Addressing Current Criticism Regarding the Value of Self-Report Dietary Data\textsuperscript{1,2}

Amy F Subar,\textsuperscript{3*} Laurence S Freedman,\textsuperscript{5} Janet A Tooze,\textsuperscript{6} Sharon I Kirkpatrick,\textsuperscript{7} Carol Boushey,\textsuperscript{8} Marian L Neuhouser,\textsuperscript{9} Frances E Thompson,\textsuperscript{3} Nancy Potischman,\textsuperscript{4} Patricia M Guenther,\textsuperscript{10} Valerie Tarasuk,\textsuperscript{11} Jill Reedy,\textsuperscript{3} and Susan M Krebs-Smith\textsuperscript{3}

Divisions of 3Cancer Control and Population Sciences and 4Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 5Biostatistics Unit, Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel; 6Wake Forest School of Medicine, Winston-Salem, NC; 7School of Public Health and Health Systems, University of Waterloo, Waterloo, Ontario, Canada; 8Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 9Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 10Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City, UT; and 11Department of Nutritional Sciences, University of Toronto, Toronto, Ontario, Canada

Abstract

Recent reports have asserted that, because of energy underreporting, dietary self-report data suffer from measurement error so great that findings that rely on them are of no value. This commentary considers the amassed evidence that shows that self-report dietary intake data can successfully be used to inform dietary guidance and public health policy. Topics discussed include what is known and what can be done about the measurement error inherent in data collected by using self-report dietary assessment instruments and the extent and magnitude of underreporting energy vs. other nutrients and food groups. Also discussed is the overall impact of energy underreporting on dietary surveillance and nutritional epidemiology. In conclusion, 7 specific recommendations for collecting, analyzing, and interpreting self-report dietary data are provided: 1) continue to collect self-report dietary intake data because they contain valuable, rich, and critical information about foods and beverages consumed by populations that can be used to inform nutrition policy and assess diet-disease associations; 2) do not use self-reported energy intake as a measure of true energy intake; 3) do use self-reported energy intake for energy adjustment of other self-reported dietary constituents to improve risk estimation in studies of diet-health associations; 4) acknowledge the limitations of self-report dietary data and analyze and interpret them appropriately; 5) design studies and conduct analyses that allow adjustment for measurement error; 6) design new epidemiologic studies to collect dietary data from both short-term (recalls or food records) and long-term (food-frequency questionnaires) instruments on the entire study population to allow for maximizing the strengths of each instrument; and 7) continue to develop, evaluate, and further expand methods of dietary assessment, including dietary biomarkers and methods using new technologies. J Nutr doi: 10.3945/jn.115.219634.

Keywords: dietary assessment, dietary surveillance, measurement error, underreporting, nutritional epidemiology, energy intake

Introduction

Recent reports have asserted that, because of energy underreporting, dietary self-report data suffer from measurement error so great that findings from all dietary surveillance and observational studies are useless for informing public health policy or investigating diet-health relations (1–5). The collection of self-report dietary intake data has been called "pseudoscience" (1, 2) and interpretations of these data have been implicated in the development of “misguided national and local health-care advice to individuals” (4). Archer was quoted in the popular media as saying, “To say they [the data] are imperfect is the equivalent of saying the Titanic had a floatation problem or a buoyancy problem. These data should not be used (6),” and that nutritional epidemiology is a fraud “far greater than any fraud perpetrated in the private sector (e.g., the Enron and Madoff scandals)” (7). Such statements do not consider the amassed evidence that shows that self-report dietary intake data can be successfully used to inform dietary guidance and public health policy. Therefore, the purposes of this commentary are as follows:

1. Describe the issue of measurement error in general and as it applies to dietary intake data.
2. Confirm that self-report data provide an inaccurate measure of energy intake (EI) and that such values should not be used to evaluate energy balance.

3. Describe the magnitude of underreporting of energy vs. other nutrients in self-reported intakes and comment on the miscalculation of implausible energy reporters in Archer et al. (1).

4. Discuss the implications of measurement error and energy underreporting for dietary surveillance and nutritional epidemiology to 1) establish what can and cannot be expected from self-report data and 2) show the importance, value, and utility of self-report dietary intake data.

5. Provide 7 specific recommendations for appropriately collecting, analyzing, and interpreting self-report dietary data.

