39	Women	Web-based FFQ	Processed food scores	24h DR	2 wk	ICC = 0.53 r = 0.72 BA acceptable	Acceptable processing-level validity
Cardamone et al., [25] 2024	Italy	Validate the NOVA-FFQ	Protocol	≥436	Adults	NOVA-FFQ	UPF-energy intake (%, g/d, kcal/d)	WDR	3–10 mo	NA	NA
Correa-Madrid et al., [26] 2023	Colombia	Validate short NOVA-UPF score	Cross-sectional	203	Women	NOVA-UPF score	UPF-energy intake (%)	24 h DR	Concurrent	PABAK = 0.75	Tracks UPF-energy intake (%)
Costa et al., [27] 2021	Brazil	Assess the validity of the NOVA-UPF score	Cross-sectional	300	Adults	NOVA-UPF score	UPF-energy intake (%)	24 h DR	Concurrent	PABAK = 0.67	Reflects UPF share
Costa et al., [28] 2024	Brazil	Validate NOVA-WPF and NOVA-UPF	Cross-sectional	1800	Adults	NOVA- WPF score NOVA-UPF score	UPF-energy intake (%)	24 h DR	Concurrent	PABAK: NOVA-WPF = 0.72 NOVA-UPF = 0.79	Diet-quality gradient
D'Abbronzo et al., [29] 2024	Italy	Compare NOVA-FFQ vs. EPIC-FFQ	Cross-sectional	130	Adults with diabetes	NOVA-FFQ EPIC-FFQ	UPF-energy intake (g/ 1000 kcal)	NA	2 wk	ICC = 0.74	EPIC underestimates UPF
Dinu et al., [31] 2021	Italy	Validate NOVA-FFQ	Cross-sectional	110	Adults	NOVA-FFQ	UPF-energy intake (% and g/d)	24 h DR	4–6 wk	ICC (test–retest): % = 0.94 g/d = 0.91 ICC (validity): % = 0.85 g/d = 0.70	Acceptable validity
Erdoğan Gövez et al., [32] 2024	Türkiye	Validate sQ-HPF screener	Cross- sectional	94	Adults	sQ-HPF	NA	DR	4 wk	ICC = 0.76 BA acceptable	Distinguishes HPF
Fangupo et al., [33] 2019	New Zealand	Validate EAT5 FFQ	Cross-sectional	100	Children	EAT5 FFQ	UPF-energy intake (% and KJ)	WDR	4 wk	ICC: % = 0.64 KJ = 0.76 BA acceptable	Acceptable preschool validity
Freire et al., [34] 2025	Ecuador	Validate the adapted NOVA score	Cross-sectional	3510	Adults	NOVA-27 score	UPF-energy intake (%)	24h DR	Concurrent	NA	National adaptation
Kébé et al., [39] 2024	Senegal	Adapt and validate the NOVA score	Cross-sectional	301	Adults	NOVA-UPF score	UPF-energy intake (%)	24h DR	Concurrent	PABAK = 0.84	Surveillance
Loftfield et al., [41] 2025	United States	Evaluate performance for FFQ	Cross-sectional	1311	Adults	NOVA-FFQ	UPF-energy intake (%, kcal/d and g/d)	24h DR	Concurrent	r ≥0.52	Moderate relative validity for UPF intake
Martínez-Pérez et al., [13] 2021	Spain	Compare processing classification systems	Cross-sectional	5636	Older adults with MetS	FFQ + NOVA/ IARC/IFIC/UNC classifications	Energy intake	NA	NA	ICC = 0.51	Associations differ by system; classification choice matters
Martínez-Pérez et al., [42] 2022	Spain	Develop sQ-HPF screener	Cross-sectional	4400	Adults	sQ-HPF	Energy intake (%)	DR	1 y and 2 y	Cronbachc's alfa = 0.67 Weighted κ = 0.36–0.88	Captures change in HPF/UPF
Motta et al., [43] 2021	Brazil	Develop the FFQ by processing the food level	Cross-sectional	7516	Adults	FFQ	NA	NA	NA	NA	Enables processing analysis
Müller et al., [44] 2025	Brazil	Validate child UPF screener	Cross-sectional	365	Early childhood	UPF score	UPF-distribution (%)	24h DR	Concurrent	PABAK = 0.65	Reflects child UPF share

(*continued on next page*)

TABLE 1 (*continued*)

Reference	Country	Objective	Design	*n*	Population	Instrument/ construct	UPF unit/metric	Reference method	Time interval between assessments	Psychometrics/validity	Main UPF-related finding
Oviedo-Solís et al., [45] 2022	Mexico	Evaluate the relative validity of SFFQ	Cross-sectional	382	Children and adolescents	SFFQ	UPF-energy intake (% and kcal/d)	24h DR	Concurrent and 2 d	ICC: %: Children = 0.53 Adolescents = 0.57 kcal/d: Children = 0.55 Adolescents = 0.46 r: %: Children = 0.56 Adolescents = 0.54 kcal/d: Children = 0.55 Adolescents = 0.47 BA acceptable	Ranks UPF-energy intake (%) intake reasonably
Oviedo-Solís et al., [46] 2022	Mexico	Evaluate the relative validity of SFFQ	Cross-sectional	226	Adults	SFFQ	UPF-energy intake (%)	24h DR	Concurrent	ICC = 0.67 r = 0.62	Ranks UPF-energy intake (%) intake reasonably
Sarbagili-Shabat et al., [47] 2020	Israel	Develop and validate PFQ	Cross-sectional	86	Adults with IBD	PFQ	Portions day	FD	2 wk	ICC = 0.75–0.88 r = 0.50 κ = 0.35	Valid for processed/ UPF in IBD

Abbreviations: BA, Bland–Altman; DR, dietary record; EPIC-FFQ, European prospective investigation into cancer and nutrition—food frequency questionnaire; FD, food diary; FFQ, food frequency questionnaire; HPF, high processed food; IARC, International Agency for Research on Cancer; IBD, inflammatory bowel disease; ICC, intraclass correlation coefficient; IFIC, International Food Information Council; MetS, metabolic syndrome; NA, not available; PABAK, prevalence-adjusted bias-adjusted kappa; PFQ, processed food questionnaire; SFFQ, semiquantitative food frequency questionnaire; sQ-HPF, short screening questionnaire of highly processed food; UNC, University of North Carolina; UPF, ultraprocessed food; WDR, weighed dietary record; 24h DR, 24-h dietary recalls; κ, Cohen's κ; WPF, whole plant foods.

TABLE 2
Characteristics of studies including biomarkers, diet quality, and behavioral tools related to ultraprocessed foods.

Reference	Country	Objective	Design	*n*	Population	Instrument/ construct	UPF unit/metric	Reference method	Time interval between assessments	Psychometrics/validity	Main UPF-related finding
Abar et al., [12] 2025	United States	Develop metabolomic UPF biomarkers	Cross-sectional	718	Adults	Metabolomic score	NR	Energy intake (%)	12 mo	r ≥0.47	Biomarkers track UPF intake
Gabe et al., [35] 2022	Brazil	Convergent validity of the dietary guideline tool	Cross-sectional	1309	Adults	eGuia score	Score	Dietary guidelines for the Brazilian population	NA	r = −0.51	Inverse association
Lemke et al., [40] 2024	Brazil	Develop a meal quality index	Cross-sectional	6399	Children	Meal quality scale	Score	No intake reference	NA	—	Processing proxy
Santos et al., [48] 2021	Brazil	Psychometric validation of the diet-quality scale (ESQUADA)	Cross-sectional	2059	Adolescents	Diet-quality scale (ESQUADA)	Score	No intake reference	NA	empirical reliability = 0.70	Quality scale

Abbreviations: ESQUADA, Brazilian food consumption and dietary practices assessment; NA, not available; UPF, ultraprocessed food; NR, not reported.

information on study characteristics, reference methods, or measurement properties was unclear or not explicitly reported in the original publication, the corresponding fields were coded as not reported. No assumptions, imputations, or extrapolations were made beyond the information provided in the published articles. Consequently, synthesis and interpretation were restricted to explicitly reported data, in accordance with the PRISMA 2020 recommendations.

Risk of bias/methodological quality

The risk of bias was assessed in duplicate (IC-R, AS-L) via the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) risk of bias checklist, which is tailored to studies of measurement properties [21]. For each instrument and property, we rated the relevant COSMIN boxes: content validity (relevance/comprehensiveness), structural validity (Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA); measurement invariance, i.e., whether the instrument measures the same construct equivalently across groups or over time, which is typically tested via configural/metric/scalar invariance), internal consistency (Cronbach's α/McDonald's ω consistent with the factor structure), cross-cultural validity/invariance**,** reliability (test–retest; ICC), measurement error (SE measurement, Smallest Detectable Change (SDC), Bland–Altman limits), criterion validity [against 24-h recalls/diaries/weighed dietary record (WDR)], hypothesis testing for construct validity (correlations with %E-UPF, known-group differences), and responsiveness when applicable. Each item was graded as "very good," "adequate," "doubtful," or "inadequate," applying the worst-score-counts rule per box. For synthesis, we mapped box-level judgments to low (very good/adequate), unclear (doubtful), or high (inadequate) risk of bias. COSMIN box-level ratings were summarized in a heatmap via the worst-score-counts rule per box (very good/adequate/doubtful/inadequate mapped to low/unclear/high risk) to provide a parallel view of content/structural validity, reliability, measurement error, agreement, and responsiveness. Disagreements were resolved by consensus or arbitration by a third reviewer (AD).

