Practice of Epidemiology

The Sedentary Time and Activity Reporting Questionnaire (STAR-Q): Reliability and Validity Against Doubly Labeled Water and 7-Day Activity Diaries

Ilona Csizmadi*, Heather K. Neilson, Karen A. Kopciuk, Farah Khandwala, Andrew Liu, Christine M. Friedenreich, Yutaka Yasui, Rémi Rabasa-Lhoret, Heather E. Bryant, David C. W. Lau, and Paula J. Robson

* Correspondence to Dr. Ilona Csizmadi, Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Richmond Road Diagnostic and Treatment Centre, 1820 Richmond Road SW, Calgary, Alberta T2T 5C7, Canada (e-mail: ilona.csizmadi@albertahealthservices.ca).

Initially submitted November 8, 2013; accepted for publication May 19, 2014.

We determined measurement properties of the Sedentary Time and Activity Reporting Questionnaire (STAR-Q), which was designed to estimate past-month activity energy expenditure (AEE). STAR-Q validity and reliability were assessed in 102 adults in Alberta, Canada (2009–2011), who completed 14-day doubly labeled water (DLW) protocols, 7-day activity diaries on day 15, and the STAR-Q on day 14 and again at 3 and 6 months. Three-month reliability was substantial for total energy expenditure (TEE) and AEE (intraclass correlation coefficients of 0.84 and 0.73, respectively), while 6-month reliability was moderate. STAR-Q-derived TEE and AEE were moderately correlated with DLW estimates (Spearman’s ρs of 0.53 and 0.40, respectively; P < 0.001), and on average, the STAR-Q overestimated TEE and AEE (median differences were 367 kcal/day and 293 kcal/day, respectively). Body mass index-, age-, sex-, and season-adjusted concordance correlation coefficients (CCCs) were 0.24 (95% confidence interval (CI): 0.07, 0.36) and 0.21 (95% CI: 0.11, 0.32) for STAR-Q-derived versus DLW-derived TEE and AEE, respectively. Agreement between the diaries and STAR-Q (metabolic equivalent-hours/day) was strongest for occupational sedentary time (adjusted CCC = 0.76, 95% CI: 0.64, 0.85) and overall strenuous activity (adjusted CCC = 0.64, 95% CI: 0.49, 0.76). The STAR-Q demonstrated substantial validity for estimating occupational sedentary time and strenuous activity and fair validity for ranking individuals by AEE.

motor activity; physical activity; reproducibility of results; sedentary lifestyle; questionnaires; validation studies

Abbreviations: AEE, activity energy expenditure; CCC, concordance correlation coefficient; DLW, doubly labeled water; ICC, intraclass correlation coefficient; MET, metabolic equivalent of task; STAR-Q, Sedentary Time and Activity Reporting Questionnaire; TEE, total energy expenditure.

As the need for comprehensive assessments of daily physical activity has grown, the limitations of existing methods have become apparent (1, 2). Many physical activity questionnaires have been developed and validated, but most are not well suited for the study of new behavioral parameters of interest, such as activity energy expenditure (AEE). In a review of physical activity questionnaires that were validated against doubly labeled water (DLW)—the gold standard for estimating total energy expenditure (TEE) in free-living individuals—we reported that overall, questionnaire performance in estimating AEE was poor when compared with DLW (3). However, none of the questionnaires were originally designed to estimate AEE.

While objective measures continue to evolve in precision and accuracy, they are limited as “stand-alone” measures when contextual information pertaining to habitual behavior is desired (4). It is generally agreed that, for at least the foreseeable future, there will continue to be a need for logistically feasible, cost-effective tools with known validity that can assess a broad range of habitual behaviors in large populations (4, 5).

We designed the Sedentary Time and Activity Reporting Questionnaire (STAR-Q), a past-month activity questionnaire,
to address the need for a self-administered questionnaire that could assess all types and intensities of physical activity, including sedentary behavior, across all activity domains for the purpose of ascertaining AEE among adults in large-scale epidemiologic studies. Following several phases of development and cognitive testing (6), the next step was to determine the STAR-Q’s validity and reliability. We assessed validity by comparing STAR-Q-derived AEE and TEE estimates with those derived using DLW. We also compared activities reported in the STAR-Q with intensity- and domain-specific activities reported prospectively in a 7-day activity diary. In addition, we assessed the reliability of the STAR-Q for estimating AEE by administering it on 3 occasions.

