minireview

Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake

JILLIAN TRABULSI AND DALE A. SCHOELLER

Department of Nutritional Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53706-1571

> Trabulsi, Jillian, and Dale A. Schoeller. Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. Am J Physiol Endocrinol Metab 281: E891-E899, 2001.-Epidemiological studies of diet and disease rely on the accurate determination of dietary intake and subsequent estimates of nutrient exposure. Although methodically developed and tested, the instruments most often used to collect self-reported intake data are subject to error. It had been assumed that this error was only random in nature; however, an increasing body of literature suggests that systematic error in the reporting of true dietary intake exists as well. Here, we review studies in which dietary intake by self report was determined while energy expenditure was simultaneously measured using the doubly labeled water (DLW) method. In seeking to establish the relative accuracy of each instrument to capture true habitual energy intake, we conclude that none of the self-reported intake instruments demonstrates greater accuracy against DLW. Instead, it is evident that the physical and psychological characteristics of study participants play a significant role in the underreporting bias observed in these studies. Further research is needed to identify underreporters and to determine how to account for this bias in studies of diet and health.

dietary intake; underreporting; deuterium; oxygen-18

IN SEEKING TO UNDERSTAND the relationship between diet and health, investigators have often relied on selfreported dietary intake data to measure nutrient intake. Not surprisingly, investigators have created a wide range of intake assessment instruments that are the products of careful development and testing. It is readily acknowledged that all methods used to assess self-reported dietary intake are subject to measurement error, but it is often assumed that this error is random. Random error, such as writing mistakes or processing variations, attenuates true associations between diet and disease, resulting in the generation of false-negative conclusions. Self-reported dietary intake, however, may also be subject to systematic error. A systematic error, such as the underreporting of true intake by certain population groups, could bias nutri-

Address for reprint requests and other correspondence: D. A. Schoeller, Univ. of Wisconsin-Madison, 1415 Linden Drive, Madison, WI 53706-1571 (E-mail: dschoell@nutrisci.wisc.edu).

ent intake estimates and result in misleading conclusions with regard to diet and disease (34).

The accuracy of dietary intake assessment instruments, in the true analytic sense, has rarely been established. This is because there are very few criterion methods against which self-reported intake can be validated. Validation has involved comparison of the new assessment instrument against another, more established self-report method. As such, it is actually the relative validity of the new instrument that is being tested. This type of validation will fail to detect true reporting bias, however, if both the new and established instruments have correlated error. Alternatively, the new instrument can be validated against an external criterion such as a biomarker. The doubly labeled water (DLW) technique is an example of a biomarker that can be used to validate self-reported energy intake. The error of the DLW technique is independent of self-reported intake error, thus allowing true reporting bias to be detected.

E892 MINIREVIEW

Established for use in animals (17) and later validated for use in humans (25), the DLW technique is used to measure energy expenditure in free-living subjects. This method involves the administration of water containing enriched quantities of the stable isotopes deuterium (²H) and oxygen-18 (¹⁸O). The term "doubly" comes from the fact that both the hydrogen and the oxygen of water are labeled. The oxygen-18 is eliminated from the body in the form of carbon dioxide $(C^{18}O_2)$ and water $(H_2^{18}O)$, and the deuterium is eliminated in water (²H₂O). The difference in elimination rate between these two isotopes is a measure of carbon dioxide production. Carbon dioxide production can then be used to calculate energy expenditure by use of standard equations for indirect calorimetry. The DLW technique has been shown to be accurate to 1%, with a coefficient of variation of 2–12% (25), and is currently considered the "gold standard" for the measurement of total energy expenditure in humans. From a subject's viewpoint, it is quite simple to use. A baseline specimen of blood or urine is collected, and a loading dose of labeled water is taken by mouth. Additional specimens are collected on the dose day and up to 2 wk later to determine the initial dilution space and the isotopic elimination rates. The technique is quite robust with regard to accuracy but is subject to error if some of the dose water is lost, if the subject switches to a new source of water that has a different background isotope abundance, or if the specimens are contaminated with ambient water or mislabeled with regard to time and date (25). DLW also requires calculating energy expenditure from CO₂ production, and thus an error in the assumed macronutrient composition of the diet can introduce a bias. For example, if it is assumed that 40% of dietary energy is derived from fat, but it really is only 20%, then the energy expenditure will be underestimated by 6%. The precision of the technique, however, is potentially quite variable between analytic centers, because a small random error in measuring the isotope enrichments of the physiological specimens can introduce random errors in the calculated energy expenditure of 3-15% (25). Because of this, it is important to determine the repeatability of the analyses in any given laboratory.

The First Law of Thermodynamics states that energy is conserved, meaning that it cannot be created or destroyed. The energy put into a system is therefore equal to the energy used and/or stored by the system. In terms of human metabolism, metabolizable energy intake is equal to energy expenditure plus any change in body stores. Thus, in the weight-stable individual who is not in an active stage of growth, such as childhood or pregnancy, energy expenditure will be equal to habitual energy intake.

In large population studies, the most common instruments employed to assess dietary intake at the individual level include food diaries or records, repeated 24-h recalls, diet history, and food frequency questionnaires. These can crudely be separated into two categories, i.e., prospective and retrospective methods. In

addition, the techniques can be classified on the basis of the degree of quantitative detail that is provided.

