Traditional Self-Reported Dietary Instruments Are Prone to Inaccuracies and New Approaches Are Needed

Michele N. Ravelli[1] and Dale A. Schoeller[2]*

[1] Department of Neurology, University of Wisconsin, Madison, WI, United States, [2] Nutritional Sciences and Biotechnology Center, University of Wisconsin, Madison, WI, United States

Background: Diet is a modifiable behavior that influences an individual's health. Because of this, diet assessment is an important component of public health surveillance, evaluating response to community health interventions, and monitoring individual compliance to medical interventions. Diet assessments are usually performed using one of three basic methods: diet recall, diet diaries, or food frequency questionnaires. Although these three assessment instruments have displayed a strong agreement between themselves, when reported intake is compared with intake measured using quantitative nutrient biomarkers, investigators have identified systematic misreporting errors for all three of these self-reported dietary instruments.

Aims: This work aims to summarize the state of knowledge regarding misreporting and why it impedes diet–health research and to introduce advances in the collection and the treatment of dietary data.

Methods: This work reviews and summarizes published data on misreporting and the recent efforts to reduce such errors.

Results: The evidence demonstrates a strong and consistent systematic underreporting of energy intake (EIn) across adults and children studies. Underreporting of EIn has been found to increase with body mass index (BMI), and the differences between macronutrient reports indicate that not all foods are underreported equally. Protein is least underreported, but which specific foods are commonly underreported are not known.

Conclusions: Because energy underreporting varies as a function of BMI, self-reported EIn should not be used for the study of energy balance in the study of obesity. The between-individual variability in the underreporting of self-reported intake of energy and other nutrients attenuates diet–disease relationships. Recent efforts to correct for underreporting have reduced misreporting of diet outcomes, but improvements have been incremental in nature and more research is needed to validate and extend these efforts.

Keywords: diet records, self-report, energy intake, underreporting, doubly labeled water method

INTRODUCTION

Investigations into the role of diet in the development of disease in humans are viewed as difficult but important because diet is one of the behaviors that individuals may employ to maintain or improve health (1). For example, human feeding studies performed or dictated by external disruptions of food supplies have clearly demonstrated that diet plays a central role in the negative impacts on health resulting from energy, vitamin, and mineral deficiencies in controlled studies conducted for at least 75 years (2). At the same time, the Minnesota Semi-starvation Study demonstrated the difficulty and the ethical issues of performing month-long feeding trials in order to investigate the quantitative nature of diet–disease relationships using only an inpatient paradigm (3).

A major source of difficulty has been documenting what individuals consume as their typical diet due to dietary measurement error (4). Because controlled and thus more accurate feeding studies are costly and difficult to perform, a far more common approach to the study of the relationship between diet and disease has been to perform studies of free-living participants. These studies usually rely on assessing diet using self-report instruments to access diet and this introduces diet measurement error into the study (5). The most common dietary assessment instruments are diet recall surveys, in which the participants report from memory each item of food consumed for the previous day, weeks, or months; food frequency questionnaires (FFQ), which have the aim of assessing the food consumption during some specified period of time or over a period such as adolescence; and diet diary methods where the subjects record dietary intake for each eating event for a period of days or weeks.

Perhaps the best documented evidence of dietary instrument measurement error is that of self-reported energy intake (EIn) being often less than that of the individual's measured energy expenditure. The aim of this short review is to summarize the development of that evidence and how misreporting impedes diet—health research. Because of the evidence, there has been a renewal of efforts to improve or develop alternatives to these instruments. These recent advances are introduced here as part of a series of reviews in this issue.

DIETARY MISREPORTING

Over the past 50 years, many traditional diet assessment instruments have undergone modification of content or structure, sometimes to the degree of being considered a new instrument, and then were validated by comparison against an older version or previously evaluated instrument. In most instances, these comparisons demonstrated a moderate or strong agreement between the basic types of dietary assessment instruments and the tool was considered to be of reasonable accuracy and precision (6). A few investigators, however, performed studies comparing self-reported dietary intake against a biomarker of dietary intake, such as urinary nitrogen, which provides an objective measure of dietary protein intake (7). Such comparisons against a biomarker often did not find

these self-report instruments to be accurate. For example, Warnold et al. (8) reported that self-reported protein intake underestimated protein consumption by 47% compared to protein intake measured using urinary nitrogen outputs among women undergoing a weight loss treatment. Studies comparing self-report against biomarkers to test the accuracy of traditional diet instruments, however, were infrequent and had only a modest influence on the growth of the use of self-report dietary assessments in the study of diet–disease relationships.

