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

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

Author Affiliations:

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

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

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

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

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

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

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

Abstract: 299 words

Manuscript: 4,098 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 and Oncology, 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

BACKGROUND: Dietary intake misreporting is widely acknowledged in nutritional epidemiology. The Goldberg method has been proposed as an option for identifying under-reporters of energy intake (EI) in large-scale epidemiologic studies. However, its implementation remains limited by challenges associated with estimating physical activity levels (PAL), a critical component.

OBJECTIVE: To quantify the accuracy of the Goldberg method to classify EI misreporting using Sedentary Time and Activity Reporting Questionnaire (STAR-Q)-derived $PAL_{STAR\text{-}Q}$ compared with doubly-labeled water (DLW)-derived total energy expenditure (TEE_{DLW}).

DESIGN: Between 2009 and 2011, 99 weight-stable men and women (mean [SD]: 48 [8] years) completed a two-week DLW protocol, the Canadian Diet History Questionnaire I, and the STAR-Q, a comprehensive past-month activity questionnaire. TEE_{DLW} was the criterion measure of EI to determine the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy (proportion correctly classified as under-reporters or acceptable reporters) of Goldberg cut-points based on the $PAL_{STAR\text{-}Q}$. A global PAL of 1.55 frequently used in the original Goldberg method was also studied.

RESULTS: The Goldberg method using $PAL_{STAR\text{-}Q}$ classified 58% of men and women as under-reporters, compared with 60% of men and 56% of women identified with TEE_{DLW}. Among men, values for sensitivity, specificity, PPV, NPV and accuracy of the Goldberg method and $PAL_{STAR\text{-}Q}$ were estimated as: 88%, 87%, 91%, 81%, and 87%, respectively. Among women, these values were: 79%, 69%, 77%, 72%, and 75%, respectively. Substituting a PAL of 1.55 for the $PAL_{STAR\text{-}Q}$ reduced the proportion of men and women classified as under-reporters to 35% and 19%, respectively, lowered sensitivity to 54% and 33%, and increased specificity to 93% and 100%.

CONCLUSIONS: Population-specific validation sub-studies using the Goldberg method and questionnaire-derived PALs can be informative for understanding EI misreporting to ultimately improve estimates of diet-disease associations in large-scale epidemiologic 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 inadequacies in nutritional surveys and potentially biases estimates of diet-disease associations in etiologic studies impeding progress in nutritional epidemiology (1). Statistical approaches that mitigate the effects of dietary intake misreporting have been suggested, but have not been routinely implemented, indicating a need for further study (2). Prerequisites for undertaking measurement error correction include having knowledge of *true* dietary intake and an understanding of the underlying systematic errors associated with self-report (3). Ideally, misreporting could be detected using 'gold standard' objective biomarkers of dietary intake; however, these methods are limited in number and often too labor-intensive and costly to implement in large-scale epidemiologic studies (4). A practical alternative approach specifically for EI, may be to classify individuals according to the plausibility of their reported EI (rEI), compared with a method validated to estimate energy requirements (5).

The Goldberg method (6) was originally proposed as a strategy for identifying EI misreporting in community-dwelling populations and 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 and TEE/BMR approximates PAL (7). Further, the method proposes estimating cut-point boundaries, or 95% confidence intervals (95% CI), for the ratio EI:TEE which accounts for the variability of PAL and additional energy-related factors. From this we can infer that in weight-stable individuals, a habitual rEI below or above these boundaries represent under and over-reported EI.

While conceptually tenable, the method is logistically challenged by the difficulty associated with the estimation of PAL in free-living populations (8). As such, in the absence of population-specific objective measures a low activity PAL of 1.55 has frequently been assumed based on World Health Organization predictions of energy requirements (9) and the original Goldberg paper (10). Indeed, in a study of middle-aged men and women Tooze *et al*. (11) reported that the Goldberg method and a PAL of 1.55 had acceptable accuracy compared with doubly-labelled water (DLW), the 'gold standard' for estimating TEE (12, 13), although alternative PALs were not considered. In contrast, Black *et al*. (14) demonstrated that the proportion of misclassified men and women could be reduced by assigning low, medium, or high group-level PALs (e.g., 1.55, 1.75 or 1.95, respectively), based on DLW results. More recently, a large survey of Australian adults Grech et al (15) examined the associations between EI underreporting, diet quality and socioeconomic status using the Goldberg method with a PAL of 1.55 versus prediction equations that estimated individual TEE levels and reported comparable results. Nonetheless, based on historical DLW studies (16) and the recent Dietary Reference Intake (DRIs) for Energy report (median adult PAL = 1.68) (7), a global PAL of 1.55 may be unrealistically low for substantial segments of the population leading to active EI under-reporters misclassified as acceptable.

To our knowledge, the Goldberg method has not been evaluated using PALs derived from individual self-reports of physical activity despite their appeal given the potential to capture habitual behaviors of greatest relevance in chronic disease etiologic research. 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 global activity questionnnaire that we previously validated against DLW (17). In our primary analysis, we determined the accuracy of the Goldberg method with

STAR-Q-derived PALs compared with DLW-derived TEE_{DLW} to classify EI reporting status on a food frequency questionnaire (FFQ). In secondary analyses we assumed the classic Goldberg PAL of 1.55. In addition, in sensitivity analyses we studied the effect of modifying the FFQ variability assumption in the estimation of Goldberg cut-points as previously studied (11).

METHODS

Participants in this analysis were those enrolled in the original STAR-Q DLW validation study (17). The STAR-Q was conceptualized to capture 24-hour movement behaviours (18). It was therefore uniquely designed to ascertain self-reported time spent in sleep, sedentary behavior, and all types and intensities of physical activity (e.g., light, moderate and vigorous) within all domains (e.g., occupation, leisure, household, and transportation) (17, 19).

Subject Eligibility and Ethics

Participant recruitment has previously been described (17). 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 (20, 21). ATP participants who met eligibility criteria for the DLW validation study – body mass index (BMI) ≤ 35, weight stabile (± 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; the lower age limit for the urban Calgary sample was 30 years, and for ATP participants it was 35 years.

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 the ethical standards outlined by the two governing bodies.

Doubly-labeled water study

Details of the 14-day DLW protocol have been published (17). 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). We collected post-dose saliva samples at 3 and 4 hours to measure 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. The post-dose saliva sample collection protocol was followed as on Day 0. To measure the decline in isotope enrichment, we batch analyzed samples 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). We performed data processing using IonVantage (2012) software for Isoprime (Cheadle, United Kingdom).

