Leveraging questionnaire-based physical activity levels (PAL) to identify energy intake misreporting using the Goldberg method: A doubly labeled water validation study

Heather K Neilson[1], Shervin Asgari[1], Janet A. Tooze[2], Farah Khandwala[1], Anita Koushik[3], Rémi Rabasa-Lhoret[4], Karen A. Kopciuk[1,5,6*], Ilona Csizmadi[6*]✉

Author Affiliations:

[1]Department of Cancer Epidemiology and Prevention Research, Cancer Care Alberta, Calgary, Alberta, Canada (HKN, SA, FK, KAK)

[2]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Wake Forest University, Winston-Salem, NC, USA (JAT)

[3]Department of Oncology, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada (AK)

[4]Departments of Medicine and Nutrition, Université de Montreal, and Montreal Clinical Research Institute, Montreal, Quebec, Canada (RRL)

[5]Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada (KAK)

[6]Departments of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada (KAK, IC)

* Co-senior authors.

Authors' Last Names: Neilson, Asgari, Tooze, Khandwala, Koushik, Rabasa-Lhoret, Kopciuk, Csizmadi

Abstract: 199 words

Manuscript: 3,992 words

Short Running Head: Energy intake misreporting – a doubly labeled validation study.

ORCiD:

Ilona Csizmadi: 0000-0002-4895-1310

Heather K. Neilson: 0000-0002-6803-2800

Karen A. Kopciuk: 0000-0001-7499-1861

Anita Koushik: 0000-0001-5304-7660

Janet A. Tooze: 0000-0002-5160-0407

Rémi Rabasa-Lhoret: 0000-0003-4706-5170

Sources of Support: Funding for this study was provided to institutions by the Canadian Institutes of Health Research (Grant MOP-86632) and the Alberta Cancer Research Institute (Grant 24265). **Role of funding agencies:** The funders had no role in the design of this study, the analysis, interpretation of results or drafting of the manuscript.

✉ **Corresponding Author:**

Dr Ilona Csizmadi, Departments of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

Email: icsizmad@ucalgary.ca

Abbreviations:

BMR_{EST}: estimated basal metabolic rate; **C-DHQI**: Canadian Diet History Questionnaire I; **DLW**: doubly labeled water; **EI**: energy intake; **NPV**: negative predictive values; **PAL**: physical activity level; **PPV**: positive predictive value; **rEI**: self-reported energy intake; **STAR-Q**: Sedentary Time and Activity Reporting Questionnaire; **TEE**: total energy expenditure; **TBW**: total body water

ABSTRACT

The Goldberg method has been suggested for identifying energy intake (EI) under-reporting in nutritional epidemiology. Its implementation, however, is limited by challenges associated with estimating physical activity levels (PAL). We quantified the accuracy of the Sedentary Time and Activity Reporting Questionnaire (STAR-Q) derived PAL (PAL_{STAR-Q}) combined with the Goldberg method (Goldberg-PAL_{STARQ}) to identify EI misreporting as compared with doubly labeled water (DLW) derived total energy expenditure (TEE_{DLW}). Between 2009 and 2011, 99 men and women completed a two-week DLW protocol, a food frequency questionnaire, and the STAR-Q. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy of the Goldberg-PAL_{STAR-Q} were determined. Fifty-eight percent of men and women were classified as under-reporters by Goldberg-PAL_{STAR-Q} compared with 60% of men and 56% of women by TEE_{DLW}. Among men, sensitivity, specificity, PPV, NPV and accuracy and 95% confidence intervals were 88% (68%-97%), 87% (61%-98%), 91% (72%-99%), 81% (56%-94%), and 87% (72%-95%), respectively; and among women 79% (62%-90%), 69% (50%-84%), 77% (60%-88%), 72% (52%-86%), and 75% (62%-84%), respectively. Validated individual level PALs used with the Goldberg method can be informative in sensitivity analyses to gain insight into EI misreporting in nutritional epidemiology studies lacking in objective EI measures.

Keywords: energy intake, accuracy, misreporting, doubly labeled water, Goldberg method, physical activity level, adults, under-reporting, 24-hour movement paradigm

INTRODUCTION

Under-reporting of energy intake (EI) on dietary questonnaires is a widely recognized problem that distorts estimates of dietary intake in nutrition surveys and etiologic studies – thus impeding progress in diet-related research (1). Various statistical approaches have been suggested to identify dietary intake misreporting, but none have satisfactorily addressed the problem (2). Regression calibration of dietary intake is a promising approach for nutrients and dietary patterns (3-5); however, the method relies on the knowledge of *true* dietary intake and an understanding of the underlying systematic errors associated with self-report which are generally unknown (6). Ideally, misreporting could be detected using 'gold standard' objective biomarkers of dietary intake, but, these methods are limited in number and usually too labor-intensive and costly to feasibly implement in large-scale epidemiologic studies (7). A practical alternative, specifically for EI is the classification of individuals according to the plausibility of their reported EI (rEI) compared with their energy requirements (8).

The Goldberg method (9), originally proposed for identifying EI misreporting in community-dwelling populations, is based on the principle that under long-term, weight-stable conditions, EI is approximately equal to total energy expenditure (TEE). In addition, the method leverages the well-established relations between EI, TEE, basal metabolic rate (BMR) and physical activity level (PAL), where EI/BMR approximates TEE/BMR which in turn represents PAL (10, 11). Further, the method proposes the estimation of Goldberg cut-point boundaries, or 95% confidence intervals (95% CI), against which ratios of rEI:BMR are compared to identify under and over-reporting.

While conceptually tenable, the method is logistically challenged by difficulties associated with estimating PAL in free-living populations (11, 12). In the absence of population-specific

objective measures, a low-activity PAL of 1.55 based on World Health Organization energy requirement prediction equations (13) was applied in the original Goldberg paper (14). The Goldberg method combined with a globally applied PAL of 1.55 was subsequently compared with doubly labelled water (DLW), the 'gold standard' for estimating TEE (15, 16), and reported to have acceptable accuracy in a study of middle-aged men and women; however, alternate PALs were not tested (17). More recently, contemporary DLW studies included in the report from the Dietary Reference Intake (DRI) for Energy (10), point to a median adult PAL of 1.68, suggesting that a global PAL of 1.55 may be unrealistically low for substantial segments of the population potentially resulting in very active under-reporters misclassified as acceptable.

To our knowledge, the Goldberg method has not been evaluated for accuracy using PALs derived from comprehensive individual self-reports of activity despite their value in capturing habitual behaviors of relevance in nutritional epidemiology. With that in mind, we appraised the accuracy of the Goldberg method for classifying EI reporting status using individual-level PALs derived from the Sedentary Time and Activity Reporting Questionnaire (STAR-Q), a comprehensive all-inclusive activity questionnnaire previously validated against DLW (18). In our primary analysis, we determined the accuracy of the Goldberg method using STAR-Q-derived PALs, hereafter referred to as Goldberg-$PAL_{STAR\text{-}Q}$, in comparison with DLW-derived TEE_{DLW} for identifying EI misreporting on a food frequency questionnaire (FFQ). In secondary analyses we similarly tested the commonly used PAL of 1.55 and in sensitivity analyses we modified the assumption for FFQ variability in estimating of 95% CIs (17).

METHODS

Participants in this analysis were those enrolled in the original STAR-Q DLW validation study (18). The STAR-Q was conceptualized to capture 24-hour movement behaviours (19). It was therefore uniquely designed to ascertain self-reported time spent in sleep, sedentary behavior, and all types and intensities (e.g., light, moderate and vigorous) and domains of activity (e.g., occupation, leisure, household, and transportation) and took a median time of 34 minutes to complete (18, 20).

Subject Eligibility and Ethics

Participant recruitment has previously been described (18). Briefly, urban residents living in Calgary, Alberta, Canada were recruited from the community or from the Alberta Tomorrow Project (ATP). The ATP is a large geographically dispersed cohort established in 2001 in Alberta, to study links between lifestyle behaviors and chronic disease risk (21, 22). ATP participants who met eligibility criteria for the DLW validation study – body mass index (BMI) ≤ 35, weight stable (± 2.5 kg for at least 3 months), not planning to gain or lose weight, not pregnant or breastfeeding, not diagnosed with metabolic disorders, and not taking medications that modify water balance – were actively recruited by mail. Non-ATP participants were recruited by community-based advertisements. The upper age limit for eligibility was 60 years, and the lower age limit was 30 and 35 years for urban Calgary and ATP participants, respectively.