The number of dietary instrument validations against a biomarker increased dramatically following the development of the doubly labeled water (DLW) method to measure total energy expenditure (TEE) in humans (9). This method, developed by Lifson, is based on the difference in the elimination kinetics of two stable isotopes in water, namely, deuterium (^{2}H) and ^{18}O (10). The difference in the elimination rate of ^{2}H and ^{18}O is proportional to carbon dioxide production (10). The latter is the end-product of oxidative phosphorylation, and TEE can be calculated using standard indirect calorimetric equations (11). The human validations that included conditions of weight stability, overfeeding, underfeeding, intravenous feeding, and heavy exercise have been summarized by Speakman et al. (12–15). These validations have displayed an average accuracy of TEE of 1 to 2% and an individual precision of 7%, which support its use as a biomarker for use in a criterion method against which one may test the accuracy and the precision of self-reported energy intake.

The development of the DLW method for the measurement of TEE created an opportunity to validate diet assessment instruments against an objective energy expenditure based on the first law of thermodynamics. The first law of thermodynamics states that energy cannot be created nor destroyed, and thus EIn equals energy expenditure plus or minus the change in body energy stores during the measurement interval. Moreover, when body energy stores are unchanged over time, the energy storage term falls to zero and then EIn equals energy expenditure. Among pregnant women, infants, and children, change in energy stores over time is expected. At 1 month of age, the average daily energy storage is about 40% of EIn, but this decreases to 3% by 1 year of age and continues to decrease to 1% at 10 years of age (16). Among pregnant women, the average increase in body energy stores is about 190 kcal/day or 8% of EIn during the 3rd trimester (17). Among most non-pregnant adults, weight gain is not recommended, but it is common. The rate of weight gain is 0.2–0.7 kg/year or about 1–2 g/day (18). Assuming that the gain is adipose tissue with 20% fat-free mass and 80% fat mass, this stores only about 8 to 16 kcal/day or about 0.3–0.6% of EIn (19) and thus energy expenditure almost equals EIn, and TEE is an excellent biomarker of EIn. There are exceptions to this essential near-equality assumption in adulthood. These include periods of voluntary weight loss, loss of appetite during illness, or periods of holiday feasts when energy intake can be quite different from expenditure and TEE will fail to be a quantitative biomarker of actual EIn. Under habitual conditions outside of these short periods, EIn roughly equals TEE. Thus, TEE is termed as a biomarker of habitual dietary energy intake rather than one of actual EIn. This is an important distinction because while TEE is generally a good measure of habitual energy intake for the study

of diet and health, it fails as a measure of actual energy intake under the above mentioned short-term conditions.

Another consideration in using TEE as a biomarker of habitual EIn is that energy intake can be expressed in one of three ways. The first is gross energy. This is the total energy available when foods are combusted to dioxide carbon, water, and nitrogen gas using bomb calorimetry (20). Not all the gross energy, however, is available to the body for metabolism. About 8% of the gross energy is not absorbed and thus lost from the body as waste products in feces (21, 22). The second expression is the absorbed energy. Finally, not all the absorbed energy is available to the body for energy production because some compounds that still contain chemical energy are lost as waste products in urine. The third expression is the portion of absorbed energy retained by the body that is available for energy production, which is the metabolizable energy. Metabolizable energy, as the name implies, is available for use in oxidative phosphorylation. This is the energy value listed in the food handbooks and tables. Metabolizable energy is thus the energy value used for calculating the dietary EIn by dietary assessment instruments.