Data synthesis

Given the heterogeneity in study designs, instruments, and reference methods, a quantitative meta-analysis was not feasible. Instead, a structured narrative synthesis was conducted, which was explicitly organized according to the 2 review objectives. Evidence related to objective 1 was synthesized by measurement property and instrument type, whereas evidence relevant to objective 2 was summarized separately for studies comparing multiple processing taxonomies applied to the same dietary data. The results were synthesized by the type of measurement property and instrument category, prioritizing commonly reported statistics such as ICCs [1,2], correlation coefficients, κ and PABAK statistics, and internal consistency metrics such as Cronbach's α or McDonald's ω. Figures and tables were used to enhance transparency and facilitate comparisons across studies rather than to infer pooled effect sizes. The measurement properties were interpreted via a standardized glossary and framework (Supplementary Table 3) to harmonize heterogeneous validation metrics across studies.

No data conversions or transformations were applied for synthesis. The measurement properties, agreement statistics, and UPF metrics were extracted and presented via the original units, scales, and statistical summaries reported in each study. As a result, synthesis was qualitative and descriptive, without rescaling or pooling of estimates.

In addition to tabular summaries (Tables 1–3) [12,13,22–49], we constructed a psychometric evidence map to visually synthesize the

TABLE 3
Characteristics of studies including contextual/auxiliary instruments related to ultraprocessed foods.

Reference	Country	Objective	Design	*n*	Population	Instrument/construct	UPF unit/metric	Reference method	Time interval between assessments	Psychometrics/ validity	Main UPF-related finding
Borges et al., [23] 2021	Brazil	Develop and validate a retail food environment audit tool based on NOVA	Cross-sectional	650	Retail outlets	AUDITNOVA/CFEHS	Environment score	NA	NA	$\alpha = 0.91$	Environment scoring by NOVA groups
Calvo-Porral et al., [24] 2024	Spain	Develop and validate a scale measuring consumer acceptance of UPF	Cross-sectional	478	Adults	UPF acceptance Scale	Aceptance Score	NA	NA	$\alpha = 0.76–0.93$	Acceptance relates to UPF intentions
Dey et al., [30] 2019	Mexico	Develop and validate a questionnaire on the intake of foods containing Allura Red AC	Cross-sectional	197; pilot $n = 28$	Mothers of 10–13 y	Questionnaire on foods with Allura Red AC	Food types	Additive intake score	NA	$\alpha = 0.86$	Frequent intake of Allura Red foods
Gabe et al., [36] 2025	Brazil	Adapt and validate a scale assessing knowledge of food processing	Cross-sectional	1245	Adults	NOVA-Conhecimento (processing knowledge)	Knowledge score (unitless)	24hDR	NA	NA	Knowledge inversely association with UPF Energy intake %
Hoil Sosa et al., [37] 2024	Mexico	Validate attitudes toward the UPF consumption scale	Cross-sectional	277	University students	Attitudes toward UPF scale	Attitude score	NA	NA	$\omega = 0.97$	Strong reliability (attitudes)
Hu et al., [38] 2024	China	Validate scale measuring UPF withdrawal-like symptoms	Cross-sectional	244	Adults	mProWS	Score	ProWS YFAS 2.0 TFEQ QEWP-R	6 mo	$r = 0.61$	Behavioral dependence/ withdrawal
Warkentin et al., [49] 2016	Brazil	Translate and validate the feeding practices questionnaire related to UPF exposure	Cross-sectional	402	Parents of preschoolers	CFPQ	Score	FFQ	2 wk	$\alpha = 0.74–0.88$	Contextual practices for UPF exposure

Abbreviations: CFEHS, consumer food environment healthiness score; CFPQ, comprehensive feeding practices questionnaire; mProWS, modified highly processed food withdrawal scale; NA, not available; UPF, ultraprocessed food; α, Cronbach's α; ω, McDonald's ω; AUDITNOVA, Adherence to the NOVA classification-based dietary index.

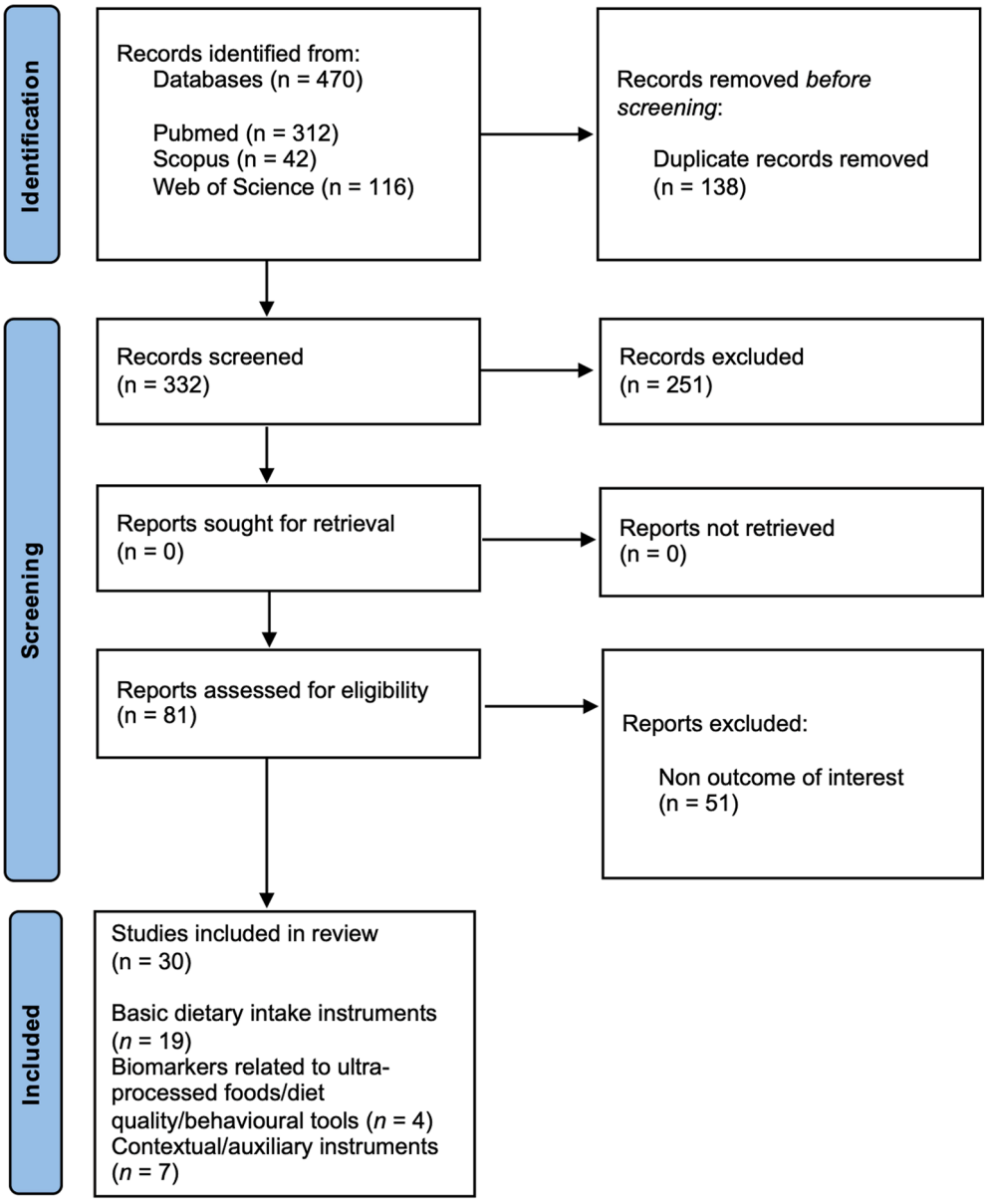

FIGURE 1. PRISMA 2020 flow chart.

measurement properties reported across the included studies. This map displays, for each instrument category, the range and clustering of key psychometric indicators, including test–retest reliability (ICCs), criterion or convergent validity coefficients (correlations with 24-h dietary recalls, WDRs, or food diaries), agreement indices (κ and PABAK), and internal consistency measures where applicable.

Protocol deviations

This review was conducted in accordance with the PROSPERO-registered protocol (CRD420251135087). Given the degree of heterogeneity across instruments, reference methods, and psychometric metrics, all findings were synthesized narratively, and no quantitative pooling was undertaken. Published study protocols identified through database searches were retained for descriptive purposes but excluded from psychometric synthesis because of the absence of results. No other substantive deviations from the registered protocol were identified.

Results

Characteristics of the included instruments

The systematic search identified 470 records. After the removal of duplicates ($n = 138$), 332 records were screened by title and abstract, and 81 full-text articles were assessed for eligibility. Fifty-one studies were excluded at the full-text stage and presented no outcome of interest (Supplementary Table 4), resulting in 30 studies included in the qualitative synthesis (Figure 1) [12,13,22–49].

Among the included studies, 19 evaluated basic dietary intake instruments, including FFQs, short screening tools, and NOVA-based scores [13,22,25–29,31–34,39,41–47]. In addition, 4 studies examined biomarker-based or indirect UPF-related tools [12,35,40,48], and 7 studies focused on contextual or auxiliary instruments related to food environments, attitudes, knowledge, or behavioral constructs [23,24,30,36–38,49].