We calculated TEE_{DLW} in kilocalories per day using a modified Weir equation (22) and an assumed respiratory quotient of 0.85. We estimated BMR using the Schofield Equation (23) and assumed 10% for the thermic effect of food (12), to derive individual level PALs from DLW as follows:

$$PAL_{DLW} = 0.9TEE_{DLW} / BMR_{EST}$$

122

123 **Health and lifestyle measurements**

124 Height and weight were measured (to the nearest 0.1 cm and 0.1 kg, respectively) on Days
125 0 and 14 of the DLW protocol using standard procedures (24). The TBF-310 Tanita Body
126 Composition Analyzer and Scale (TheCompetitiveEdge.com, Preston, Washington) was used to
127 measure body weight. BMI was determined as body weight divided by height-squared (kg/m^2).

128 The past-year Canadian Diet History Questionnaire I (C-DHQ I) was administered once,
129 on Day 0 of the DLW protocol to estimate reported rEI. The C-DHQ I is a comprehensive FFQ
130 that assesses usual dietary intake over the previous 12 months, modified for Canadian populations
131 from the U.S. National Cancer Institute's DHQ (25, 26). Several large epidemiologic studies use
132 the C-DHQ I in Canada including the ATP (20, 21, 27, 28).

133 The STAR-Q was completed on Day 14 of the DLW study to estimate energy expenditure
134 from all domains, types and intensities of past-month physical activity, sedentary behavior, and
135 sleep. STAR-Q-derived TEE and activity energy expenditure were highly reproducible over three
136 months, with intraclass correlation coefficients of 0.84 and 0.73, respectively (17). Compared with
137 DLW, the STAR-Q demonstrated moderate validity for ranking individuals by TEE and activity
138 energy expenditure (Spearman's rho=0.53 and 0.40, respectively) (17).

139 To derive TEE from the STAR-Q, we first assigned an activity code and metabolic
140 equivalents of task (MET) value (29) to each reported activity. If the total time reported for
141 physical activity, sedentary behavior, and sleep was less than 24 hours/day, then we assigned a
142 MET of 1.2 for unaccounted time. The reported frequency, duration and assigned MET values
143 were used to estimate energy expenditure for each activity in units of MET-hours/day, which were

144 then summed to derive the total average MET-hours/day. The STAR-Q-derived TEE ($TEE_{STAR\text{-}Q}$)
145 was then estimated as follows:
146
147 $$TEE_{STAR\text{-}Q} = [[0.9\ MET \times hours\ of\ sleep/d] + [MET\text{-}hours/d \times weight\ (kg)]]*1.1$$
148
149 – where $TEE_{STAR\text{-}Q}$ from reported hours of sleep per day on the STAR-Q was multiplied by 0.9
150 METs (29, 30) and then added to the total average MET-hours per day in awake time multiplied
151 by baseline body weight. We applied the energy equivalence of 1 MET = 1 kcal/kg/hr to estimate
152 the energy cost of activities and added 10% for the thermic effect of food (12).
153 The calculation for individual PALs derived from self-report was as follows:
154
155 $$PAL_{STAR\text{-}Q} = 0.9\ TEE_{STAR\text{-}Q}/BMR_{EST}$$
156
157 – where BMR_{EST} was estimated using the Schofield equation based on age, sex, weight and
158 height (23).
159
160 **Classification of misreporting**
161 For comparability with a previous DLW study, we applied methods described by Tooze *et*
162 *al*. (11) to classify misreporting. Each participant was classified as an under-reporter (UR),
163 acceptable reporter (AR), or over-reporter (OR) depending on his or her individual ratio [rEI:TEE]
164 relative to group-level cut-points described below (31). Participants with a ratio less than the lower
165 cut-point were deemed URs; participants with ratios greater than the upper cut-point were deemed

ORs; and participants with ratios between the lower and upper cut-points were deemed ARs. Cut-points were calculated separately for $TEE_{STAR\text{-}Q}$ and TEE_{DLW}.

In our primary analysis to evaluate the Goldberg method and $PAL_{STAR\text{-}Q}$ for classifying individual misreporting, we calculated cut-points for the log of rEI:$TEE_{STAR\text{-}Q}$ as follows:

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

– where CV_{EI} was the within-subject coefficient of variation for EI, *d* was the number of days of dietary assessment, CV_{BMR} was the precision of estimated versus measured BMR, and CV_{PAL} was the total variation in PAL. We assigned one day for the number of days of dietary assessment (*d*=1) since the C-DHQ I was only administered once. The CV_{EI} was assumed to be 18.6% for men and 19.8% for women based on repeatability of the US DHQ in a previous study (11) and CV_{BMR} was assumed to be 8.5% for men and women as previously suggested (31). The CV_{PALs} calculated in our three-month STAR-Q repeatability study were 17.5% for men and 15.1% for women (17).

In secondary analyses, we applied the classic Goldberg method (10) assuming a constant PAL of 1.55 for participants (using [rEI:$TEE_{1.55}$] derived from [rEI:(BMR_{EST}*1.55)]). Goldberg cut-points were calculated as described above for rEI:$TEE_{STAR\text{-}Q}$, except CV_{PAL} was assumed to be 15% for both men and women as previously reported (31).

To derive gold standard classifications of misreporting, we calculated DLW cut-points based on the log of the ratio [rEI:TEE_{DLW}] as:

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

190 – where CV_{EI} was the within-subject coefficient of variation for EI, *d* was the number of days of
191 dietary assessment, and CV_{DLW} was the within-subject coefficient of variation for TEE_{DLW}. We
192 assumed CV_{rEI} to be 18.6% for men and 19.8% for women and CV_{DLW} to be 5.1% for both men
193 and women as previously reported (11).

194 Finally in sensitivity analyses, for comparison with a previous study (11), we substituted
195 $d=\infty$ for $d=1$ for dietary assessment in each of the cut-point equations (above). This alternate
196 estimate of '*d*' reflects the intended use of the past year C-DHQ I to ascertain habitual long-term
197 diet. Given a cut-point denominator of $d=\infty$, the CV^2_{EI} term was then eliminated.