Many of the human studies using DLW performed during the 10 years between 1982 and 1992 included an assessment of dietary EIn using traditional instruments. One of the first studies was conducted by Prentice et al. (23), in which it was observed that EIn, assessed using a 7-days food diary, was 34% ($P < 0.05$) less than TEE measured by DLW in young adult obese women (32.9 ± 4.6 kg/m^2), but there was no difference detected between EIn and TEE (2%, NS) in lean women. The authors also found that half of the EIn vs. TEE difference was due to underrating as assessed by weight loss during the dietary diary period. These findings of low self-reported EIn were confirmed in a later review (24), which included papers which found that underreporting of dietary EIn was observed in women with anorexia nervosa, who perceive that they have excess body fat, and also in individuals with measured excess body weight, who are concerned about actual excess fat. Thus, underreporting was associated with individuals likely to be concerned about excess weight and not just with actual weight status (body mass index, BMI) itself (24). Thus, as early as 1990, it was found that underreporting of dietary EIn was common among adults and linked to concerns regarding excess body weight or fat (24). Even these early studies found that the degree of underreporting was of similar magnitude regardless of whether intake was assessed using retrospective instruments such as diet recalls or histories or with instruments such as food diaries (24). Based on these observations, it was concluded that dietary assessment instruments were subject to errors that increased with the individual's concern regarding their relative weight, which would result in a correlation of increased underreporting with increased BMI (24). Because of this, it was strongly recommended that self-reported EIn should not be used as a primary assessment instrument to measure EIn in investigations into the role of EIn in weight regulation as early as 1990 (24).

Most of these early studies employing DLW as a quantitative biomarker of dietary EIn were conducted in cohorts with sample sizes categorized as small to medium and, in many cases, by investigators without extensive experience on the use of dietary assessment instruments. Based on anecdotal evidence provided by questions from the audience following oral presentations, some investigators in the audience suggested that the finding of underreporting may have been an artifact and that it might not occur if experienced investigators performed studies in large cohorts. This hypothesis, however, was not supported by the results from one study and soon thereafter by four more studies that were performed by investigators who had extensive experience on the use of dietary assessment instruments and which included cohorts with several hundreds of adult subjects each as summarized by Freedman et al. (25). The combined results of these studies confirmed that underreporting of habitual EIn in the United States was common as it was observed in each of the five studies which, when combined, involved over 2,000 participants (25). The 24-h recall (24HR) exhibited an EIn underreporting compared to the DLW-measured TEE which averaged −16% (range, −10 to −28%), and the FFQ was subject to an even larger reporting error than was 24HR (range, −26 to −32%) (25). The combined number of participants in these five studies ($n = 2,265$) permitted sub-analyses, and it was found that those having a BMI of >30 kg/m^2 underreported EIn by 7% more than those of a BMI in the healthy range, but there was no difference between men and women or adult age groups when centered on ages 50–59 years (25). One of these five studies (26) found that the administration of up to eight 24HRs on different days of the week did not eliminate the average reporting error, thus demonstrating that the underreporting was not simply due to day-to-day variation in actual EIn. The underreporting did decrease when two 24HRs were averaged (−11%) relative to that when only one 24HR was employed (−15%), but the percent error changed only a little when more than two 24HRs were averaged. Even when six 24HR data were collected and averaged, the bias dropped to only −9%. Thus, dietary data were more consistent when two recalls were employed in each participant, but little was gained by further replication.

The findings from the combination of the five large studies discussed above have been confirmed and extended through a systematic review conducted by Burrows et al. (27). The review identified an additional 59 studies that included 6,298 adults, including the abovementioned five studies and the 2,265 participants in the abovementioned summary by Freedman (25). The studies employed a mixture of diet instruments, including 24HR; the food diaries include weighed food records and FFQs. The degree of underreporting relative to habitual EIn as measured by DLW varied over a wide range. This included two studies that reported group averages displaying over-reporting (7 and 8%), but the vast majority identified cohort average underreporting by between 1 and 38%, and the plurality of studies found an average under-reporting between 20 and 30%. A comparison of methods indicated that the most misreporting was observed for the FFQ and the least for 24HR, but all three methods displayed underreporting errors. Studies that included advanced technology such as photography, hand-held personal digital assistants, or oral recordings did reduce the underreporting slightly compared to non-technology-assisted instruments but were still found to be subject to underreporting. Included in that review were studies conducted in countries

other than the United States, including Australia, Brazil, Canada, Finland, Germany, Japan, New Zeeland, Norway, Sweden, and the United Kingdom, demonstrating that underreporting with regard to EIn was a global problem.

