RESEARCH
Monograph

Best Practices for Conducting and Interpreting Studies to Validate Self-Report Dietary Assessment Methods

Sharon I. Kirkpatrick, PhD, RD*; Tom Baranowski, PhD; Amy F. Subar, PhD, RD; Janet A. Tooze, PhD; Edward A. Frongillo, PhD

ARTICLE INFORMATION

Article history:
Submitted 17 October 2018
Accepted 12 June 2019
Available online 11 September 2019

Keywords:
Dietary assessment
Validation
Validity
Reliability
Psychometric

Supplementary materials:
PowerPoint presentation available at www.jandonline.org

https://doi.org/10.1016/j.jand.2019.06.010

*Registered with the College of Dietitians of Ontario, Canada.

ABSTRACT

Careful consideration of the validity and reliability of methods intended to assess dietary intake is central to the robustness of nutrition research. A dietary assessment method with high validity is capable of providing useful measurement for a given purpose and context. More specifically, a method with high validity is well grounded in theory; its performance is consistent with that theory; and it is precise, dependable, and accurate within specified performance standards. Assessing the extent to which dietary assessment methods possess these characteristics can be difficult due to the complexity of dietary intake, as well as difficulties capturing true intake. We identified challenges and best practices related to the validation of self-report dietary assessment methods. The term *validation* is used to encompass various dimensions that must be assessed and considered to determine whether a given method is suitable for a specific purpose. Evidence on the varied concepts of validity and reliability should be interpreted in combination to inform judgments about the suitability of a method for a specified purpose. Self-report methods are the focus because they are used in most studies seeking to measure dietary intake. Biomarkers are important reference measures to validate self-report methods and are also discussed. A checklist is proposed to contribute to strengthening the literature on the validation of dietary assessment methods and ultimately, the nutrition literature more broadly.
J Acad Nutr Diet. 2019;119(11):1801-1816.

DIETARY ASSESSMENT METHODS DATE BACK decades. The earliest reference to dietary assessment published in this *Journal* was in 1947 by Burke,[1] who described the dietary history. A PubMed search for dietary assessment conducted in May 2019 returned more than 43,000 hits for articles published in peer-reviewed journals, more than 14,000 of which were published during the past 5 years. This suggests broad interest in measuring food consumption among human beings. Dietary risk factors have been recognized globally as key contributors to morbidity and mortality,[2] and there is continued interest in characterizing their relationships with health outcomes.[3] There is unprecedented attention globally to interventions such as product labeling and taxation to support healthy eating and reduce disease risk among populations,[4-6] with a concomitant need to monitor intake. Although commonly used dietary assessment methods have been subject to criticism based on their reliance on self-report,[7,8] there is more interest than ever in measuring what people eat and drink.[9] This speaks to the need for robust efforts to understand whether a given method or measure is well suited to a given purpose and context.

Dietary intake is a complex human exposure, characterized by multidimensionality and dynamism,[10] as well as variability across populations and cultures.[11,12] When assessing intake, interest may be solely in whether consumption occurred. There often is also interest in quantifying the amounts consumed of foods, beverages, food groups,[13,14] nutrients, and other dietary constituents such as phytochemicals and contaminants[15,16] and in characterizing dietary patterns[10,17-19] and eating behaviors, such as meal and snacking patterns.[20,21] For each of these exposures, interest is typically in habitual, or usual, intake over some period of time,[22-24] whereas foods and beverages consumed generally vary across days, seasons, and the lifecycle.[23,25,26] When making decisions regarding methods to assess dietary intake, critical considerations include the degree to which the data

captured correspond to the dietary exposures of interest and how close the values yielded by the method are to estimates of true usual intake over a given period of time among members of a specified population.

Because there are few known objective measures of dietary intake and these are infeasible for most studies, self-report dietary assessment methods are typically used.[27,28] True usual intake is not observable in most studies,[27] underscoring the need for robust validation to understand whether a given method yields estimates of intake that are reasonable for providing data for a given purpose, context, and population.[29] Understanding the properties of an assessment method can also help to identify analytic methods and study design considerations to mitigate error in the measurement of dietary exposures.[30] We identified complexities and recommended best practices related to the validation of self-report dietary assessment methods. The emphasis is on applications of dietary assessment methods to estimate dietary intake among groups, not among individual persons as may be the case in clinical settings.

CONCEPTS OF VALIDITY, RELIABILITY, AND VALIDATION

"Validation is the process of determining whether a measure or indicator is suitable for providing useful analytical measurement for a given purpose and context. A measure or indicator is valid if each of six criteria are met: 1) its construction is well-grounded in theory; 2) its performance is consistent with that theory; it is 3) precise, 4) dependable, and 5) accurate within specified performance standards; and 6) its accuracy is attributable to the well-grounded theory for that purpose and context ."[29] This holistic conceptualization indicates that distinctions among different types of validity are somewhat arbitrary and that data from a method are only useful when they represent the construct of interest.[31]

Two conceptual systems for validation within the field of nutrition sciences include the biometric, in which truth is typically observable,[32] and the psychometric,[33-35] in which measures of an underlying construct, for example, depression or self-esteem, typically depend on self-reported subjective assessments and truth is not observable. Dietary intake is theoretically observable (eg, using objective methods such as observation) in line with the biometric system, but in most studies, measures of true intake are not feasible. As a result, psychometric methods have often been adopted in the validation of dietary assessment tools.[33] In the psychometric system, the six concepts noted above and explored in detail elsewhere by Frongillo and colleagues[29] are often described as face or content validity (1), construct validity (2), test–retest reliability (3 and 4), and criterion validity (5 and 6). Validity typically takes precedence[31,35-37] over reliability because even in the event that a method provides consistent measurements, those measurements are useful only when they are sufficiently accurate for a specified purpose. Concepts related to measurement error, which refers to the difference between a true and measured value, are also pertinent.[38] In terms of characteristics of dietary data, the terms systematic error and bias generally relate to validity, whereas random error relates to reliability[39] (although the use of unreliable data for analytic purposes can lead to biased results).

RESEARCH SNAPSHOT

Research Question: What are key challenges and best practices related to the validation of self-report dietary assessment methods?

Key Findings: Assessing the validity and reliability of dietary assessment methods is difficult due to the complexity of dietary intake and challenges in capturing true intake. The study design must be carefully aligned with the purpose of a given validation. Evidence on the varied concepts of validity and reliability should be interpreted in combination and in a nuanced manner to inform judgements about the suitability of a method for a specified purpose.

We use the term *validation* to encompass various dimensions that must be assessed and considered to determine whether or not a given method is suitable for a specific purpose, with the explicit recognition that reliability is embedded within a more holistic conceptualization of validity. Evidence on the varied concepts of validity and reliability should be interpreted in combination to inform judgments about the suitability of a method for a given purpose. There is also interest in the responsiveness of methods for detecting change that truly occurs, as well as whether the administration of the same method (or adaptations of it) yield comparable data across populations or settings. Feasibility of a given method for use with a given population in a given setting is another consideration but is not addressed here.

We present complexities in dietary assessment followed by a summary of common self-report assessment methods and of objective measures that may be used to provide reference data for validation purposes, and a description of sources of error in data from self-report methods. Limitations in the current validation literature are briefly highlighted, leading into the discussion of best practices for validation of self-report dietary assessment methods.

THE EXPOSURES OF INTEREST IN DIETARY ASSESSMENT

Diet is multidimensional[10] in that people consume various foods and drinks, and potentially supplements, each of which has its own profile of nutrients, phytochemicals, and other exposures (eg, contaminants). People may eat these foods and drinks in different combinations and patterns in terms of meal frequency and timing. Multidimensionality has been described as "the numerous attributes of dietary intake and the inherent complexities of interdependence and synergy."[10] Given the complexity that multidimensionality entails, when considering the properties of a dietary assessment tool, investigators must clearly articulate which dimensions or "layers"[10] of diet are of interest. For example, the desire to obtain an accurate estimate of total absolute intake of a single dietary component, such as sugars or a given vitamin, suggests the need for a method that captures frequency of consumption and amounts consumed of all foods and beverages contributing the dietary component of interest with a high degree of validity. To assess whether a change in intake of one dietary component is accompanied by changes in other

components, such as in response to an intervention, the method must comprehensively capture the relevant foods and beverages and be responsive to change that truly occurs.

Diet may be measured at a single time point in many studies,[10] but for surveillance, epidemiology, and intervention research, usual intake over time among the population is of primary interest. This is because nutrient requirements and food group recommendations are intended to be met over time,[40] it is generally long-term diet and not what is eaten on a given day that influences health and disease risk (although acute exposures can be critical for components such as alcohol or contaminants[41]), and interventions usually seek to promote sustained changes in eating patterns that will contribute to long-term nutrition-related and overall health. Estimating usual intake is challenging. First, diet is dynamic.[10] Day-to-day variation in intake can be substantial for dietary components consumed episodically by most persons in a given population.[42] For example, whereas intake of ubiquitously consumed dietary components, such as total sugars or refined grains, do not show high day-to-day variation in the context of current Western eating patterns, this is not true for more episodically consumed foods, such as fish, green vegetables, and whole grains.[13] Intake may also vary seasonally, although recent research considering intake of macronutrients, micronutrients, and food groups among a sample of US adults suggested seasonality in intake was not substantial.[43] Second, dietary exposures at critical periods, such as in utero, adolescence, or pregnancy, can influence human health even before conception[44,45] and across the lifecycle.[10] There is thus an emerging focus on temporal patterns in relation to health and disease outcomes.[10,24,46,47]

Because eating patterns, including the degree of multidimensionality and dynamism, vary across populations,[11,12] the nature of the dietary exposures of interest and methods suitable for capturing them may differ across settings and populations. For example, a study to monitor protein intake over a given period among young children living in a lower- or middle-income country with a relatively monotonous diet might require a different method than a study examining protein intake among adults in a high-income country. Comparability, or equivalence, of methods across populations with different dietary exposures is thus of interest.