Measurement Error in Self-Report Data

“There will always be error in dietary assessments. The challenge is to understand, estimate, and make use of the error structure during analysis” (8). This statement, written in 1997 by the late George Beaton, a scientist who spent his career conducting nutrition and dietary assessment research, captures the essence of how to approach self-report dietary assessment data.

Measurement error is the difference between the observed or measured value and the true value. Such error is an inherent part of the measurement process. It is well accepted that, at a minimum, random error is inherent in all measures, be they self-report, laboratory-based, or clinical. What is worrisome, however, is not random error, which can usually be mitigated through large-enough sample sizes and repeat measures, but systematic error, which can lead to erroneous findings. What has been asserted by critics of self-report dietary data is that the presence of systematic error renders these data useless (5). The science does not support this assertion, and such statements do little to promote constructive dialog and advancement in our field.

Collecting self-report data always has been part of medical and population research. Such data have laid the foundation for numerous important discoveries. For example, self-report of the intensity and duration of cigarette smoking allowed researchers to detect important associations between smoking and chronic disease risk despite biomarker evidence that cigarette smoking is underreported (9). It was self-report data on intakes of foods and supplements before and during pregnancy that revealed the potential association, later shown to be causal, between low folic acid intake and neural tube defects in offspring (10). Self-report data on exercise and physical activity have established a clear association between increased activity and multiple health benefits (11).

In addition, lest we think that other types of measures are without error, consider clinical measures. Blood pressure values include multiple sources of error, including bias that occurs with improper procedures, faulty equipment, and the “white coat effect” (elevated blood pressure due to anxiety over medical procedures) (12). We would not, however, consider disregarding these values in assessing the effect of blood pressure on cardiovascular disease outcomes.

Recent criticisms have suggested that the nutrition community has mostly ignored the issue of measurement error in self-reported diet and that attempts to adjust for it are “statistical machinations” (1, 5). To the contrary, since the 1970s, nutrition scientists, in collaboration with epidemiologists and statisticians, have been proactive in acknowledging error and developing methods to mitigate it in self-report intake data (13–37). Such work has resulted in recommendations for reducing error through appropriate study design, improvement in dietary assessment instruments, selection of assessment instruments, and statistical methods (28, 35, 38, 39). This body of literature acknowledges the limitations of dietary data, while also recognizing their value, and proposes and tests new methods to improve both data collection and analysis in the contexts of dietary surveillance and nutritional epidemiology.

EI Is Inaccurately Measured by Self-Report Data

A major criticism of self-report data is the extent of error in energy measurement. This error is real and significant, based on consistent findings from comparisons of self-reported EIs to total energy expenditure. These comparisons use doubly labeled water (DLW), which is considered to be an unbiased biomarker for EI (40), or formulas that estimate energy requirements (14, 32, 37, 41–47). Although energy expenditure estimates based on DLW are also affected by both random error (48, 49) and error based on assumptions that individuals are in energy balance, they nonetheless provide useful insights into the degree of misreporting in self-report data. It is critically important to the field that investigators acknowledge that EIs based on self-report dietary assessment instruments are generally not well measured and that the severity of systematic bias associated with underreporting varies by type of instrument and population characteristics (32, 33, 50–53).

Given the current focus on the error in measuring EIs (1, 41), it is instructive to consider its source. EIs are derived from reporting of foods and beverages by using 1 of 3 instruments: FFQs, 24-h dietary recalls (24HRs), and food records (FRs). FFQs are not meant to measure EIs (54), a point that is sometimes overlooked but seems obvious given the following: 1) the finite list of foods and beverages, 2) limited specificity collected on food preparation and types, and 3) the application of nutrient-composition databases that represent composites of similar food/beverage items. 24HRs suffer primarily from limitations associated with memory and difficulty in estimating quantities. FRs have the potential to foster accurate reporting in real time (especially with the weighing of foods and beverages), but they are subject to reactivity that may lead to changes in intake and underreporting on reporting days—which is an important reason they are used to encourage changes in eating in behavioral interventions. Moreover, because some participants complete FRs at the end of the day, rather than in real time, FRs can take on limitations similar to those for 24HRs. In addition, each of these methods is associated with nutrient-composition databases that contain error and are unlikely to precisely represent the energy content of the foods and beverages actually consumed.