Two systematic reviews concluded that underreporting of EIn was also an issue among children (28, 29). There have not been as many studies performed in children (aged 3–18 years) as have been performed in adults, but the results were similar with those in adults. Underreporting was common in children, and children with excess weight (overweight or obese) underreported more than those having a BMI in the healthy weight range. It was found that underreporting was reduced when parents assisted their children for ages <11 years. Unlike what was observed in adults, age was a significant modifier of misreporting, and underreporting was greater in adolescents than it was in younger children.

Because of the significant underreporting of EIn observed in the studies discussed above, one of the vital next steps for research directed at studying the phenomenon of EIn underreporting is to identify whether underreporting of EIn is due to a failure to accurately report specific foods or is a general underreporting of all foods. Addressing this issue is difficult because it means one has to measure something that is not reported rather than what is reported. We speculate that one means of accomplishing this would be to include multiple biomarkers in a study of self-reported dietary intake. As evidence, studies that have included DLW as a biomarker for EIn and urinary nitrogen as a biomarker for protein intake have shown that energy is underreported by a larger percentage than protein. For example, the abovementioned study combining the results from five large dietary intake studies (25) found that while energy was misreported by −16% (range, −6 to −28% using 24HR), protein was misreported by only −5% (range, −21 to +20%), indicating that protein was not as underreported as carbohydrate and/or fat. We speculate that a cluster analysis using multiple quantitative and possibly semi-quantitative biomarkers will provide vital insight into the foods that are misreported. The value of identifying what foods were being underreported as well as the difficulty of doing so without using biomarkers is illustrated by a Brazilian study performed in obese women prior to bariatric surgery (30). The study found that the under-reporters reported lower intakes of foods with high energy density but with similar intakes of calories provided by healthy foods (fruits, leafy vegetables, and vegetables) compared to those of plausible reporters. This reporting behavior influenced the determination of dietary patterns by exploratory factor analysis, in which the principal component analysis with VARIMAX rotation was applied for the selection of food groups that composed the matrix and then used for dietary pattern interpretation (30). By combining diet factor analysis with biomarker data on energy, protein, sugar, sodium, and potassium, it should be possible to infer if these differences were due to actual dietary intake differences.

Misreporting of energy and protein intake when assessing diet by self-report is well-documented and recognized by many as a major limitation to the investigation of the effects of diet on health. The problem of underreporting, particularly because of the inter-individual variation in misreporting, dramatically attenuates diet–disease relationships. Kipnis et al. (31) modeled the effects of misreporting of protein and EIn in the OPEN study and concluded that the variation in the degree of misreporting using an FFQ would severely attenuate the relative risk between true protein or EIn and disease from a true value of 2.0 to an apparent relative risk of <1.1. Even worse, it may even reverse the association between diet and disease as had occurred in an analysis of energy balance using self-reported EIn and physical activity by Kromhout et al. (32). The data from these investigators indicated that energy balance and BMI were negative and becoming more negative with increasing BMI, a result that they considered implausible and a possible artifact of underreporting EIn.

In summary, the problem of misreporting of dietary intake is limiting the ability of investigators to study diet–disease relationships (31). Investigators are, therefore, performing studies of novel approaches that may either reduce misreporting or adjust the self-reported data using *post hoc* techniques that may reduce the effect of such misreporting on study outcomes (33). These include the development of advanced technology to reduce the reporting errors themselves, adjustment of reported nutrient intake using calibration against a nutrient biomarker, statistical approaches that provide novel analyses of data from traditional self-reported dietary instruments, or direct use of dietary biomarkers to assess intake (34).

APPROACHES TO REDUCE MISREPORTING

Advanced technological tools include digital photography with on-line submission, movement monitors on the wrist or eating utensils to detect feeding, microphones to detect chewing, and scales to monitor the disappearance of food from a plate. Photographic methods provide the most detailed information about foods consumed, but they are still prone to underreporting (27) and they require a large amount of technical support (35). The other methods listed above have demonstrated the ability to detect eating events, but they provide only partial quantitative and qualitative information regarding the foods being consumed (34).