SELF-REPORT METHODS FOR ASSESSING DIETARY INTAKE

Data collection and analytic strategies within dietary assessment are generally intended to yield estimates of usual intake. Usual intake may be queried directly during a specified time period or estimated using statistical methods, possibly by combining data from tools that capture short-term and long-term intake.[26,27,38,48,49]

Short-term dietary assessment methods capture a detailed accounting of all foods and beverages, and sometimes supplements, consumed over a short period of time (eg, days). These methods include 24-hour dietary recalls (24HR) and food records (FR), which can capture multidimensionality because all foods and beverages consumed over a given number of days are queried.[27,38] To estimate mean usual intake among a group, a single administration of a 24HR or FR is sufficient.[38,50] Characteristics beyond mean intake, such as the proportion of a group with intakes below or above a threshold, may also be of interest and cannot be estimated based on data for a single day because of the need to account for day-to-day variation.[26] To estimate distributions of usual intake among a group, many repeats of 24HR or FR can be collected and averaged. In practice, it is usually not possible to collect a sufficient number of repeats to average out the effects of day-to-day variation.[26,50] Nevertheless, data from a small number of nonconsecutive 24HR or FR administrations can be used to estimate usual intake through analytic strategies that account for within-person variation (mostly contributed by day-to-day variation), a form of random error.[26] Nonconsecutive administrations (eg, of a 3-day or 7-day FR) are recommended as intake on consecutive days may be affected by autocorrelation (eg, through leftover and compensation effects), precluding accurate estimation of day-to-day variation.[38] Depending on the study design, assumptions are often made regarding the period to which usual intake pertains. For example, in analyses drawing upon surveillance systems such as the National Health and Nutrition Examination Survey,[51] in which data are typically collected across the sample over a year, usual intake among populations and population subgroups may be conceptualized as reflecting the survey year. Methods to estimate distributions of usual intake among groups do not allow characterization of usual intake for a given individual, for whom a substantial number of repeat 24HR or FR administrations is needed to reliably estimate usual intake over a specified period.[22]

Methods such as 24HR and FR can be administered in different ways, with potential implications for the validity of the resulting data. Adaptations of the methods continue to be developed as researchers seek to minimize the limitations of existing tools, such as by leveraging technology.[52-56] Approaches such as the Automated Multiple-Pass Method[57-59] have been developed to improve the validity of data collected using interviewer-administered recalls. Automated self-administered recalls and records[52,55] have also been developed, as have mobile FR applications.[54] Aids, such as household measures and digital images, may be used to enhance accuracy of portion size estimation, with the format and presentation of the aids potentially influencing the degree of accuracy of reported data.[60]

Data from 24HR and FR are theoretically comparable across populations because they capture all foods and beverages consumed,[38] although the extent to which those foods and beverages can be linked to food composition databases may vary. When comparability across populations or settings is relevant to a given use of 24HR or FR, it is also important to ensure that questionnaires, or interfaces in the case of technology-enabled tools, used to collect data are understood similarly and include appropriate food and beverage items.[61]

Long-term dietary assessment methods are those that aim to measure habitual intake of foods and beverages, and sometimes supplements, over a long period of time, such as a month or year. Such methods include food frequency questionnaires (FFQ) and frequency-based screeners.[27,28] Frequency questionnaires and to a lesser extent, screeners, may enable consideration of multidimensionality when sufficient foods and beverages are included to capture varied aspects of the diet, but the food groupings in the questionnaire will influence the granularity with which facets of diet can be examined. Because these tools directly ask about frequency of consumption over a long period of time, the data from them

are not believed to be highly influenced by within-person random variation in intake. There can be some variation across administrations spread over time, but within-person variation tends to be smaller than for short-term measures.[62]

Suitability of a given FFQ for use with different populations may be limited.[63] For example, a questionnaire tailored to one population may not sufficiently capture the foods and beverages commonly consumed in another, leading to misestimation of intake of food groups, nutrients, or other dietary components. Differences in data captured by different FFQ (even in the same population) can also reflect differences in design, such as variations in frequency response categories, methods for portion size estimation, nutrient and food group databases, and the underlying algorithms and procedures to determine mean daily intake values. Mode of administration may also be relevant.

OBJECTIVE MEASURES OF DIETARY INTAKE

Objective measures of intake may be used to provide reference data against which self-report methods are compared to assess validity. True intake can be comprehensively quantified through detailed observation or in feeding studies in which intake is documented for one or more eating occasions.[60,64,65] These data are likely to be influenced by some sources of error, such as observer error that results in missed foods or drinks, inaccurate calibration of scales for weighing plate waste, or reactivity on the part of respondents in response to being monitored in the case that it is not sufficiently unobtrusive. These can be minimized through training and quality control procedures. The data collected using observation and feeding studies do not typically reflect usual intake because it is challenging to implement these methods over the long term in naturalistic settings.

Biomarkers are also relevant to validation of dietary assessment methods.[50,66] Biomarker data do not provide information on what was actually consumed by individuals, and thus do not provide substantial insights into factors such as the multidimensionality of diet.[67] Further, the collection of biomarker data is burdensome, expensive, and usually feasible for only a short period of time in small samples of community-dwelling populations.[50] Biomarker data are useful for assessing the extent to which data from self-report dietary assessment methods reflect truth. Specifically, recovery biomarkers, which are biological products considered markers of true intake because they are directly related to intake and not influenced by homeostasis or substantial differences in metabolism between persons,[50,66] have been identified for a small number of dietary components. For example, doubly-labeled water represents energy expenditure over a 10-14 day period assuming individuals are in energy balance,[68] and 24-hour nitrogen level in urine provides a measure of protein intake over 24 hours.[69] These recovery biomarkers are believed to provide nearly unbiased estimates of energy and protein intake, respectively, with some potential for bias due to incomplete collection of 24-hour urine output. Recovery biomarkers have also been recognized for potassium and sodium, again based on 24-hour urine collection.[70] Like short-term dietary assessment methods, recovery biomarkers are affected by within-person variation[68] and thus, repeat administrations are needed to estimate usual intake among a group by characterizing and adjusting for within-person random error.

Predictive biomarkers,[71,72] like recovery biomarkers, show a stable dose–response relationship to intake and offer promise for mitigating error inherent in self-report data, but few have been identified (eg, 24-hour urinary fructose and sucrose as markers of sugar intake[71]). Thus, their use in validation efforts has been limited. The distinction between recovery and predictive biomarkers is that the overall recovery is lower for those designated as predictive,[73,74] but their stable relationship with intake may nonetheless help to identify reporting errors. Concentration biomarkers do not have the stable relationship with intake that is observed for recovery biomarkers and "cannot be translated into absolute levels of intake,"[74] but correlate with intake and have been shown to perform well compared to recovery biomarkers.[75] Examples of concentration biomarkers include serum carotenoids, which may be correlated with fruit and vegetable intake. Some validation studies have compared self-report data to concentration biomarkers as an indication of validity of the self-report method.[76,77]

SOURCES OF ERROR IN DATA COLLECTED USING SELF-REPORT METHODS

All self-report dietary assessment methods are subject to measurement error[28]; that is, the difference between a true and measured value. Error can be random or systematic.[78] Random error occurs in both directions from true intake[79] and leads to unreliable estimates.[29,78] A main source of random error in estimating usual dietary intake is day-to-day variation,[25] which primarily affects data collected using short-term measures such as 24HR and FR.[50,62] Random errors may also be due to variation in data collection parameters (eg, time of day or day of week of data collection) or biological or environmental factors (eg, variation in vitamin C content in relation to storage conditions). Random error, which may vary by dietary component, can be addressed using repeated nonconsecutive days of measurement with the application of appropriate statistical techniques.[26]

Systematic error, also known as bias, results in misestimation in a particular direction.[28,78] Systematic error is related to the concepts of accuracy and validity. Sources of systematic error include recall biases (eg, consistently forgetting items consumed when completing a 24HR or FFQ of past intake[27]), reactivity biases (eg, changing eating patterns in response to real-time monitoring in the case of FR[28,80]), intake-related biases (ie, quantity of intake is related to discrepancies in reporting[81]), and social desirability biases (ie, under- or overreporting of certain foods and beverages based on perceived healthfulness[82-84] or in response to intervention messaging). Systematic error may also be due to deficiencies in the measure, such as lack of inclusion of specific foods that contribute to intake of a dietary component of interest in a given population. Poor alignment between characteristics of the method and participants' capacities, such as literacy and/or numeracy, may also lead to systematic error in reported consumption.[38] Systematic error can be mitigated only in cases in which true intake is known. This is limited to the relatively few studies that make use of objective markers of truth, such as observation or recovery biomarkers.[66]

Studies of measurement error in self-report dietary data have used recovery biomarkers[66] for energy, protein, potassium, and sodium.[62,70,85] Biomarker-based validation studies have indicated that 24HR recall data are influenced to a greater extent by random error than are FFQ data, whereas the opposite is true for systematic error.[70,85] This is consistent with our understanding of the design and administration of these methods. For example, recalls are generally used to collect data for a day or a few days and therefore do not account for day-to-day variation in intake (without statistical correction using repeat nonconsecutive administrations), resulting in random error. This does not reflect misreporting on the part of respondents, but is a source of error when interest is in usual intake rather than intake on a given day or a few days. Frequency questionnaires may omit food sources that are important to a given population subgroup and/or may impose substantial cognitive challenges in terms of requiring estimation of usual intake over a long period of time, contributing to systematic error.[38] Although 24HR data are not bias-free, they have been shown to produce data with less bias than FFQs in a range of populations.[70,85] Thus, 24HR are sometimes used as reference measures in validation studies. Data from FR may also be used as reference measures with the assumption that they perform similarly to 24HR and provide data that are closer to truth than other self-report methods such as FFQ or screeners.