All participants provided written informed consent. Ethical approval was obtained from the Alberta Cancer Research Ethics Committee of Alberta Health Services and the Conjoint Health Research Ethics Board of the University of Calgary. All study procedures were in accordance with ethical standards outlined by the two governing bodies.

Doubly labeled water study

Details of the 14-day DLW protocol have been published (18). Briefly, at baseline (Day 0), following an overnight fast, participants provided urine and saliva samples to determine background isotope levels. Each subject was then orally dosed with 2.5g 10 atom % of ^{18}O per kg total body water (TBW) and 0.18 g 99 atom % deuterium (^{2}H) per kg estimated TBW (Rotem Inc., Topsfield, Massachusetts). Post-dose saliva samples were collected at 3 and 4 hours to estimate TBW from deuterium-isotope dilution, and second-void urine samples on Days 1, 8, and 14. On Day 14 after an overnight fast, participants received 0.18 g 99 atom % ^{2}H per kg TBW. Post-dose saliva sample collection was as described for Day 0. To estimate the decline in isotope enrichment, samples were batch-analyzed in duplicate using the Isoprime Stable Isotope Ratio Mass Spectrometer and the Multiflow-Bio module for Isoprime equipped with a Gilson 222XL Autosampler (GV Instruments, Manchester, United Kingdom). Data processing was performed using IonVantage (2012) software for Isoprime (Cheadle, United Kingdom).

We calculated TEE_{DLW} in kilocalories per day using a modified Weir equation (23) and an assumed respiratory quotient of 0.85. In the absence of objective measures, BMR was estimated using the Schofield Equation (24) to derive individual level PALs from DLW as:

$$PAL_{DLW} = TEE_{DLW} / BMR_{EST}$$

Health and lifestyle measurements

Height and weight were measured (to the nearest 0.1 cm and 0.1 kg, respectively) on Days 0 and 14 of the DLW protocol using standard procedures (25). The TBF-310 Tanita Body Composition Analyzer and Scale (TheCompetitiveEdge.com, Preston, Washington) was used to measure body weight. BMI was determined as body weight divided by height-squared (kg/m^2).

112 The Canadian Diet History Questionnaire I (C-DHQ I) was administered once on Day 0 of
113 the DLW protocol to estimate rEI. The C-DHQ I is a comprehensive FFQ that assesses usual
114 dietary intake over the past-year, modified from the U.S. National Cancer Institute's DHQ for
115 Canadian populations (26, 27) and used in large epidemiologic studies including the ATP (21, 22,
116 28, 29).

117 The STAR-Q was completed on Day 14 of the DLW study to estimate average daily energy
118 expenditure from all activities including sleep during the preceding month. STAR-Q-derived TEE
119 and activity energy expenditure demonstrated high reproducibility over a three-month period, with
120 intraclass correlation coefficients of 0.84 and 0.73, respectively (18). Compared with DLW, the
121 STAR-Q demonstrated moderate validity for ranking individuals by TEE and activity energy
122 expenditure (Spearman's rho=0.53 and 0.40, respectively) (18).

123 To derive TEE from the STAR-Q, activity code and metabolic equivalents of task (MET)
124 (30) were first assigned to each reported activity. If the total time reported for physical activity,
125 sedentary behavior, and sleep was less than 24 hours/day, a MET value of 1.2 was assigned for
126 unaccounted time based on evidence that sedentary activity is more likely to be under-reported
127 than other intensity levels of activity (18, 31). Reported frequency, duration and assigned MET
128 values were used to estimate energy expenditure for each activity in units of MET-hours/day,
129 which were then summed to derive the total average MET-hours/day. The STAR-Q-derived TEE
130 ($TEE_{STAR\text{-}Q}$) was then estimated as:

131

132 $TEE_{STAR\text{-}Q}$ = [[0.9 MET×hrs sleep/d×weight(kg)] + [MET-hrs activity/d ×weight (kg)]]*1.1

133

where STAR-Q reported hours of sleep per day was multiplied by 0.9 METs (30, 32) and then added to the total average MET-hours/day in awake time multiplied by baseline body weight. The equivalence of 1 MET = 1 kcal/kg/hr was applied to estimate the energy cost of activities (33) and 10% was added for the thermic effect of food (10).

Individual PALs were derived from self-report ($TEE_{STAR\text{-}Q}$) as:

$$PAL_{STAR\text{-}Q} = TEE_{STAR\text{-}Q}/BMR_{EST}$$

where BMR_{EST} was estimated using the Schofield sex-stratified equations based on age, weight and height (24).

Classification of misreporting

EI reporting status was determined for each of the ratios rEI:MBR_{EST}, rEI:TEE_{DLW}, and rEI:TEE_{STARQ} compared with respective estimated 95% CIs based on $PAL_{STAR\text{-}Q}$, TEE_{DLW}, and $TEE_{STAR\text{-}Q}$ (calculations below). Each participant was classified as an under-reporter (UR), acceptable reporter (AR), or over-reporter (OR) according to comparisons of individual rEI ratios with 95% CIs. Participants with ratios falling below their respective 95% CIs were deemed URs, those with ratios above the upper boundary were deemed ORs and those with ratios between the boundaries were deemed ARs.

In the primary analysis, to evaluate the Goldberg-$PAL_{STAR\text{-}Q}$ for classifying individual misreporting, 95% CIs for comparison with rEI:BMR_{EST} were calculated as:

$$e^{\left(\pm 2 \times \sqrt{CV_{EI}^2/d + CV_{BMR}^2 + CV_{PAL}^2}\right)}$$

where CV_{EI} was the within-subject coefficient of variation (CV) for rEI, assigned as 18.6% for men and 19.8% for women based on repeatability of the US DHQ (17), d =1 since the C-DHQ I was administered once, CV_{BMR} was the CV for estimated BMR assumed to be 8.5% for men and women as previously suggested (11) and CV_{PAL} was the CV for $PAL_{STAR\text{-}Q}$, calculated to be 17.5% for men and 15.1% for women from our three-month STAR-Q repeatability study (18).

For the gold standard $rEI{:}TEE_{DLW}$ the 95% CIs were calculated as:

$$e^{(\pm 2 \times \sqrt{CV_{EI}^2 / d + CV_{DLW}^2})}$$

where CV_{EI} (d =1) was assumed to be the same as above and CV_{DLW} was assigned 5.1% for both men and women as previously reported (17).

To examine the magnitude of discrepancies relative to the gold standard ratio $rEI{:}TEE_{DLW}$, graphical illustrations described by Tooze *et al*. (17) were used comparing individual-level $rEI{:}TEE_{STAR\text{-}Q}$ and $rEI{:}TEE_{DLW}$ ratios to their respective 95% CIs.

The 95% CI for with which $rEI/TEE_{STAR\text{-}Q}$ was compared was calculated as:

$$e^{(\pm 2 \times \sqrt{\frac{CV_{EI}^2}{d} + CV_{STAR-Q}^2})}$$

where CV_{EI} and d were as described above. The $CV_{STAR\text{-}Q}$ was the within-subject CV for $TEE_{STAR\text{-}Q}$ calculated to be 17.7% for men and 15.2% for women based on the first and second administrations of the STAR-Q in the repeatability study (18). If an individual $rEI{:}TEE_{STAR\text{-}Q}$ ratio fell between the lower (or upper) 95% CI boundaries for $rEI{:}TEE_{STAR\text{-}Q}$ and $rEI{:}TEE_{DLW}$ ratios, the source of discrepancy was attributed to differences in the 95% CI calculations. Otherwise, it was attributed to differences in TEE estimation ($TEE_{STAR\text{-}Q}$ vs TEE_{DLW}). Furthermore, the

183 contribution of these discrepancies to sensitivity and specificity of Goldberg-PAL_{STAR-Q} was
184 determined.

185 In secondary analyses, a constant PAL 1.55 was assumed for participants ($PAL_{1.55}$) and
186 95% CIs calculated as described above for PAL_{STAR-Q}, except CV_{PAL} was assumed to be 15% for
187 both genders as previously reported (11).

188 Finally in sensitivity analyses, for comparison with a previous study (17), $d=\infty$ was
189 substituted for $d=1$, in each of the cut-point equations (above) eliminating the CV^2_{EI} term. This
190 alternate estimate of '*d*' assumes that the C-DHQ I ascertains habitual diet over an infinite time
191 period.