In addition, *post hoc* approaches that reduced the influence of misreporting have been presented. For example, Mozaffarian et al. (36) analyzed diet data obtained using an FFQ administered at 4-year intervals in a large longitudinal study. They used dietary change scores from the bracketing FFQs in place of raw intake scores from a single FFQ to identify foods that were associated with 4-years changes in body weight. This approach of using diet change and subsequent weight change is difficult to validate for dietary reporting accuracy, but the foods identified in this study as being associated with weight gain or loss were in general agreement with small, shorter intervention studies and thus extended the findings from the short-term interventions to the population level. Additionally, it provided high statistical power, but it did require a study design where the diet was assessed multiple times over a period of years in a large cohort and an outcome that was continuous. A second *post hoc*

data analysis approach was employed by Freedman et al. (37). They combined multiple 24HRs with an FFQ in order to use the quantitative information from 24HR along with the larger list of foods consumed from the FFQ against true intake as measured by dietary biomarkers. They reported an improvement in the correlation coefficients between reported and biomarker-measured true dietary intake of energy, protein, potassium, and sodium compared to the use of single 24HR by an average of 0.14. The highest correlation coefficient, however, was 0.64 for potassium in women, and thus the variance explained was <40%.

As an alternative to the above-discussed methods to reduce problems arising from misreporting, Prentice and Huang (38) have proposed and tested the use of a *post hoc* calibration to adjust reported intakes for misreporting identified by the use of a quantitative biomarker in the entire study cohort or a subsample of that cohort. Tasevska et al. (39) applied this approach to an analysis of self-reported sugar intake and the likelihood of developing type II diabetes or cardiovascular disease. They found that correcting reported sugar intake based on the calibration eliminated what appeared to be an implausible inverse relationship, thus avoided a false finding. The resulting positive correlation, however, was small and did not result in a significant increase in the odds ratio for disease development with increasing sugar intake during the 16-years follow-up in the Women's Health Initiative cohort of older women, thus not ending the controversy around sugar consumption and type II diabetes.

The final approach to be discussed in this review is that of Goldberg et al. (40, 41). This approach involved characterizing a self-reported intake as plausible or implausible. During a period of bodyweight stability, the ratio of $\frac{energy\ intake\ _{reported}}{resting\ metabolic\ rate}$ should correspond to the ratio of $\frac{total\ energy\ expenditure}{resting\ metabolic\ rate}$, which is identified as physical activity level (PAL). Considering the biological variability of the components of the equation, confidence limits (cutoffs) are calculated to classify the probable

accuracy of the reported EIn, and its sensibility improves when individual PAL classification is used in the cutoff points (42). A not dissimilar approach is to calculate the ratio of reported EIn to TEE from DLW (24). As an alternative to the DLW method, it may be possible to use a predicted TEE based on weight, height, age, sex and physical activity (18). The optimal method for defining the cutoff for excluding implausible reported intakes is still under debate, but the value is recognized (43).

SUMMARY AND CONCLUSION

Dietary assessment is central to the study of diet–health relationships. The most common assessment instruments are diet recalls, diet diaries, and food frequency questionnaires, and all are dependent on self-reported data. Self-reported EIn, using all of these instruments, has been shown to yield reproducible intake results. They have also been shown to yield good to strong correlations between foods consumed when compared against one another. Comparisons against quantitative biomarkers of dietary intake, however, have clearly demonstrated that self-report is prone to misreporting errors for EIn and other nutrients and that inter-individual variability in the degree of underreporting attenuates the strength of diet–disease relationships and raises questions regarding what foods are being misreported. Recent research has identified several methods for reducing many of these reporting errors. There remains, however, a need for further research to optimize the accuracy or correct for inaccuracies in self-reported dietary data because of the importance of dietary data in the prevention and the treatment of diet-induced diseases.

AUTHOR CONTRIBUTIONS

MR and DS conceptualized the manuscript. MR and DS drafted and edited the manuscript.

REFERENCES

1. U.S. Department of Health Human Services U.S. Department of Agriculture. (2015). 2015 – 2020 Dietary Guidelines for Americans. 8th. Available online at: https://health.gov/our-work/food-and-nutrition/2015-2020-dietary-guidelines/ (accessed February 20, 2020).