The food record or diary is an example of a prospective, quantitative method. The respondent is required to record all foods and beverages and the amount of each consumed on a daily basis. The recording of dietary intake is usually conducted over a period of 3-14 days. One of the strengths of the food record method is its ability to capture quantitative information, because all foods and beverages are weighed or measured before consumption. Because of this, the weighed food record has been used as the standard to which other dietary intake methodologies have been compared. Examples of retrospective dietary assessment instruments include the 24-h recall, diet history, and foodfrequency questionnaire (FFQ). These methodologies rely on the memory of the participant to recall diets eaten in the past. The 24-h recall method requires that the respondent report to an interviewer all foods and beverages and amounts consumed in the previous 24-h period. Typically, data from several 24-h recalls conducted on both weekdays and weekends are averaged together. The strength of the 24-h recall is its ability to collect detailed, qualitative information about foods consumed. The diet history method inquires about the frequency with which various foods are consumed and also gathers information about the typical content of meals. One of the strengths of the diet history is its detailed assessment of usual meal pattern in addition to the frequency of consumption data collected. Finally, the FFQ requires that respondents report the frequency of consumption of foods from a list of foods over a specific time period. A strength of the FFQ is its ability to capture the intake of infrequently consumed nutrients that are ingested with a high degree of intraindividual variability. A complete description of these dietary assessment methodologies may be found in the review by Thompson and Byers (29).

Prospective methodologies, such as the food diary, are believed to be subject to reporting bias, because the act of recording intake is thought to influence the respondent's usual food choices and alter their intake during the recording period. Indeed, some studies have reported a loss of body weight in subjects during the record-keeping period (12, 20, 21), thus highlighting a potential weakness of this methodology for defining usual or habitual intake. Retrospective methodologies are generally thought to be free of this type of bias but are subject to random error due to poor recollection of past diet as well as systematic error due to the underreporting of true intake.

The aim of this review is to assess whether self-reported dietary intake methods are subject to systematic error by collating the studies in which these methods were validated against the DLW method. As such, this review is limited to assessment of dietary energy intake. The literature review was limited to studies of 10 or more subjects that were nonathletic adults living in industrialized countries, and in which the DLW technique was used to validate reported energy intake. Potential studies were identified using a MEDLINE

search with the following key words: doubly labeled water, energy metabolism, oxygen isotopes, deuterium oxide, energy balance, energy expenditure, energy intake, diet records, food-frequency questionnaires, and diet recalls. A survey of the literature found a number of different methods used to quantify the agreement between reported energy intake and measured energy expenditure. For comparison, reporting bias (energy intake — energy expenditure/energy expenditure × 100%) was calculated for each study and is summarized in Tables 1 and 2.

THE FOOD RECORD VS. DLW METHODS

The majority of studies comparing reported energy intake to measured energy expenditure have used food records or diaries, in which the quantity of food intake is determined either by weight or common household measures. Research has shown that as the number of days of record keeping increases, the number of incomplete records increases (9). Therefore, we decided to categorize these studies into those with a recording period of 3-5 days, those that use 7-day records, and those with a recording period of ≥ 14 days.

Food records of 3-5 days (6 studies). Several studies have explored the relationship between reported en-

ergy intake and measured energy expenditure in older adults. Goran and Poehlman (11) recruited six women and seven men (mean age $67 \pm 6 \text{ yr}$) to complete a 3-day food record. Reporting accuracy differed significantly by gender (P = 0.031), with women underestimating their intake by 32% and men underestimating their intake by 13%. In all subjects, no significant change in body mass occurred during the study period; thus the difference between reported intake and expenditure was not due to changes in energy balance. Another study of older adults found similar results. Reilly et al. (22) recruited 10 healthy women (mean age 73 yr) to complete a 3-day weighed food record. Although energy balance was maintained throughout the study period (indicated by no change in weight or skinfold thickness measurements), reported energy intake was 27% less than measured energy expenditure. In the largest study of elderly adults, Tomoyasu et al. (30) also quantified the degree of underreporting. Their sample consisted of 82 men and women, ≥55 yr of age, whose energy intake was determined from a 3-day weighed food record. The magnitude of underreporting of total energy intake for all subjects was 20% and, in contrast to the findings of Goran and Poehlman (11), was comparable between the two genders. Body mass

Table 1. Summary of studies comparing reported EI by a single dietary instrument to TEE: reporting bias and correlation coefficient between EI and TEE