Given that FFQs have a finite list of foods/portions with little detail, 24HRs rely on memory, and FRs are reactive, it is not surprising that study participants often omit some foods and beverages consumed and underreporting is more likely to occur than overreporting. Add to this a potential desire to present oneself positively (social desirability bias) and a social environment in which obesity is prevalent and stigmatized, and some

---

12 Abbreviations used: BMR, basal metabolic rate; DLW, doubly labeled water; EI, energy intake; FR, food record; 24HR, 24-h recall.
degree of underreporting bias is likely to occur, particularly among those who are overweight or obese. Finally, and perhaps most important, energy, unlike any other nutrient, is contained in nearly everything we eat and drink. Small or large errors for each individual food and beverage reported will add up, leading to errors in EI that are likely larger in magnitude than for other nutrient and food components. Although participants provide important and useful information about their dietary intakes, it comes with some degree of error and limitations that must be borne in mind when the data are analyzed and interpreted.

Magnitude of the Measurement Error in Dietary Self-Report Data

Let’s consider the accuracy of the view that the magnitude of energy underreporting is so large as to render all self-report data worthless by examining the extent of the measurement error problem for energy as well as for other nutrients. Our highest-quality evidence comes from studies that have used DLW to assess true EIs. In a pooling study that combined data from 5 of the largest US recovery biomarker studies in healthy adults, for FFQs—an instrument not intended to capture energy—EI was underreported by 24–33% for both men and women relative to DLW (32, 33). However, the average EI underreported on 24HRs compared with DLW was lower: 12–13% for middle-aged men and 6–16% for young and middle-aged women and 25% for elderly women (flawed memory may be a more significant limitation of 24HRs in this age group). These data also showed that the percentage of underreporting on 24HRs was much lower for absolute intakes of some nutrients than for recovery biomarkers: 5% for protein and 3% for potassium. Furthermore, findings from a recent controlled feeding study indicated that reported vs. known intakes were not different for a variety of foods and nutrients (55). Such findings indicate that self-report dietary data contain very low levels of underreporting for many foods and dietary constituents.

In the absence of DLW, estimates of energy requirements are based on basal energy expenditure estimated from formulas that take into account sex, age, weight, and height, usually assuming constant sedentary activity. The Goldberg method (56, 57), which calculates the ratio of reported EI to estimated basal metabolic rate (BMR; EI:BMR) and assumes a constant physical activity level, is commonly used to categorize individuals into categories of plausible or implausible reporters. The original method suggested 2 cutoffs: an absolute value of 1.35 for EI:BMR (cutoff 1), predicated on the assumptions that BMR had been measured rather than predicted and that the recorded EI represented habitual intake, and another based on number of days of self-report, CVs for EI, estimated BMR, physical activity level, and sample size (cutoff 2). Black (58) later cautioned that “cut-off 1 should be abandoned” and recommended the use of cutoff 2 at the individual level for evaluating plausible reporting because it accounts for intraindividual variation in EI and energy expenditure. Despite this, Archer et al. (1), using single 24HRs across multiple NHANESs, used cutoff 1 at the group level to conclude that 67.3% of women and 58.7% of men reported EIs that were physiologically implausible. For NHANES III alone, these estimates were 63% of women and 51% of men (1). In contrast, in an earlier analysis of NHANES III, Briefel et al. (14) concluded that 27.7% of women and 18.1% of men were implausible reporters using cutoff 2 at the individual level, values that are consistent with the prevalence of implausible reporters on 24HRs estimated from large DLW studies (21–24%) (44, 57) and other studies of plausible reporting on 24HRs (14, 46, 59–63). Although Archer and Blair (64) defended their use of cutoff 1, it was clearly inappropriate.

Unfortunately, the study by Archer et al. (1) that showed overestimates of implausible energy reporting generated a significant amount of press coverage (6, 65–77), only a few reports of which included any counterarguments (71, 74, 76). In contrast, careful responses and criticisms in the scientific community (78–80) have received little attention. As exemplified by the continuing crisis with regard to parents refusing to vaccinate their children, one inaccurate study linking vaccines to autism, later retracted (81), can take years to negate. The bottom line is that the magnitude of underreporting is not as abysmal as Archer et al. (1) estimate, and the preponderance of scientific evidence with regard to energy underreporting does not justify the conclusions that all self-reported foods and dietary constituents are misreported to the same extent and that self-report data are worthless.