Funding source information was reported in 20 of the included studies [12,13,22–25,27–29,32,33,36,39–42,44–46,49]. Three studies

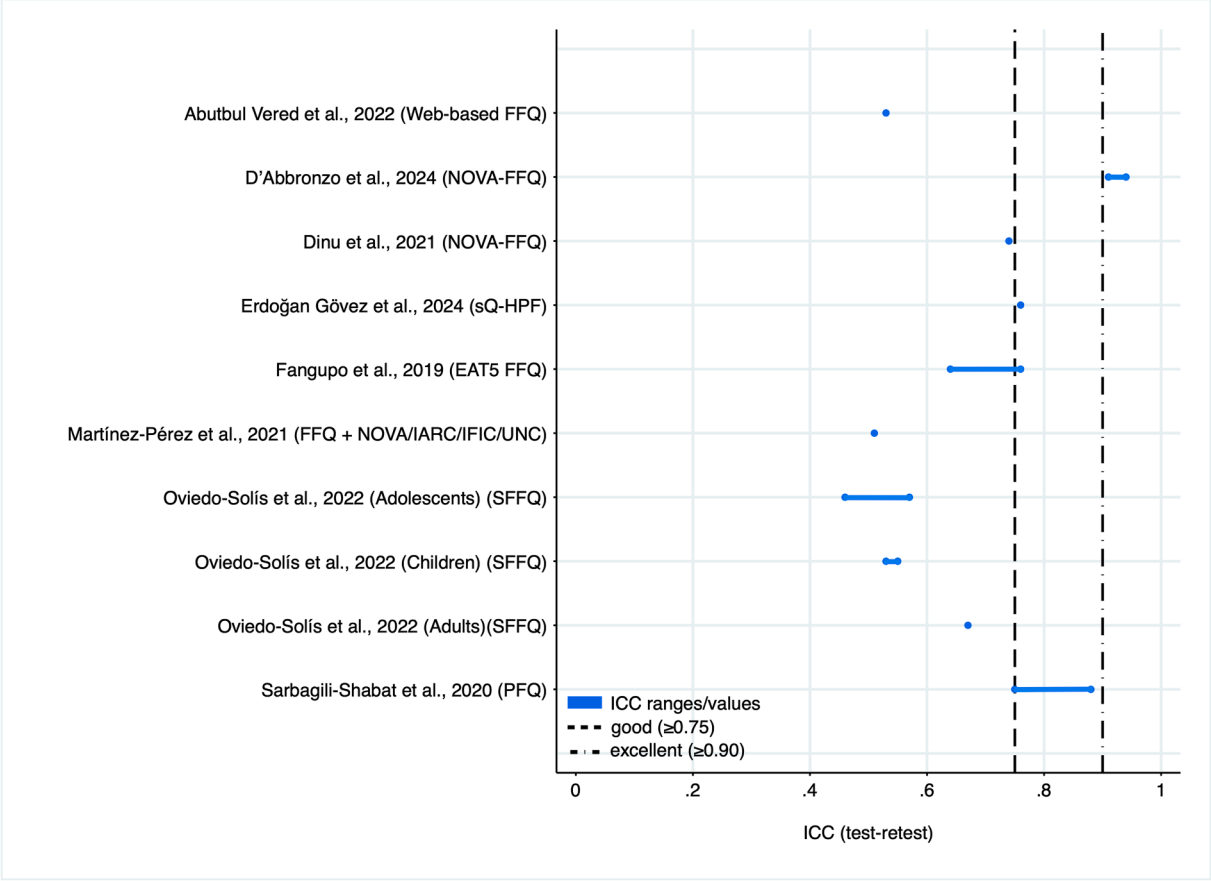

FIGURE 2. Summary of test–retest reliability of UPF-related intake tools. Points indicate study-level central statistics; whiskers show observed ranges across subscales/strata (not confidence intervals). FFQ, food frequency questionnaire; IARC, International Agency for Research on Cancer; ICC, intraclass correlation; IFIC, International Food Information Council; PFQ, processed food questionnaire; SFFQ, semiquantitative food frequency questionnaire; sQ-HPF, short questionnaire on highly processed foods; UNC, University of North Carolina; UPF, ultraprocessed food.

explicitly reported no external funding [25,33,45], whereas funding information was not reported in 10 studies [26,30,31,34,35,37,38,43,47,48] (Supplementary Table 5). The main characteristics of the included studies and instruments, including study design, population, age range, country, and UPF exposure definition, are summarized in Table 1.

The studies were conducted across diverse geographical settings, including Europe [13,24,25,29,31,42], the United States [12,41], Latin America [23,26–28,34,40,44–46,48,49], Africa [39], the Middle East [22,38,47], and Oceania [33]. Populations ranged from early childhood to older adults with cardiometabolic conditions.

UPF exposure was most frequently operationalized as the percentage of total energy intake (%E-UPF) [26–28,31,33,34,39,45,46], although several studies also reported absolute intake metrics such as grams or kilocalories per day [29,31,41,45] or unitless index scores, depending on the instrument used [23,24,27,28,40,48].

Studies reporting validation protocols without available results were retained to reflect ongoing methodological development in the field and are clearly identified as "protocol, results pending" [25]. These studies were not included in the synthesis of psychometric outcomes. To facilitate interpretation and practical use of the reviewed instruments, we provide a supplementary "toolbox" table summarizing the key characteristics of each tool, including country of use, instrument type, and the specific UPF-related construct measured (Supplementary Table 6).

Reference methods used for validation (objective 1)

Across the 30 included studies (Tables 1–3), the validation approaches used varied substantially. Among studies evaluating basic dietary intake instruments [22,27,28,31,33,45,46], the most frequently used reference methods were 24-h dietary recalls, followed by WDRs and food diaries. These reference methods were used to validate FFQs, short UPF screening tools, and NOVA-based scores.

Recall-based reference methods predominated, whereas WDRs and food diaries, which are used in fewer studies [33,47], generally provide more detailed information on food preparation and processing. Studies relying on higher-burden reference methods tended to report stronger agreement metrics.

For studies evaluating biomarker-based or indirect UPF-related tools [12,35,40,48], reference standards differ conceptually and include controlled feeding designs, dietary guideline-adherence indices, or no direct intake reference, reflecting their complementary role rather than direct estimation of UPF intake.

Reliability outcomes (objective 1)

Reliability was most frequently assessed for basic dietary intake instruments (Table 1), including FFQs, short screening tools, and NOVA-based scores [13,22,29,31–33,45–47]. Test–retest reliability, typically expressed as ICCs, ranged from ~0.51 to ~0.94, indicating

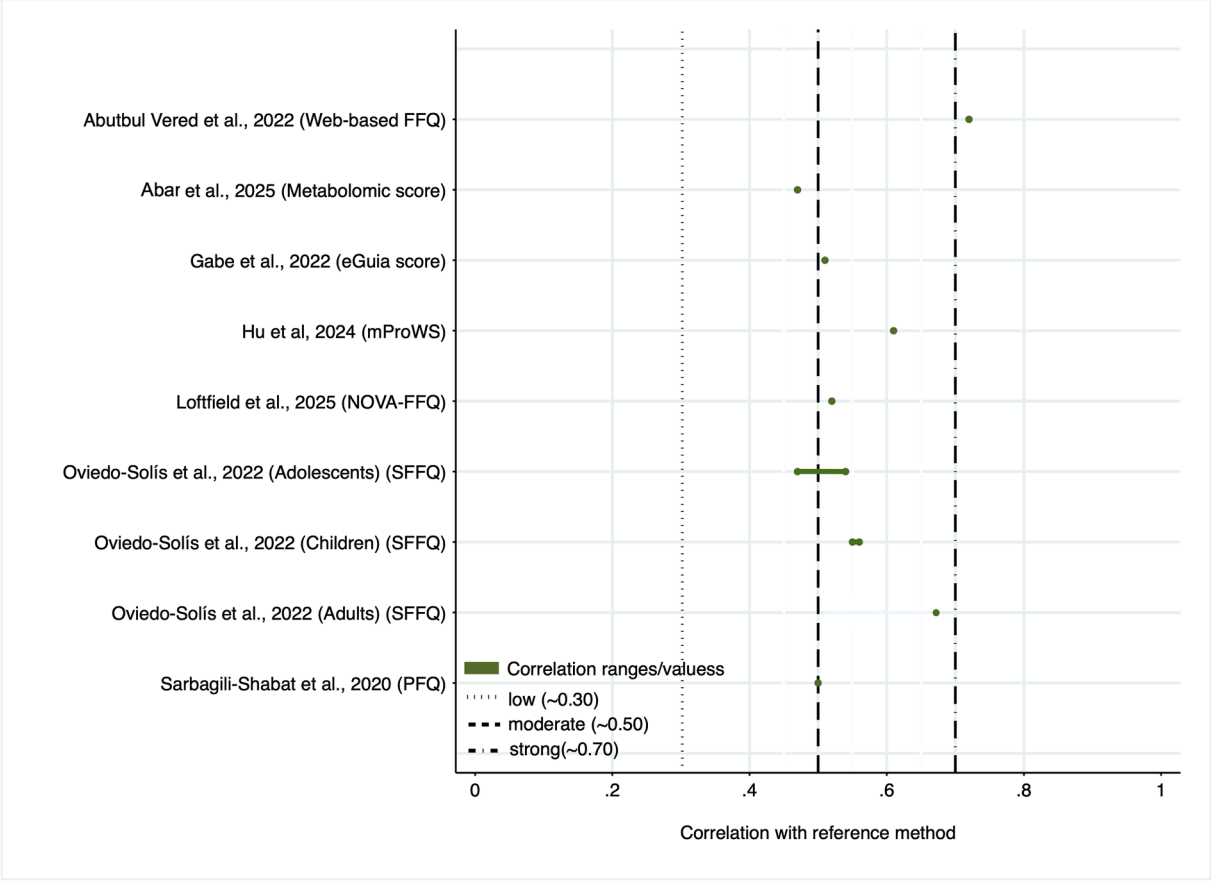

FIGURE 3. Summary of the correlation of UPF-related intake tools with reference methods. Points indicate study-level central statistics; whiskers show observed ranges across subscales/strata (not confidence intervals). FFQ, food frequency questionnaire; mProWS, modified processed food withdrawal scale; PFQ, processed food questionnaire; SFFQ, semiquantitative food frequency questionnaire.

moderate–to-high short-term stability across instruments, populations, and age groups.