198

199 **Statistics**

200 Means (SD) were determined for continuous sociodemographic characteristics,
201 frequencies and percentages for categorical characteristics, and medians (Q1, Q3) for EI and
202 energy expenditure. In all analyses, DLW classifications were assumed to be the 'gold standard'
203 or *true* classifications. All analyses were stratified by gender. To appraise the validity of Goldberg
204 classifications using $PAL_{STAR\text{-}Q}$, and $PAL_{1.55}$, we estimated the proportion of participants correctly
205 classified and misclassified in terms of UR, AR and OR. Sensitivity, specificity, positive predictive
206 value (PPV), negative predictive value (NPV), and the accuracy of UR and AR classifications
207 excluded the sole *true* OR (n=1) identified by TEE_{DLW}. Sensitivity was the proportion of true URs
208 correctly classified using the Goldberg method. Specificity was the proportion of true ARs
209 correctly classified using the Goldberg method. PPV was the proportion of Goldberg URs (i.e.,
210 classified as such using the Goldberg method) that were true URs, interpreted as the probability of
211 being a true UR if classified an UR by the Goldberg method. NPV was the proportion of Goldberg
212 ARs that were true ARs, interpreted as the probability of being a true AR if classified as an AR by

213 the Goldberg method. Accuracy of UR and AR was the proportion of [true URs + true ARs]
214 classified correctly with the Goldberg method.

215 To appraise the validity of $PAL_{STAR\text{-}Q}$ for estimating the magnitude of misreporting (UR
216 and OR) in kilocalories per day, we compared the within-person differences of $TEE_{STAR\text{-}Q}$-rEI
217 versus TEE_{DLW}-rEI using Pearson correlation coefficients and the Wilcoxon signed-rank test.

218 Furthermore, we used graphical illustrations described by Tooze *et al*. (11) to examine the
219 magnitude of discrepancies relative to the gold standard ratio rEI:TEE_{DLW}. If an individual-level
220 rEI:$TEE_{STAR\text{-}Q}$ ratio fell between the Goldberg ($TEE_{STAR\text{-}Q}$) and TEE_{DLW} lower (or upper) cut-
221 points, we attributed the source of discrepancy to differences in cut-point calculations. Otherwise,
222 we attributed the discrepancy to differences in TEE estimation (STAR-Q vs DLW). In addition,
223 we estimated the contribution of these discrepancies to the resulting sensitivity and specificity
224 attributes of the Goldberg method ($TEE_{STAR\text{-}Q}$).

225 All analyses were done using SAS Institute Software, Gary, NC (Version 9.4). We
226 considered a *P*-value < 0.05 statistically significance.

227

228 **RESULTS**

229 Participants aged 30-60 years (n=106) were recruited to the original DLW study between
230 July 2009 and July 2010 (17). Seven participants were excluded from the present analysis for
231 reasons outlined in **Figure 1,** leaving 99 participants of whom 60% were women. Compared with
232 men, women were somewhat younger (mean: 46.2 vs 50.6 years of age), had a lower BMI (mean:
233 23.5 vs 25.9 kg/m^2), and were less likely to complete post-secondary education (45% vs 51%)
234 **Table 1**. Most of the participants were Caucasian and had paid or volunteer work. Median EI,
235 TEE_{DLW}, $TEE_{STAR\text{-}Q}$, and BMR_{EST} were lower for women (1676 kcal, 2604 kcal, 2889 kcal and

236 1362 kcal, respectively) compared with men (2036 kcal, 3306 kcal, 3620 kcal, and 1866 kcal,
237 respectively). Median values for PAL_{DLW} and $PAL_{STAR\text{-}Q}$, were 1.8 and 2.0 for men and, 1.9 and
238 2.1 , for women (Table 1), consistent with 'active' and 'very active' PALs (7).

239 Results from the primary, secondary and sensitivity analyses are presented in **Table 2**. In
240 the primary analysis, 60% of men and 56% of women were classified as URs by TEE_{DLW}, and
241 58% of men and women by $TEE_{STAR\text{-}Q}$. The $TEE_{STAR\text{-}Q}$ misclassified fewer men (15%) than
242 women (25%), resulting in sensitivity, specificity, PPV, NPV and accuracy >80% among men,
243 whereas for women these estimates were <80%. In secondary analyses, $PAL_{1.55}$ resulted in fewer
244 participants classified as URs (35% and 19% for men and women, respectively) than with TEE_{DLW}
245 and $TEE_{STAR\text{-}Q}$ which led to high specificity and PPVs (both 93% for men and 100% for women),
246 but low sensitivity and NPVs (all <60% for men and women).

247 In sensitivity analyses, where we assumed $d=\infty$ for the FFQ, a greater proportion of men
248 and women were misclassified than with $d=1$. Seven were deemed to be true OR in contrast to 1
249 true OR when $d=1$, attributed to narrower cut-points when $d=\infty$. Otherwise, when $d=\infty$, results
250 were generally similar (or less favorable) to assumed $d=1$ except for specificity and PPV, which
251 were 100% among men.

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

259 men and 7 women) while the magnitude of EI under-reporting was typically smaller for
260 misclassified ARs (<1,000 kcal/d approximately; 2 men and 8 women).

261 In **Figures 3A and 3B**, we show whether discrepant classifications arose from different
262 cut-points or from different TEE estimation methods. Among 40 men (**Figure 3A**), five
263 misclassifications were due to TEE estimation (one classified as an OR), and one was due to cut-
264 point estimation. Among 59 women (**Figure 3B**), more misclassifications were due to differences
265 in TEE estimation than to cut-off estimation. Among *true* URs misclassified as ARs using the
266 $PAL_{STAR\text{-}Q}$ method (reflecting sensitivity), 2 of 3 men and 3 of 7 women were misclassified owing
267 to TEE estimation. Among true ARs misclassified as URs using the $PAL_{STAR\text{-}Q}$ method (reflecting
268 specificity), all misclassifications (2 men and 8 women) were due to differences in TEE estimation.

269

270 **DISCUSSION**

271 Using DLW, the 'gold-standard' reference biomarker for TEE and unbiased proxy for
272 habitual EI in weight-stable individuals (2), we determined the accuracy of STAR-Q-derived PALs
273 and the Goldberg method to classify the reporting status of FFQ-derived EI in community dwelling
274 middle-aged men and women. We found that the $PAL_{STAR\text{-}Q}$ generally improved the sensitivity
275 and overall accuracy of the Goldberg method for discriminating between URs and ARs compared
276 with the Goldberg method based on a global PAL of 1.55. We also identified gender-based
277 differences in the performance of the Goldberg-$PAL_{STAR\text{-}Q}$. In men, the method discriminated
278 between UR and AR with an overall high level of accuracy at 87%, while in women it exhibited
279 fair performance, with an overall classification accuracy of 75%.