Study	Subjects	Dietary Methodology	Energy Intake, MJ/day	Energy Expenditure, MJ/day	Reporting Bias, %	r
Goran and Poehlman (11)	6 F (elderly)	3-day Food record	5.99	8.75	-32	0.77*
	7 M (elderly)	·	9.73	11.19	-13	0.77*
Reilly et al. (22)	10 F (elderly)	3-day Food record	6.71	9.21	-27	NA
Taren et al. (28)	37 F	3-day Food record	8.25	9.33	-12	$0.18\dagger$
Tomoyasu et al. (31)	28 M (African-American)	3-day Food record	9.75	11.32	-14	NA
	36 F (African-American)		7.89	8.74	-10	NA
Tomoyasu et al. (30)	39 M (older)	3-day Food record	8.73	11.30	-23	NA
	43 F (older)		6.87	8.35	-18	NA
Kaczkowski et al. (14)	53 F	4-day Food record	7.50	10.40	-28	NA
Clark et al. (7)	6 F (large eaters)	5-day Food record	10.49	8.49	+24	NA
	6 F (small eaters)		5.89	11.27	-48	NA
Goris et al. (12)	30 M (obese)	7-day Food record	10.40	16.70	-38	0.22
Livingstone et al. (18)	16 M	7-day Food record	11.21	14.23	-21	NA
	15 F		8.00	9.93	-19	NA
Martin et al. (19)	29 F	7-day Food record	6.98	9.00	-22	$0.46 \ddagger$
Prentice et al. (21)	9 F (obese)	7-day Food record	6.73	10.22	-34	NA
	13 F (lean)		7.85	7.99	-2	NA
Seale and Rumpler (27)	14 F	7-day Food record	7.88	9.57	-18	NA
	10 M		8.96	12.91	-31	NA
Lichtman et al. (16)	10 M/F (diet resistant)	14-day Food record	4.30	10.32	-59	NA
	6 M/F (control group)		7.08	11.07	-36	NA
Platte et al. (20)	10 F (control group)	14-day Food record	9.62	9.86	-2	NA
	10 F (obese)		8.43	15.51	-46	NA
Tuschl et al. (33)	11 F (unrestrained eaters)	14-day Food record	9.45	9.62	-2	NA
	12 F (restrained eaters)		8.18	8.61	-5	NA
Buhl et al. (5)	10 F (obese, diet resistant)	14-day Food record	NA	NA	-58	0.86§
Black et al. (3)	18 F	16-day Food record	8.30	9.50	-13	0.45
	27 M		10.06	11.67	-14	0.28
Rothenberg et al. (23)	12 M/F	Diet history	8.62	9.90	-13	0.27
Tran et al. (32)	35 F	24-h Recall (4 days)	9.26	11.06	-16	NA

Reporting bias = $\frac{\text{reported energy intake (EI) - measured total energy expenditure (TEE)}}{\text{measured TEE}} \times 100. \text{ A negative value for reporting bias}$

indicates underreporting of EI. F, female; M, male; NA, not available. *Correlation coefficient determined using both genders. $\dagger P < 0.05$; $\dagger P < 0.01$; $\S P < 0.001$.

E894 MINIREVIEW

Table 2. Summary of studies comparing reported EI by multiple dietary instruments to TEE: reporting bias and correlation coefficient between EI and TEE

Study	Subjects	Dietary Methodology	Energy Intake, MJ/day	Energy Expenditure, MJ/day	Reporting Bias, %	r
Bathalon et al. (1)	26 F (restrained eaters)	7-day Food record	7.10	8.96	-21	0.13
	, , , , , , , , , , , , , , , , , , , ,	24-h Recall (3 days)	6.60	8.95	-26	0.06
		FFQ	6.50	9.07	-28	0.06
	34 F (unrestrained eaters)	7-day Food record	8.30	9.49	-13	0.13
		24-h Recall (3 days)	7.60	9.42	-19	0.06
		FFQ	7.40	9.61	-23	0.06
Black et al. (4)	16 F	4-day Food record	8.99	9.42	-12	0.48
		Diet history	8.27	9.42	-5	0.11
Kroke et al. (15)	28 M/F	24-h Recall (12 days)	8.35	11.23	-26	NA
		FFQ	9.05	11.23	-19	NA
Sawaya et al. (24)	10 F (young)	7-day Food record	7.96	9.82	-19	NS
		24-h Recall (2 days)	9.56	9.82	-3	NS
		Block FFQ	7.08	9.82	-28	0.66*
		Willet FFQ	8.09	9.82	-18	0.63
	10 F (old)	7-day Food record	5.85	7.52	-22	NS
		24-h Recall (2 days)	6.42	7.52	-15	NS
		Block FFQ	6.41	7.52	-15	NS
		Willet FFQ	7.98	7.52	+6	NS

Reporting bias = $\frac{\text{reported EI} - \text{measured TEE}}{\text{measured TEE}} \times 100$. A negative value for reporting bias indicates underreporting of EI. FFQ, food-frequency questionnaire. NS, not significant (actual correlation coefficient not provided). *P < 0.05.

index, waist circumference, and fat mass were found to be significant correlates of underreporting, with the magnitude of underreporting increasing as these parameters increased (30). In another study by this same group (31), 64 African-American men and women 52-84 yr old completed a 3-day weighed food record. The magnitude of misreporting was not affected by gender and, on average, these men and women underestimated their intake by 11%. The authors speculate that a decreased preoccupation with body size and image in these older African-American men and women may be responsible for the more modest underreporting found in this later study (31). Kaczkowski et al. (14) explored the use of a 4-day multimedia diet record (MMDR). Fifty-three women 50-93 yr old were provided with standard household measures, a ruler, a 35-mm camera, and a microcassette tape recorder and were instructed to measure and record all foods and beverages consumed as well as left over. It was expected that the MMDR would decrease subject burden, thus reducing the omission of foods consumed but not recorded. The subjects in this study, however, were found to have underreported their energy intake by 27%, a bias greater than or similar to that found in other studies. Acknowledging this, the authors attributed their findings to the fact that the underlying principle of the MMDR remains the same as other dietary recording methodologies, namely the competence and the willingness of the subject to complete the record. Taren et al. (28) studied 37 women 30-60 yr old who completed a 3-day food record. Mean reported energy intake was found to be 12% lower than measured energy expenditure. Finally, in a study using a slightly longer recording period (5 days), Clark et al. (7) investigated the energy metabolism of 12 women categorized as either

"large" or "small" eaters. Placement into these categories was based on self-reported intake from studies in which these women had participated previously. Interestingly, large eaters overreported their energy intake by 24%, whereas small eaters underreported their intake by 48%. Neither group demonstrated a significant change in body weight during the study.