LIMITATIONS IN THE CURRENT DIETARY ASSESSMENT VALIDATION LITERATURE

Frongillo and colleagues[29] described limitations in validation studies in the field of nutrition more generally. An issue specific to dietary assessment is that claims of validity often rely on suboptimal methods. The collection of true intake data to which to compare self-report methods is a challenge given the paucity of and burdens associated with objective markers. As a result, research described as validation often entails the comparison of error-prone measures to one another. Although often unavoidable, such studies are often presented erroneously as indicating that measures are valid, rather than the extent to which they agree with another error-prone measure.

In addition, a range of statistical tests are used within validation studies and these may capture agreement, association, or bias at the level of groups or individuals. Lombard and colleagues[30] noted that there is no consensus on the ideal type and number of tests and found that among 60 validation studies, 21 different combinations of statistical tests were used. Although different tests yield different insights, they must be appropriately interpreted in combination and in relation to the validation study design to guide conclusions regarding validity and reliability and future uses of a given method. In their review of validation research, Lombard and colleagues[30] found that "all studies concluded that the test dietary assessment method was valid for use in the respective populations." Such statements suggest a dichotomy in terms of whether a method is valid or not,[31] which is not appropriate because all self-report measures capture intake with some degree of error or inaccuracy. Investigators seeking to make use of a particular method may be motivated to describe it as "validated," potentially with little to no information on the degree of validity or reliability observed, for what dietary components, or in what population(s), making it difficult to interpret the suitability of the method for their purpose.

BEST PRACTICES IN DIETARY ASSESSMENT VALIDATION

This section describes overarching considerations for the design of validation studies, corresponding to the presented concepts of validity, reliability, and validation. Although not necessarily exhaustive given that there are large literatures on relevant concepts, key statistical considerations and critical issues related to the validation of dietary assessment measures are noted.

A checklist (see the Figure) is proposed for use in developing and reviewing articles describing the validation of dietary assessment methods. Some aspects of the checklist reflect general overarching considerations and are adapted from the Strengthening the Reporting of Observational studies in Epidemiology Statement-nutritional epidemiology extension,[86] which is aimed at supporting consistent reporting of the details of nutritional epidemiology studies. Other considerations are specific to the design and implementation of validation studies. Specifications are also provided on how best to report and interpret results.

Overarching Considerations

The purpose of the measure being validated and the dietary constructs of interest should be clearly identified in all facets of an article reporting on a validation study, including the title and abstract. Given the complexity of diet, it is essential to fully define the dietary exposures, with careful consideration to the dietary behaviors (eg, meal patterns and social context) or constituents (eg, foods, food groups, nutrients, patterns, or other food components) and the time period (eg, acute intake over a short period of time such as a day or usual intake over some longer period such as a month or a year). The ways in which the method will be used to characterize these dietary behaviors or intakes (eg, capture occurrence of consumption, rank intake among a group of persons, or estimate absolute intake) should be characterized *a priori*. These considerations will inform design decisions such as the number and timing of administrations of the method being validated, the selection of criterion or comparison measures(s) (if any), and appropriate statistical techniques and their interpretation.

The intended target populations and settings of interest are pivotal for informing study design, including the sampling frame and eligibility/exclusion criteria, as well as appropriate interpretation. The method to be validated and its intended purpose should be placed in the context of what is already known about the properties of similar methods proposed for similar purposes in similar and different populations.

The measure that is being validated, as well as any criterion or comparison measures when applicable, should be described in detail, including how the measure and the criterion or comparison measures were developed and/or adapted, the format and mode of administration, associated food composition databases, and the results of any prior validation studies or other testing. In situations involving repeat administrations of a method, the potential for contamination between administrations, changes to usual

Topic	Description
Title and abstract	*Indicate the study's design and purpose, including the construct of interest (ie, dietary components of interest and over what time frame) and dietary assessment measure undergoing validation, in the title and/or abstract.* *Provide an informative and balanced summary of what was done and what was found.* Avoid summary statements that do not reflect the totality of the findings and that treat validity and/or reliability as dichotomous rather than properties operating on a continuum of low to high.
Introduction	
Background and rationale	*Explain the scientific background and rationale for the validation study.* Place the study in the context of existing research on dietary assessment methods, and build justification for the specific focus of the validation study.
Objectives	*State specific objectives, including any prespecified hypotheses.* Objectives should explicitly identify the aim of the study in terms of the properties assessed to examine suitability of the measure to address a given research purpose in a specified population and setting.
Methods	
Study design	*Present key elements of the study design early.*
Measure	Describe the measure undergoing validation in detail, including how it was developed/adapted, its format, the method and location of administration, and salient characteristics such as the associated food composition database(s) and whether supplement intake is assessed. Describe the intended use of the measure (ie, to capture occurrence of consumption (or not), to rank intake among a group of persons, or to estimate absolute intake), as well as the dietary components (foods, food groups, nutrients, patterns, or other food components) and time frame of interest.
Settings	*Describe the setting, locations, and relevant dates, including periods of recruitment and data collection. Describe any characteristics of the study setting that might influence participants' dietary intake.*
Participants	*Provide the eligibility criteria and sources and methods of selection of participants.* Justify the sample size. Describe the representativeness of the sample to the target population and discuss response rates. *Report characteristics related to nutrition, dietary intake, and physiology considered in defining the eligibility criteria.*
Procedures	**Is the measure well-constructed and grounded in an understanding of the underlying phenomenon of interest? (face and content validity)** • Provide a detailed presentation of the procedures employed to gather feedback to assess whether items reflect the construct of interest (eg, reviews of the relevant literature(s), surveys of lay persons and/or relevant experts). Address approaches employed to ensure the items included, for example, in a frequency-based questionnaire, are comprehensive and reflect the most commonly-consumed sources of the dietary components of interest, as well as relevant portion sizes (if applicable). **Does the measure perform in a manner consistent with the theory underlying its construction? (construct validity)** • Describe the methods used to examine whether or not the measure assesses the intended construct (eg, contrasted groups) and the theory that underlies this particular approach to assessing this aspect of validity. Include a detailed description of statistical procedures used.

(continued on next page)

Figure. Checklist for authors and reviewers when describing studies to validate dietary assessment measures. Italicized text is adapted from: Lachat C, Hawwash D, Ocké MC, et al. Strengthening the reporting of observational studies in epidemiology—Nutritional epidemiology (STROBE-nut): An extension of the STROBE statement. *PLOS Med*. 2016;13(6). 2016;e1002036.[86] (NOTE: Information from this figure is available at www.jandonline.org as part of a PowerPoint presentation.)

Topic	Description
	Is the measure accurate within specified performance standards? (criterion or relative validity) • Describe criterion or comparison reference measure(s) (eg, biomarkers, observation, feeding studies, other self-report tools) used to validate the measure of interest in detail, including potential biases. Avoid language such as gold standard to describe reference measures. o Justify the use of particular reference measure(s) in terms of the dietary components and time frame of interest. o Address whether the reference measure(s) and the measure undergoing validation assess intake over the same period of time. Describe considerations regarding number and timing of administration(s) of both the measure to be evaluated and the reference measure(s). • Explicitly describe and justify the statistical procedures used, including multiple tests as appropriate. **Does the measure produce data that are precise and dependable? (reliability)** • Describe the particular aspects of reliability of interest; for example, precision (test–retest reliability) or interrater reliability. If test–retest reliability, describe the relevant period of time over which reliability is of interest. • Discuss whether dietary intake could be expected to change over the relevant time period due to confounding factors and how this was addressed. • Explicitly describe the statistical tests used. Use multiple tests as appropriate. **Is the measure responsive to change?** • Describe procedures used to assess responsiveness to meaningful change over time and to assess the smallest detectable change for given dietary components. • Comment on issues related to statistical power related to responsiveness. **Does the measure produce data that are equivalent or comparable across populations?** • Discuss approach to comparability in terms of adapting measures for different contexts or identifying particular variables captured using different measures but that can be theoretically harmonized.
Results	
Participants	*Report the numbers of individuals at each stage of the study and give reasons for nonparticipation at each stage. Consider the use of a flow diagram to illustrate.* Report the results of each procedure implemented for each dietary construct of interest.
Descriptive data	*Give information on the study participants (eg, demographic characteristics)* Provide salient information regarding dietary intake data (eg, low or high values, avoidance of certain foods)
Discussion	
Key results	Make use of all statistical tests and/or procedures to objectively summarize the key results with reference to the study objectives. Discuss the degree of validity, reliability, sensitivity to change, and/or equivalence of the evaluated measure as appropriate to the study, rather than referring to these properties as present or absent. When comparing error-prone measures to one another, consider the contribution of correlated error to measures of association. Avoid overstating the level of validity or reliability based on the available data.
Limitations	Describe study limitations that may affect conclusions. This may include the reference measures in studies of validity or recruitment methods (eg, paid volunteers) in any study.