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

282 approaches to identify the optimal PAL for estimating EI misreporting (11, 14, 32). Two of these
283 studies examined EI misreporting on food records and/or diet histories and found under-reporting
284 in 14% (32) and 34% (14) of adult participants. Similar to our results these studies reported high
285 levels of specificity but relatively low sensitivity for identifying URs when a global PAL of 1.55
286 was assumed. Additionally, Black (14) leveraged the strength of a large database comprised of 21
287 DLW studies (n=429) to test the effect of various global PALs (from 1.55 to 1.95) and reported
288 incremental increases in sensitivity, with increasing PALs; however, this improvement coincided
289 with decreases in specificity. In a small sample of adults (n=14), Livingstone *et al* (32) tested the
290 performance of gender-specific PALs to detect URs based on individual heart-rate monitors
291 matched to low, medium and high activity categories of PAL (9). Compared with $PAL_{1.55}$,
292 sensitivity increased substantially from 50% to 100% with negligible decreases in specificity.
293 Subsequently, Tooze *et al* (11) reported favorable results for the Goldberg method with a $PAL_{1.55}$
294 to detect DLW-classified URs on 24-hour dietary recalls (~20%) and an FFQ (~50% UR) in the
295 Observing Protein and Energy Nutrition (OPEN) Study (n=451). Specificity was 88% for the FFQ
296 in both men and women. The sensitivity to detect URs was high for the FFQ at 96% when usual
297 intake was assumed for the FFQ ($d=\infty$), but substantially lower when OPEN study FFQ within-
298 person variability was applied (72% sensitivity for men and 62% for women).

299 The results reported by Livingstone *et al* (32) and Black (14) are consistent with our results
300 in supporting the premise that a global $PAL_{1.55}$ may be underestimating the activity levels of some
301 adult populations – and hence, more individualized PALs may improve the overall accuracy of the
302 Goldberg method to identify EI under-reporting. While our findings contrast with those of Tooze
303 *et al* (11), it is worth noting that the prevalence of true UR (50%) was lower than in our population,
304 which affects PPV and NPV values. In addition, participants in the OPEN study were on average

305 less active and slightly older than in the STAR-Q validation study. Median values for DLW-
306 derived TEE and PAL were 2,813 kcal and 1.69 for men and 2,283 kcal and 1.75 for women in
307 OPEN (33, 34) and 3,306 kcal and 1.8 for men and 2604 kcal and 1.9 for women in the STAR-Q
308 study, with ages ranging from 40 to 69 years in OPEN (33, 34) versus 30 to 60 years in the STAR-
309 Q study.

310 Statistical approaches have been explored with respect to handling URs once identified,
311 e.g., exclusion of misreporters, stratification by reporting status and adjustment for reporting status
312 in regression analyses (35-39). While the optimal approach may be study-specific, based on study
313 design, method of data collection and population characteristics, there is consensus that exclusion
314 of misreporters is undesirable given the loss of statistical power (36) and loss of valuable
315 information inherent in the characteristics of those excluded (35). As such, it is essential that
316 investigators understand the potential impact of measurement error on study results prior to
317 selecting the approach with the greatest potential to mitigate biasing effects of misreporting (39).

318 While the STAR-Q was designed to capture a comprehensive profile of daily activities
319 over a 24-hour period, 50% of true URs misclassified as ARs and all true ARs misclassified as
320 URs, were likely misclassified owing to errors in the estimation of $TEE_{STAR\text{-}Q}$ **(Figure 3a and 3b).**
321 Hence, both sensitivity and specificity which are important characteristics in the assessment of
322 exposure validity in epidemiologic studies (40), could be strengthened. For example, measuring
323 rather than estimating participant BMR (23), and using updated MET (energy expenditure cost)
324 values (41, 42), could improve future $TEE_{STAR\text{-}Q}$ estimation. In addition, the STAR-Q relies on the
325 reporting of activities during the previous month, a task not without cognitive challenges, though
326 less so than reporting past-year activities (19). Even less cognitively challenging is the Activities
327 Completed over Time in 24 hours (ACT24), an internet-based recall of previous day activities that

328 has been shown to estimate mean TEE and PAL values within 5% of DLW-based estimate and is
329 feasible to administer in epidemiologic studies (43, 44). The accuracy of short-term recall
330 combined with longer-term measures that reflect habitual lifestyle behaviors is an intriguing
331 approach that has been suggested for improving dietary intake measures (45) and may also have
332 potential to improve self reports of activities for TEE and PAL estimation. Moreover, new
333 wearable device-based technologies may in the foreseeable future more accurately estimate PALs
334 and TEE in large populations to complement the contextual data collected by traditional
335 questionnaires (46). Hence, it is anticipated that methodologies may be developed to combine
336 device-derived estimates of energy expenditure and body composition to validate self-reported EI
337 (47-49); though these approaches are still viewed as 'works in progress' (50).

338 A noteworthy strength of this study is the validation of the Goldberg-STAR-Q method
339 against DLW – and the unique use of individual-level PALs. Furthermore, as noted in the recent
340 report on DRIs for Energy, of 128 DLW studies representing 25 countries in the International
341 Atomic Energy Agency DLW Database, none were Canadian (7) highlighting a gap in this area
342 of Canadian research that this study begins to address. Nonetheless, our study has some limitations
343 that merit consideration. First, we used prediction equations for BMR, which can account for a
344 substantial proportion of TEE (51) and could be a source of error in the estimation of $PAL_{STAR\text{-}Q}$.
345 Second, we cannot be certain that some of the under-reporting on the past-year C-DHQ I did not
346 actually reflect undereating during the previous year, though we aimed to minimize this occurrence
347 by recruiting generally weight stable participants to the DLW study. Third, we caution that we
348 cannot assume that misreporting occurs proportionally across all nutrients similar to EI. Indeed,
349 while still relatively under-studied, evidence is accumulating to suggest that foods and nutrients
350 vary in the accuracy with which they are reported (3, 52-54). On a final note, our population was

351 largely Caucasian middle-aged adults; hence, our results may not be generalizable to older (55,
352 56) or more racially diverse populations and those with other characteristics that may affect
353 reporting and health status (57-60).

354 Overall, our DLW-validated results lend credibility to the thesis that detailed
355 comprehensive questionnaire-based activities may be employed with the Goldberg method to
356 explore EI misreporting in large-scale diet studies. Importantly, such strategies create opportunities
357 to carry out sensitivity analyses to better understand the impact of EI misreporting on diet-disease
358 associations in nutritional epidemiology.