FFQ-based instruments and short UPF screening tools showed comparable reproducibility in adult and pediatric samples, supporting their use for ranking individuals according to UPF intake over short intervals.

For the contextual and auxiliary instruments (Table 3), the reliability assessment focused on internal consistency. Cronbach's α and McDonald's ω coefficients generally exceeded 0.80 [23,24,30,37], with values approaching 0.97 at some scales [37], indicating a strong internal structure.

Relative validity outcomes (objective 1)

Relative validity estimates for dietary intake instruments varied by instrument type and reference method (Table 1). The correlations between UPF assessment tools and dietary reference methods [22,41,45–47] ranged from approximately r = 0.47 to r = 0.72, reflecting modest to moderate agreement.

The validity tended to be greater when UPF intake was expressed as a compositional measure (e.g., percentage of total energy intake) and when more detailed reference methods were used. Evidence from calibration-based analyses [41] further indicates that analytic approaches, including measurement error modeling, materially influence the interpretation of relative validity estimates, even when the underlying FFQ classifications are similar.

Agreement between processing classification systems (objective 2)

Across studies comparing multiple food processing classification systems applied to the same dietary data [13,28], the agreement was generally moderate. The ICCs for the cross-system comparisons clustered between 0.46 and 0.91, whereas the PABAK and weighted κ values [26–28,39,44] ranged from ~0.35 to ~0.84.

When reported, Bland–Altman analyses [22,32,33,45] indicated acceptable agreement for %E-UPF estimates, although wider limits of agreement were observed for absolute intake metrics.

Emerging patterns across studies

The synthesis results across Tables 1–3 reveal several consistent patterns. First, most instruments demonstrated acceptable reproducibility but only moderate relative validity, particularly when validated against recall-based reference methods. Second, measurement performance was strongly dependent on reference method quality, with higher agreement observed in studies using weighed records or food diaries.

Biomarker-based and indirect tools [12,40,48] provide promising complementary evidence but are typically expressed as unitless indices or relative concentrations, limiting direct comparability with dietary intake measures. Contextual and auxiliary instruments [23,24,36,38] offered robust psychometric performance for their intended constructs but did not directly quantify UPF intake.

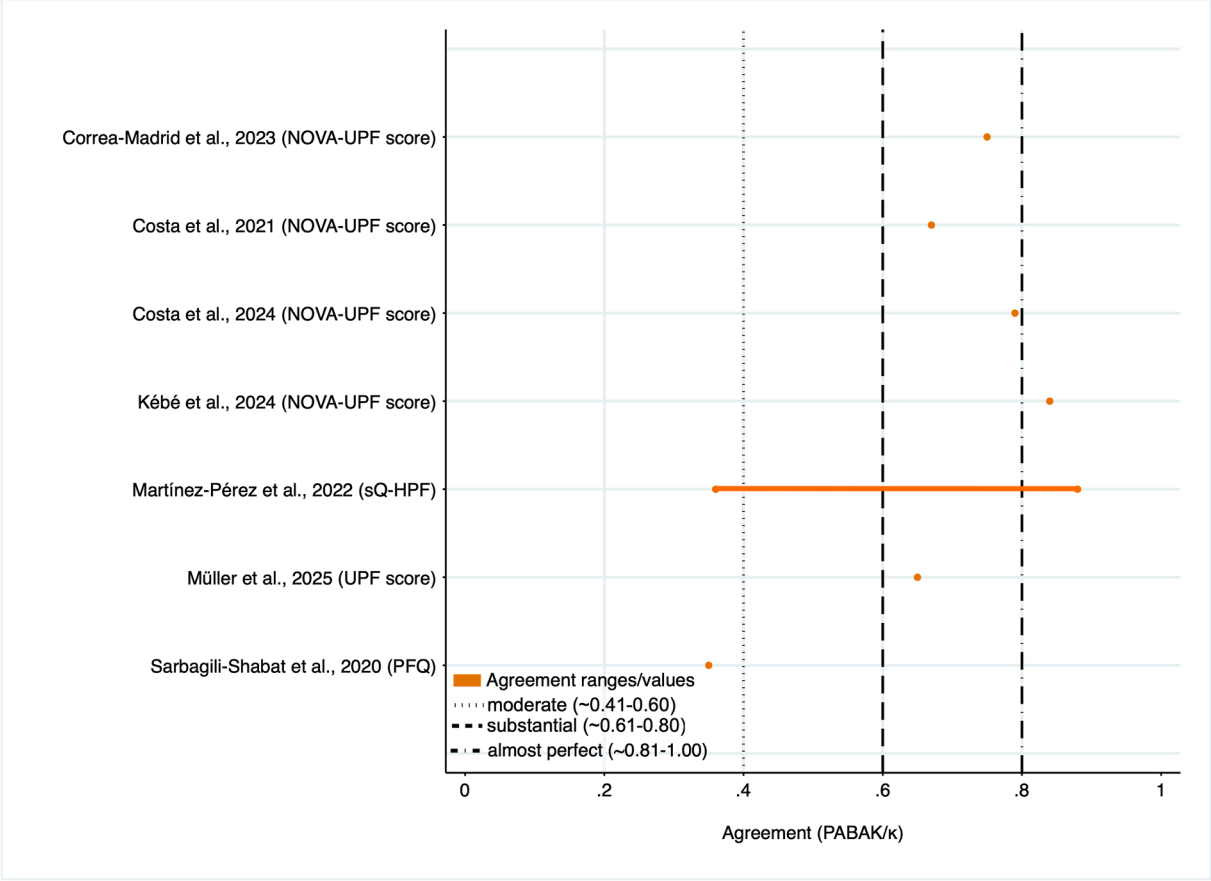

FIGURE 4. Summary of classification agreement for UPF-related intake tools. Points indicate study-level central statistics; whiskers show observed ranges across subscales/strata (not confidence intervals). FFQ, food frequency questionnaire; PABAK, prevalence-adjusted bias-adjusted kappa; PFQ, processed food questionnaire; sQ-HPF, short questionnaire on highly processed food; UPF, ultraprocessed food; κ, Cohen's kappa.

Psychometric evidence map

Figures 1 and **3–5** present a visual synthesis of the psychometric evidence across instrument types disaggregated by measurement properties to improve readability: test–retest reliability (Figure 2), criterion/convergent validity (Figure 3), classification agreement (Figure 4), and internal consistency (Figure 5).

Only studies reporting sufficient quantitative information for a given psychometric property were included in the corresponding figure. Instruments that did not report plottable estimates (ICCs, correlation coefficients, κ and PABAK, or α/ω) were retained in the narrative synthesis and summary tables and contributed to the overall interpretation of measurement performance by instrument type.

Across basic dietary intake instruments, test–retest reliability estimates clustered predominantly in the moderate-to-high range, with ICC values ranging from ~0.46 to ~0.94. The estimates of criterion/convergent validity, expressed as correlation coefficients, were more variable, ranging from approximately $r \approx 0.47$ to $r \approx 0.88$, whereas the agreement metrics (κ and PABAK) generally fell between 0.36 and 0.84, indicating fair-to-substantial agreement depending on the instrument and population.

In contrast, contextual and auxiliary instruments consistently demonstrated strong internal consistency, with Cronbach's α and McDonald's ω values typically ≥0.74 and frequently exceeding 0.90, with the highest values approaching $\omega \approx 0.97$.

Risk of bias and methodological quality

Using the COSMIN risk of bias checklist, 15 of the 30 studies (50.0%) were rated as having a low risk of bias, whereas the remaining 15 (50.0%) were classified as having an unclear risk; none were rated as having a high risk. Content validity was assessed in 29 studies, structural validity in 27, and criterion validity in 19. Hypothesis testing for construct validity confirmed a priori expectations in 86.6% of the studies.

The interrater agreement for the COSMIN ratings was high (κ: 0.90; 95% confidence interval: 0.86, 0.95). Detailed property-level assessments are provided in Supplementary Table 7, with a summary heatmap displayed in Figure 6.

Discussion

This systematic review addressed 2 complementary objectives: first, to evaluate the measurement performance of instruments used to assess NOVA-defined UPF intake, and second, to examine the agreement between food processing classification systems when applied to the same dietary data. Together, the findings provide insight into both instrument-level performance and system-level sources of variability. Across the literature, substantial heterogeneity was observed in validity, reliability, and agreement estimates across instruments, settings, and populations, reflecting both methodological advances and persistent challenges in the measurement of UPF exposure [49–51].