In only two of these studies did the authors report the correlation between intake and expenditure. Goran and Poehlman (11) found a strong association (r =0.77, P = 0.02) between self-reported energy intake and measured energy expenditure; however, a consistent negative bias was observed, with the regression line falling below the line of identity. Conversely, Taren et al. (28) found only a weak correlation (r =0.18, P < 0.05). Thus the underreporting did not appear to be a uniform error, but rather one in which the amount of underreporting varied between subjects. To better understand the determinants of underreporting, Taren et al. also investigated associations between reporting bias and physiological and psychological characteristics of participants. They found that reporting accuracy was inversely related to age and percent total body fat. Moreover, a high degree of social desirability was negatively associated with reporting accuracy $(r = -0.36, P \le 0.05)$.

Food records of 7 days (5 studies). One of the earliest studies to report a discrepancy between reported energy intake and measured energy expenditure was conducted by Prentice et al. (21). In this study, 13 lean and 9 obese women completed a 7-day weighed food record, with the obese group completing an additional 7-day record. Lean subjects were found to have underestimated their dietary intake by 2%, whereas the obese subjects underestimated their intake by 34%.

Additionally, the obese group tended to lose weight during the study, suggesting that the low self-reported intake was due to both undereating and underreporting. Consistent with these findings are the results of a recent study by Goris et al. (12), in which 30 obese males completed a 7-day dietary record. Reported energy intake was 37% less than measured energy expenditure, and a mean body mass loss of 1.0 kg occurred during the recording period. Livingstone et al. (18) published one of the first studies that showed a bias in the reporting of habitual energy intake in participants of a large population study. Subjects included 31 males and females (mean age 36 yr) from the Northern Ireland diet and health study. Average reported energy intakes were found to be significantly lower than measured expenditure in the overall group, with men and women underreporting energy intake by 21 and 19%, respectively. Another study of similar sample size recruited 24 normal body weight males and females (27). In this study, reported energy intake from diet records was 18% lower in females and 31% lower in males. Finally, a subset of participants from a long-term dietary intervention trial recorded their dietary intake for 7 consecutive days. The 29 women (mean age 49 yr) underreported their dietary intake by 22% (19). Again, most authors did not report the correlation between intake and expenditure. The one study that did, however, reported a modest correlation (r = 0.46, P < 0.01) between energy intake as reported on the food record and measured total energy expenditure (19).

Diet records of 14 days (5 studies). Longer diet records have the potential to reduce the effect of day-to-day variation in dietary energy intake and thus should improve the precision of reported energy intake. The longer recording period, however, may result in recording fatigue and a larger underreporting of dietary energy intake. They also provide a time period over which weight changes help to identify whether any discrepancy between reported intake and expenditure is due to underreporting or undereating.

The best correlation between intake and expenditure was reported by Buhl et al. (5) among a group of 10 women (r = 0.86, P < 0.001), but reporting accuracy was dismal (underreporting of 58%). The large reporting bias, however, probably reflects the fact that the women were selected from among obese individuals who had a history of being resistant to weight loss therapy despite consistently reporting low energy intakes (<1,200 kcal/day). Similar findings of very poor accuracy were reported by Lichtman et al. (16). Subjects in that study were divided into two groups. Group 1 [mean body mass index (BMI) = 33.8] consisted of 9 women and 1 man with a history of repeated failure to lose weight despite low caloric self-reported intake (subjects termed "diet resistant" by the authors). Group 2 (mean BMI = 36.4) consisted of 80 men and women without a history of diet resistance. Energy intake was measured using a 14-day weighed food record. Energy expenditure was measured in all group 1 subjects and in a subset of 6 group 2 subjects. The diet-resistant members of group 1 underreported energy intake by 58%, whereas group 2 subjects underreported intake by 36%. The characteristics of group 1 subjects may have played a role in these findings. Group 1 subjects were more likely to be using thyroid medication, had a stronger belief that their body weight was due to a genetic predisposition, had a series of failed diet attempts (some >20), and reported less hunger and more cognitive restraint than did group 2 subjects.