192

193 **Statistics**

194 All analyses were stratified by gender. Means (SD) were determined for continuous
195 sociodemographic characteristics, frequencies and percentages for categorical characteristics, and
196 medians (Q1, Q3) for rEI and energy expenditure. To appraise the validity of Goldberg-PAL_{STAR-}
197 $_{Q}$ and $PAL_{1.55}$ classifications, the estimated proportions of participants correctly classified as URs,
198 ARs and ORs were compared with DLW classifications assumed to be *true* classifications.
199 Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the
200 accuracy of UR and AR classifications were estimated (34), excluding ORs identified by
201 $TEE_{DLW,d=1}$ (n=1) and $TEE_{DLW,d=\infty}$ (n=7). The sensitivity was the proportion of true URs correctly
202 classified by the Goldberg method (PAL_{STAR-Q} or $PAL_{1.55}$) and specificity was the proportion of
203 true ARs correctly classified. PPV was the proportion of Goldberg method URs that were true
204 URs, interpreted as the probability of being a true UR if classified as UR by Goldberg. NPV was
205 the proportion of Goldberg ARs that were true ARs, interpreted as the probability of being a true

206 AR if classified as AR by the Goldberg method. Accuracy was determined as the proportion of
207 true URs plus true ARs classified correctly with the Goldberg method.

208 To appraise the validity of $TEE_{STAR\text{-}Q}$ for estimating the magnitude of misreporting (UR
209 and OR) in kilocalories per day, the within-person differences of $TEE_{STAR\text{-}Q}$ - rEI were compared
210 with TEE_{DLW} - rEI using Pearson correlation coefficients and Wilcoxon signed-rank tests.

211 Agresti Coull 95% CIs were calculated for diagnostic metrics using Python (Version
212 3.12.12; Statsmodels Version 0.14.5). All other analyses were performed using SAS Institute
213 Software, Gary, NC (Version 9.4). A P-value < 0.05 was considered statistically significant.

214

215 **RESULTS**

216 Participants aged 30-60 years (n=106) were recruited to the original DLW study between
217 July 2009 and July 2010 (18). Seven participants were excluded from the present analysis for
218 reasons outlined in **Figure 1,** leaving 99 participants of whom 60% were women. Compared with
219 men, women were somewhat younger (mean: 46.2 vs 50.6 years of age), had a lower BMI (mean:
220 23.5 vs 25.9 kg/m^2), and were less likely to have completed post-secondary education (45% vs
221 51%) (**Table 1)**. Most of the participants were Caucasian and had paid or volunteer work. Median
222 EI, TEE_{DLW}, $TEE_{STAR\text{-}Q}$, and BMR_{EST} were lower for women (1676 kcal, 2604 kcal, 2889 kcal and
223 1362 kcal, respectively) compared with men (2036 kcal, 3306 kcal, 3620 kcal, and 1866 kcal,
224 respectively). Median values for PAL_{DLW} and $PAL_{STAR\text{-}Q}$, were 1.8 and 2.0 for men and, 1.9 and
225 2.1 for women **(Table 1)**, consistent with 'active' and 'very active' PALs (10).

226 Results from the primary, secondary and sensitivity analyses are presented in **Table 2**. In
227 the primary analysis, 60% of men and 56% of women were classified as URs by TEE_{DLW}, and
228 58% of men and women by the Goldberg-$PAL_{STAR\text{-}Q}$. The Goldberg-$PAL_{STAR\text{-}Q}$ misclassified

229 fewer men (15%) than women (25%), resulting in sensitivity, specificity, PPV, NPV and accuracy
230 >80% among men, whereas for women these estimates were <80%.

231 In secondary analyses, Goldberg-$PAL_{1.55}$ resulted in fewer participants classified as URs
232 (35% and 19% for men and women, respectively) than with TEE_{DLW} and Goldberg-$PAL_{STAR\text{-}Q}$
233 which led to high specificity and PPVs (both 93% for men and 100% for women), but low
234 sensitivity and NPVs (all <60% for men and women).

235 In sensitivity analyses, where $d=\infty$ was assumed for the FFQ, a larger proportion of
236 participants were classified as URs with both TEE_{DLW} and Goldberg-$PAL_{STAR\text{-}Q}$ than with the
237 assumption that $d=1$. Seven were deemed to be true ORs in contrast to one, when $d=1$, attributable
238 to a narrower 95% CI with $d=\infty$ following elimination of the variability term. As such, less
239 favorable results were seen for the identification of URs with the Goldberg-$PAL_{STAR\text{-}Q}$ except for
240 specificity and PPV, which were 100% among men. However, the assumption that the FFQ
241 captures habitual intake representing an infinite time period may be an overestimate of this
242 attribute (35).

243 **Figures 2A and 2B** illustrate individual-level agreement and magnitude of misreporting in
244 kilocalories per day (i.e., $TEE_{STAR\text{-}Q}$ - rEI relative to TEE_{DLW} - rEI). Pearson correlation
245 coefficients were higher for men (r=0.76, $P<0.0001$; classification agreement in 34 of 40 men)
246 than for women (r=0.37, $P=0.0042$; classification agreement in 44 of 59 women). Nonetheless, the
247 Wilcoxon signed-rank test indicated statistically significant within-person differences between the
248 two methods (P = 0.0021 and P < 0.0001 for men and women, respectively). The magnitude of EI
249 under-reporting tended to be larger for misclassified URs (e.g., URs classified as ARs,
250 approximately 1,000 to 1,500 kcal/d; 3 men and 7 women) while the magnitude of EI under-

251 reporting was typically smaller for misclassified ARs (e.g., ARs classified as URs, <1,000 kcal/d
252 approximately; 2 men and 8 women).

253 **Figures 3A and 3B**, show whether discrepant classifications arose from different 95% CI
254 cut-points or from different TEE estimation. Among 40 men **(Figure 3A)**, five misclassifications
255 were due to TEE estimation, and one was due to interval cut-point. Among 59 women (**Figure**
256 **3B)**, 11 misclassifications were due to differences in TEE estimation and 4 due to cut-point
257 estimation. Among *true* URs misclassified as ARs using the $PAL_{STAR\text{-}Q}$ (reflecting sensitivity), 2
258 of 3 men and 3 of 7 women were misclassified owing to TEE estimation. Among *true* ARs
259 misclassified as URs using the $PAL_{STAR\text{-}Q}$ method (reflecting specificity), all misclassifications (2
260 men and 8 women) were due to differences in TEE estimation.

261

262 **DISCUSSION**

263 Using DLW, the 'gold-standard' biomarker for TEE and an unbiased proxy for habitual EI
264 in weight-stable individuals (2), we determined the accuracy of the Goldberg method using STAR-
265 Q-derived PALs to classify EI reporting status among community dwelling middle-aged men and
266 women following completion of an FFQ. The $PAL_{STAR\text{-}Q}$ generally improved the sensitivity and
267 overall accuracy of the Goldberg method for discriminating between URs and ARs compared with
268 a global PAL of 1.55. We also identified gender-based differences in the performance of the
269 Goldberg-$PAL_{STAR\text{-}Q}$. In men, the method discriminated between UR and AR with an overall high
270 level of accuracy at 87%, while in women it exhibited fair performance, with an overall
271 classification accuracy of 75%.

272 A key determinant of the accuracy of the Goldberg method is the degree to which PALs
273 represent populations under study. Three previous DLW validation studies have explored

274 approaches to identify the optimal PAL for estimating EI misreporting (17, 36, 37). Two studies
275 examined EI misreporting on food records and/or diet histories and found under-reporting in 14%
276 (37) and 34% (36) of adult participants. Consistent with our results, the studies reported high levels
277 of specificity but relatively low sensitivity for identifying URs when a global PAL of 1.55 was
278 assumed. Additionally, Black (36) leveraged the strength of a large database comprised of 21 DLW
279 studies (n=429) to test the effect of various global mean activity PALs (from 1.55 to 1.95) and
280 reported incremental increases in sensitivity, with increasing PALs; however, this improvement
281 coincided with decreases in specificity. In a small sample of adults (n=14), Livingstone *et al* (37)
282 tested the performance of gender-specific PALs to detect URs based on individual heart-rate
283 monitors matched to low, medium and high activity categories of PAL (13). Compared with
284 $PAL_{1.55}$, sensitivity increased substantially from 50% to 100% with negligible decreases in
285 specificity. Subsequently, Tooze *et al* (17) reported favorable results for the Goldberg method with
286 a $PAL_{1.55}$ to detect DLW-classified URs on 24-hour dietary recalls (~20%) and an FFQ (~50%
287 UR) in the Observing Protein and Energy Nutrition (OPEN) Study (n=451). Specificity was 88%
288 for the FFQ in both genders. The sensitivity to detect URs on the FFQ was high at 96% when usual
289 intake was assumed for the FFQ ($d=\infty$), but substantially lower when the OPEN study FFQ within-
290 person variability was applied (72% sensitivity for men and 62% for women).