The measurement performance strongly depends on the reference method quality

A central finding of this review is that the interpretation of relative validity depends critically on the quality of the reference or

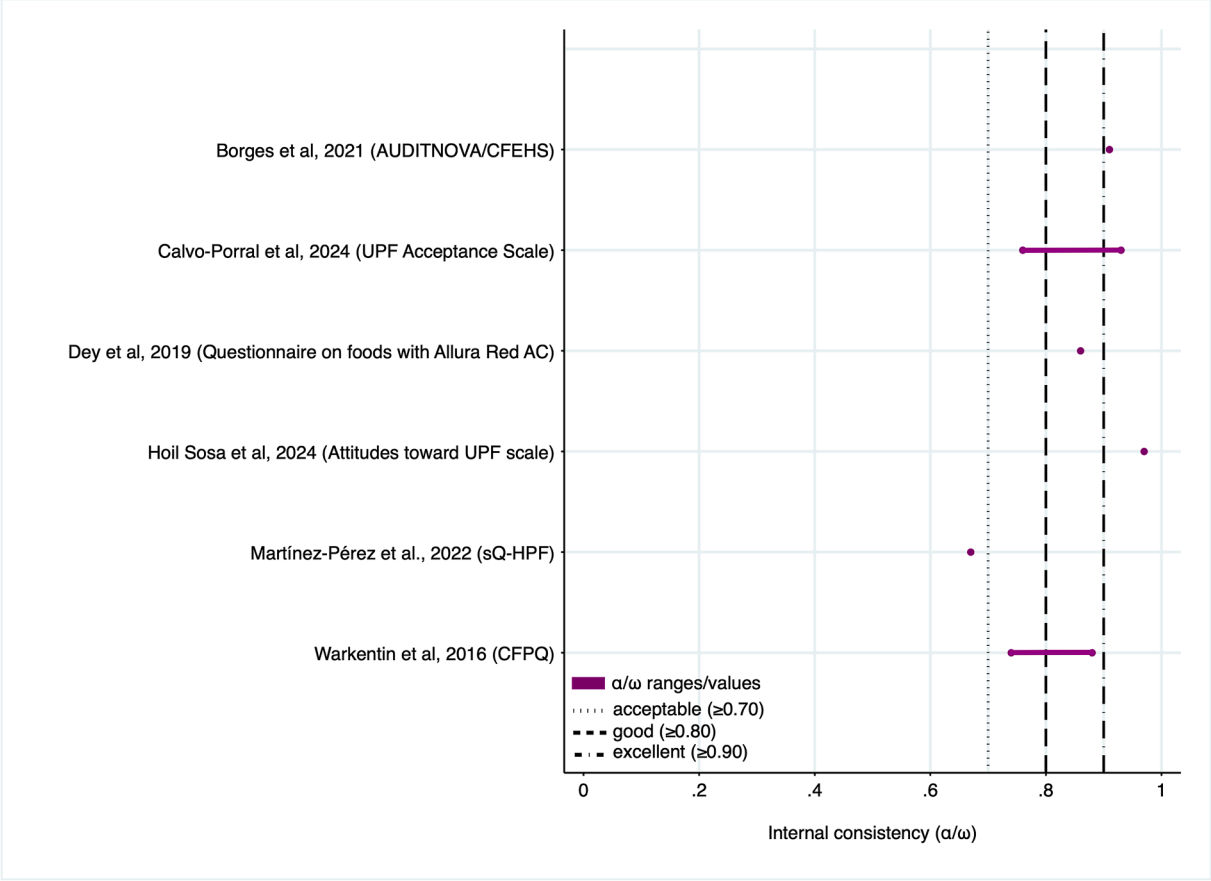

FIGURE 5. Summary of the internal consistency of UPF-related intake tools. Points indicate study-level central statistics; whiskers show observed ranges across subscales/strata (not confidence intervals). CFEHS, consumer food environment healthiness score; CFPQ, comprehensive feeding practices questionnaire; sQ-HPF, short questionnaire on highly processed food; UPF, ultraprocessed food; α, Cronbach's α; ω, McDonald's ω.

comparator method. In the included evidence base, recall-based comparators predominated, whereas fewer studies relied on higher-burden methods such as WDRs or multiday food diaries, which typically capture more detail relevant for processing classification (e.g., preparation and product specificity) [12,22,25,33,45,46,52]. This dependency on reference method quality likely explains a substantial proportion of the variability observed across studies and should be regarded as a defining characteristic of the current evidence base rather than a methodological flaw [50,51,53].

In contrast, validation against recalls that do not systematically capture processing-relevant details, such as brand names, industrial formulations, or ingredient lists, likely underestimates classification error and attenuates validity coefficients, particularly when UPF exposure is operationalized as ratios that incorporate total energy intake (e.g., %E-UPF) [50,51,54]. When viewed through this lens, the literature suggests that many instruments perform as well and can reasonably be expected, given the limitations of available reference standards [50,51,53]. Recognizing that "validity depends on the reference" therefore reframes a commonly cited limitation into a key methodological insight with implications for future validation studies and the interpretation of observed associations between UPF intake and health outcomes [50,51,55].

Reliability and responsiveness: gaps in reporting

Across the studies with available estimates, test–retest reliability for basic dietary intake instruments ranged from moderate-to-high (ICC ≈ 0.46–0.94), supporting short-term stability for NOVA-focused FFQs and UPF screening tools across adult and pediatric populations [22,31–33,45–47]. However, reproducibility over short intervals primarily reflects consistent reporting and scoring and can coexist with substantial systematic and random errors in self-reported dietary assessment [50,51,53].

In contrast, measurement error indices (e.g., SE measurement/SDC and Bland–Altman limits of agreement) and responsiveness to change were reported for only a small minority of instruments. This limits the suitability of most tools for intervention studies or longitudinal monitoring, where sensitivity to dietary change is essential [56–59]. Future validation work should prioritize responsiveness analyses and formal measurement error estimation alongside traditional reliability metrics, including designs that better characterize within-person variability and attenuation [55,59,60].

Sources of disagreement in UPF classification

In addition to quantitative performance metrics, recurring qualitative sources of disagreement in UPF classification, including mixed dishes requiring recipe decomposition, brand-specific formulation variability, incomplete ingredient information, and subjective definitional elements, were identified. These challenges align with broader critiques of processing-based taxonomies and their operationalization [61,62]. In studies explicitly comparing multiple processing classification systems applied to the same dietary data, agreement was typically fair-to-substantial (κ and PABAK ≈ 0.36–0.84), with

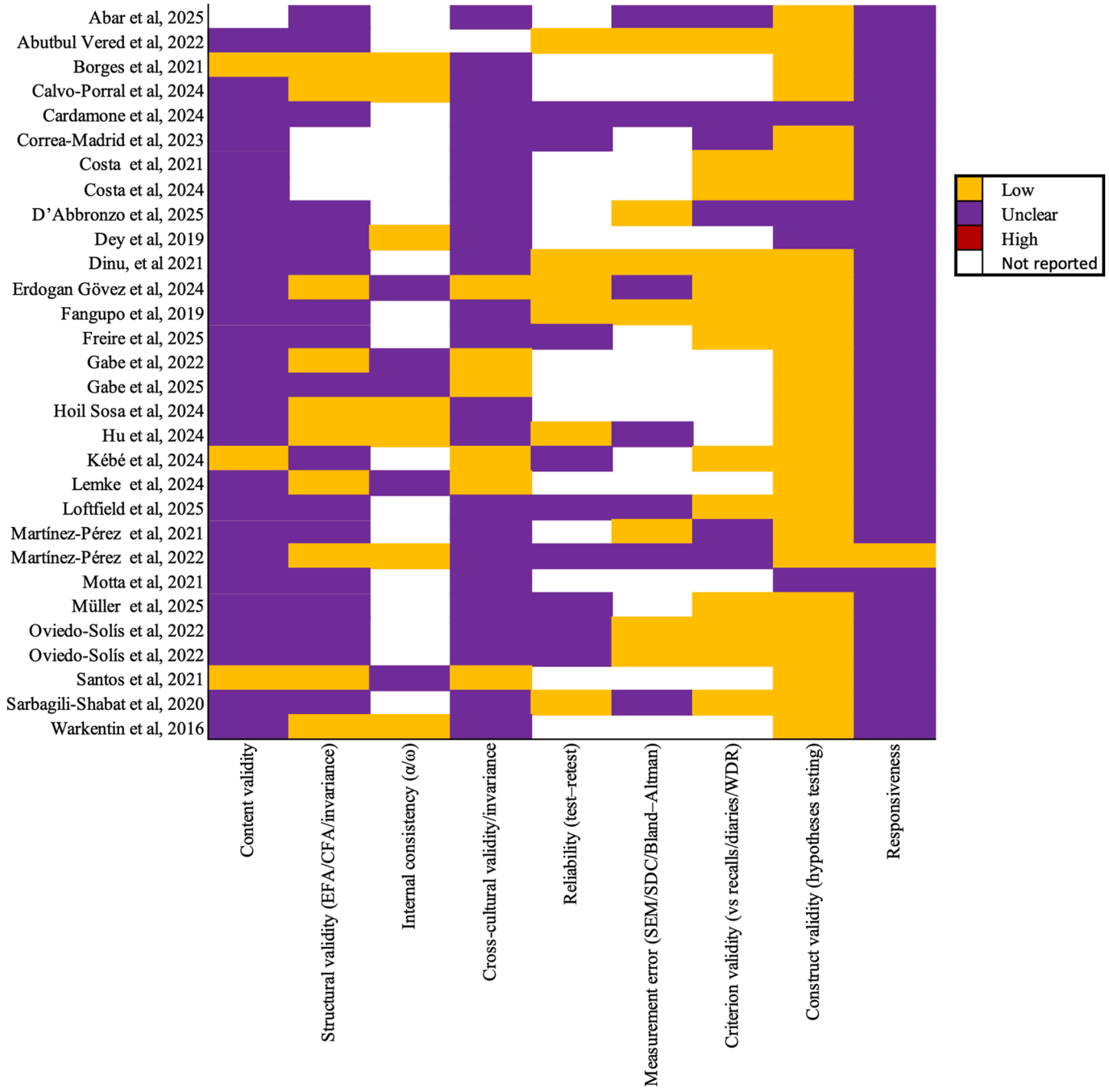

FIGURE 6. Measuring ultraprocessed food exposure: validity, reliability, and risk of bias assessed with COSMIN. SEM, standard error measurement; WDR, weighed dietary record; α, Cronbach's α; ω, McDonald's ω; CFA: Confirmatory Factor Analysis; COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments; EFA: Exploratory Factor Analysis; SDC: Smallest Detectable Change.

weaker concordance for borderline foods and composite items [13,26–28,39,44].