In a larger, more general population, Black et al. (3) validated reported energy intakes in 56 subjects. Subjects included 18 middle-aged women, 27 retired men, and 11 postobese persons. All subjects completed a 16-day weighed food intake record, the exception being postobese subjects who completed a 21-day record. The magnitude of energy intake underreporting was 13% in women, 14% in men, but 26% in postobese subjects. The correlations between intake and expenditure were 0.45 and 0.28 in the middle-aged women and retired men, respectively. The large underreporting in the studies by Buhl et al. (5) and Lichtman et al. (16) were therefore not likely to be a result of the longer recording period and recording-related fatigue but rather of the special selection criteria. In this regard, the larger underreporting in the postobese women supports the idea that the underreporting may be related to dieting behaviors. Consistent with this suggestion, the lowest magnitude of underreporting was found in a study conducted by Tuschl et al. (33). In this study, 23 normal-weight women completed a 14-day diet record as well as the Three-Factor Eating Questionnaire, used to classify subjects as either restrained or unrestrained eaters. The reported energy consumption of unrestrained eaters was found to be 2% lower than measured expenditure, whereas that of the restrained eaters was only 5% lower. Furthermore, restrained eaters had a significantly greater BMI than unrestrained

Continuing with the obesity and underreporting relationship, Platte et al. (20) recruited 10 normal-weight (BMI = 24) and 10 obese (BMI = 33) women to complete a 14-day diet record. The obese group showed a larger magnitude of energy intake underreporting compared with normal-weight women, 46 vs. 2%, respectively. Additionally, the obese subjects lost 1 kg of body weight during the recording period. The difference between reported energy intake and measured expenditure indicates that the obese group should have lost a total of 3.1 kg during the 2-wk recording period, suggesting that this group not only underate but also underrecorded their true intake.

Whether a longer period of dietary record keeping improves the reported energy intake for an individual appears to be a minor issue compared with the impact that subject characteristics have on such measures. In subjects defined as not diet resistant (16), normal weight (20), unrestrained eaters (33), or subjects from the general population (3), a longer dietary recording period does indeed appear to result in improved reported energy intake. However, in subjects defined as diet resistant (16), obese (20), or both obese and diet

E896 MINIREVIEW

resistant (5), any precision gained through the longer dietary recording period was outweighed by the largest magnitudes of underreporting bias of all studies included in this review. Therefore, the physical and psychological characteristics of subjects should be carefully considered.

THE DIET HISTORY VS. DLW METHODS

We found only one study, with ≥10 subjects, in which the diet history method was used to determine energy intake while energy expenditure was simultaneously measured using the DLW method. In this study, Rothenberg et al. (23) recruited 12 elderly subjects from the Gerontologic and Geriatric population studies in Sweden. The 9 females and 3 males were described as healthy and active 73 yr olds. Reported energy intake measured by diet history was 13% lower than measured energy expenditure. Additionally, reported energy intake from the diet history was not significantly correlated with measured total energy expenditure.

THE 24-H RECALL VS. DLW METHODS

Similarly, with regard to the 24-h dietary recall method, only one study met our criteria for this review. In this study, 35 weight-stable women (mean age 30 yr) completed four multiple-pass 24-h recalls (2 in person, 2 by telephone). On average, these subjects underreported their energy intake by 16%. Additionally, there was no significant difference in mean daily energy intake reported via telephone vs. in-person recalls (32).

On the basis of all these studies, there was no evidence that the accuracies of prospective or retrospective dietary techniques differed. Rather, there were strong indications that the physical and psychological characteristics of the subjects were determinants of underreporting. First, there appears to be a tendency for obese subjects to reduce their intake and subsequently their body mass when food records are used to assess habitual dietary intake. Goris et al. (12) demonstrated this phenomenon by defining a method to differentiate between undereating and underrecording. Undereating was defined as a consumption of less than the usual amount of food during the recording period and was calculated from the loss of body weight during the study. Underrecording of food intake assumes that the accuracy of recording water intake is reflective of the accuracy in recording food intake; it was calculated using the water balance technique. In this study of 30 obese males, undereating occurred in 26% of the subjects, whereas underrecording occurred in 12% of the subjects. The tendency for obese subjects to lose weight during the recording period has also been reported by others (21). Although not limited to obese participants, other studies have similarly reported either underrecording (24) or undereating (19) during the period in which food records are kept. Second, the behavioral and psychological profiles of study participants may also provide insight into the reporting accuracy of dietary methodologies. Restrained and unrestrained eaters studied by Bathalon et al. (1) completed a 7-day diet record, three 24-h recalls, and an FFQ. Compared with unrestrained eaters, restrained eaters consistently underreported intake to a larger degree, regardless of the dietary methodology employed. Taren et al. (28) also explored the influence of psychological characteristics on reporting bias and found that social desirability was found to be negatively associated with reporting accuracy. Additionally, body dissatisfaction and the concept of a smaller body size than one's own as being more healthy were also associated with a lower reporting accuracy (28). This inverse association between BMI and the degree of underreporting of energy intake has been demonstrated repeatedly, as exemplified by three studies performed in three different countries (16, 21, 33). Although beyond the scope of this review, it should be noted that the inverse relationship between BMI and reporting accuracy does not hold true for all segments of the population. For example, studies conducted in elite athletes have found large underreporting despite the fact that the subjects had low or average weights for height. Schulz et al. (26) studied energy balance in nine elite female runners of normal body weight (mean BMI = 20.4). Results from the 6-day food intake records compared with energy expenditure determined via DLW indicated that these athletes underreported their energy intake by 22%. Edwards et al. (8) also compared energy intake from a 7-day food record with energy expenditure measured by DLW in nine female distance runners. Energy intake was underreported by 32%, and a negative correlation (r = -0.74) was observed between reported energy intake and body weight. It should be noted, however, that although underreporting increased with increasing body weight, the weights of these subjects were in the range that is associated with accurate reporting among nonathletes. Additionally, Edwards et al. did find that the larger athletes had greater body image dissatisfaction, possibly related to athletic performance anxiety. This leads to the speculation that the observation, that underreporting increases with BMI or body weight, may in fact be related to body image dissatisfaction. This speculation is further supported by documented underreporting of energy intake by underweight women diagnosed with anorexia nervosa (6).