291 The results reported by Livingstone *et al* (37) and Black (36) are consistent with our results
292 and support the premise that a global $PAL_{1.55}$ may be underestimating the activity levels of some
293 adult populations – and hence, more individualized PALs may improve the overall accuracy of the
294 Goldberg method to identify EI under-reporting. While our findings contrast those of Tooze *et al*
295 (17), we note that the prevalence of true UR (50%) was lower in the OPEN study than in our
296 population which affects PPV and NPV values. In addition, participants were on average less

active and slightly older than in the STAR-Q validation study. Median values for DLW-derived TEE and PAL were 2,813 kcal and 1.69 for men and 2,283 kcal and 1.75 for women in OPEN (38, 39) and 3,306 kcal and 1.8 for men and 2604 kcal and 1.9 for women in the STAR-Q study, with ages ranging from 40 to 69 years in OPEN (38, 39) versus 30 to 60 years in the STAR-Q study.

Statistical approaches, such as the exclusion of misreporters, and stratification and adjustment for reporting status in regression analyses (40-45), have been explored to handle identified URs. Consensus exists only with respect to an agreement that exclusion of misreporters is undesirable given loss of statistical power (41) and valuable information inherent in excluded data (40). While the optimal approach may be study-specific (e.g., health outcomes of interest, study design, method of data collection and population characteristics), corrective methods cannot be expected to fully eliminate bias associated with measurement error. Nonetheless, the identification of implausible EI reporting using the Goldberg method has been demonstrated to significantly reduce, albeit not completely eliminate, misreporting-related bias for some EI-outcome associations (44, 45). This suggests an important role for the method in sensitivity analyses aiming to better understand the potential impact of EI misreporting on study results. Of note, the aforementioned studies used group level PALs – potentially limiting the extent to which bias due to misreporting could be mitigated.

While the STAR-Q was designed to capture a comprehensive profile of daily activities over 24-hours, the misclassification of 50% of true URs as ARs and all true ARs as URs, was likely due to errors in the estimation of $TEE_{STAR\text{-}Q}$ **(Figure 3a and 3b).** Using updated MET values (33, 46) for energy expenditure, could improve future TEE and PAL estimations. In addition, the reporting of activities during the previous month are not without cognitive challenges, though less so than past-year reporting (20). The Activities Completed over Time in 24 hours (ACT24), an

320 internet-based recall of previous day activities is even less cognitively challenging and has been
321 shown to estimate mean TEE and PAL values within 5% of DLW-based estimates (47, 48). The
322 accuracy of short-term recall combined with longer-term measures that reflect habitual lifestyle
323 behaviors is an intriguing approach that has been suggested for improving dietary intake measures
324 (49) and may also improve TEE and PAL estimations. Moreover, new wearable device-based
325 technologies may in the foreseeable future accurately estimate PALs and TEE in large populations,
326 to complement the contextual data collected by traditional questionnaires (50). Hence, it is
327 anticipated that methodologies combining device-derived estimates of energy expenditure and
328 body composition may be developed to validate self-reported EI (51-53); though these approaches
329 are still viewed as 'works in progress' (54).

330 A noteworthy strength of this study is the validation of the Goldberg-$PAL_{STAR\text{-}Q}$ method
331 against DLW – and the unique use of individual-level PALs not previously studied. Furthermore,
332 as noted in the recent DRIs for Energy report, of 128 DLW studies representing 25 countries in
333 the International Atomic Energy Agency DLW Database, none were Canadian (10) highlighting a
334 gap in this area of Canadian research that this study begins to address. Nonetheless, our study has
335 limitations that merit consideration. First, we used prediction equations for BMR (24), which
336 accounts for a substantial proportion of TEE (55) and could be a source of error in the estimation
337 of $PAL_{STAR\text{-}Q}$. Second, we cannot be certain that some of the under-reporting was not true
338 undereating, though we aimed to minimize this occurrence by recruiting generally weight stable
339 participants to the DLW study. Third, we caution that we cannot assume that EI misreporting is
340 reflective of misreporting of other nutrients. Indeed, though still relatively under-studied, evidence
341 is accumulating to suggest that foods and nutrients vary in the accuracy with which they are
342 reported (6, 56-59). On a final note, our population was largely Caucasian middle-aged adults;

343 hence, our results may not be generalizable to older (60, 61), racially diverse populations and those
344 with other characteristics that may affect reporting and health status (62-65).

345 Overall, our DLW-validated results lend credibility to the thesis that detailed
346 comprehensive questionnaire-based activities may be effectively employed with the Goldberg
347 method to describe and explore EI misreporting in large-scale diet studies. Importantly, such
348 strategies could create opportunities to carry out sensitivity analyses that provide critical insights
349 into the impact of EI misreporting on diet-disease associations in nutritional epidemiology.
350 Ongoing effort to improve upon this approach with cohort validation sub-studies and innovative
351 study designs integrated with new technologies need to remain as research priorities in nutritional
352 epidemiology.
353

354 **STATEMENTS AND DECLARATIONS**

355 **Data availability**: Data described in the manuscript, code book, and analytic code will be made
356 available upon reasonable request to the corresponding author. The data are not publicly available
357 due to privacy and/or ethical restrictions.

358 **Conflicts of Interest**: IC and KAK were awarded research funding paid to their institutions from
359 the Canadian Institutes of Health Research (CIHR: Grant MOP-86632) and the Alberta Cancer
360 Research Institute (Grant 24265) to support this investigator-initiated study. *Support during past*
361 *36 months*: IC has received honoraria from the American Institute of Cancer Research and Genome
362 Quebec for her role as a grant reviewer. AK has been awarded research funding paid to her
363 institution from CIHR to support investigator-initiated research unrelated to this study. AK has
364 also received an honorarium from the Université Laval for her role as an epidemiology program
365 evaluator. JAT discloses her role as Committee Member on the *Dietary Reference Intakes for*
366 *Energy*, National Academies of Sciences, Engineering, and Medicine, 2023 (no remuneration
367 received). All other co-authors have no conflicts or competing interests to declare.

368 **Author contributions:**

369 *Concept and design*: HKN, KAK and IC; *Acquisition of data and/or lab analysis*: HKN, KAK,
370 RRL and IC; *Statistical analysis*: SA, FK and IC; *Interpretation of data*: All authors; *Drafting the*
371 *manuscript*: HKN and IC; *Critical revision of manuscript and important intellectual contribution*:
372 All authors; *Obtaining funding*: IC and KAK; *Administrative, technical or material support*: HKN,
373 KAK and IC; *Supervision*: IC; *Primary responsibility for the final content*: IC.

374

375

376

377 **FIGURE LEGENDS**

378

379 **Figure 1.** Participant recruitment in the STAR-Q validation study (2009-2011), Alberta, Canada.

380

381 **Figure 2.** Correlation between the degree of energy intake misreporting based on Goldberg
382 classifications (i.e., individual differences: $TEE_{STAR\text{-}Q}$ – rEI) and the degree of energy intake
383 misreporting based on doubly labeled water classifications (i.e., individual differences: TEE_{DLW}-
384 rEI) in kilocalories per day. **Figure 2A** shows the correlation among 40 men and **Figure 2B** shows
385 the correlation among 59 women in the STAR-Q validation study (2009-2011), Alberta, Canada.
386 AR, acceptable reporters; DLW, doubly labeled water; UR, under-reporters.

387

388 **Figure 3.** Sources of discrepancy between the Goldberg and doubly labeled water methods for
389 classifying energy intake misreporting. **Figure 3A** illustrates sources of discrepancy in men and
390 **Figure 3B** illustrates sources of discrepancy in women. The gold standard doubly labeled water
391 classifications of reporting are represented with shaded circles rEI:TEE_{DLW}. The Goldberg
392 classifications of misreporting, based on the ratio rEI:$TEE_{STAR\text{-}Q}$, are represented with open circles
393 (agreement between methods), squares (discrepant because of TEE estimation) or triangles
394 (discrepant because of different cut-offs). The gray solid line represents the lower cut-off for gold
395 standard doubly labeled water classifications (0.68 for men; 0.66 for women). The black solid line
396 represents the lower cut-off for Goldberg classifications (0.58 for men; 0.59 for women).