Agreement tended to be greater when operationalized coding rules and trained raters were used and lower when classification required subjective interpretation, underscoring the need for transparent decision rules and standardized protocols when processing-based taxonomies were applied [57,61,62].

Use-case–specific implications

Taken together, the evidence supports a differentiated view of instrument suitability by use case. For population surveillance, short NOVA-based screening tools appear feasible and sufficiently reproducible for ranking individuals, particularly when they are calibrated against repeated recalls or diaries [22,33,45–47]. For etiologic research, FFQs with explicit NOVA coding rules and validation against higher-quality reference methods showed the strongest overall performance [31–33,41,45,46]. For intervention monitoring, however, the evidence base remains limited, as few tools have reported both test–retest reliability and responsiveness to change [59].

Most instruments operationalize UPF intake as a compositional measure (e.g., %E-UPF), which is consistent with the conceptualization of UPFs as part of overall dietary patterns. Our findings do not

support replacing this approach with absolute intake metrics. Rather, compositional and absolute measures capture related but distinct dimensions of exposure and are best interpreted as complementary. Although compositional metrics support pattern-based analyses and substitution interpretations, absolute metrics may be advantageous for calibration studies and measurement error modeling and may be less sensitive to ratio-related bias arising from total energy misreporting [50,51,53–55].

Definitional harmonization and modular instrument design

Given the observed measurement uncertainty and system dependence, definitional harmonization may represent an important prerequisite for the development of next-generation UPF assessment tools. In this context, the joint Food and Drug Administration /USDA request for information (July 2025) aimed at establishing a uniform definition of UPFs reflects growing scientific and policy recognition of the need for greater consistency. Nevertheless, NOVA-based tools remain highly relevant because of their widespread application in the literature. Our results do not imply that UPF assessment instruments must be NOVA-based or that NOVA is superior to alternative frameworks; rather, they reflect the dominance of NOVA in current research [61,62].

Recommendations regarding modular UPF assessment tools should be interpreted as inherently population- and context-specific. A modular core–extension structure, combining a standardized core with context-specific modules, may balance cross-study comparability with local relevance and reduce avoidable misclassification, provided that content validity, cognitive testing, and measurement invariance are explicitly evaluated [52,63–68].

Biomarkers and controlled feeding conditions

Biomarker-based approaches represent promising complementary strategies. The controlled feeding context provides a stronger reference standard for UPF exposure than do observational self-reports alone [12,69]. However, the proposed biomarkers have not been systematically evaluated against established dietary biomarker validation frameworks, such as the Food Biomarkers Alliance (FoodBALL) criteria [70–73]. Accordingly, future biomarker research should align validation efforts with recognized methodological standards, including robust cross validation, independent replication, and transparent reporting [72–74].

Generalizability, limitations, and strengths

The generalizability of these findings should be interpreted with caution. Most validated instruments were developed and tested in high- or middle-income settings, and performance may differ in contexts with distinct food environments and product markets. Cross-cultural adaptation and age-specific validation have been inconsistently evaluated, which can introduce differential misclassification and limit external validity [53,57,68].

Additional limitations include post hoc PROSPERO registration after the literature search due to an evolving review scope; substantial heterogeneity across instruments, reference methods, classification systems, and statistical summaries, which precluded meta-analysis and necessitated narrative synthesis. Reference methods are themselves imperfect measures of habitual intake, so validity estimates likely reflect error in both the index instrument and the comparator [50,51,53,60]. The reporting of key measurement properties was uneven, and publication bias cannot be excluded.

The strengths of this review include adherence to the PRISMA 2020 guidelines, explicit separation of instrument performance from the taxonomy agreement, and structured synthesis across complementary tool categories. Importantly, the review demonstrates that apparent differences in UPF instrument validity are strongly conditioned by reference method quality, providing a critical framework for interpreting inconsistencies in the literature and guiding future validation efforts [50,51,55].

In summary, existing tools for assessing UPF intake demonstrate acceptable reproducibility but variable validity, which is strongly influenced by reference method quality and classification rules. Progress in this field will require investment in higher-quality reference data, transparent coding protocols, and validation frameworks integrating dietary, contextual, and biological measures. These advances are essential for robust estimation of UPF exposure and for reliable inference in nutritional epidemiology and policy evaluation.

Acknowledgments

We thank Consorcio Centro de Investigación Biomédica en Red (CIBER) de la Fisiopatología de la Obesidad y Nutrición (CIBER-Obn), Instituto de Salud Carlos III (Spain), for the initiative on ultra-processed foods (UPF) research.

Author contributions

The authors' responsibilities were as follows – IC-R, AS-L, M-CLdLH, AD: designed the research; IC-R, AS-L: conducted the research and analyzed data; IC-R: wrote the paper; CB, AdS-L, GC-B, JK, CP, FF-A, MJM-A, LD, JAM, JS-S, MPP, ET, MG, DC, IM-I, M-CLdLH; revised critically and edited the paper; IC-R, AD: had primary responsibility for the final content; and all authors: read and approved the final manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

No AI tool was used.

Conflict of interest

The authors report no conflicts of interest.

Funding

This research was partially funded by MCIN/AEI/10.13039/501100011033 through European Regional Development Fund (ERDF), A Way of Making Europe (PID2022-143100OB-I00, and PID2023-150666OA-I00) and by the Ramon Areces foundation (CIVP21S13273); by Comunidad de Madrid EXOHEP2-CM (S2022/BMD-7409); and by the Carvascare Research Group, Universidad de Castilla-La Mancha (2025-GRIN-38419). FF-A and JS-S are partially supported by Catalan Institution for Research and Advanced Studies (ICREA) under the ICREA Academia program. JK received funding from Instituto de Salud Carlos III (ISCIII) through a Miguel Servet grant (CP24/00089), co-funded by the European Union. M-CLdLH is a recipient of a Ramón y Cajal grant (RYC2022-037626-I). The Consorcio Centro de Investigación Biomédica en Red (CIBER) de la Fisiopatología de la Obesidad y Nutrición is an initiative of the Instituto de Salud Carlos III, Spain.

Data availability

The data described in the manuscript, code book, and analytic code will be made available upon request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ajcnut.2026.101263.

References

[1] M.M. Lane, E. Gamage, S. Du, D.N. Ashtree, A.J. McGuinness, S. Gauci, et al., Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses, BMJ 384 (2024) e077310, https://doi.org/10.1136/bmj-2023-077310.

[2] M. Marino, F. Puppo, C. Del Bo, V. Vinelli, P. Riso, M. Porrini, et al., A systematic review of worldwide consumption of ultra-processed foods: findings and criticisms, Nutrients 13 (8) (2021) 2778, https://doi.org/10.3390/nu13082778.

[3] M.J. Gibney, C.G. Forde, Nutrition research challenges for processed food and health, Nat. Food. 3 (2) (2022) 104–109, https://doi.org/10.1038/s43016-021-00457-9.

[4] Y. Park, K.W. Dodd, V. Kipnis, F.E. Thompson, N. Potischman, D. A. Schoeller, et al., Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers, Am. J. Clin. Nutr. 107 (1) (2018) 80–93, https://doi.org/10.1093/ajcn/nqx002.

[5] T.P. De Araújo, M.M. De Moraes, C. Afonso, C. Santos, S.S. Rodrigues, Food processing: comparison of different food classification systems, Nutrients 14 (4) (2022) 729, https://doi.org/10.3390/nu14040729.

[6] M. Reichenheim, J.L. Bastos, What, what for and how? Developing measurement instruments in epidemiology, Rev. Saude. Publica. 55 (2021) 40, https://doi.org/10.11606/s1518-8787.2021055002813.

[7] J. Cade, R. Thompson, V. Burley, D. Warm, Development, validation and utilisation of food-frequency questionnaires – a review, Public Health Nutr 5 (4) (2002) 567–587, https://doi.org/10.1079/PHN2001318.

[8] B. Swinburn, G. Sacks, S. Vandevijvere, S. Kumanyika, T. Lobstein, B. Neal, et al., INFORMAS (International Network for Food and Obesity/non-communicable diseases Research, Monitoring and Action Support): overview and key principles, Obes. Rev. 14 (Suppl 1) (2013) 1–12, https://doi.org/10.1111/obr.12087.

[9] S.I. Kirkpatrick, J. Reedy, E.N. Butler, K.W. Dodd, A.F. Subar, F. E. Thompson, et al., Dietary assessment in food environment research: a systematic review, Am. J. Prev. Med. 46 (1) (2014) 94–102, https://doi.org/10.1016/j.amepre.2013.08.015.