MULTIPLE DIETARY INSTRUMENTS VS. THE DLW METHOD

Because physical and psychological characteristics of the subjects have been shown to be determinants of underreporting, the lack of homogeneity among the studies considered in this review makes a comparison between dietary assessment techniques performed in different subjects very difficult. The most appropriate relatively accurate comparison of different dietary assessment instruments would be based on a comparison of studies drawn from similar subject pools in terms of age, gender, race, BMI, psychological profiles, and the like.

We found four studies in which the relative accuracy of several dietary methodologies could be evaluated within the same study population. All of these studies included at least one method that was retrospective and one that was prospective. Two of these studies included 24-h recalls and 7-day food records. Sawaya et al. (24) studied 10 young and 10 elderly adult females. In both groups, the 24-h recalls from 2 days of intake were more accurate than the 7-day diaries. The 24-h recalls indicated underreporting of 3 and 15% in the young and elderly women, respectively, whereas the 7-day records indicated underreporting of 19 and 22%, respectively. Conversely, Bathalon et al. (1) recently reported just the opposite results. While also investigating the association of psychological measures of eating behavior with the accuracy of reported energy intake, these authors classified 60 subjects as either restrained eaters (n = 26) or unrestrained eaters (n =34). For both restrained and unrestrained eaters, the 7-day weighed records showed a lower magnitude of underreporting, 21 and 12%, whereas the 24-h recalls displayed underreporting of 26 and 19%. Unfortunately, the one agreement between the two studies was that the correlations between intake and expenditure were either nonsignificant or very weak.

Black et al. (4) sought to validate retrospective dietary intakes measured using two methodologies, the diet history and a 4-day weighed food record. Sixteen middle-aged (mean age 58 yr) females participated in the study. Reported energy intakes, derived from the diet history method and the 4-day food record, were found to be 5 and 12% lower, respectively, than measured energy expenditure. In contrast to the studies by Sawaya et al. (24) and Bathalon (1), however, the intake by 4-day record was modestly correlated with expenditure (r = 0.48), although no differences between the methods were identified that could explain the different findings.

Three of the studies compared 24-h recalls with an FFQ. Kroke et al. (15) compared reported energy intake over a 1-yr period. Data from twelve 24-h recalls and an FFQ were obtained from a subgroup of 28 participants of the European Prospective Investigation into Cancer (EPIC) study. The magnitude of underreporting was 19% for the FFQ and 26% for the twelve 24-h recalls. Bathalon et al. (1) compared nutrient intakes derived from three 24-h dietary recalls and a FFQ and found that the FFQ showed the largest magnitude of underreporting. There was no significant association between reported energy intake by any method and total energy expenditure in either restrained or unrestrained eaters, separately or combined. Finally, Sawaya et al. (24) compared two 24-h recalls to the Block FFQ and the Willett FFQ in 10 young and 10 older women and found even more contrasts. In young women, the 24-h recall gave mean energy intakes that were closest to measured energy expenditure, underestimating intake by only 3%. In older women, the Willet FFQ yielded reported energy intakes closest to measured values (overreporting by 6%). The Block FFQ displayed more underreporting than the Willet FFQ in both young and old. The only case, however, in which reported energy intake was significantly correlated ($r=0.66,\,P<0.04$) to measured total energy expenditure was the use of the Block FFQ in young women.

SUMMARY OF DIETARY METHODOLOGIES VS. THE DLW METHOD

By using the DLW technique as a biomarker of habitual energy intake, the studies presented in this review demonstrate a variable occurrence of underreporting among different groups and types of individuals within a population. There was not an obvious relationship between reported energy intake and reporting accuracy. Neither prospective nor retrospective methods were consistently more accurate than the other, not even when the comparisons were performed in the same individuals. Perhaps such an expectation is naïve, because for each methodology, the exact questions, format, or nutrient database may be modified from study to study. These modifications allow the investigator to tailor the type of data gathered to the nutrient of interest; however, at least with current understanding, each modified version of the methodology developed may require validation of the data collected with respect to both traditional criteria (such as those described by I. M. Buzzard in 1994) and now underreporting. Such research requires the use of benchmark methods, such as the DLW technique, against which the accuracy of a self report can be assessed.