Table 1. Sociodemographic characteristics, energy expenditure and energy intake for 99 participants in the STAR-Q validation study (2009-2011), Alberta, Canada.

	Men, *n* = 40		Women, *n* = 59	
Sociodemographic characteristics	Mean (SD)	%	Mean (SD)	%
Age, years	50.6 (7.0)		46.2 (8.5)	
Body mass index, kg/m^2	25.9 (3.1)		23.5 (2.9)	
Education				
Less than post-secondary education		34		45
At least post-secondary education		51		45
Missing		15		10
Ethnicity				
Caucasian		73		85
Asian		3		5
Other		3		-
Missing		20		10
Paid or volunteer work		93		93
Energy intake and expenditure	Median [Q1, Q3]		Median [Q1, Q3]	
Energy intake, kcal/d	2036 [1627, 2634]		1676 [1317, 2033]	
Self-reported activity (STAR-Q), MET-h/d[1]	28.8 [24.6, 32.6]		30.6 [25.6, 39.5]	
$TEE_{DLW,}$ kcal/d	3306 [2843, 3727]		2604 [2226, 2854]	
$TEE_{STAR\text{-}Q}$, kcal/d	3620 [3178, 4019]		2889 [2420, 3387]	
BMR_{EST}, kcal/d	1866 [1744, 1924]		1362 [1331. 1438]	
PAL (TEE_{DLW}:BMR_{EST})	1.8 [1.6, 2.0]		1.9 [1.7, 2.1]	
PAL ($TEE_{STAR\text{-}Q}$:BMR_{EST})	2.0 [1.8, 2.1]		2.1 [1.8, 2.4]	

[1] The sum of MET-h/d from all types of physical activity including light, moderate, and vigorous intensity and sedentary activities (sleep is excluded).

BMR_{EST}, estimated basal metabolic rate; MET, metabolic equivalent of task; PAL, physical activity level; STAR-Q, Sedentary Time and Activity Reporting Questionnaire; TEE, total energy expenditure.

Table 2. Validity of the Goldberg method compared with the 'gold standard' doubly labeled water method for classifying energy intake misreporting, for 40 men and 59 women[1] in the STAR-Q validation study (2009-2011), Alberta, Canada.

Analysis	Source of TEE/PAL	Gender	Cut-offs		UR[2]	Mis-classified[2]	Sensitivity[1]	Specificity[1]	PPV[1]	NPV[1]	Accuracy of UR & AR[1]
			Lower	Upper	(%)	(%)					
Primary & Secondary	$DLW_{d=1}$	M	0.68	1.47	60	-	-	-	-	-	-
		W	0.66	1.51	56	-	-	-	-	-	-
Primary (PAL_{STAR-Q})[3]	STAR-Q	M	0.60	1.67	58	15	88 (68-97)	87 (61-98)	91 (72-99)	81 (56-94)	87 (72-95)
		W	0.61	1.65	58	25	79 (79-90)	69 (50-84)	77 (60-88)	72 (52-86)	75 (62-84)
Secondary ($PAL_{1.55}$)[3]	Assumed 1.55	M	0.60	1.66	35	33	54 (35-72)	93 (68-100)	93 (67-100)	56 (37-73)	69 (54-82)
		W	0.59	1.69	19	37	33 (20-51)	100 (85-100)	100 (70-100)	54 (40-67)	63 (50-74)
Sensitivity Analysis ($d=\infty$) [4]	$DLW_{d=\infty}$	M	0.90	1.11	85	-	-	-	-	-	-
		W	0.90	1.11	83	-	-	-	-	-	-
	STAR-Q	M	0.70	1.43	60	35	71 (54-83)	100 (29-100)	100 (84-100)	17 (4-46)	72 (56-84)
		W	0.73	1.36	76	32	78 (64-87)	29 (8-65)	88 (75-95)	15 (3-44)	71 (58-82)

[1] Calculations of sensitivity, specificity, PPV, NPV, and accuracy (expressed as percent with Agresti-Coull 95% confidence intervals) - by definition - exclude true overreporters identified with DLW. In our Primary and Secondary analyses (where d=1 in the DLW cut-off formula), one participant was a true over-reporter (therefore n=98; 39 men, 59 women). In Sensitivity analysis (where d=∞ in the DLW cut-off formula), seven participants were true over-reporters (therefore n=92; 36 men, 56 women). Doubly labeled water classifications of

misreporting, based on the ratio [rEI:TEE_{DLW}], served as the 'gold standard' (true) classification. Sensitivity is the proportion of true under-reporters who were correctly classified using the Goldberg method. Specificity is the proportion of true acceptable reporters who were correctly classified using the Goldberg method. Positive predictive value (PPV) is the proportion of Goldberg under-reporters (i.e., classified as under-reporters using the Goldberg method) that were true under-reporters. Negative predictive value (NPV) is the proportion of Goldberg acceptable reporters (i.e., classified as acceptable reporters using the Goldberg method) that were true acceptable reporters. Accuracy of UR & AR is the proportion of [true under-reporters (UR) + true acceptable reporters (AR)] classified correctly using the Goldberg method.

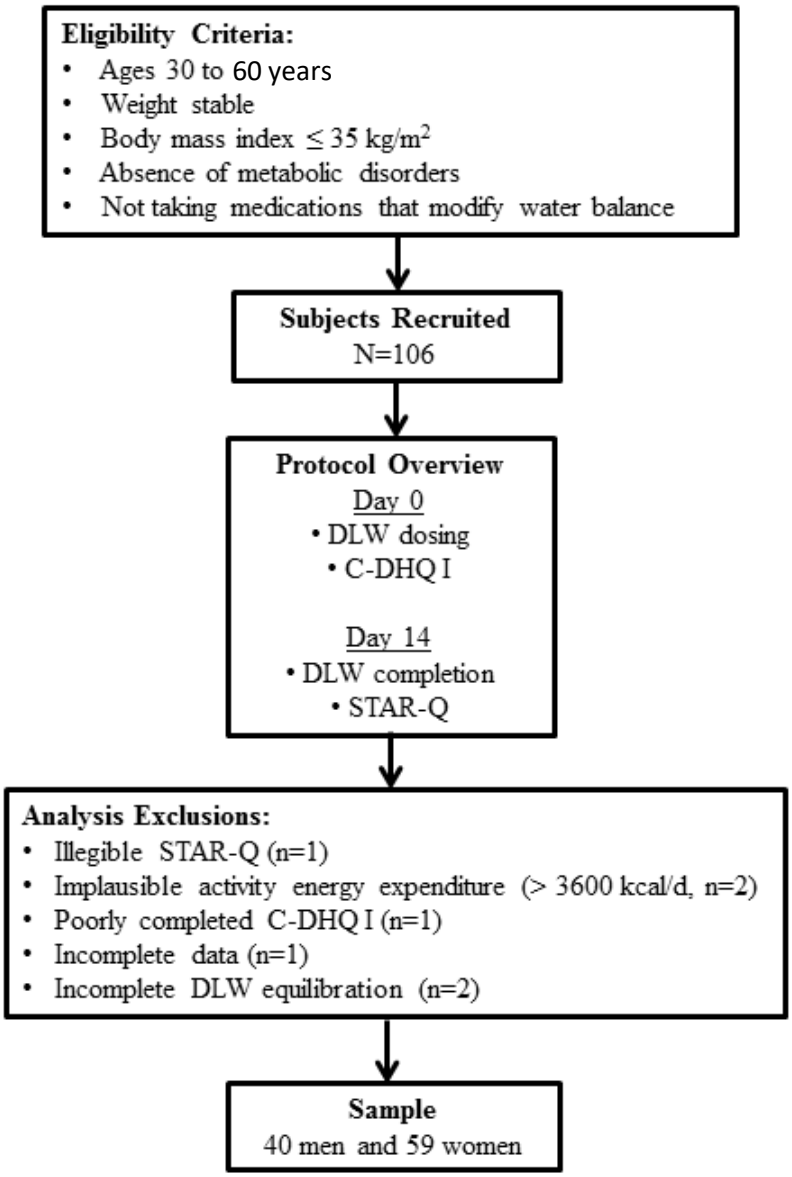
[2] Calculations based on the entire study population (n=99; 40 men, 59 women), including overreporters. Doubly labeled water classifications of misreporting, based on the ratio [rEI:TEE_{DLW}], served as the 'gold standard' (true) classification. UR was the proportion classified as an under-reporter. Misclassified was the proportion classified incorrectly by the Goldberg method.