[10] L.E. O'Connor, K.D. Hall, K.A. Herrick, J. Reedy, S.T. Chung, M. Stagliano, et al., Metabolomic profiling of an ultraprocessed dietary pattern in a domiciled randomized controlled crossover feeding trial, J. Nutr. 153 (8) (2023) 2181–2192, https://doi.org/10.1016/j.tjnut.2023.06.003.

[11] D. Su, J. Chen, S. Du, H. Kim, B. Yu, K.E. Wong, et al., Metabolomic markers of ultra-processed food and incident CKD, Clin. J. Am. Soc. Nephrol. 18 (3) (2023) 327–336, https://doi.org/10.2215/CJN.0000000000000062.

[12] L. Abar, E.M. Steele, S.K. Lee, L. Kahle, S.C. Moore, E. Watts, et al., Identification and validation of poly-metabolite scores for diets high in ultra-processed food: an observational study and post-hoc randomized controlled crossover-feeding trial, PLOS Med 22 (5) (2025) e1004560, https://doi.org/10.1371/journal.pmed.1004560.

[13] C. Martinez-Perez, R. San-Cristóbal, P. Guallar-Castillon, M.Á. Martínez-González, J. Salas-Salvadó, D. Corella, et al., Use of Different Food Classification Systems to Assess the Association between Ultra-Processed Food Consumption and Cardiometabolic Health in an Elderly Population with Metabolic Syndrome (PREDIMED-Plus Cohort), Nutrients 13 (7) (2021) 2471, https://doi.org/10.3390/nu13072471.

[14] C.B. Terwee, C.A. Prinsen, A. Chiarotto, M.J. Westerman, D.L. Patrick, J. Alonso, et al., COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study, Qual. Life Res. 27 (5) (2018) 1159–1170, https://doi.org/10.1007/s11136-018-1829-0.

[15] D.E. Beaton, C. Bombardier, F. Guillemin, M.B. Ferraz, Guidelines for the process of cross-cultural adaptation of self-report measures, Spine 25 (24) (2000) 3186–3191, https://doi.org/10.1097/00007632-200012150-00014.

[16] M.B. Livingstone, P.J. Robson, Measurement of dietary intake in children, Proc. Nutr. Soc. 59 (2) (2000) 279–293, https://doi.org/10.1017/s0029665100000318.

[17] J.L. Walker, S. Ardouin, T. Burrows, The validity of dietary assessment methods to accurately measure energy intake in children and adolescents who are overweight or obese: a systematic review, Eur. J. Clin. Nutr. 72 (2) (2018) 185–197, https://doi.org/10.1038/s41430-017-0029-2.

[18] F. Visioli, F. Marangoni, V. Fogliano, D. Del Rio, J.A. Martinez, G. Kuhnle, et al., The ultra-processed foods hypothesis: a product processed well beyond the basic ingredients in the package, Nutr. Res. Rev 36 (2) (2023) 340–350, https://doi.org/10.1017/S0954422422000117.

[19] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C. D. Mulrow, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, Br. Med. J. (Clin. Res. Ed.). 372 (2021) n71, https://doi.org/10.1136/bmj.n71.

[20] J.P.T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M.J. Page, et al., Cochrane Handbook for Systematic Reviews of Interventions, 2nd ed., Wiley, Chichester, UK, 2019 https://doi.org/10.1002/9781119536604.

[21] L.B. Mokkink, E.B. Elsman, C.B. Terwee, COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0, Qual. Life Res. 33 (11) (2024) 2929–2939, https://doi.org/10.1007/s11136-024-03761-6.

[22] S. Abutbul Vered, C. Shani Levi, G.A. Rozen, I. Solt, G.S. Rozen, Development and validation of a computerized web-based quantitative food frequency questionnaire, Clin. Nutr. ESPEN. 52 (2022) 169–177, https://doi.org/10.1016/j.clnesp.2022.10.012.

[23] C.A. Borges, K.T. Gabe, P.C. Jaime, Consumer food environment healthiness score: development, validation, and testing between different types of food retailers, Int. J. Environ. Res. Public Health. 18 (7) (2021) 3690, https://doi.org/10.3390/ijerph18073690.

[24] C. Calvo-Porral, S. Rivaroli, J. Orosa-González, Proposal and validation of a measurement scale of the acceptance of ultra-processed Food Products, Foods 13 (10) (2024) 1481, https://doi.org/10.3390/foods13101481.

[25] E. Cardamone, F. Iacoponi, F. Fiori, M. Marinoni, U. Agrimi, M. Silano, et al., The development of a food frequency questionnaire for the assessment of ultra-processed food consumption in the Italian adult population: protocol for a validity and reproducibility study, Nutrients 16 (22) (2024) 3896, https://doi.org/10.3390/nu16223896.

[26] M.C. Correa-Madrid, N. Correa Guzmán, G. Bergeron, S.L. Restrepo-Mesa, G. Cediel, Validation of the NOVA score for the consumption of ultra-processed foods by young women of Medellín, Colombia, Ann. N Y Acad. Sci. 1528 (1) (2023) 69–76, https://doi.org/10.1111/nyas.15054.

[27] C.D. Costa, F.R. Faria, K.T. Gabe, I.F. Sattamini, N. Khandpur, F.H. Leite, et al., Nova score for the consumption of ultra-processed foods: description and performance evaluation in Brazil, Rev. Saude Publica. 55 (2021) 13, https://doi.org/10.11606/s1518-8787.2021055003588.

[28] C.D. Costa, F.S. Santos, K.T. Gabe, E.M. Steele, F.H. Marrocos-Leite, N. Khandpur, et al., Description and performance of two diet-quality scores based on the Nova classification, Rev. Saude Publica. 58 (2024) 47, https://doi.org/10.11606/s1518-8787.2024058006470.

[29] G. D'Abbronzo, C. Quaglia, G. Di Costanzo, R. Testa, R. Giacco, G. Riccardi, et al., Assessing the consumption of ultra-processed foods in people with diabetes: is a specific NOVA questionnaire always necessary? Nutrients 17 (1) (2024) 1, https://doi.org/10.3390/nu17010001.

[30] R. Dey, G. Linares, R. Munguía, E. Chávez, Construction and validation of an instrument for evaluating the consumption of food with Red Allura, Inf. Tecnol. 30 (3) (2019) 219–226, https://doi.org/10.4067/S0718-07642019000300219.

[31] M. Dinu, M. Bonaccio, D. Martini, M.P. Madarena, M. Vitale, G. Pagliai, et al., Reproducibility and validity of a food-frequency questionnaire (NFFQ) to assess food consumption based on the NOVA classification in adults, Int. J. Food Sci. Nutr. 72 (6) (2021) 861–869, https://doi.org/10.1080/09637486.2021.1880552.

[32] N. Erdoğan Gövez, E. Köksal, C. Martinez-Perez, L. Daimiel, Validity and reliability of the Turkish version of the screening questionnaire of highly processed food consumption (sQ-HPF), Nutrients 16 (15) (2024) 2552, https://doi.org/10.3390/nu16152552.

[33] L.J. Fangupo, J.J. Haszard, C. Leong, A.M. Heath, E.A. Fleming, R.W. Taylor, Relative validity and reproducibility of a food frequency questionnaire to assess energy intake from minimally processed and ultra-processed foods in young children, Nutrients 11 (6) (2019) 1290, https://doi.org/10.3390/nu11061290.

[34] W.B. Freire, B. Tello, P. Belmont Guerrón, Validation of NOVA 27 ultra-processed food screener: adaptation and performance in Ecuador, Public Health Nutr 28 (1) (2025) e105, https://doi.org/10.1017/S1368980025100475.

[35] K.T. Gabe, P.C. Jaime, Convergent validity and invariance analysis of a scale to measure adherence to eating practices recommended by the Dietary Guidelines for the Brazilian Population [Validade convergente e análise de invariância de uma escala de adesão a práticas alimentares recomendadas pelo Guia Alimentar para a População Brasileira], Rev. Bras. Epidemiol. [Braz. J.