ALTERNATE MEANS OF ASSESSING REPORTED ENERGY INTAKE

The expense of the DLW technique has generally restricted its use to relatively small-scale studies of energy intake vs. expenditure. An alternate approach for identifying accurate and nonaccurate reporters of dietary intake is to test for unphysiologically low reported energy intakes. An example of such a test, the "Goldberg cut-off" (10), is based on the fact that energy expenditure can be represented as the basal metabolic rate (BMR) times an activity factor. For light activity, this factor is commonly taken as 1.55. A ratio of reported energy intake to calculated BMR (EI/BMR_{est}) that is less than this activity factor is suggestive of underreporting, or the cut-off under which reported intake is not valid. Recently, the sensitivity and specificity of the Goldberg cut-off for identifying diets of poor validity were tested. Black et al. (2) compiled a database of 429 subjects from 22 studies in which reported energy intake and energy expenditure (EE) determined via the DLW method were measured simultaneously. For comparison against Goldberg cut-off classification, the ratio of energy intake to energy expenditure (EI/EE) was used to define "true" underreporters (UR, EI/EE < 0.76), acceptable reporters (AR, EI/EE = 0.76-1.24), and overreporters (OR, EI/EE>1.24). The sensitivity was calculated as the proportion of UR, correctly identified as such with the GoldE898 MINIREVIEW

berg cut-off method. Similarly, the specificity was calculated as the proportion of non-UR correctly identified. When a cut-off factor for physical activity of 1.55 was used for men and women, the sensitivity was 0.50 and 0.52 and the specificity 1.00 and 0.95, respectively. Use of a higher physical activity factor cut-off of 1.95 in men and women increased the sensitivity (0.76 and 0.85, respectively) but resulted in decreased specificity (0.87 and 0.78, respectively). When subjects were classified into low, medium, and high activity levels (according to their individual EE/BMR ratio) and different cut-offs were applied for each activity level, there was a sensitivity of 0.74 and 0.67, with a specificity of 0.97 and 0.98, respectively, in men and women. The authors concluded that this last approach offered a reasonable compromise between sensitivity and specificity. This approach, however, relies on the ability of the investigator to choose suitable physical activity factor values for each activity level.

CONCLUSION

This review has demonstrated that, regardless of the dietary methodology employed to assess habitual diet, true energy intake is consistently underreported by certain groups within the population. As the ability to identify underreporters of dietary energy intake improves, it may seem that the next logical step would be to address what to do about underreporters of true intake. Proposed ideas to account for the reporting error have focused around two methods: 1) removing the underreporters from the data set, or 2) statistically correcting "reported" intake to represent true intake. Although the merits of these approaches are being investigated, it is our opinion that which foods are being underreported is of even greater importance than the magnitude of underreporting. For example, if the observed underreporting of energy intake on dietary surveys is due to a random omission of foods or an underestimation of portion size, techniques to account for this error, such as energy adjustment, should improve nutrient intake estimates. If, however, only certain types of foods are consistently underreported by some population groups, suggestive of a systematic bias, misleading conclusions will be drawn with regard to diet and disease relationships. Indeed, a recent study found that underreporters of energy intake reported consuming fewer foods rich in fat and sugar and reported higher vitamin C and fiber intakes than participants with a reported energy intake in the normal range (13). Goris et al. (12) found similar results with respect to fat intake.

In conclusion, further research is needed to identify underreporters of intake and to determine how to account for this bias in studies of diet and health. Additionally, it would be prudent to determine whether a reporting bias exists in the types of foods consumed by underreporters of energy intake. Such research would require the use of benchmark methods such as the

DLW technique or other established biomarkers of habitual dietary intake.

REFERENCES

- Bathalon GP, Tucker KL, Hays NP, Vinken AG, Greenberg AS, McCrory MA, and Roberts SB. Psychological measures of eating behavior and the accuracy of 3 common dietary assessment methods in healthy postmenopausal women. Am J Clin Nutr 71: 739-745, 2000.
- Black AE. The sensitivity and specificity of the Goldberg cut-off for EI:BMR for identifying diet reports of poor validity. Eur J Clin Nutr 54: 395–404, 2000.
- 3. Black AE, Bingham SA, Johansson G, and Coward WA. Validation of dietary intakes of protein and energy against 24 h urinary N and DLW energy expenditure in middle-aged women, retired men, and post-obese subjects: comparisons with validation against presumed energy requirements. Eur J Clin Nutr 51: 405–413, 1997.
- 4. Black AE, Welch AA, and Bingham SA. Validation of dietary intakes measured by diet history against 24 h urinary nitrogen excretion and energy expenditure measured by the doubly-labelled water method in middle-aged women. Br J Nutr 83: 341–354, 2000.
- Buhl KM, Gallagher D, Hoy K, Matthews DE, and Heymsfield SB. Unexplained disturbance in body weight regulation: diagnostic outcome assessed by doubly labeled water and body composition analyses in obese patients reporting low energy intakes. J Am Diet Assoc 95: 1391–1400, 1995.
- Casper RC, Schoeller DA, Kushner R, Hnilicka J, and Gold ST. Total daily energy expenditure and activity level in anorexia nervosa. Am J Clin Nutr 53: 1143–1150, 1991.
- Clark D, Tomas F, Withers RT, Chandler C, Brinkman M, Phillips J, Berry M, Ballard FJ, and Nestel P. Energy metabolism in free-living "large-eating" and "small-eating" women: studies using ²H₂¹⁸O. Br J Nutr 72: 21-31, 1994.
- Edwards JE, Lindeman AK, Mikesky AE, and Stager JM. Energy balance in highly trained female endurance runners. Med Sci Sports Exerc 25: 1398-1404, 1993.
- Gersovitz M, Madden JP, and Smiciklas-Wright H. Validity
 of the 24-hr dietary recall and seven day record for group comparisons. J Am Diet Assoc 73: 48-55, 1978.
- 10. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, and Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-reporting. Eur J Clin Nutr 45: 569–581, 1991.
- 11. **Goran MI and Poehlman ET.** Total energy expenditure and energy requirements in healthy elderly persons. *Metabolism* 41: 744–753, 1992.
- 12. Goris AH, Westerterp-Plantenga MS, and Westerterp KR. Undereating and underrecording of habitual food intake in obese men: selective underreporting of fat intake. *Am J Clin Nutr* 71: 130–134, 2000.
- Johansson L, Solvoll K, Bjorneboe G, and Drevon C. Under- and over-reporting of energy intake related to weight status and lifestyle in a nationwide sample. Am J Clin Nutr 68: 266–274, 1998.
- Kaczkowski CH, Jones PH, Feng J, and Bayley HS. Four day multimedia diet records underestimate energy needs in middle-aged and elderly women as determined by doubly labeled water. J Nutr 130: 802–805, 2000.
- 15. Kroke A, Klipstein-Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R, and Boeing H. Validation of a self-administered food frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-hour dietary recall methods. Am J Clin Nutr 70: 439–447. 1999.
- 16. Lichtman SW, Pisarska K, Berman ER, Pestone M, Dowling H, Offenbacher E, Weisel H, Heshka S, Matthews DE, and Heymsfield SB. Discrepancy between self-reported and