[3] The formula for the Goldberg cut-offs included d=1 for the number of days of dietary assessment. We assigned d=1 because the C-DHQ I was administered once.

[4] The formula for Goldberg and DLW cut-offs included $d=\infty$, representing an infinite number of days of dietary assessment. We assigned $d=\infty$ for an extreme interpretation of 'd', as a sensitivity analysis, because the past year C-DHQ I is intended to measure habitual long-term diet.

AR, acceptable reporters; C-DHQ I, Canadian Diet History Questionnaire version I; DLW, doubly-labeled water; NPV, negative predictive value; PAL, Physical Activity Level; PPV, positive predictive value; rEI, self-reported energy intake; STAR-Q, Sedentary Time and Activity Reporting Questionnaire; TEE_{DLW}, total energy expenditure estimated from doubly labeled water; UR, underreporters.

Figure 1


Eligibility Criteria:
• Ages 30 to 60 years
• Weight stable
• Body mass index ≤ 35 kg/m2
• Absence of metabolic disorders
• Not taking medications that modify water balance
Subjects Recruited
N=106
Protocol Overview
Day 0
• DLW dosing
• C-DHQ I
Day 14
• DLW completion
• STAR-Q
Analysis Exclusions:
• Illegible STAR-Q (n=1)
• Implausible activity energy expenditure (> 3600 kcal/d, n=2)
• Poorly completed C-DHQ I (n=1)
• Incomplete data (n=1)
• Incomplete DLW equilibration (n=2)
Sample
40 men and 59 women

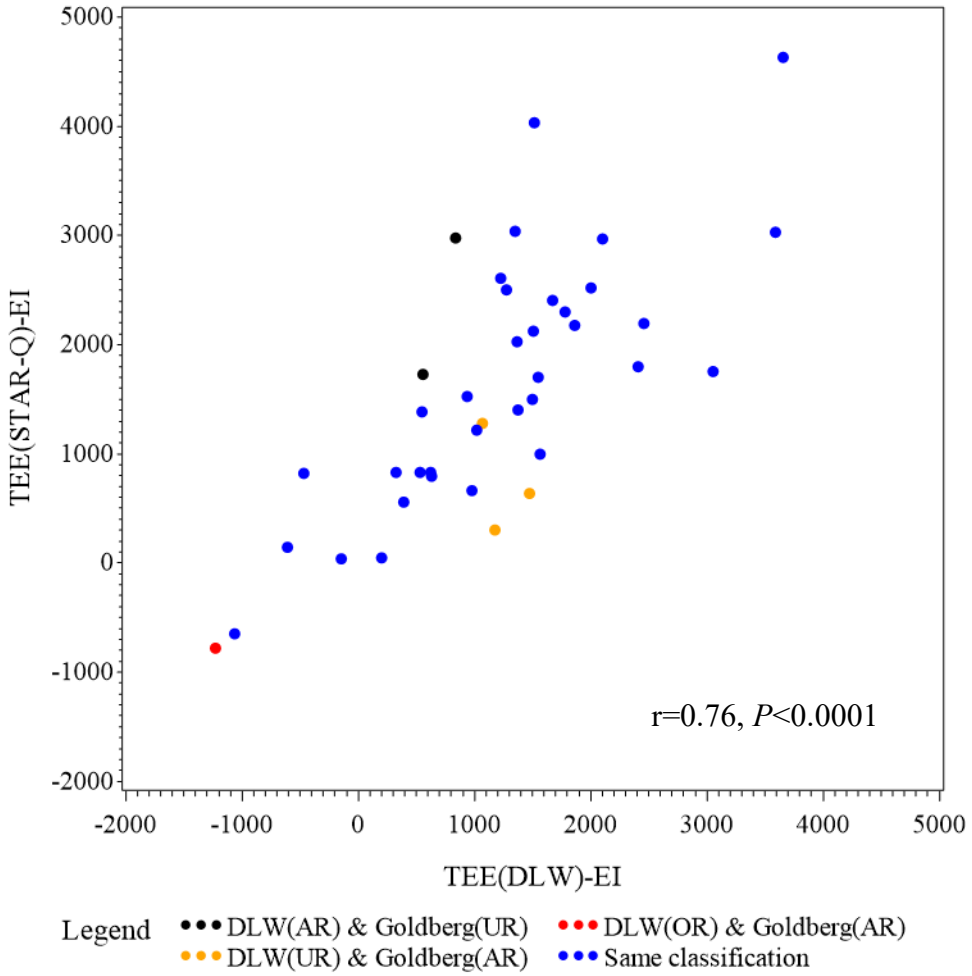


Figure 2A (men)

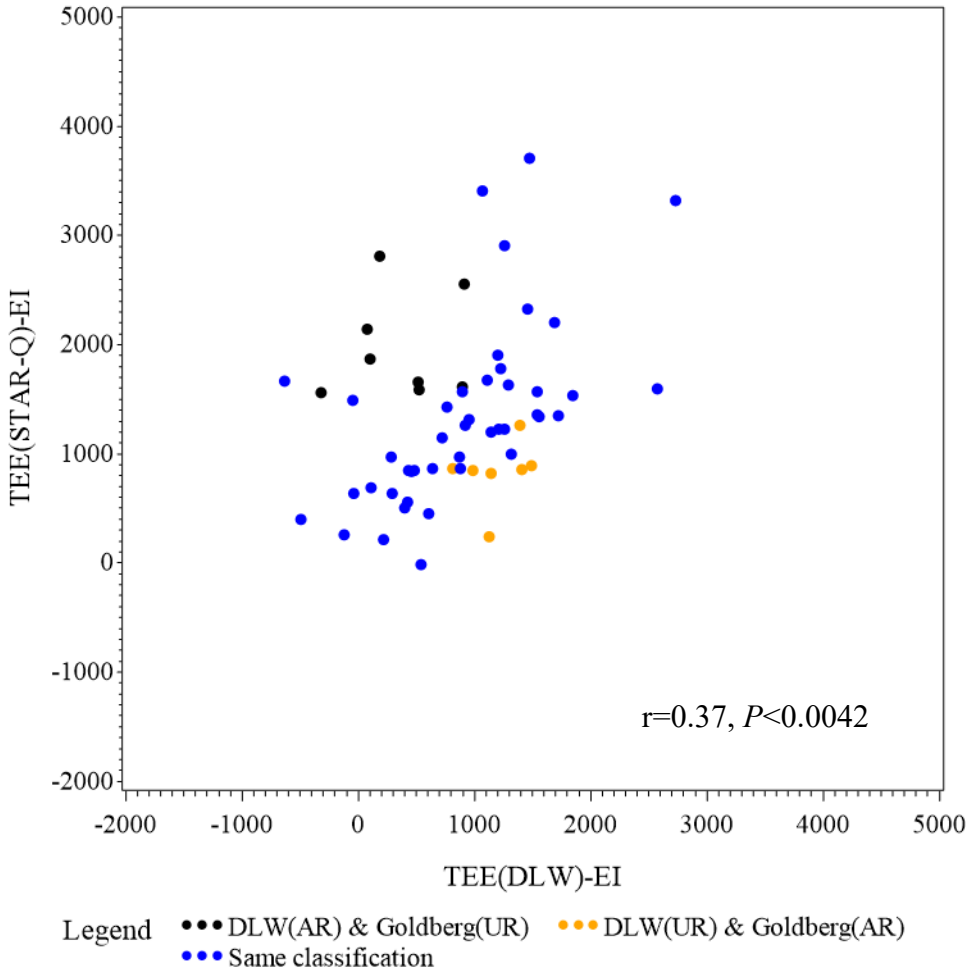


Figure 2B (women)