Epidemiol.] 25 (2022) e220009, https://doi.org/10.1590/1980-549720220009.2.
[36] K.T. Gabe, G. Bassetto Junior, P.C. Jaime, Adaptation and validation of a tool for assessing food knowledge based on the Nova classification for the Brazilian context, Epidemiol. Serv. Saude. 34 (2025) e20240335, https://doi.org/10.1590/S2237-96222024v34e20240335.en.
[37] C.G. Hoil Sosa, G.A. Suárez Lara, L.E. Velazquez Melken, E.F. Peña Torres, N.Y. Torres Soto, B. Martínez Ramírez, et al., Diseño y validación de la escala de actitudes hacia el consumo de alimentos ultraprocesados en estudiantes de Ciencias de la Salud de la Universidad Autónoma del Estado de Quintana Roo, Acta. Univ. 34 (2024) 1–17, https://doi.org/10.15174/au.2024.4126.
[38] S. Hu, A.N. Gearhardt, E.M. LaFata, Development of the modified Highly Processed Food Withdrawal Scale (mProWS), Appetite 198 (2024) 107370, https://doi.org/10.1016/j.appet.2024.107370.
[39] S.D. Kébé, A. Diouf, P.M. Sylla, K. Kane, C. Dos Santos Costa, F.H. Leite, et al., Assessment of ultra processed foods consumption in Senegal: validation of the Nova-UPF screener, Arch Public Health 82 (1) (2024) 4, https://doi.org/10.1186/s13690-024-01239-y.
[40] S. Lemke, D.F. de Andrade, P. de Fragas Hinnig, S.A. da Silva, S.L. Vincenzi, D.M. Roberto, et al., Development and application of the Meal and Snack Assessment (MESA) quality scale for children and adolescents using item response theory, Nutr. J. 23 (1) (2024) 50, https://doi.org/10.1186/s12937-024-00948-y.
[41] E. Loftfield, P. Zhang, C.P. O'Connell, L.L. Kahle, K. Herrick, L. Abar, et al., Performance of a food frequency questionnaire for estimating ultraprocessed food intake according to the Nova classification system in the United States NIH-AARP diet and health study, J. Nutr. 155 (7) (2025) 2376–2384, https://doi.org/10.1016/j.tjnut.2025.04.029.
[42] C. Martinez-Perez, L. Daimiel, C. Climent-Mainar, M.Á. Martínez-González, J. Salas-Salvadó, D. Corella, et al., Integrative development of a short screening questionnaire of highly processed food consumption (sQ-HPF), Int. J. Behav. Nutr. Phys. Act. 19 (1) (2022) 6, https://doi.org/10.1186/s12966-021-01240-6.
[43] V.W. Motta, S.C. Lima, D.M. Marchioni, C.O. Lyra, Food frequency questionnaire for adults in the Brazilian Northeast region: emphasis on the level of food processing, Rev. Saude Publica. 55 (2021) 51, https://doi.org/10.11606/s1518-8787.2021055002473.
[44] A. Müller, C.D. Costa, R. Buffarini, J.D. Vaz, M.R. Domingues, P.C. Hallal, et al., Description and evaluation of an ultra-processed food consumption score for children, Rev. Saude Publica. 59 (2025) e5, https://doi.org/10.11606/s1518-8787.2025059005816.
[45] C.I. Oviedo-Solís, E.A. Monterrubio-Flores, G. Cediel, E. Denova-Gutiérrez, S. Barquera, Relative validity of a semi-quantitative food frequency questionnaire to estimate dietary intake according to the NOVA classification in Mexican children and adolescents, J. Acad. Nutr. Diet. 122 (6) (2022) 1129–1140, https://doi.org/10.1016/j.jand.2021.11.002.
[46] C.I. Oviedo-Solís, E.A. Monterrubio-Flores, S. Rodríguez-Ramírez, G. Cediel, E. Denova-Gutiérrez, S. Barquera, A semi-quantitative food frequency questionnaire has relative validity to identify groups of NOVA food classification system among Mexican adults, Front Nutr 9 (2022) 737432, https://doi.org/10.3389/fnut.2022.737432.
[47] C. Sarbagili-Shabat, S. Zelber-Sagi, N. Fliss Isakov, Y. Ron, A. Hirsch, N. Maharshak, Development and validation of processed foods questionnaire (PFQ) in adult inflammatory bowel diseases patients, Eur. J. Clin. Nutr. 74 (12) (2020) 1653–1660, https://doi.org/10.1038/s41430-020-0632-5.
[48] T.S. Santos, P.H. Araújo, D.F. Andrade, M.L. Louzada, M.A. Assis, B. Slater, Two validity evidences of the ESQUADA and Brazilians' dietary quality levels, Rev. Saude Publica. 55 (2021) 39, https://doi.org/10.11606/s1518-8787.2021055002397.
[49] S. Warkentin, L.A. Mais, M. do Rosário Dias de Oliveira Latorre, S. Carnell, J. A. Taddei, Validation of the comprehensive feeding practices questionnaire in parents of preschool children in Brazil, BMC Public Health 16 (2016) 603, https://doi.org/10.1186/s12889-016-3282-8.
[50] V. Kipnis, D. Midthune, L. Freedman, S. Bingham, N.E. Day, E. Riboli, et al., Bias in dietary-report instruments and its implications for nutritional epidemiology, Public Health Nutr 5 (6A) (2002) 915–923, https://doi.org/10.1079/PHN2002383.
[51] A.F. Subar, L.S. Freedman, J.A. Tooze, S.I. Kirkpatrick, C. Boushey, M. L. Neuhouser, et al., Addressing current criticism regarding the value of self-report dietary data, J. Nutr. 145 (12) (2015) 2639–2645, https://doi.org/10.3945/jn.115.219634.
[52] B.B. Reeve, R.D. Hays, J.B. Bjorner, K.F. Cook, P.K. Crane, J.A. Teresi, et al., Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS), Med. Care 45 (5) (2007) S22–S31, https://doi.org/10.1097/01.mlr.0000250483.85507.04. Suppl 1.
[53] W. Willett, Nutritional epidemiology, 3rd ed., Oxford University Press, 2012 https://doi.org/10.1093/acprof:oso/9780199754038.001.0001.
[54] A.E. Black, Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations, Int. J. Obes. Relat. Metab. Disord. 24 (9) (2000) 1119–1130, https://doi.org/10.1038/sj.ijo.0801376.
[55] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement error in nonlinear models: a modern perspective, 3rd ed., Chapman and Hall/CRC, 2006.
[56] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (2) (1999) 135–160, https://doi.org/10.1177/096228029900800204.
[57] H.C. De Vet, C.B. Terwee, L.B. Mokkink, D.L. Knol, Measurement in medicine: a practical guide, Cambridge University Press, 2011, https://doi.org/10.1017/CBO9780511996214.
[58] C.B. Terwee, S.D. Bot, M.R. de Boer, D.A. Van der Windt, D.L. Knol, J. Dekker, et al., Quality criteria were proposed for measurement properties of health status questionnaires, J. Clin. Epidemiol. 60 (1) (2007) 34–42, https://doi.org/10.1016/j.jclinepi.2006.03.012.
[59] J.A. Husted, R.J. Cook, V.T. Farewell, D.D. Gladman, Methods for assessing responsiveness: a critical review and recommendations, J. Clin. Epidemiol. 53 (5) (2000) 459–468, https://doi.org/10.1016/s0895-4356(99)00206-1.
[60] G.H. Beaton, J. Milner, P. Corey, V. McGuire, M. Cousins, E. Stewart, et al., Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation, Am. J. Clin. Nutr. 32 (12) (1979) 2546–2559, https://doi.org/10.1093/ajcn/32.12.2546.
[61] C.R. Sadler, T. Grassby, K. Hart, M. Raats, M. Sokolović, L. Timotijevic, Processed food classification: conceptualisation and challenges, Trends Food Sci. Technol. 112 (2021) 149–162, https://doi.org/10.1016/j.tifs.2021.02.059.
[62] R. Bleiweiss-Sande, K. Chui, E.W. Evans, J. Goldberg, S. Amin, J. Sacheck, Robustness of food processing classification systems, Nutrients 11 (6) (2019) 1344, https://doi.org/10.3390/nu11061344.
[63] C.A. Prinsen, L.B. Mokkink, L.M. Bouter, J. Alonso, D.L. Patrick, H.C. de Vet, et al., COSMIN guideline for systematic reviews of patient-reported outcome measures, Qual. Life Res. 27 (5) (2018) 1147–1157, https://doi.org/10.1007/s11136-018-1798-3.
[64] G.B. Willis, Cognitive interviewing: A tool for improving questionnaire design, 1st ed., Sage Publications, 2004.
[65] G.O. Boateng, T.B. Neilands, E.A. Frongillo, H.R. Melgar-Quiñonez, S. L. Young, Best practices for developing and validating scales for health, social, and behavioral research: a primer, Front Public Health 6 (2018) 149, https://doi.org/10.3389/fpubh.2018.00149.
[66] T.A. Brown, Confirmatory factor analysis for applied research, 2nd ed., Guilford Publications, 2015.
[67] G.W. Cheung, R.B. Rensvold, Evaluating goodness-of-fit indexes for testing measurement invariance, Struct. Equ. Model. 9 (2) (2002) 233–255, https://doi.org/10.1207/S15328007SEM0902_5.
[68] D.L. Putnick, M.H. Bornstein, Measurement invariance conventions and reporting: the state of the art and future directions for psychological research, Dev. Rev. 41 (2016) 71–90, https://doi.org/10.1016/j.dr.2016.06.004.
[69] R.L. Prentice, Y. Huang, L.H. Kuller, L.F. Tinker, L.V. Horn, M.L. Stefanick, et al., Biomarker-calibrated energy and protein consumption and cardiovascular disease risk among postmenopausal women, Epidemiol. (Camb. Mass). 22 (2) (2011) 170–179, https://doi.org/10.1097/EDE.0b013e31820839bc.
[70] A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L.O. Dragsted, J. Draper, et al., The food metabolome: a window over dietary exposure, Am. J. Clin. Nutr. 99 (6) (2014) 1286–1308, https://doi.org/10.3945/ajcn.113.076133.
[71] M. Guasch-Ferré, S.N. Bhupathiraju, F.B. Hu, Use of metabolomics in improving assessment of dietary intake, Clin. Chem. 64 (1) (2018) 82–98, https://doi.org/10.1373/clinchem.2017.272344.
[72] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, et al., Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), Metabolomics 3 (3) (2007) 211–221, https://doi.org/10.1007/s11306-007-0082-2.
[73] J.T. Leek, R.B. Scharpf, H.C. Bravo, D. Simcha, B. Langmead, W.E. Johnson, et al., Tackling the widespread and critical impact of batch effects in high-throughput data, Nat. Rev. Genet. 11 (10) (2010) 733–739, https://doi.org/10.1038/nrg2825.
[74] J.A. Westerhuis, H.C. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J. van Velzen, et al., Assessment of PLSDA cross validation, Metabolomics 4 (1) (2008) 81–89, https://doi.org/10.1007/s11306-007-0099-6.