(continued on next page)

Figure. *(continued)* Checklist for authors and reviewers when describing studies to validate dietary assessment measures. Italicized text is adapted from: Lachat C, Hawwash D, Ocké MC, et al. Strengthening the reporting of observational studies in epidemiology—Nutritional epidemiology (STROBE-nut): An extension of the STROBE statement. *PLOS Med*. 2016;13(6). 2016;e1002036.[86] (NOTE: Information from this figure is available at www.jandonline.org as part of a PowerPoint presentation.)

Topic	Description
Interpretation	Limit interpretations about validity, reliability, responsiveness, and/or equivalence of the evaluated measure to the specific populations and contexts evaluated, as well as the particular objectives (eg, interpretations of an evaluation of validity should be limited to validity and not reliability). Base interpretations on the totality of the evidence, including all tests and comparisons conducted, as well as results from similar studies. Place the findings in the context of other literature.
Generalizability	Describe potential appropriate and inappropriate uses of the measure given the study design and findings. Describe features of the measure that may influence the design of studies proposing to make use of it; for example, sample size calculations to account for loss of power due to biased measurement of dietary intake.
Other information	
Funding	*Give the source of funding and the role of the funders in the validation study and, when applicable, in prior studies on which the present validation study is based.*
Ethics	*Describe the procedures for consent and study approval from ethics committee(s).*

Figure. *(continued)* Checklist for authors and reviewers when describing studies to validate dietary assessment measures. Italicized text is adapted from: Lachat C, Hawwash D, Ocké MC, et al. Strengthening the reporting of observational studies in epidemiology—Nutritional epidemiology (STROBE-nut): An extension of the STROBE statement. *PLOS Med*. 2016;13(6). 2016;e1002036.[86] (NOTE: Information from this figure is available at www.jandonline.org as part of a PowerPoint presentation.)

dietary intake, and declining data quality over time (eg, due to learning effects or fatigue among study participants) should be considered.

As in any study, the setting, location, and relevant time periods for recruitment and data collection should be detailed, as should study characteristics that might influence either actual or reported dietary intake. For example, in a validation substudy embedded within a larger intervention study, the intervention itself may elicit changes in intake and/or reporting biases.[87] In addition, participants' sociodemographic characteristics should be summarized, and information on response rates and generalizability of the sample to the target population provided.

Is the Measure Well Constructed and Grounded in an Understanding of the Underlying Phenomenon of Interest?

A valid measure is well constructed and grounded in an understanding of the underlying phenomenon of interest, or shows face and content validity, also referred to as translational validity.[35] Face validity refers to whether a measure appears, "on the face of it,"[35] to provide the information that is sought.[34,35] It is a subjective[35] rather than empirical assessment of whether a measure meets expectations set by the assessor, and may be examined among lay people and experts who determine, among other things, whether or not a screener, for example, appears to capture intake of a given food or food groups. Face validity provides insights into how participants might interpret and respond to items, not whether the tool does indeed measure the construct of interest[35] (necessitating assessment of other forms of validity).

Content validity is also subjective but relies upon experts and feedback mechanisms (such as the Delphi technique[88]) to develop and evaluate items to capture the salient features or dimensions of a construct.[35] For example, although a lay person may be able to list vegetables that should be included in a screener, a content expert is needed both to ensure commonly consumed vegetables within the population are represented and to capture frequency of consumption reflective of intake patterns. Content validity is particularly relevant to tools such as screeners targeted to certain foods or food groups within a given population. Content validity may also pertain to the way in which a 24HR is administered; for example, using multiple passes in approaches such as the Automated Multiple-Pass Method[57-59] to capture all foods and beverages consumed within a short period. Examinations of face and content validity may also inform the design of online interfaces, with attention to the potential of probes to exacerbate or reduce social desirability biases, for instance.

Articles reporting on face and content validity should provide information on procedures used to ensure that the items reflect the domain of interest, including with whom methods were tested, how participants were recruited, and how feedback was solicited.

Does the Measure Perform in a Manner Consistent with the Theory Underlying its Construction?

A measure that performs in a manner consistent with the theory underlying its construction, or measures what it is intended to measure, is said to have construct validity.[35] A measure with high construct validity is consistent with related constructs and distinct from unrelated constructs, often measured by correlation. For example, intake of sugary beverages as captured by a screener might be expected to be positively correlated with the frequency of access that children have to those beverages. In addition to correlation, construct validity can be assessed by comparing groups that are expected to perform differently using two-group *t* tests or analysis of variance.[35] For example, the Healthy Eating Index-2015 has been found to differentiate the diet quality of subgroups with known differences (those who do and do not

smoke, consistent with other literature suggesting that smoking is related to lower diet quality), supporting the assertion that index scores reflect dietary quality.[89] Reports of studies of construct validity should indicate the methods used to examine whether or not the measure assesses the intended construct, the sample characteristics, and statistical procedures employed.

Is the Measure Accurate within Specified Performance Standards?

Studies that examine the accuracy of a particular measure assess criterion validity or "evidence of a relationship between the attributes in a measurement tool with its performance on some other variable."[35] In the strictest sense, assessing criterion validity entails examining whether the measure produces data that are representative of the true value by comparing to data reflecting true intake. Given challenges inherent in obtaining true definitive measures of dietary intake influenced by little or no error, the comparison measure used in dietary assessment may or may not be unbiased. Studies using unbiased criterion measures of true intake to validate a given method are first described, followed by studies comparing the method to be evaluated to another biased or error-prone measure.

Studies Comparing a Method to an Unbiased Reference Measure. Studies of criterion validity can make use of observation, feeding studies, or recovery biomarkers.[60,62,64,65,70,85] For the reference measure to be a truly useful criterion, it must measure intake of the same dietary component over the same period of time as the method being assessed. Studies using unbiased reference measures may assess the accuracy of a self-report tool for capturing true short-term intake on a specific day or number of days. For instance, studies examining the validity of 24HR and FR data have been conducted, comparing the self-report data to true short-term intake based on observational and feeding studies.[60,64,65] These types of studies are typically not used to evaluate FFQs because observational and feeding studies are not feasible over the time periods to which FFQs typically pertain (eg, usually a month or a year). Recovery biomarkers can also be used to assess the validity of short-term tools for capturing true short-term intake of energy, protein, potassium, and sodium, particularly in the case that the data are collected for the same period.[62,70,85] Predictive biomarkers may also be used for this purpose, for instance, to provide insights into a method's capability to capture intake of sugars.[71]

Alternatively, studies may examine the accuracy of a tool for capturing true usual intake in comparison to recovery biomarkers.[62,70,85] In this case, the time periods to which the tool being validated and the criterion data are collected or pertain may not be identical, necessitating the estimation of usual intake using repeat administrations of the measures among at least a subsample to characterize and adjust for within-person random error. For short-term dietary instruments, repeat measures allow for adjustment for day-to-day variation using statistical modeling,[26] whereas for long-term dietary instruments and biomarker data, the repeat measures enable accounting for random variation across administrations.[85] The time period over which repeat measures should be spread depends on the time period of interest (eg, repeat measures spread across months are likely needed when the time period of interest is a year or more).

Studies using unbiased criterion measures yield insights into the absolute error with which a method captures intake through analyses focused on the correspondence or agreement between truth and self-reported consumption. These insights may be gleaned in different ways. When evaluating short-term tools using observational or feeding study data, one may examine agreement in terms of the proportion of foods and beverages accurately reported or excluded among participants, as well as phantom foods (or intrusions) that are reported, but were not consumed.[64,90,91] Given information on the total diet, it is possible to examine whether some types of foods or beverages (eg, perceived to be more or less healthy, single-unit vs liquids, or main dishes vs additions and condiments) are reported to a greater or lesser extent than others by study participants. In addition, differences between true and estimated intakes of foods, food groups, and nutrients, as well as portion sizes (with differentiation as to types of foods such as shaped vs amorphous or shapeless vs liquids[60,92]) can be examined. Regression modeling can be used to examine the relationship between personal characteristics such as age, sex, income, education, and body mass index, and the accuracy of reporting. The influence of study design characteristics, such as the time frame for which dietary intake was reported (eg, midnight to midnight the prior day or the most recent 24 hours) or the number of foods offered or truly consumed, may also be of interest.