2. Keys A, Brožek J, Henschel A, Mickelsen O, Taylor HL. *The Biology of Human Starvation, Vols. 1 and 2.* Minneapolis; London: University of Minnesota Press; Geoffrey Cumberlege, Oxford University Press (1950). doi: 10.5749/j.ctv9b2tqv

3. Kalm LM, Semba RD. They starved so that others be better fed: remembering Ancel Keys and the Minnesota experiment. *J Nutr.* (2005) 135:1347–52. doi: 10.1093/jn/135.6.1347

4. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst.* (2011) 103:1086–92. doi: 10.1093/jnci/djr189

5. Dao MC, Subar AF, Warthon-Medina M, Cade JE, Burrows T, Golley RK, et al. Dietary assessment toolkits: an overview. *Public Health Nutr.* (2019) 22:404–18. doi: 10.1017/S1368980018002951

6. Kirkpatrick SI, Baranowski T, Subar AF, Tooze JA, Frongillo EA. Best practices for conducting and interpreting studies to validate self-report

dietary assessment methods. *J Acad Nutr Diet.* (2019) 119:1801–16. doi: 10.1016/j.jand.2019.06.010

7. Bingham S, Wiggins HS, Englyst H, Seppanen R, Helms P, Strand R, et al. Methods and validity of dietary assessments in four Scandinavian populations. *Nutr Cancer.* (1982) 4:23–33. doi: 10.1080/01635588209513735

8. Warnold I, Carlgren G, Krotkiewski M. Energy expenditure and body composition during weight reduction in hyperplastic obese women. *Am J Clin Nutr.* (1978) 31:750–63. doi: 10.1093/ajcn/31.5.750

9. Schoeller DA, van Santen E. Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol Respir Environ Exerc Physiol.* (1982) 53:955–9. doi: 10.1152/jappl.1982.53.4.955

10. Lifson N. Theory of use of the turnover rates of body water for measuring energy and material balance. *J Theor Biol.* (1966) 12:46–74. doi: 10.1016/0022-5193(66)90185-8

11. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol.* (1949) 109:1–9. doi: 10.1113/jphysiol.1949.sp004363

12. Speakman JR. The history and theory of the doubly labeled water technique. *Am J Clin Nutr.* (1998) 68:932s−8s. doi: 10.1093/ajcn/68.4.932S

13. Schoeller DA, Kushner RF, Jones PJH. Validation of doubly labeled water for measuring energy expenditure during parenteral

nutrition. *Am J Clin Nutr.* (1986) 44:291–8. doi: 10.1093/ajcn/44.2.291

14. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jequier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *Am J Physiol.* (1986) 250(5 Pt 2):R823–30. doi: 10.1152/ajpregu.1986.250.5.R823

15. Ravussin E, Harper IT, Rising R, Bogardus C. Energy expenditure by doubly labeled water: validation in lean and obese subjects. *Am J Physiol.* (1991) 261(3 Pt 1):E402–9. doi: 10.1152/ajpendo.1991.261.3.E402

16. Institute of Medicine. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids Washington, DC: The National Academies Press (2005).

17. Butte NF, King JC. Energy requirements during pregnancy and lactation. *Public Health Nutr.* (2005) 8:1010–27. doi: 10.1079/PHN2005793

18. Yanovski JA, Yanovski SZ, Sovik KN, Nguyen TT, O'Neil PM, Sebring NG. A prospective study of holiday weight gain. *N Engl J Med.* (2000) 342:861–7. doi: 10.1056/NEJM200003233421206

19. Thomas DM, Schoeller DA, Redman LA, Martin CK, Levine JA, Heymsfield SB. A computational model to determine energy intake during weight loss. *Am J Clin Nutr.* (2010) 92:1326–31. doi: 10.3945/ajcn.2010.29687

20. Miles CW, Kelsay JL, Wong NP. Effect of dietary fiber on the metabolizable energy of human diets. *J Nutr.* (1988) 118:1075–81. doi: 10.1093/jn/118.9.1075

21. Atwater WO, Bryant AP. *The Availability and Fuel Values of Food Materials.* Connecticut (Storrs). Agricultural Experiment Station 12th Annual Report, 1899 (1900).

22. Food and Agriculture Organization. *Food Energy - Methods of Analysis and Conversion Factors.* Rome: Food and Agriculture Organization of the United Nations (2003).