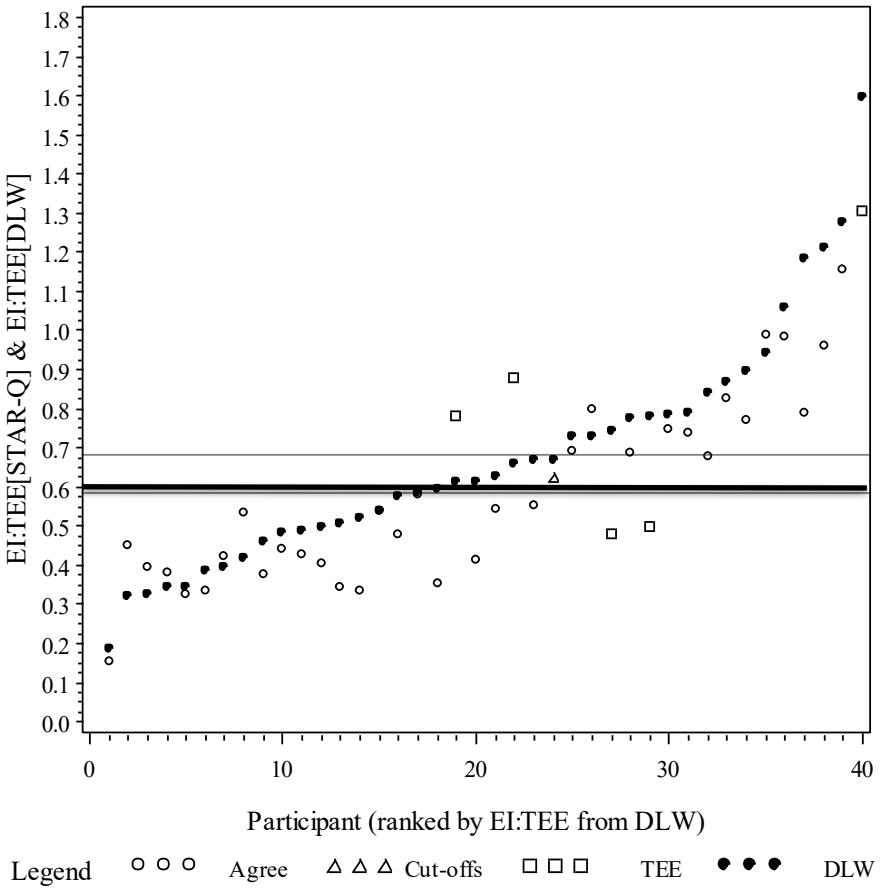


Figure 3A (men)

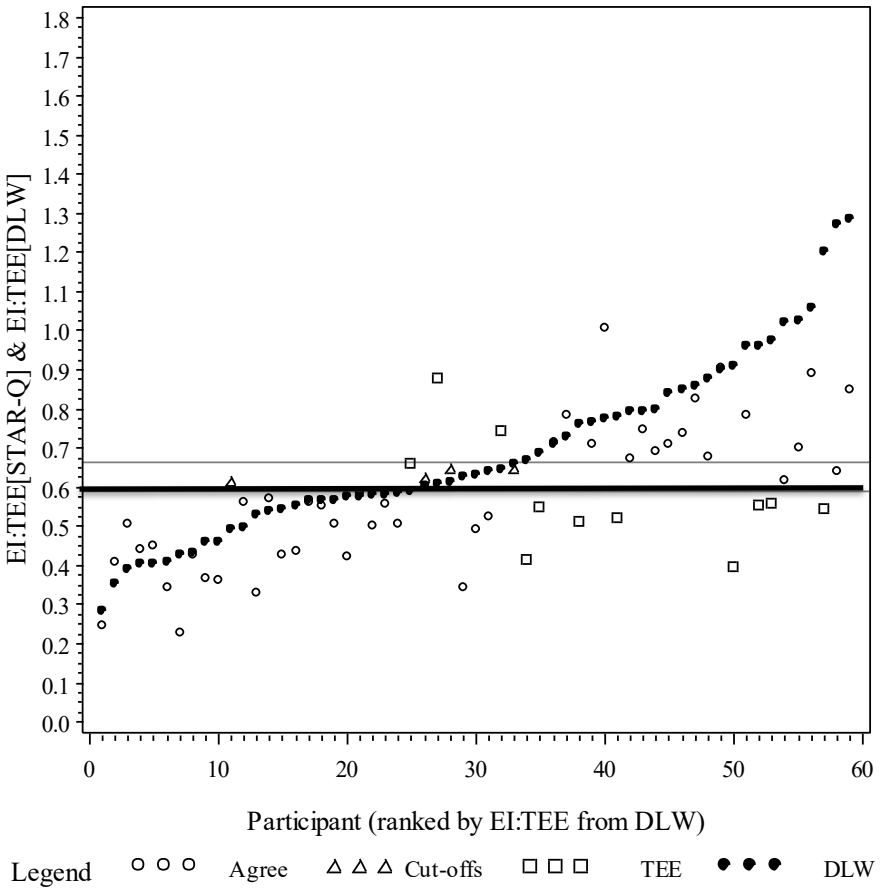


Figure 3B (women)

REFERENCES:

1. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* 2014;180(2):172–88.
2. Ravelli MN, Schoeller DA. Traditional Self-Reported Dietary Instruments Are Prone to Inaccuracies and New Approaches Are Needed. *Front Nutr* 2020;7(90):90.
3. Zhang X, Vasan S, Zheng C, et al. Calibrated Dietary Patterns and Cancer Risk in the Women's Health Initiative Cohorts. *American Journal of Epidemiology* 2025.
4. Prentice RL, Pettinger M, Tinker LF, et al. Regression calibration in nutritional epidemiology: example of fat density and total energy in relationship to postmenopausal breast cancer. *Am J Epidemiol* 2013;178(11):1663–72.
5. Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol* 2008;167(10):1247–59.
6. Prentice RL. Intake Biomarkers for Nutrition and Health: Review and Discussion of Methodology Issues. *Metabolites* 2024;14(5).
7. Neuhouser ML, Prentice RL, Tinker LF, et al. Enhancing Capacity for Food and Nutrient Intake Assessment in Population Sciences Research. *Annu Rev Public Health* 2022.
8. Banna JC, McCrory MA, Fialkowski MK, et al. Examining Plausibility of Self-Reported Energy Intake Data: Considerations for Method Selection. *Frontiers in Nutrition* 2017;4(45).

9. Black AE, Goldberg GR, Jebb SA, et al. 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(12):583–99.

10. National Academies of Sciences, Engineering, and Medicine 2023. *Dietary Reference Intakes for Energy*. Washington (DC): National Academies Press (US) https://doi.org/10.17226/26818.

11. 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 : journal of the International Association for the Study of Obesity* 2000;24(9):1119–30.

12. Neilson HK, Robson PJ, Friedenreich CM, et al. Estimating activity energy expenditure: how valid are physical activity questionnaires? *Am J Clin Nutr* 2008;87(2):279–91 [PMID: 18258615].

13. Energy and protein requirements. Report of a joint FAO/WHO/UNU Expert Consultation. *World Health Organ Tech Rep Ser* 1985;724:1–206.

14. Goldberg GR, Black AE, Jebb SA, 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(12):569–81.

15. Speakman JR, Pontzer H. Quantifying physical activity energy expenditure based on doubly labelled water and basal metabolism calorimetry: what are we actually measuring? *Curr Opin Clin Nutr Metab Care* 2023;26(5):401–8.

16. Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. *J Nutr* 1999;129(10):1765–8.

17. Tooze JA, Krebs-Smith SM, Troiano RP, et al. The accuracy of the Goldberg method for classifying misreporters of energy intake on a food frequency questionnaire and 24-h recalls: comparison with doubly labeled water. *Eur J Clin Nutr* 2012;66(5):569–76.

18. Csizmadi I, Neilson HK, Kopciuk KA, et al. The Sedentary Time and Activity Reporting Questionnaire (STAR-Q): reliability and validity against doubly-labeled water and 7-day activity diaries. *Am J Epidemiol* 2014;180(4):424–35 – Editor's Choice.

19. Šuc A, Einfalt L, Šarabon N, et al. Validity and reliability of self-reported methods for assessment of 24-h movement behaviours: a systematic review. *The international journal of behavioral nutrition and physical activity* 2024;21(1):83 [PMID: 39095778].

20. Neilson HK, Ullman R, Robson PJ, et al. Cognitive testing of the STAR-Q: insights in activity and sedentary time reporting. *J Phys Act Health* 2013;10(3):379–89 [PMID: 22820674].

21. Robson PJ, Solbak NM, Haig TR, et al. Design, methods and demographics from phase I of Alberta's Tomorrow Project cohort: a prospective cohort profile. *CMAJ Open* 2016;4(3):E515–e27.