In studies comparing a method to an unbiased criterion reference, a measurement error framework[62] may be used to differentiate random error from systematic bias. Attenuation factors and correlation coefficients are relevant to the estimation of diet and health or disease relationships.[85] The attenuation factor represents bias or shrinkage in the estimated effect of self-reported vs true dietary intake on a health outcome.[85] Correlation coefficients are often generated to describe associations between dietary intakes estimated using two measures and are sometimes referred to as validation coefficients. When two measures share common bias (eg, two self-report measures), simple correlation coefficients reflect correlation with both true intake and with bias,[38] and, although a correlation=1 is considered perfect, it may reflect in part bias and therefore not be indicative of an ideal instrument. Measurement error models that use an unbiased reference instrument separate sources of error to enable quantification of the correlation between reported and true intake.[62] The correlation coefficient is informative because it provides an indication of the loss of statistical power to observe associations between diet and health or disease outcomes when reported intake is used in place of true intake,[85] and indicates the potential for bias in this estimated association. In studies that conduct statistical modelling to remove within-person variation using repeat administrations of a method, de-attenuated coefficients may be presented. These may also be energy-adjusted.[70,85]

Studies Comparing a Method to a Biased or Error-Prone Reference Measure. Given the paucity of unbiased criterion measures and challenges associated with implementing them in real-world studies,[74] validation studies often compare the method being evaluated to another measure assumed to capture intake with less bias. For example,

FFQ or screener data may be compared with data from multiple recalls or records. The comparison measure should be demonstrated to have some degree of validity[30] in prior research, ideally in comparison to an unbiased reference. Concentration biomarkers may also be used to provide insights into how well data from a self-report measure aligns with another marker of intake,[76,77] although not of true intake as is the case with recovery biomarkers. Biased measures used in place of true criterion measures are sometimes referred to as imperfect or error-prone reference measures, and studies using them as relative validations.[38]

In such studies, the method to be evaluated and the error-prone reference measure are administered in a sample and the estimates from the two measures compared. As with studies using unbiased references, the two measures should capture the same construct over the same period.[33] Thus, in cases in which short-term tools such as multiple 24HR or FR are used as the reference for an FFQ designed to assess usual intake over a month or year, careful consideration must be given to the number and timing of recall or record administrations so that usual intake over that same period can be estimated.

Correlation coefficients are often used in studies that compare two error-prone measures. They do not provide information on the absolute error, either random or systematic, that characterizes data collected using a given method, Thus, they have been suggested to be inappropriate as the sole strategy to assess a method against another error-prone method.[30] Not only true intake, but also error, may be correlated across methods. Thus, correlation coefficients describing the association between two error-prone methods may be inflated due to correlated errors. For example, sources of systematic error such as recall or social desirability biases may be correlated between instruments, resulting in overly optimistic estimates of the degree to which intake estimates based on the two methods are associated with one another. Correlated error may be reduced to some extent, such as by assessing methods with different sources of error such as a self-report FFQ against a weighed food record.[30]

The association between intake estimates from two error-prone methods can be examined using the Bland-Altman method, which was established for comparison of agreement between a new method and an existing method.[93,94] It examines the mean difference between methods across the mean of the two methods and constructs limits of agreement, reflecting the presence, direction, and extent of bias.[30] A plot of the mean difference versus the mean of the two methods helps to visualize the bias and can show estimated intervals within which 95% of the differences of the second method compared with the first one are expected to fall.[95] Acceptable limits must be defined *a priori*, and may be based on the desired use of the dietary assessment method (eg, to categorize individuals as low vs high consumers compared with estimating absolute intake). Lombard and colleagues[30] noted challenges in identifying acceptable limits for the Bland-Altman method given that these may differ from nutrient to nutrient. Further, these authors found among the validation studies they reviewed that "not one … that included Bland Altman analyses considered the clinical importance of the width of limits of agreement in their discussion and conclusions regarding the validity of the method being tested." They indicated that such considerations be identified *a priori*, considering the research question and population.[30]

In addition to the Bland-Altman technique and correlation coefficients, Lombard and colleagues[30] identified four other statistical techniques used in the dietary assessment validation literature. Two assess agreement at the group level; these include paired *t* tests or Wilcoxon signed-rank tests and examination of the mean percent difference between the method being evaluated and the reference measure. It is beneficial to use these tests combined with a method such as a Bland-Altman plot to assess bias. For example, one method could consistently overestimate intake among persons with low intakes and underestimate intake among persons with high intakes; in this case, a paired *t* test might not suggest significant differences in estimates between the two tools, but this does not indicate that the method is accurate. When the goal of the method is to order participants, cross-classification can be used to examine the concordance of classification by investigating the proportion of observations falling in the same or opposite terciles according to two methods. This reflects agreement between two methods at the individual level.[30] In addition, weighted κ statistics are commonly used for data ranked into categories and groups, with a weighting used to account for the degree of disagreement between methods.[30]

Used in combination, these tests can provide insight into the properties of the method being evaluated against another error-prone measure. Lombard and colleagues[30] conducted multiple statistical techniques using a test data set and found that interpretation was challenging. The results of one test may suggest a high degree of agreement whereas another may not. This is possible because agreement at the group level may be good whereas agreement at the individual level may be poor.[30] Further, findings may differ across dietary components. Interpretation must, therefore, consider how well the method measures the dietary construct that was defined *a priori*, and may be supported by the prior determination of content validity of the measure being tested as well as criterion validity of the reference measure.

Articles reporting on studies of comparisons with criterion and error-prone measures should detail and justify the reference measure, address the time periods over which intake was assessed by both the reference measure and the method being validated, describe associated considerations regarding the number and timing of method administrations, specify the sample characteristics and context of administration, and describe and justify the statistical tests used. When the method to be evaluated is compared with another error-prone measure, this should be justified with a description of prior work supporting the use of the latter measure as a reference.

Does the Measure Produce Data that Are Precise and Dependable? Test–retest reliability (which may be referred to as repeatability or reproducibility) is related to the concept of precision (also referred to as the technical error of measurement[29]) and refers to the extent to which repeated measurements yield the same value. A common method for assessing the reliability of a method such as an FFQ or screener is to administer it two (or more) times to the same participants, with administrations separated by a short interval. Each administration should be independent of the

other and the results of the second administration should not be influenced by the first.[33] Two weeks is considered sufficient to ensure participants do not simply recall their original responses, but not so extensive that usual dietary intake might be expected to change.[35] The intraclass correlation, calculated as the ratio of between-person variation to total variation,[39] indicates relative reliability, or the degree to which measurements vary among individuals. The values of the intraclass correlation theoretically range from zero to one, with those closer to one indicating stronger reliability.[39] Interpretation of the intraclass correlation is context-specific because the magnitude depends on between-person variability.[39] A Bland-Altman plot can display the association between the estimates from the two administrations.

Data from short-term tools such as recalls or records have large day-to-day variation (referred to as undependability[29]); therefore, by definition, their test–retest reliability is low. This does not mean that the methods are poor, but rather that averaging over repeat administrations is needed to estimate usual intake. Statistical modeling drawing upon a small number of replicate measures (at least two for dietary constituents consumed daily and potentially more for episodically consumed foods) is used to diminish the effects of within-person variation on the estimate of usual intake.[26] Estimates of reliability are useful in determining how many repeat administrations are needed when the mean will be used to estimate usual intake, as well as the extent to which the estimation of the tails of distributions of usual intake will be influenced by random measurement error.[96]

Another aspect of reliability relates to consistency within a measure. Within measures, Cronbach's α is commonly used to assess how well individual items are related to one another (called internal consistency reliability),[35] representing inter-item correlation. In a screener with multiple items intended to capture related constructs, such as intake of different varieties or forms of fruit and vegetables, high internal consistency is expected.[31] When a measure captures multiple dimensions of diet, such as a diet quality index,[89] Cronbach's α may be lower than that observed for unidimensional measures.

Interrater reliability may be used to assess consistency across observers in a study using observation to collect true intake data, or between coders of 24HR or FR. For categorical variables, the weighted κ coefficient is often used to assess interrater reliability.

For studies of reliability, authors should note the aspects of reliability of interest and why, the sample and context of assessment, and the protocols and statistical tests used. For test–retest reliability, the time period over which the measure was administered should be explained, as should any considerations regarding the likelihood that intake changed during that period due to variation in diet or confounding factors.

Is the Measure Responsive to Change? Whether a measure is sensitive to change has been called the responsiveness of a measure, first raised by Guyatt and colleagues[97] in regard to sensitivity of measures to clinically meaningful change over time. Responsiveness has not been frequently addressed in regard to nutrition- and diet-related measures.[98] It is possible that the lack of detectable change observed in dietary change interventions may relate to nonresponsiveness of the assessment method rather than ineffectiveness of the intervention. Outside of dietary assessment and drawing from psychosocial constructs, some measures have been shown to be responsive to change,[99-101] whereas others have not.[102] When true change is not known, one method of assessing responsiveness to change is to estimate the smallest detectable difference,[97,103] which is the average change in a measure over a specific time divided by the standard error of the measurement error times the square root of 2. This provides the smallest difference that the measure can detect. Methods with large smallest detectable differences would be relatively nonresponsive or insensitive to detecting meaningful change. For example, some measures otherwise shown to have good validity have been shown to be responsive to change, whereas others are not.[104] Measured reliability may not be related to responsiveness.[105]

Some scales have been shown to be responsive to change in longitudinal studies, but not in randomized clinical trials.[102] Responsiveness also appears to relate to the number of items included in a measure; in one study, responsiveness was acceptable when the number of items in a quality of life scale was reduced from five to two, but dropped precipitously when reduced to one item,[106] with potential implications for brief measures intended to capture dietary constructs such as fruit and vegetable intake. For measures of physical activity, objective measures appear to be more responsive than self-report methods, but only slightly so.[107] With respect to diet, one study indicated that some folate biomarkers were responsive to change below a certain level, but not above it.[108]

Further research is needed within the field of dietary assessment to assess methods for responsiveness and to identify the smallest detectable difference to inform the design of longitudinal and experimental studies. Articles reporting responsiveness of a measure should specify the measure, the nutrient or food group, the smallest detectable difference, relevant characteristics of the sample, and any contextual factors that may influence the measure (eg, whether a trained dietitian or researcher guided the participants in providing the responses).