359

360 **STATEMENTS AND DECLARATIONS**

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

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

374 **Author contributions:**

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

380

381

382

383 **FIGURE LEGENDS**

384

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

386

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

393

394 **Figure 3.** Sources of discrepancy between the Goldberg and doubly-labeled water methods for
395 classifying energy intake misreporting. **Figure 3A** illustrates sources of discrepancy in men and
396 **Figure 3B** illustrates sources of discrepancy in women. The gold standard doubly-labeled water
397 classifications of misreporting are represented with shaded circles. The Goldberg classifications
398 of misreporting, based on the ratio [rEI:$TEE_{STAR\text{-}Q}$], are represented with open circles (agreement
399 between methods), squares (discrepant because of TEE estimation) or triangles (discrepant
400 because of different cut-offs). The gray solid line represents the lower cut-off for gold standard
401 doubly-labeled water classifications (0.68 for men; 0.66 for women). The black solid line
402 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 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	0.58	1.71	58	15	88	87	91	81	87
		W	0.59	1.69	58	25	79	69	77	72	75
Secondary ($PAL_{1.55}$)[3]	Assumed 1.55	M	0.60	1.66	35	33	54	93	93	56	69
		W	0.59	1.69	19	37	33	100	100	54	63
Sensitivity ($d=\infty$) [4]	$DLW_{d=\infty}$	M	0.90	1.11	85	-	-	-	-	-	-
		W	0.90	1.11	83	-	-	-	-	-	-
	STAR-Q	M	0.68	1.48	60	35	71	100	100	38	75
		W	0.71	1.42	76	32	78	30	84	21	69

[1] Calculations of sensitivity, specificity, PPV, NPV, and accuracy - 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 our Sensitivity analysis (where $d=\infty$ 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.

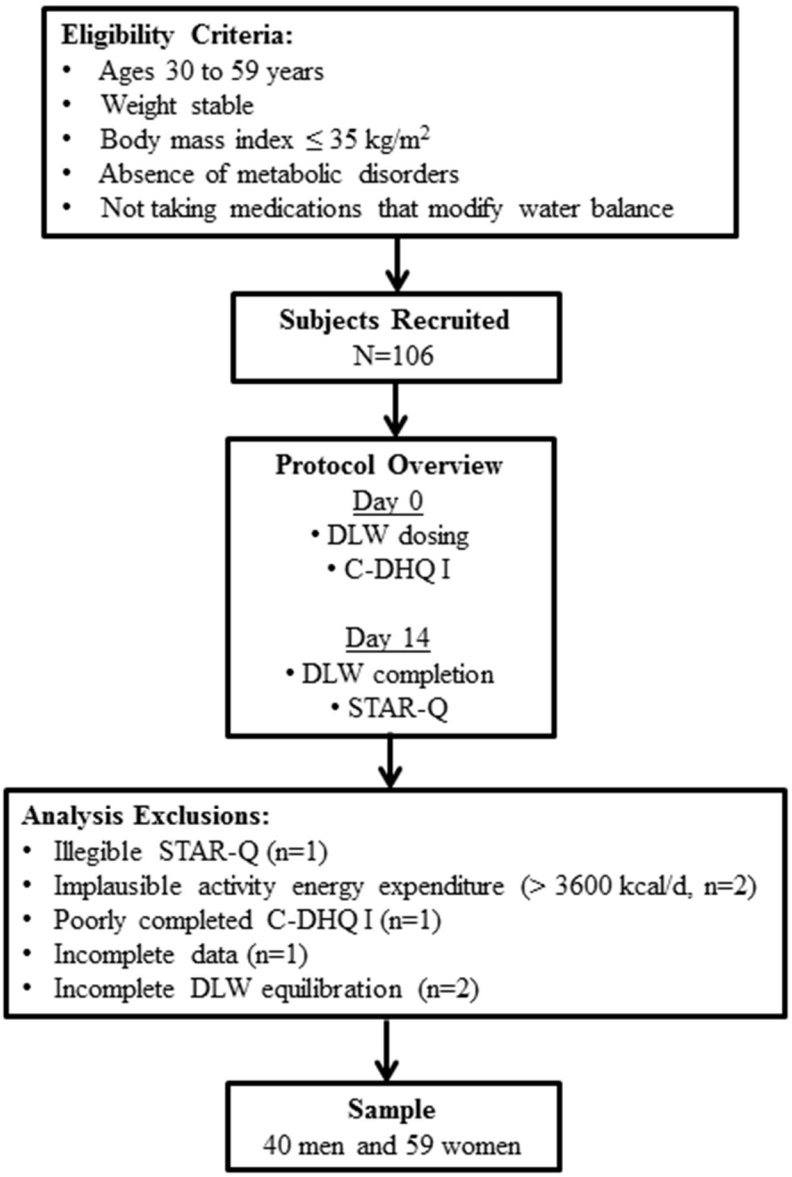
[2] Calculations based on the entire study population (n=99; 40 men, 59 women), including overreporters. Doubly-labeled water classifications of misreporting, based on the ratio [rEI:TEE_{DLW}], served as the 'gold standard' (true) classification. UR was the proportion classified as an under-reporter. 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*=∞, representing an infinite number of days of dietary assessment. We assigned d=∞ 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 59 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