- actual caloric intake and exercise in obese subjects. $N\ Engl\ J\ Med\ 327:\ 1893-1898,\ 1992.$
- Lifson N and McClintock R. Theory of use of the turnover rates of body water for measuring energy and maternal balance. J Theor Biol 12: 46–74, 1966.
- Livingstone MB, Prentice AM, Strain JJ, Coward WA, Black AE, Barker ME, McKenna PG, and Whitehead RG. Accuracy of weighed dietary records in studies of diet and health. Br Med J 300: 708-712, 1990.
- 19. Martin LJ, Su W, Jones PJ, Lockwood GA, Tritchler DL, and Boyd NF. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary-intervention trial. Am J Clin Nutr 63: 483–490, 1996.
- Platte P, Pirke KM, Wade SE, Trimborn P, and Fichter MM. Physical activity, total energy expenditure, and food intake in grossly obese and normal weight women. *Int J Eat Disord* 17: 51–57, 1995.
- Prentice AM, Black AE, Coward WA, Davies HL, Goldberg GR, Murgatroyd PR, Ashford J, Sawyer M, and Whitehead RG. High levels of energy expenditure in obese women. Br Med J 292: 983–987, 1986.
- Reilly JJ, Lord A, Bunker VW, Prentice AM, Coward WA, Thomas AJ, and Briggs RS. Energy balance in healthy elderly women. Br J Nutr 69: 21–27, 1993.
- 23. **Rothenberg E, Bosaeus I, Lernfelt B, Landahl S, and Steen B.** Energy intake and expenditure: validation of a diet history by heart rate monitoring, activity diary, and doubly labeled water. *Eur J Clin Nutr* 52: 832–838, 1998.
- 24. Sawaya AL, Tucker K, Tsay R, Willett W, Saltzman E, Dallal GE, and Roberts SB. Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure. Am J Clin Nutr 63: 491–499, 1996.

- Schoeller DA. Measurement of energy expenditure in freeliving human beings by using doubly labeled water. J Nutr 118: 1278–1289, 1988.
- 26. Schulz LO, Alger S, Harper I, Wilmore JH, and Ravussin E. Energy expenditure of elite female runners measured by respiratory chamber and doubly labeled water. J Appl Physiol 72: 23–28, 1992.
- Seale JL and Rumpler WV. Comparison of energy expenditure measurements by diet records, energy intake balance, doubly labeled water, and room calorimetry. Eur J Clin Nutr 51: 856– 863, 1997.
- Taren DL, Tobar M, Hill A, Howell W, Shisslak C, Bell I, and Ritenbaugh C. The association of energy intake bias with psychological scores of women. Eur J Clin Nutr 53: 570–578, 1999.
- 29. **Thompson FE and Byers T.** Dietary assessment resource manual. *J Nutr* 124: 2245S–2317S, 1994.
- 30. **Tomoyasu NJ, Toth MJ, and Poehlman ET.** Misreporting of energy intake in older men and women. *J Am Geriatr Soc* 47: 710–715, 1999.
- 31. **Tomoyasu NJ, Toth MJ, and Poehlman ET.** Misreporting of total energy intake in older African Americans. *Int J Obes Relat Metab Disord* 24: 20–26, 2000.
- 32. Tran KM, Johnson RK, Soultanakis RP, and Matthews **DE.** In-person vs telephone-administered multiple-pass 24-hour recalls in women: validation with doubly labeled water. *J Am Diet Assoc* 110: 777–783, 2000.
- 33. Tuschl RJ, Platte P, Laessle RG, Stichler W, and Pirke K. Energy expenditure and everyday eating behavior in healthy young women. Am J Clin Nutr 52: 81–86, 1990.
- 34. **Wacholder S.** When measurement errors correlate with truth: surprising effects of nondifferential misclassification. *Epidemiology* 6: 157–161, 1995.