Does the Measure Produce Data that Are Equivalent or Comparable across Populations? Considerations regarding equivalence of dietary assessment methods and the data they produce relate to the ability to compare findings and synthesize evidence across studies in the nutrition surveillance, epidemiology, and intervention literature. In a given study, there may a desire to collect comparable data in multiple settings, such as across countries or other regions that might be different regarding the composition of the population in terms of culture or cuisine. In such a study, the focus may be on creating a method suitable for use across settings. For instance, for the purposes of the European Prospective Investigation into Cancer and Nutrition, a standardized protocol was developed to conduct 24HR.[109] The protocol allowed for the use of country-specific databases reflecting differences in dietary patterns.[110] Similarly, the Automated Self-Administered 24-Hour Dietary Assessment Tool was initially developed in English for use in the United States and has been translated to Spanish for use with US populations and adapted for Canada (English and French versions) and Australia. The adaptations include the

integration of foods and beverages commonly consumed in each country, appropriate portion size measures, and a nutrient database aligned with each country's food supply. Given the similarity in data collection processes, the different interfaces for the Automated Self-Administered 24-Hour Dietary Assessment Tool are expected to collect data that are largely comparable while allowing for country-specific eating patterns, but this has not been formally tested. Research is also needed to examine whether different populations react to a given interface differently based on variations in understanding of the wording, social desirability bias, or other factors.

Tailoring an FFQ to different populations can make it difficult to compare findings. This is because the foods and beverages queried are likely appropriately different, and there may be differences in the questionnaires' design, such as whether and how portion size information is collected, how underlying databases are constructed, and/or how the measure is administered.[63] Studies with multiple FFQs may require substantial harmonization efforts to arrive at comparable estimates for purposes of pooling data.

Articles reporting on assessments of comparability of measures should describe the approach used in detail, including information about the foods and portion sizes queried, available response categories, the nutrient and food group databases and how they were determined including algorithms for missing items or inconsistent responses, and characteristics of the sample and context of assessment.

Interpreting and Reporting the Findings of Validation Studies. Interpretations of validity and reliability necessitate the application of judgment, based on *a priori* decisions regarding the performance of the method being validated. The following guidelines are intended to help with transparency and clarity in the interpretation of dietary validation studies:

1. Inferences should be limited to the specific objectives of the validation study, and the specific populations and contexts in which it was conducted, in addition to differentiating findings across dietary components. Blanket statements regarding validity and reliability should be avoided. For example, it is not appropriate to conclude that a measure is valid and reliable based on a study focused only on reliability. Likewise, it is not appropriate to conclude that a measure has high validity for measuring dietary intake when findings differ according to different dietary components.
2. Validity is "viewed as a carefully structured argument assembling evidence from a variety of sources to support or refute proposed interpretations" of data from a method.[31] Thus, the totality of evidence, including all tests and comparisons conducted, should be weighed in relation to the purpose of the measure and of the validation study, as well as the study's strengths and limitations. This includes considerations regarding the reference measure used to assess validity. For example, in the case that a biased comparison measure is used to assess validity, the only conclusion that can be reached is the extent to which the data collected correspond to that collected using the reference, not the extent to which the evaluated tool captures accurate data. When an unbiased criterion measure is used, stronger conclusions about the potential value of the tool are warranted, subject to caveats related to the study's other strengths and weaknesses. Kelly and colleagues[34] refer to purpose and context validity, referring to whether all assessments conducted indicate that the measure is "suitable for the proposed use and likely to allow the research question to be answered" and the extent to which the measure will provide useful information given the proposed context.
3. Statistical tests should be interpreted with attention to not only statistical significance but also the meaningfulness of the results. Although cut-points, for example, for correlation coefficients and κ values have been proposed, their application to determine that a tool is valid or reliable should be used cautiously with consideration of the totality of the evidence and the intended uses of the measure. For example, for studies that assess criterion validity, attenuation and correlation values below 0.4 (reflective of estimation of a true relative risk of 2 as 1.32, in the context of a diet–disease study) have been viewed as undesirable, but this is not a sharp cut-point.[85] In addition, with large samples, correlations may be statistically significant, but have no practical significance.
4. Inferences should be nuanced, recognizing that constructs such as validity and reliability operate on continuums from low to high, and degrees in between may be appropriate depending on the research objective. For example, a dietary assessment measure may capture intake accurately enough to allow differentiation of high from low consumers, but not to compare intake with sufficient accuracy for comparison to nutrient requirements or food group recommendations.
5. Inferences should reflect the characteristics of the measure validated. Although a particular iteration of a 24HR or FR may have been shown to have high validity for capturing intake of a given dietary construct in a given population over a specific period of time, this may not be true for other variations of the method; for example, using different modes of administration.
6. Findings should be placed in the context of those from similar studies to assess the potential value and uses of the method compared with other available methods. The links between different systems of validity provided here and in Frongillo and colleagues[29] are intended to ensure that terms are clearly explained and related to other terms that might be used in articles reporting on other validation studies. This should support appropriate interpretation and synthesis of the literature.
7. Based on the study design and findings, authors should describe potentially appropriate and inappropriate uses of the measure and features that may influence the design of studies proposing to use it. For example, methods that capture diet with substantial random error will generally lead to attenuation of associations within epidemiologic research. Thus, such tools may not be useful for this purpose unless this error can be mitigated (eg, through the collection

of repeat measures and statistical modelling to adjust for the random error[26]). In addition, findings from validation studies can be useful for estimating the loss of power due to measurement error (as error increases, the required sample size to achieve the same level of statistical power also increases).[111] Authors should also be clear about what their findings do not imply. For example, as noted, a comparison of two error-prone measures cannot be used to conclude that a method is valid but rather how it compares with the measure that served as the reference.

8. Researchers making use of the literature to select and justify a given measure should critically evaluate the available validation studies and clearly and transparently convey the findings in subsequent publications. For example, stating that a method has been shown to be valid or previously validated is insufficient for readers to assess whether the method is suitable for the given purpose and context.

Developers of new measures of dietary intake invest substantial time, energy, and ability in the belief their measure is remedying an important need, and they may wish for their measure to be used by others. There is, thus, the potential for a high level of self-interest by maximizing the reporting of accuracy and reliability while minimizing limitations of a new method. For example, studies of comparative validation have reported that correlations of 0.2 or 0.3 were statistically significant and "validated" the new measure. Although this tendency can be easily understood at the human level, it does not advance nutrition science. The envelope of data points around a regression line for two measures at these levels of correlation is enormous.[29] Thus, using the new measure may give very misleading findings. As a result, all reports of a validation study should report the phenomenon measured, the criterion used, the relevant sample characteristics, contextual factors of assessment, all tests conducted, and the findings. For example, one might report: This FFQ for measuring total energy intake, self-completed in a school classroom, was significantly correlated at 0.52 against doubly-labeled water among Hispanic children, aged 12-15 years. This detailed statement gives indications of both the level of confidence one might have in the new measure, and the groups to which the findings generalize. The findings so reported enable a user to consider whether limitations may be in the new measure, or in the participants' ability to respond (or both). Completeness in reporting the results of validation studies will enhance the science, especially among future users of the method who will now have more context by which to consider which available measure best meets their needs.

CONCLUSIONS

The validation of self-report dietary assessment measures is challenging but the application of the best practices described here can advance our knowledge of the suitability of different methods for specified uses. As noted by Freedman and colleagues,[85] "dietary self-reporting is currently indispensable for population surveillance of dietary intake, many studies of interventions to modify dietary intake, and most studies of diet-health outcome relationships… Knowledge of the measurement properties of self-report instruments is required to interpret the results of studies that rely on such instruments." Better characterization and understanding of the properties of tools also has the potential to support improved measure selection and implementation, thereby strengthening the overall nutrition literature and efforts to support healthy eating patterns and overall health.