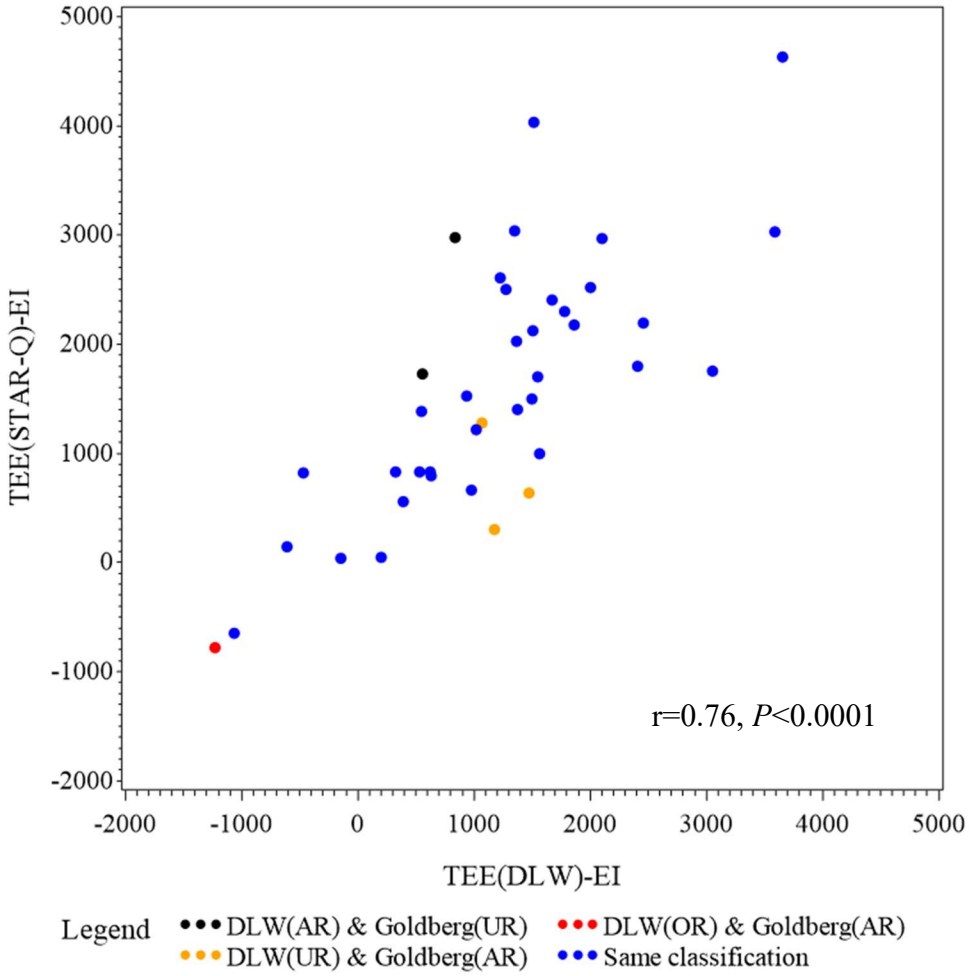


Figure 2A (men)

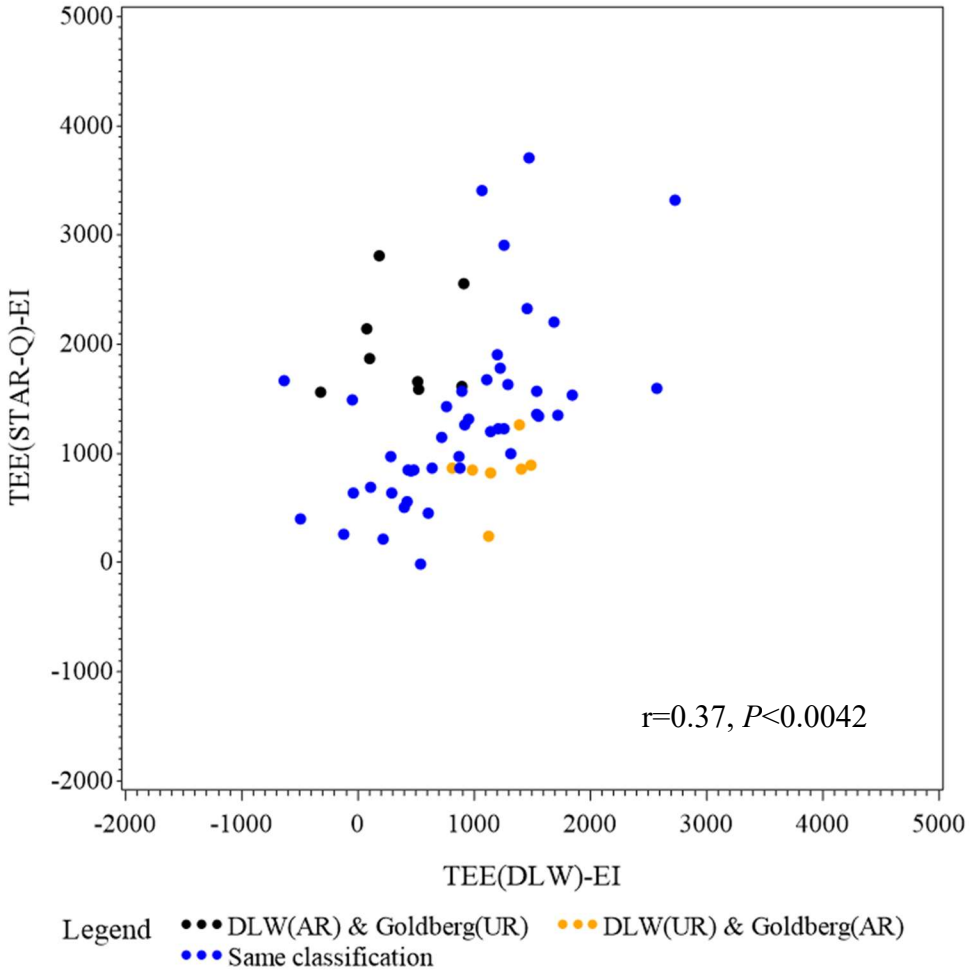


Figure 2B (women)

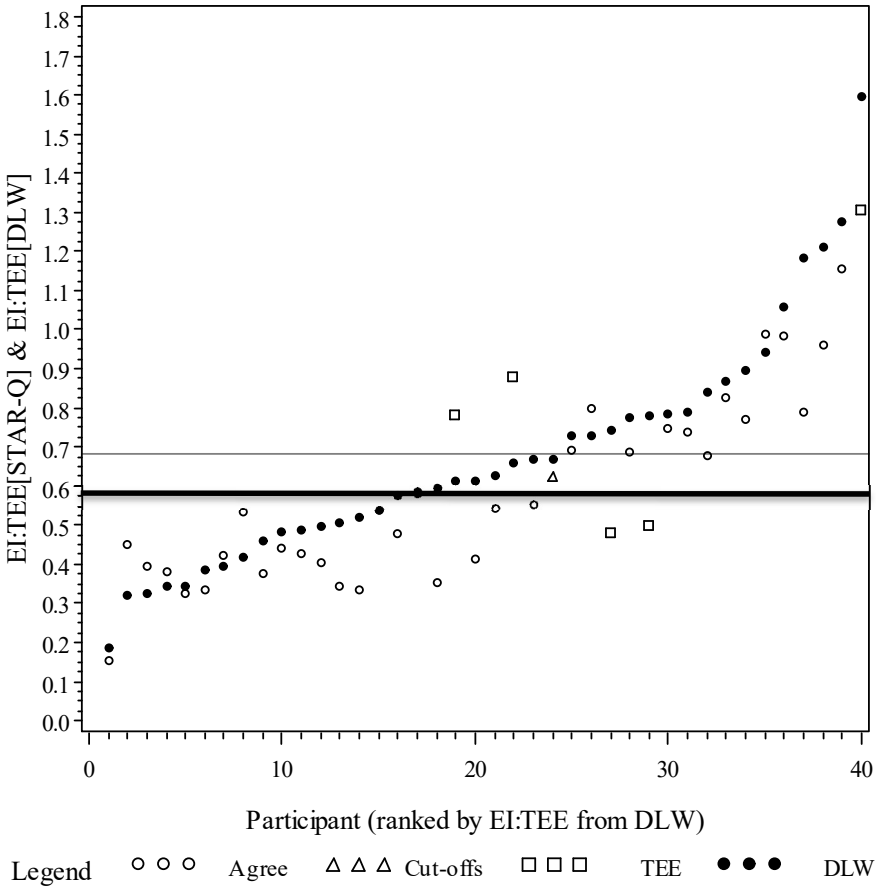


Figure 3A (men)