Implications of Measurement Error for Dietary Research

Implications for dietary surveillance. What are the implications of measurement error for dietary surveillance? First, with respect to energy, population distributions based on 24HRs are biased and shifted to the left, with means, quantiles, and percentage of persons above and below a cutoff affected (82). Trends, too, are affected, especially because methods of collecting recall data have improved and, therefore, levels of misreporting may have changed over time. Furthermore, because measurement error varies by population subgroup (32, 51–53), any attempt at correction for bias needs to consider factors known to differentially affect reporting, such as BMI and education. In summary, estimated population distributions of EI from dietary surveillance data, therefore, are biased and difficult to interpret. Currently, the optimal method for estimating EI distributions at the population level is to administer DLW in at least a subset representative of the population to permit measurement error adjustment. At present, however, this has been considered too costly and impractical for large-scale population surveillance. Furthermore, the DLW method only measures energy expenditure over a time span of ~2 wk; multiple assumptions must be made as to weight stability and whether this represents long-term usual EI.

If the scientific questions under investigation relate to obesity or energy balance, then a measure of weight status, ideally controlling for body composition, is a far better strategy than attempting to capture accurate estimates of EI and output. Without such data, however, self-report dietary intake data could be valuable in answering a number of key obesity-related questions. These include characterizing the types of foods consumed by individuals with obesity vs. those of normal weight as well as contextual factors such as when and with whom they eat or whether they prepare their meals or often eat away from home. If the scientific questions under investigation concern how EI is related to obesity or body composition, then controlled clinical feeding studies may be the best option. Such questions require tight control of energy expenditure, EI, food composition, and perhaps other metabolic measurements.

In surveillance, some primary interests are the assessment of both nutrient adequacy (83) and diet quality, which can be assessed by examining the amounts of various dietary components per 1000 kcal, as is done with the Healthy Eating Index (84). Estimates of energy based on self-report data are used in creating these energy-adjusted density variables, which provide...
an indication of the mix of foods or dietary patterns, and this in
turn can be evaluated in relation to dietary guidelines. The use of
self-reported energy in this manner leads to overestimates of the
densities of foods more accurately reported and underestimate
of the densities of foods more highly underreported. Research on
this topic indicates that low energy reporters tend to report
lower intakes of foods high in fats and sweets (62, 85–89);therefore, the population’s intake is likely worse than food-
density variables and the Healthy Eating Index would suggest.

Dietary surveillance research provides a wealth of useful data
to guide nutrition policy as indicated by the following examples: 1) findings that added-sugar consumption alone far exceeds
recommendations supported the 2015 Dietary Guidelines Advisory Committee’s conclusion that intakes should be limited to
10% of total EI (90), 2) data indicating poor overall dietary quality led to changes in the Dietary Guidelines for Americans
(91) and the development of diet-related national health objectives
(92), and 3) results showing how reported intakes of fruits and
vegetables among students led to modified standards for school
lunch programs (93). To the extent that social desirability bias
exists in dietary reporting, one would expect less underreporting
of fruits, vegetables, and whole grains and more underreporting of
added sugars and saturated fats, suggesting dietary imbalances
may be even worse than reported. Nonetheless, the data as
presented are sufficiently grim to warrant corrective action and are
sufficiently valuable to be useful in informing nutrition policy.

**Implications for nutritional epidemiology.** What are the
implications of measurement error for nutritional epidemiol-
yogy? For those dietary constituents for which there are recovery
biomarkers, such as energy, protein, potassium, and sodium, and
predictive biomarkers, such as urinary sucrose (94, 95), collections of such biomarkers in at least a subset of the study
population for the purpose of conducting measurement error
adjustment would advance the field, as would the discovery of new recovery and predictive biomarkers (28, 50). Because of the
expense and logistical considerations of administering and
analyzing DLW and collecting 24-h urine samples, however, such adjustments are impractical for most epidemiologic studies.
In summary, EI estimates derived solely from self-report dietary
assessment instruments should not be considered in any epide-
imologic study as the exposure variable.