23. Prentice AM, Black AE, Coward WA, Davies HL, Goldberg GR, Murgatroyd PR, et al. High levels of energy expenditure in obese women. *Br Med J.* (1986) 292:983–7. doi: 10.1136/bmj.292.6526.983

24. Schoeller DA, Bandini LG, Dietz WH. Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method. *Can J Physiol Pharmacol.* (1990) 68:941–9. doi: 10.1139/y90-143

25. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, 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) 180:172–88. doi: 10.1093/aje/kwu116

26. Arab L, Wesseling-Perry K, Jardack P, Henry J, Winter A. Eight self-administered 24-hour dietary recalls using the Internet are feasible in African Americans and Whites: the energetics study. *J Am Diet Assoc.* (2010) 110:857–64. doi: 10.1016/j.jada.2010.03.024

27. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of dietary assessment methods when compared to the method of doubly labeled water: a systematic review in adults. *Front Endocrinol (Lausanne).* (2019) 10:850. doi: 10.3389/fendo.2019.00850

28. Walker JL, Ardouin S, Burrows T. The validity of dietary assessment methods to accurately measure energy intake in children and adolescents who are overweight or obese: a systematic review. *Eur J Clin Nutr.* (2018) 72:185–97. doi: 10.1038/s41430-017-0029-2

29. Burrows TL, Martin RJ, Collins CE. A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc.* (2010) 110:1501–10. doi: 10.1016/j.jada.2010.07.008

30. Ravelli MN, Sartori MMP, Corrente JE, Rasera I, de Souza NPP, de Oliveira MRM. The under-reporting of energy intake influences the dietary pattern reported by obese women in the waiting list for bariatric surgery. *Rev Nutr.* (2018) 31:235–49. doi: 10.1590/1678-98652018000200009

31. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol.* (2003) 158:14–21; discussion 2–6. doi: 10.1093/aje/kwg091

32. Kromhout D, Saris WH, Horst CH. Energy intake, energy expenditure, and smoking in relation to body fatness: the Zutphen Study. *Am J Clin Nutr.* (1988) 47:668–74. doi: 10.1093/ajcn/47.4.668

33. Bennett DA, Landry D, Little J, Minelli C. Systematic review of statistical approaches to quantify, or correct for, measurement error in a continuous exposure in nutritional epidemiology. *BMC Med Res Methodol.* (2017) 17:146. doi: 10.1186/s12874-017-0421-6

34. Schoeller DA, Westerterp M (editors). *Advances in the Assessment of Dietary Intake.* 1st ed. Boca Raton, FL: CRC Press (2017). doi: 10.1201/9781315152288

35. Martin CK, Kaya S, Gunturk BK. Quantification of food intake using food image analysis. *Conf Proc IEEE Eng Med Biol Soc.* (2009) 2009:6869–72. doi: 10.1109/IEMBS.2009.5333123

36. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. *N Engl J Med.* (2011) 364:2392–404. doi: 10.1056/NEJMoa1014296

37. Freedman LS, Midthune D, Arab L, Prentice RL, Subar AF, Willett W, 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) 187:2227–32. doi: 10.1093/aje/kwy126

38. Prentice RL, Huang Y. Nutritional epidemiology methods and related statistical challenges and opportunities. *Stat Theory Relat Fields.* (2018) 2:2–10. doi: 10.1080/24754269.2018.1466098

39. Tasevska N, Pettinger M, Kipnis V, Midthune D, Tinker LF, Potischman N, et al. Associations of biomarker-calibrated intake of total sugars with the risk of type 2 diabetes and cardiovascular disease in the women's health initiative observational study. *Am J Epidemiol.* (2018) 187:2126–35. doi: 10.1093/aje/kwy115

40. Goldberg G, Black A, Jebb S, Cole T, Murgatroyd P, Coward W, et al. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr.* (1991) 45:569–81.

41. Black AE, Goldberg GR, Jebb SA, Livingstone MB, Cole TJ, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 2. Evaluating the results of published surveys. *Eur J Clin Nutr.* (1991) 45:583–99.

42. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *International journal of obesity and related metabolic disorders. J Int Assoc Study Obesity.* (2000) 24:1119–30. doi: 10.1038/sj.ijo.0801376

43. Banna JC, McCrory MA, Fialkowski MK, Boushey C. Examining plausibility of self-reported energy intake data: considerations for method selection. *Front Nutr.* (2017) 4:6. doi: 10.3389/fnut.2017.00045

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.