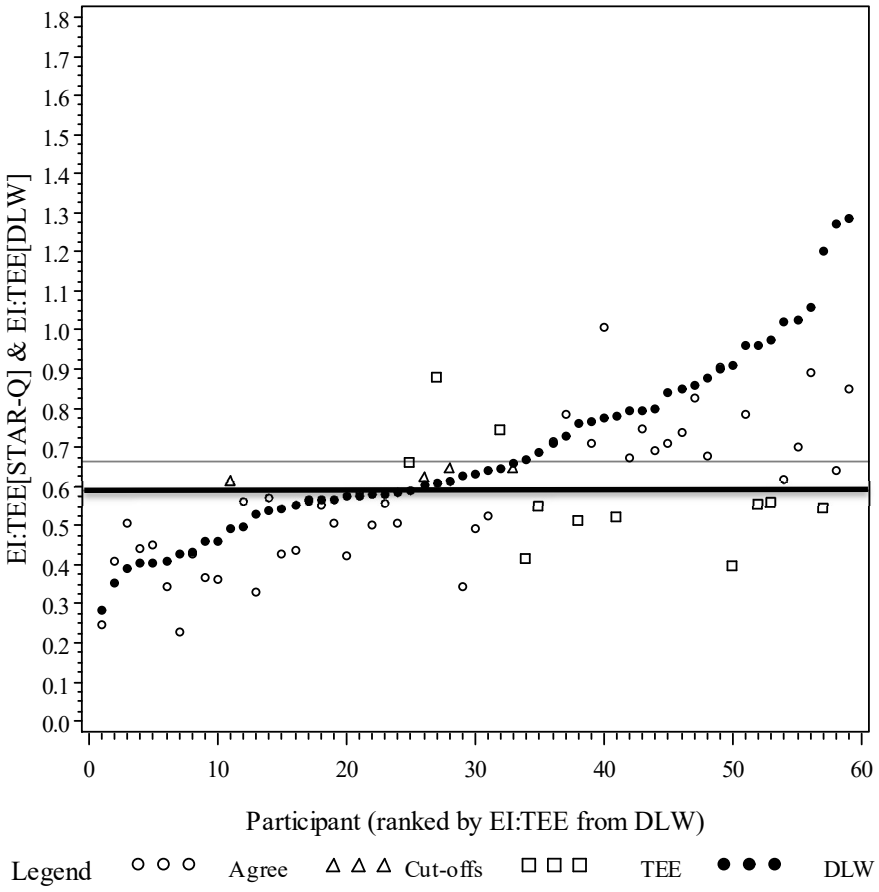


Figure 3B (women)

REFERENCES:

1. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. Am J Epidemiol. 2014;180(2):172-88. doi: 10.1093/aje/kwu116.
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. doi: 10.3389/fnut.2020.00090.
3. Prentice RL. Intake Biomarkers for Nutrition and Health: Review and Discussion of Methodology Issues. Metabolites. 2024;14(5). doi: 10.3390/metabo14050276.
4. Neuhouser ML, Prentice RL, Tinker LF, Lampe JW. Enhancing Capacity for Food and Nutrient Intake Assessment in Population Sciences Research. Annu Rev Public Health. 2022. doi: 10.1146/annurev-publhealth-071521-121621.
5. Banna JC, McCrory MA, Fialkowski MK, Boushey C. Examining Plausibility of Self-Reported Energy Intake Data: Considerations for Method Selection. Frontiers in Nutrition. 2017;4(45). doi: 10.3389/fnut.2017.00045.
6. Black AE, Goldberg GR, Jebb SA, Livingstone MB, Cole TJ, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 2. Evaluating the results of published surveys. Eur J Clin Nutr. 1991;45(12):583-99.
7. 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.

8. Neilson HK, Robson PJ, Friedenreich CM, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? Am J Clin Nutr. 2008;87(2):279-91 [PMID: 18258615]. doi: 10.1093/ajcn/87.2.279.

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

10. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, Prentice AM. 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.

11. Tooze JA, Krebs-Smith SM, Troiano RP, Subar AF. 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. doi: 10.1038/ejcn.2011.198.

12. 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. doi: 10.1097/mco.0000000000000937.

13. Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. J Nutr. 1999;129(10):1765-8. doi: 10.1093/jn/129.10.1765.

14. 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. doi: 10.1038/sj.ejcn.1600971.

15. Grech A, Hasick M, Gemming L, Rangan A. Energy misreporting is more prevalent for those of lower socio-economic status and is associated with lower reported intake of discretionary foods. British Journal of Nutrition. 2021;125(11):1291-8. doi: 10.1017/S0007114520003621.

16. Black AE, Coward WA, Cole TJ, Prentice AM. Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. Eur J Clin Nutr. 1996;50(2):72-92.

17. Csizmadi I, Neilson HK, Kopciuk KA, Khandwala F, Liu A, Friedenreich CM, 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. doi: 10.1093/aje/kwu150.

18. Šuc A, Einfalt L, Šarabon N, Kastelic K. 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]. doi: 10.1186/s12966-024-01632-4.

19. Neilson HK, Ullman R, Robson PJ, Friedenreich CM, Csizmadi I. Cognitive testing of the STAR-Q: insights in activity and sedentary time reporting. J Phys Act Health. 2013;10(3):379-89 [PMID: 22820674]. doi: 10.1123/jpah.10.3.379.

20. Robson PJ, Solbak NM, Haig TR, Whelan HK, Vena JE, Akawung AK, 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. doi: 10.9778/cmajo.20160005.

21. Ye M, Robson PJ, Eurich DT, Vena JE, Xu JY, Johnson JA. Cohort Profile: Alberta's Tomorrow Project. Int J Epidemiol. 2017;46(4):1097-8l. doi: 10.1093/ije/dyw256.

22. Racette SB, Schoeller DA, Luke AH, Shay K, Hnilicka J, Kushner RF. Relative dilution spaces of 2H- and 18O-labeled water in humans. Am J Physiol. 1994;267(4 Pt 1):E585-90. doi: 10.1152/ajpendo.1994.267.4.E585.

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

24. Douketis JD, Paradis G, Keller H, Martineau C. 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. doi: 10.1503/cmaj.045170.

25. Csizmadi I, Kahle L, Ullman R, Dawe U, Zimmerman TP, Friedenreich CM, Bryant H, Subar AF. 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]. doi: 10.1017/s1368980007184287.