If an investigator requires accurate EI estimates, the optimal
method is to use DLW in at least a subset of the study population
to allow for measurement error adjustment of self-reported
EIs (37). If energy expenditure estimates based on DLW are
available, however, they should not be used to energy-adjust
other self-reported dietary constituents. Although it might seem
counterintuitive, flawed self-reported EI estimates (and not
dLW energy expenditure estimates) are quite helpful in adjusting
for measurement error of other self-reported dietary constitu-
ents (27) because the error in energy reporting is correlated with
error in the reported intakes of all foods and beverages. In
particular, FFQ-based nutrient densities (nutrient intakes per
total EI) correlate more closely with true intakes than do abso-
lute intakes. Furthermore, it is recommended that in multivari-
ate models of disease risk, self-reported energy be controlled for
by including it in the model even when nutrient density variables
are included because doing so further reduces bias; however, the
coefficients for energy should never be used to make inferences
about EI and disease outcomes (54).

Epidemiologic studies should be designed with translation in
mind. That is, when studying the associations between diet and
health outcomes, results should be interpretable to allow for

**The Future of Dietary Data Collection and Analysis**

Improving self-report dietary assessment tools is an endeavor
worth pursuing for both surveillance and epidemiology. Technol-
ogy may hold promise but must be carefully evaluated to establish
respondent preferences and improvements over existing methods
(97). In addition, the design of future studies should always
consider the use of the least biased and optimal combination of
dietary assessment tools as a means to improve the quality of
dietary data. For evaluating EIs and energy expenditure, admin-
istering DLW to all participants would be useful, but self-report
data, despite limitations, are critical for providing valuable infor-
mation about dietary patterns and diet quality to evaluate ques-
tions such as whether intakes are consistent with recommendations
(91) or associated with health outcomes. In the meantime, more
accurate and affordable methods for both EI and energy expendi-
ture would be welcome. This includes new biomarkers for EI or
technologies that make the administration of DLW more practical
and affordable. Finally, thoughtfully interpreting the data we do
have, given our knowledge of measurement error, is critical.

**Dietary surveillance.** To improve estimates of mean intake and
population distributions for energy, collecting DLW values in at
least a subsample of the population would be useful to adjust for
measurement error in reported EIs. For the study of energy
balance, accurate estimates of both EI and energy expenditure
would be optimal. Such data collection, however, is unlikely to be
practical or possible in the near future. Underreporting, partic-
ularly among those who are obese or overweight, poses a chal-
lenge (44, 51, 52). Yet, as discussed above, weight fluctuation
is the best measure of energy balance, and BMI coupled with waist
circumference is the best measure of energy overconsumption.
**Conclusions**

Measurement error is inherent in all types of data. The errors in self-report dietary intake data are well documented. On the basis of current knowledge, we recommend that investigators:

1. continue to collect self-report dietary intake data because they contain valuable, rich, and critical information about foods and beverages consumed by populations that can be used to inform nutrition policy and assess diet-disease associations;
2. not use self-reported EI as a measure of EI;
3. use self-reported EI for energy adjustment of other self-reported dietary constituents to improve risk estimation in studies of diet-health associations;
4. acknowledge the limitations of self-report dietary data and analyze and interpret them appropriately;
5. design studies and conduct analyses that allow adjustment for measurement error;
6. design new epidemiologic studies to collect dietary data from both short-term (24HRs or FRs) and long-term (FFQs) instruments on the entire study population to allow for maximizing the strengths of each instrument; and
7. continue to develop, evaluate, and further expand methods of dietary assessment, including dietary biomarkers and methods using new technologies.

Self-report dietary data provide information on food intake, food behaviors, and eating patterns that is not possible to obtain from a comprehensive set of biomarkers. To guide people in how to eat more healthfully, asking them what they are currently eating is imperative and should not be abandoned. Assessing total EI via self-report will probably always be difficult, but energy is only one of hundreds of dietary constituents of interest. Its precise measurement is not required for self-report data to be useful for informing nutrition policy and for elucidating the associations between diet and disease.

**Acknowledgments**

We thank Magdalena Wilson for her careful editing and assistance with references. AFS designed the research, wrote the manuscript, and had primary responsibility for final content; LSE, JAT, and SMK-S designed the research and wrote the manuscript; and SIK, CB, MLN, FET, NP, PMG, VT, and JR provided substantial intellectual input and contributed to writing the manuscript. All authors read and approved the final manuscript.

**References**