References

1. Burke BS. The dietary history as a tool in research. *J Am Diet Assoc*. 1947;23:1041-1046.
2. Afshin A, Sur PJ, Fay KA, et al. Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2019;393(10184):1958-1972.
3. World Cancer Research Fund. Cancer prevention recommendations. Accessed, https://www.wcrf.org/dietandcancer/cancer-prevention-recommendations. Accessed May 25, 2019.
4. Kumanyika S. INFORMAS (International network for food and obesity/non-communicable diseases research, monitoring and action support): Summary and future directions. *Obes Rev*. 2013;14(suppl 1):157-164.
5. Vandevijvere S, Swinburn B. Towards global benchmarking of food environments and policies to reduce obesity and diet-related non-communicable diseases: Design and methods for nation-wide surveys. *BMJ Open*. 2014;4:e005339.
6. Hawkes C, Jewell J, Allen K. A food policy package for healthy diets and the prevention of obesity and diet-related non-communicable diseases: The NOURISHING framework. *Obes Rev*. 2013;14:159-168.
7. Archer E. Opinion: A wolf in sheep's clothing. *The Scientist*. https://www.the-scientist.com/opinion/opinion-a-wolf-in-sheeps-clothing-38534. Published October 22, 2013. Accessed August 8, 2019.
8. Archer E, Marlow ML, Lavie CJ. Controversy and debate: Memory-based methods paper 1: The fatal flaws of food frequency questionnaires and other memory-based dietary assessment methods. *J Clin Epidemiol*. 2018;104:113-124.
9. Stanhope KL, Goran MI, Bosy-Westphal A, King JC, Schmidt LA, Schwarz JM, Stice E, Sylvetsky AC, Turnbaugh PJ, Bray GA, Gardner CD. Pathways and mechanisms linking dietary components to cardiometabolic disease: Thinking beyond calories. *Obes Rev*. 2018;19(9):1205-1235.
10. Reedy J, Subar AF, George SM, Krebs-Smith SM. Extending methods in dietary patterns research. *Nutrients*. 2018;10(5):571.
11. Jerome NW, Kandel RF, Pelto GH. *Nutritional Anthropology: Contemporary Approaches to Diet and Culture*. New York, NY: Redgrave Publishing; 1980.
12. Beardsworth A, Keil T. *Sociology on the Menu*. London, UK: Routledge; 1996.
13. Krebs-Smith SM, Guenther PM, Subar AF, Kirkpatrick SI, Dodd KW. Americans do not meet federal dietary recommendations. *J Nutr*. 2010;140:1832-1838.
14. Kirkpatrick SI, Dodd KW, Reedy J, Krebs-Smith SM. Income and race/ethnicity are associated with adherence to food-based dietary guidance among US adults and children. *J Acad Nutr Diet*. 2012;112(5):624-635.
15. Melnyk LJ, Wang Z, Li Z, Xue J. Prioritization of pesticides based on daily dietary exposure potential as determined from the SHEDS model. *Food Chem Toxicol*. 2016;96:167-173.
16. Kant AK, Graubard BI. Energy density of diets reported by American adults: Association with food group intake, nutrient intake and body weight. *Int J Obes (Lond)*. 2005;29(8):950-956.
17. Krebs-Smith SM, Subar AF, Reedy J. Examining dietary patterns in relation to chronic disease: Matching measures and methods to questions of interest. *Circulation*. 2015;132(9):790-793.
18. Moeller SM, Reedy J, Millen AE, et al. Dietary patterns: Challenges and opportunities in dietary patterns research. *J Acad Nutr Diet*. 2007;107(7):1233-1239.
19. Ocké MC. Evaluation of methodologies for assessing the overall diet: Dietary quality scores and dietary pattern analysis. *Proc Nutr Soc*. 2013;72(2):191-199.

20. Kant AK, Graubard BI. Secular trends in patterns of self-reported food consumption of adult Americans: NHANES 1971-1975 to NHANES 1999–2002. *Am J Clin Nutr*. 2006;84:1215-1223.
21. Piernas C, Popkin BM. Snacking increased among U.S. adults between 1977 and 2006. *J Nutr*. 2010;140:325-332.
22. Basiotis PP, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. *J Nutr*. 1987;117(9):1638-1641.
23. Tarasuk V, Beaton GH. Statistical estimation of dietary parameters: Implications of patterns in within-subject variation—a case study of sampling strategies. *Am J Clin Nutr*. 1992;55(1):22-27.
24. Eicher-Miller HA, Khanna N, Boushey CJ, Gelfand SB, Delp EJ. Temporal dietary patterns derived among the adult participants of the National Health and Nutrition Examination Survey 1999-2004 are associated with diet quality. *J Acad Nutr Diet*. 2016;116(2):283-291.
25. Tarasuk V, Beaton GH. The nature and individuality of within-subject variation in energy intake. *Am J Clin Nutr*. 1991;54(3):464-470.
26. Dodd KW, Guenther PM, Freedman LS, et al. Statistical methods for estimating usual intake of nutrients and foods: A review of the theory. *J Am Diet Assoc*. 2006;106(10):1640-1650.
27. Thompson FE, Subar AF. Dietary Assessment Methodology. In: Coulston AM, Boushey CK, Ferruzzi MG, Delahanty LM, eds. *Nutrition in the Prevention and Treatment of Disease*. 4th edition. Academic Press; 2017:5-48.
28. Thompson FE, Kirkpatrick SI, Krebs-Smith SM. The National Cancer Institute's dietary assessment primer: A resource for diet research. *J Acad Nutr Diet*. 2015;115(12):1986-1995.
29. Frongillo EA, Baranowski T, Subar AF, Tooze JA, Kirkpatrick SI. Establishing validity and cross-context equivalence of measures and indicators. *J Acad Nutr Diet*. 2018 November 20. Epub ahead of print.
30. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. *Nutr J*. 2015;14:40.
31. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med*. 2006;119(2):166.e7-166.e16.
32. Habicht J-P, Yarbrough C, Martorell R. Anthropometric field methods: Criteria for selection. In: Jelliffe DB, Jelliffe EFP, eds. *Nutrition and Growth (A Comprehensive Treatise)*. Vol 2. Boston, MA: Springer; 1979:365-387.
33. Gleason PM, Harris J, Sheean PM, Boushey CJ, Bruemmer B. Publishing nutrition research: Validity, reliability, and diagnostic test assessment in nutrition-related research. *J Am Diet Assoc*. 2010;110(3):409-419.
34. Kelly P, Fitzsimons C, Baker G. Should we reframe how we think about physical activity and sedentary behaviour measurement? Validity and reliability reconsidered. *Int J Behav Nutr Phys Act*. 2016;13:32.
35. DeVon HA, Block ME, Moyle-Wright P, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh*. 2007;39(2):155-164.
36. Barry AE, Chaney EH, Stellefson ML, Chaney JD. So you want to develop a survey: Practical recommendations for scale development. *Am J Health Stud*. 2011;26(2):97-105.
37. Barry AE, Chaney B, Piazza-Gardner AK, Chavarria EA. Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Educ Behav*. 2014;41(1):12-18.
38. National Cancer Institute. Dietary assessment primer. https://dietassessmentprimer.cancer.gov. Accessed May 25, 2019.
39. Weir JP. Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.
40. *Nutrition and Your Health: 2015-2020 Dietary Guidelines for Americans*. 8th edition. Washington, DC: US Government Printing Office; 2015.
41. Birch RJ, Bigler J, Rogers JW, Zhuang Y, Clickner RP. Trends in blood mercury concentrations and fish consumption among U.S. women of reproductive age, NHANES, 1999–2010. *Environ Res*. 2014;133:431-438.
42. Tooze JA, Midthune D, Dodd KW, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc*. 2006;106(10):1575-1587.
43. Bernstein S, Zambell K, Amar MJ, Arango C, Kelley RC, Miszewski SG, Tryon S, Courville AB. Dietary intake patterns are consistent across seasons in a cohort of healthy adults in a metropolitan population. *J Acad Nutr Diet*. 2016;116(1):38-45.
44. Barker DJ. The developmental origins of adult disease. *J Am Coll Nutr*. 2004;23(6 suppl):588S-595S.
45. Barker DJ. Fetal origins of coronary heart disease. *BMJ*. 1995;311(6998):171-174.
46. Freedman LS, Midthune D, Dodd KW, Carroll RJ, Kipnis V. A statistical model for measurement error that incorporates variation over time in the target measure, with application to nutritional epidemiology. *Stat Med*. 2015;34(27):3590-3605.
47. Liao X, Zucker DM, Li Y, Spiegelman D. Survival analysis with error-prone time-varying covariates: A risk set calibration approach. *Biometrics*. 2011;67(1):50-58.
48. Carroll RJ, Midthune D, Subar AF, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol*. 2012;175(4):340-347.
49. Freedman LS, Midthune D, Arab L, et al. Combining a food frequency questionnaire with 24-hour recalls to increase the precision of estimation of usual dietary intakes—Evidence from the validation studies pooling project. *Am J Epidemiol*. 2018;187(10):2227-2232.
50. National Cancer Institute. Measurement error webinar series. https://epi.grants.cancer.gov/events/measurement-error/. Accessed May 25, 2019.
51. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey. http://www.cdc.gov/nchs/nhanes/. Accessed May 25, 2019.
52. Subar AF, Kirkpatrick SI, Mittl B, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): A resource for researchers, clinicians, and educators from the National Cancer Institute. *J Acad Nutr Diet*. 2012;112(8):1134-1137.
53. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: A systematic review of the evidence. *J Acad Nutr Diet*. 2015;115(1):64-77.
54. Khanna N, Boushey CJ, Kerr D, Okos M, Ebert DS, Delp EJ. An overview of the Technology Assisted Dietary Assessment project at Purdue University. *ISM*:290-295.
55. Carter MC, Albar SA, Morris MA, et al. Development of a UK online 24-h dietary assessment tool: myfood24. *Nutrients*. 2015;7(6):4016-4032.
56. Amoutzopoulos B, Steer T, Roberts C, et al. Traditional methods v. new technologies— dilemmas for dietary assessment in large-scale nutrition surveys and studies: A report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015. *J Nutr Sci*. 2018;7:e11.
57. Moshfegh AJ, Rhodes DG, Baer DJ, et al. The US Department of Agriculture automated multiple-pass method reduces bias in the collection of energy intakes. *Am J Clin Nutr*. 2008;88(2):324-332.
58. Blanton CA, Moshfegh AJ, Baer DJ, Kretsch MJ. The USDA Automated Multiple-Pass Method accurately estimates group total energy and nutrient intake. *J Nutr*. 2006;136(10):2594-2599.
59. Rhodes DG, Murayi T, Clemens JC, Baer DJ, Sebastian RS, Moshfegh AJ. The USDA Automated Multiple-Pass Method accurately assesses population sodium intakes. *Am J Clin Nutr*. 2013;97(5):958-964.
60. Kirkpatrick SI, Potischman N, Dodd KW, et al. The use of digital images in 24-hour recalls may lead to less misestimation of portion size compared with traditional interviewer-administered recalls. *J Nutr*. 2016;146(12):2567-2573.
61. Huybrechts I, Casagrande C, Nicolas G, et al. Inventory of experiences from national/regional dietary monitoring surveys using EPIC-Soft. *Eur J Clin Nutr*. 2011;65(suppl 1):S16-S28.

62. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *Am J Epidemiol*. 2003;158(1):14-21.
63. Cade J, Thompson R, Burley V. Warm D. Development, validation and utilisation of food-frequency questionnaires—a review. *Public Health Nutr*. 2002;5(4):567-587.
64. Kirkpatrick SI, Subar AF, Douglass D, et al. Performance of the Automated Self-Administered 24-hour Recall relative to a measure of true intakes and to an interviewer-administered 24-h recall. *Am J Clin Nutr*. 2014;100:233-240.
65. Baxter SD, Smith AF, Guinn CH, et al. Interview format influences the accuracy of children's dietary recalls validated with observations. *Nutr Res*. 2003;23(11):1537-1546.
66. Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr*. 2002;5(6a):969-976.
67. Subar AF, Freedman LS, Tooze JA, et al. Addressing current criticism regarding the value of self-report dietary data. *J Nutr*. 2015;145(12):2639-2645.
68. Schoeller DA, Hnilicka JM. Reliability of the doubly labeled water method for the measurement of total daily energy expenditure in free-living subjects. *J Nutr*. 1996;126(1 suppl):348S-354S.
69. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. *J Nutr*. 2003;133(3 suppl):921S-924S.
70. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am J Epidemiol*. 2015;181(7):473-487.
71. Tasevska N. Urinary sugars—a biomarker of total sugars intake. *Nutrients*. 2015;7(7):5816-5833.
72. Tasevska NA, Midthune D, Potischman N, et al. Use of the predictive sugars biomarker to evaluate self-reported total sugars intake in the Observing Protein and Energy Nutrition (OPEN) study. *Cancer Epidemiol Biomarkers Prev*. 2011;20(3):490-500.
73. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prev*. 2005;14(5):1287-1294.
74. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham S. Biomarkers in nutritional epidemiology: Applications, needs and new horizons. *Hum Genet*. 2009;125(5-6):507-525.
75. Lampe JW, Huang Y, Neuhouser ML, et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. *Am J Clin Nutr*. 2016;105(2):466-475.
76. Bingham SA, Gill C, Welch A, et al. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol*. 1997;26(suppl 1):S137.
77. Burrows TL, Warren JM, Colyvas K, Garg ML, Collins CE. Validation of overweight children's fruit and vegetable intake using plasma carotenoids. *Obesity*. 2009;17(1):162-168.
78. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CA. *Measurement Error in Nonlinear Models: A Modern Perspective*. New York, NY: Chapman and Hall/CRC; 2006.
79. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ*. 2010;340:c2289.
80. Kirkpatrick SI, Midthune D, Dodd KW, Potischman N, Subar AF, Thompson FE. Reactivity and its association with body mass index across days on food checklists. *J Acad Nutr Diet*. 2012;112(1):110-118.
81. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: Current state of the art and future prospects. *F1000Res*. 2017;6:926.
82. Miller TM, Abdel-Maksoud M, Crane LA, Marcus AC, Byers TE. Effects of social approval bias on self-reported fruit and vegetable consumption: A randomized controlled trial. *Nutr J*. 2008;7:18.
83. Klesges LM, Baranowski T, Beech B, et al. Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8- to 10-year-old African-American girls: Results from the Girls Health Enrichment Multisite Studies (GEMS). *Prev Med*. 2004;38(suppl):78-87.
84. Hebert JR, Clemow L, Pbert L, Ockene IS, Ockene JK. Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *Int J Epidemiol*. 1995;24(2):389-398.
85. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol*. 2014;180(2):172-188.
86. Lachat C, Hawwash D, Ocké MC, et al. Strengthening the reporting of observational studies in epidemiology—Nutritional epidemiology (STROBE-nut): An extension of the STROBE statement. *PLOS Med*. 2016;13(6). 2016;e1002036.
87. Natarajan L, Pu M, Fan J, Levine RA, et al. Measurement error of dietary self-report in intervention trials. *Am J Epidemiol*. 2010;172(7):819-827.
88. Hsu CC, Sandford BA. The Delphi technique: Making sense of consensus. *PARE*. 2007;12(10):1-8.
89. Reedy J, Lerman JL, Krebs-Smith SM, et al. Evaluation of the Healthy Eating Index-2015. *J Acad Nutr Diet*. 2018;118(9):1622-1633.
90. Baxter SD, Royer JA, Guinn CH, Hardin JW, Smith AF. Origins of intrusions in children's dietary recalls: Data from a validation study concerning retention interval and information from school food-service production records. *Public Health Nutr*. 2009;12(9):1569-1575.
91. Guinn CH, Baxter SD, Hardin JW, Royer JA, Smith AF. Intrusions in children's dietary recalls: The roles of BMI, sex, race, interview protocol, and social desirability. *Obesity (Silver Spring)*. 2008;16(9):2169-2174.
92. Subar AF, Crafts J, Zimmerman TP, Wilson M, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *J Am Diet Assoc*. 2010;110(1):55-64.
93. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307-310.
94. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135-160.
95. Giavarina D. Understanding Bland-Altman analysis. *Biochem Med*. 2015;25(2):141-151.
96. Nelson M, Black AE, Morris JA, Cole TJ. Between-and within-subject variation in nutrient intake from infancy to old age: Estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr*. 1989;50(1):155-167.
97. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40(2):171-178.
98. McClelland JW, Keenan DP, Lewis J, et al. Review of evaluation tools used to assess the impact of nutrition education on dietary intake and quality, weight management practices, and physical activity of low-income audiences. *J Nutr Educ*. 2001;33(suppl 1):S35-S48.
99. Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax*. 1987;42(10):773-778.
100. Pouchot J, Guillemin F, Coste J, Brégeon C, Sany J. Validity, reliability, and sensitivity to change of a French version of the arthritis impact measurement scales 2 (AIMS2) in patients with rheumatoid arthritis treated with methotrexate. *J Rheumatol*. 1996;23(1):52-60.
101. Pinto-Carral A, Fernández-Villa T, Guccione AA, Cuadrado FM, Cancela JM, Molina AJ. Validity, reliability, and responsiveness of the Spanish version of the OPTIMAL instrument. *PM R*. 2018;11(3):258-269.
102. Johnston BC, Miller PA, Agarwal A, et al. Limited responsiveness related to the minimal important difference of patient-reported outcomes in rare diseases. *J Clin Epidemiol*. 2016;79:10-21.
103. van Baalen B, Odding E, van Woensel MP, Roebroeck ME. Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. *Clin Rehabil*. 2006;20(8):686-700.
104. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. *J Clin Epidemiol*. 1989;42(5):403-408.
105. Puhan MA, Bryant D, Guyatt GH, Heels-Ansdell D, Schünemann HJ. Internal consistency reliability is a poor predictor of responsiveness. *Health Qual Life Outcomes*. 2005;3:33.

106. Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol*. 2001;54(6):571-579.

107. Lee WY, Clark BK, Winkler E, Eakin EG, Reeves MM. Responsiveness to change of self-report and device-based physical activity measures in the Living Well with Diabetes Trial. *J Phys Act Heal*. 2015;12(7):1082-1087.

108. Duffy ME, Hoey L, Hughes CF, et al. Biomarker responses to folic acid intervention in healthy adults: A meta-analysis of randomized controlled trials. *Am J Clin Nutr*. 2014;99(1):96-106.

109. Slimani N, Casagrande C, Nicolas G, et al. The standardized computerized 24-h dietary recall method EPIC-Soft adapted for pan-European dietary monitoring. *Eur J Clin Nutr*. 2011;65(suppl 1):S5-S15.

110. Slimani N, Fahey M, Welch A, et al. Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public Health Nutr*. 2002;5(6B):1311-1328.

111. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst*. 2011;103(14):1086-1092.

For more information on the subject discussed in this article, see Sites in Review on page 1967.

AUTHOR INFORMATION

S. I. Kirkpatrick is an associate professor, School of Public Health and Health Systems, University of Waterloo, Waterloo, Ontario, Canada. T. Baranowski is a professor, Department of Pediatrics, Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX. A. F. Subar is retired; at the time of the study, she was acting chief of the Risk Factor Assessment Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD. J. A. Tooze is a professor, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC. E. A. Frongillo is a professor, Arnold School of Public Health, University of South Carolina, Columbia.

Address correspondence to: Sharon I. Kirkpatrick, PhD, RD, School of Public Health and Health Systems, 200 University Ave W, LHN 1713, University of Waterloo, Waterloo, ON, Canada N2L 3G1. E-mail: sharon.kirkpatrick@uwaterloo.ca

STATEMENT OF POTENTIAL CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

FUNDING/SUPPORT

There is no funding to disclose.

AUTHOR CONTRIBUTIONS

All authors conceptualized the manuscript. S. I. Kirkpatrick led the drafting of the manuscript and all authors contributed critical content and edits.