22. Ye M, Robson PJ, Eurich DT, et al. Cohort Profile: Alberta's Tomorrow Project. *Int J Epidemiol* 2017;46(4):1097–8l.

23. Racette SB, Schoeller DA, Luke AH, et al. Relative dilution spaces of 2H- and 18O-labeled water in humans. *Am J Physiol* 1994;267(4 Pt 1):E585–90.

24. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr* 1985;39 Suppl 1:5–41.

25. Douketis JD, Paradis G, Keller H, et al. Canadian guidelines for body weight classification in adults: application in clinical practice to screen for overweight and obesity and to assess disease risk. *CMAJ* 2005;172(8):995–8.

26. Csizmadi I, Kahle L, Ullman R, et al. Adaptation and evaluation of the National Cancer Institute's Diet History Questionnaire and nutrient database for Canadian populations. *Public Health Nutr* 2007;10(1):88–96 [PMID: 17212847].

27. Csizmadi I, Boucher BA, Lo Siou G, et al. Using national dietary intake data to evaluate and adapt the US Diet History Questionnaire: the stepwise tailoring of an FFQ for Canadian use. *Public Health Nutrition* 2016;19(18):3247–55.

28. Mickle AT, Brenner DR, Beattie T, et al. The Dietary Inflammatory Index® and Alternative Healthy Eating Index 2010 in relation to leucocyte telomere length in postmenopausal women: a cross-sectional study. *J Nutr Sci* 2019;8:e35.

29. Biel RK, Csizmadi I, Cook LS, et al. Risk of endometrial cancer in relation to individual nutrients from diet and supplements. *Public Health Nutr* 2011;14(11):1948–60.

30. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;32(9 Suppl):S498–504.

31. Besson H, Brage S, Jakes RW, et al. Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. *Am J Clin Nutr* 2010;91(1):106–14.

32. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc* 1993;25(1):71–80.

33. Herrmann SD, Willis EA, Ainsworth BE, et al. 2024 Adult Compendium of Physical Activities: A third update of the energy costs of human activities. *J Sport Health Sci* 2024;13(1):6–12.

34. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health* 2017;5:307.

35. Fowke JH, Schlundt D, Gong Y, et al. Impact of season of food frequency questionnaire administration on dietary reporting. *Ann Epidemiol* 2004;14(10):778–85.

36. 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* 2000;54(5):395–404.

37. Livingstone MB, Robson PJ, Black AE, et al. An evaluation of the sensitivity and specificity of energy expenditure measured by heart rate and the Goldberg cut-off for energy intake: basal metabolic rate for identifying mis-reporting of energy intake by adults and children: a retrospective analysis. *Eur J Clin Nutr* 2003;57(3):455–63.

38. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* 2003;158(1):1–13.

39. Tooze JA, Schoeller DA, Subar AF, et al. Total daily energy expenditure among middle-aged men and women: the OPEN Study. *Am J Clin Nutr* 2007;86(2):382–7.

40. Jessri M, Lou WY, L'Abbé MR. Evaluation of different methods to handle misreporting in obesity research: evidence from the Canadian national nutrition survey. *Br J Nutr* 2016;115(1):147–59.

41. Tooze JA, Freedman LS, Carroll RJ, et al. The impact of stratification by implausible energy reporting status on estimates of diet-health relationships. *Biometrical Journal* 2016;58(6):1538–51.

42. Solbak NM, Al Rajabi A, Akawung AK, et al. Strategies to Address Misestimation of Energy Intake Based on Self-Report Dietary Consumption in Examining Associations Between Dietary Patterns and Cancer Risk. *Nutrients* 2019;11(11 [PMID: 31683814]).

43. Lo Siou G, Akawung AK, Solbak NM, et al. The effect of different methods to identify, and scenarios used to address energy intake misestimation on dietary patterns derived by cluster analysis. *Nutr J* 2021;20(1):42 [PMID: 33964947].

44. Yamamoto N, Ejima K, Zoh RS, et al. Bias in nutrition-health associations is not eliminated by excluding extreme reporters in empirical or simulation studies. *Elife* 2023;12.

45. Ejima K, Brown AW, Schoeller DA, et al. Does exclusion of extreme reporters of energy intake (the "Goldberg cutoffs") reliably reduce or eliminate bias in nutrition studies? Analysis with illustrative associations of energy intake with health outcomes. *The American Journal of Clinical Nutrition* 2019;110(5):1231–9.

46. Herrmann SD, Willis EA, Ainsworth BE. The 2024 Compendium of Physical Activities and its expansion. *J Sport Health Sci* 2024;13(1):1–2.

47. Matthews CE, Kozey Keadle S, Moore SC, et al. Measurement of Active and Sedentary Behavior in Context of Large Epidemiologic Studies. *Med Sci Sports Exerc* 2018;50(2):266–76.

48. Matthews CE, Patel S, Saint-Maurice P, et al. Physical Activity Levels (PAL) in United States Adults - 2019. *Med Sci Sports Exerc* 2022.

49. Freedman LS, Midthune D, Arab L, et al. Combining a Food Frequency Questionnaire With 24-Hour Recalls to Increase the Precision of Estimation of Usual Dietary Intakes—Evidence From the Validation Studies Pooling Project. *American Journal of Epidemiology* 2018;187(10):2227–32.

50. Sampson JN, Matthews CE, Freedman L, et al. Methods to Assess Measurement Error in Questionnaires of Sedentary Behavior. *J Appl Stat* 2016;43(9):1706–21.

51. Hibbing PR, Shook RP, Panda S, et al. Predicting energy intake with an accelerometer-based intake-balance method. *Br J Nutr* 2022:1–23.

52. Hibbing PR, Welk GJ, Ries D, et al. Criterion validity of wrist accelerometry for assessing energy intake via the intake-balance technique. *International Journal of Behavioral Nutrition and Physical Activity* 2023;20(1):115.

53. Shook RP, Yeh H-W, Welk GJ, et al. Commercial Devices Provide Estimates of Energy Balance with Varying Degrees of Validity in Free-Living Adults. *The Journal of Nutrition* 2021;152(2):630–8.

54. Gebel K, Ding D. Using Commercially Available Measurement Devices for the Intake-Balance Method to Estimate Energy Intake: Work in Progress. *The Journal of Nutrition* 2022;152(2):373–4.

55. Manini TM. Energy expenditure and aging. *Ageing Res Rev* 2010;9(1):1–11.

56. Krebs-Smith SM, Graubard BI, Kahle LL, et al. Low energy reporters vs others: a comparison of reported food intakes. *Eur J Clin Nutr* 2000;54(4):281–7.

57. Lafay L, Mennen L, Basdevant A, et al. Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Santé

(FLVS) study. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* 2000;24(11):1500–6.

58. Prentice RL, Aragaki AK, Zheng C, et al. Energy intake is associated with dietary macronutrient densities: inversely with protein and monounsaturated fat and positively with polyunsaturated fat and carbohydrate among postmenopausal females. *Am J Clin Nutr* 2025.

59. Storz MA, Ronco AL. Energy Underreporting in Low-Calorie and Carbohydrate-Restrictive Diets: Epidemiological Considerations. *Current Developments in Nutrition* 2025;9(10).

60. Porter J, Ward LC, Nguo K, et al. Development and validation of age-specific predictive equations for total energy expenditure and physical activity levels for older adults. *The American Journal of Clinical Nutrition* 2024;119(5):1111–21.

61. Wong WW, Speakman JR, Ainslie PN, et al. Decline in Isotope Dilution Space Ratio Above Age 60 Could Affect Energy Estimates Using the Doubly Labeled Water Method. *J Nutr* 2024;154(12):3824–31.

62. Prentice RL, Mossavar-Rahmani Y, Huang Y, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Am J Epidemiol* 2011;174(5):591–603.

63. Pryer JA, Vrijheid M, Nichols R, et al. Who are the 'low energy reporters' in the dietary and nutritional survey of British adults? *Int J Epidemiol* 1997;26(1):146–54.

64. Prentice AM, Black AE, Coward WA, et al. Energy expenditure in overweight and obese adults in affluent societies: an analysis of 319 doubly-labelled water measurements. *Eur J Clin Nutr* 1996;50(2):93–7.

65. Karelis AD, Lavoie ME, Fontaine J, et al. Anthropometric, metabolic, dietary and psychosocial profiles of underreporters of energy intake: a doubly labeled water study among overweight/obese postmenopausal women--a Montreal Ottawa New Emerging Team study. *Eur J Clin Nutr* 2010;64(1):68–74.