26. Csizmadi I, Boucher BA, Lo Siou G, Massarelli I, Rondeau I, Garriguet D, Koushik A, Elenko J, Subar AF. 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. doi: 10.1017/S1368980016001506.

27. Mickle AT, Brenner DR, Beattie T, Williamson T, Courneya KS, Friedenreich CM. 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. doi: 10.1017/jns.2019.32.

28. Biel RK, Csizmadi I, Cook LS, Courneya KS, Magliocco AM, Friedenreich CM. Risk of endometrial cancer in relation to individual nutrients from diet and supplements. Public Health Nutr. 2011;14(11):1948-60. doi: 10.1017/s1368980011001066.

29. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc. 2000;32(9 Suppl):S498-504. doi: 10.1097/00005768-200009001-00009.

30. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Jr., Montoye HJ, Sallis JF, Paffenbarger RS, Jr. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc. 1993;25(1):71-80. doi: 10.1249/00005768-199301000-00011.

31. 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. doi: 10.1038/sj.ijo.0801376.

32. Livingstone MB, Robson PJ, Black AE, Coward WA, Wallace JM, McKinley MC, Strain JJ, McKenna PG. 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. doi: 10.1038/sj.ejcn.1601563.

33. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, 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. doi: 10.1093/aje/kwg092.

34. Tooze JA, Schoeller DA, Subar AF, Kipnis V, Schatzkin A, Troiano RP. Total daily energy expenditure among middle-aged men and women: the OPEN Study. Am J Clin Nutr. 2007;86(2):382-7. doi: 10.1093/ajcn/86.2.382.

35. 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. doi: 10.1017/s0007114515004237.

36. Tooze JA, Freedman LS, Carroll RJ, Midthune D, Kipnis V. The impact of stratification by implausible energy reporting status on estimates of diet-health relationships. Biometrical Journal. 2016;58(6):1538-51. doi: doi:10.1002/bimj.201500201.

37. Solbak NM, Al Rajabi A, Akawung AK, Lo Siou G, Kirkpatrick SI, Robson PJ. 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]). doi: 10.3390/nu11112614.

38. Lo Siou G, Akawung AK, Solbak NM, McDonald KL, Al Rajabi A, Whelan HK, Kirkpatrick SI. 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]. doi: 10.1186/s12937-021-00696-3.

39. Yamamoto N, Ejima K, Zoh RS, Brown AW. Bias in nutrition-health associations is not eliminated by excluding extreme reporters in empirical or simulation studies. Elife. 2023;12. doi: 10.7554/eLife.83616.

40. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health. 2017;5:307. doi: 10.3389/fpubh.2017.00307.

41. Herrmann SD, Willis EA, Ainsworth BE. The 2024 Compendium of Physical Activities and its expansion. J Sport Health Sci. 2024;13(1):1-2. doi: 10.1016/j.jshs.2023.09.008.

42. Herrmann SD, Willis EA, Ainsworth BE, Barreira TV, Hastert M, Kracht CL, 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. doi: 10.1016/j.jshs.2023.10.010.

43. Matthews CE, Kozey Keadle S, Moore SC, Schoeller DS, Carroll RJ, Troiano RP, Sampson JN. Measurement of Active and Sedentary Behavior in Context of Large Epidemiologic Studies. Med Sci Sports Exerc. 2018;50(2):266-76. doi: 10.1249/MSS.0000000000001428.

44. Matthews CE, Patel S, Saint-Maurice P, Loftfield E, Keadle SK, Chen KY, Brychta R, LaMunion S, Berrigan D. Physical Activity Levels (PAL) in United States Adults - 2019. Med Sci Sports Exerc. 2022. doi: 10.1249/mss.0000000000003102.

45. Freedman LS, Midthune D, Arab L, Prentice RL, Subar AF, Willett W, Neuhouser ML, Tinker LF, Kipnis V. 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. doi: 10.1093/aje/kwy126.

46. Sampson JN, Matthews CE, Freedman L, Carroll RJ, Kipnis V. Methods to Assess Measurement Error in Questionnaires of Sedentary Behavior. J Appl Stat. 2016;43(9):1706-21. doi: 10.1080/02664763.2015.1117593.

47. Hibbing PR, Shook RP, Panda S, Manoogian ENC, Mashek DG, Chow LS. Predicting energy intake with an accelerometer-based intake-balance method. Br J Nutr. 2022:1-23. doi: 10.1017/s0007114522003312.

48. Hibbing PR, Welk GJ, Ries D, Yeh H-W, Shook RP. 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. doi: 10.1186/s12966-023-01515-0.

49. Shook RP, Yeh H-W, Welk GJ, Davis AM, Ries D. 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. doi: 10.1093/jn/nxab317.

50. 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. doi: https://doi.org/10.1093/jn/nxab393.

51. Manini TM. Energy expenditure and aging. Ageing Res Rev. 2010;9(1):1-11. doi: 10.1016/j.arr.2009.08.002.

52. Krebs-Smith SM, Graubard BI, Kahle LL, Subar AF, Cleveland LE, Ballard-Barbash R. Low energy reporters vs others: a comparison of reported food intakes. Eur J Clin Nutr. 2000;54(4):281-7. doi: 10.1038/sj.ejcn.1600936.

53. Lafay L, Mennen L, Basdevant A, Charles MA, Borys JM, Eschwège E, Romon M. 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. doi: 10.1038/sj.ijo.0801392.

54. Prentice RL, Aragaki AK, Zheng C, Manson JE, Tinker LF, Schoeller DA, 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. doi: 10.1016/j.ajcnut.2025.03.011.

55. Porter J, Ward LC, Nguo K, Ward A, Davidson Z, Gibson S, Prentice R, Neuhouser ML, Truby H. 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. doi: 10.1016/j.ajcnut.2024.02.005.

56. Wong WW, Speakman JR, Ainslie PN, Anderson LJ, Arab L, Baddou I, 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. doi: 10.1016/j.tjnut.2024.10.016.

57. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SA, Caan B, 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. doi: 10.1093/aje/kwr140.

58. Pryer JA, Vrijheid M, Nichols R, Kiggins M, Elliott P. Who are the 'low energy reporters' in the dietary and nutritional survey of British adults? Int J Epidemiol. 1997;26(1):146-54. doi: 10.1093/ije/26.1.146.

59. Prentice AM, Black AE, Coward WA, Cole TJ. 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.

60. Karelis AD, Lavoie ME, Fontaine J, Messier V, Strychar I, Rabasa-Lhoret R, Doucet E. 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. doi: 10.1038/ejcn.2009.119.