METHODS

The STAR-Q

The STAR-Q (see Web Appendix 1, available at http://aje.oxfordjournals.org/) is a quantitative, self-administered past-month questionnaire designed to capture information on all types of activities and sedentary behaviors across all domains: eating, personal/medical care, occupation, transportation, household, yard work, caregiving, exercise, light leisure (e.g., television-watching, personal computer time, reading, hobbies, etc.), stair-climbing, and “other” activities. Development and cognitive testing of the STAR-Q have been previously described (6). Briefly, detailed physical activity data from 18,838 participants enrolled in the Alberta Tomorrow Project, a Canadian cohort study (7), were used to identify and group common activities with similar energy expenditures to create activity-related questionnaire items. Open-ended reporting of some activities was also allowed on the STAR-Q.

Study participants

Study participants comprised 1) urban residents of the Calgary, Alberta, Canada, area who were recruited through posters and worksite e-mail distribution lists and 2) participants in the Alberta Tomorrow Project. The Alberta Tomorrow Project is a study of a large, geographically dispersed cohort established in 2001 to examine the relationship between lifestyle and chronic diseases in Alberta (7). Invitations to participate in the current study were mailed to Project participants who met eligibility criteria for the current study, based on characteristics reported upon enrollment in the Project. Eligibility criteria were as follows: body mass index (weight (kg)/height (m)^2) ≤35, weight stability (±2.5 kg for at least 3 months) with no intention to gain/lose weight in the near future, not being pregnant or breastfeeding, absence of metabolic disorders (e.g., diabetes, thyroid), and no use of medications that modified water balance. In addition, we set the upper age limit for eligibility at 60 years to avoid incomplete collections of urine due to bladder urine retention. The minimum age of eligibility for participants recruited from the community (i.e., the urban Calgary sample) was set at 30 years, which is comparable to the Alberta Tomorrow Project (age 35 years). All participants provided written informed consent, and 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. Participants received Can$100 upon study completion.

Reliability study design

To assess the consistency of the STAR-Q for estimating energy expenditure over a longer duration, participants completed the STAR-Q 3 times over a 6-month period (Figure 1).

Figure 1. Overview of the study design for a study conducted to validate the Sedentary Time and Activity Reporting Questionnaire (STAR-Q), Alberta, Canada, 2009–2011. Numbers along the horizontal axis indicate the number of months relative to study baseline (indicated as “0”). The reliability study occurred across months 0, 3, and 6, comparing responses between the first, second, and third administrations of the STAR-Q (STAR-Q1, STAR-Q2, and STAR-Q3). AEE, activity energy expenditure; DLW, doubly labeled water; TEE, total energy expenditure.
The first STAR-Q (STAR-Q1) was completed 14 days after DLW dosing (day 0). The second and third STAR-Qs (STAR-Q2 and STAR-Q3) were completed approximately 3 and 6 months after day 14, respectively. STAR-Q1 was self-administered at the test facility, and following completion, responses were reviewed for completeness and clarity of the text. STAR-Q2 and STAR-Q3 were also self-administered but were sent to participants and returned by mail.

**AEE estimation from the STAR-Q.** Responses on the STAR-Q were used to estimate AEE\textsubscript{STAR-Q}. TEER\textsubscript{STAR-Q} and energy expenditure in specific domains, and daily numbers of hours spent in intensity-specific activities and in sedentary behavior. All self-reported activities were assigned activity codes and metabolic equivalent of task (MET) values from the Compendium of Physical Activities (8, 9). Where total reported time spent in physical activity, sedentary behavior, and sleep was less than 24 hours/day, a MET value of 1.2 was assigned to the unaccounted-for time. Reported frequency, duration, and intensity of activities were combined to obtain a single estimate of energy expenditure in MET-hours/day for each activity. Activity-specific estimates of MET-hours/day were then summed across all activities for each subject to obtain the total average energy cost in MET-hours/day. Those estimates were subsequently applied to our estimations of average past-month TEER\textsubscript{STAR-Q} and AEE\textsubscript{STAR-Q} as follows:

\[
\text{TEER}_{\text{STAR-Q}} = \left[(0.9\text{ METs} \times \text{hours of sleep}) + (\text{MET-hours/day} \times \text{weight (kg)})\right] \times 1.1
\]

and

\[
\text{AEE}_{\text{STAR-Q}} = [\text{MET-hours/day} \times \text{weight (kg)} - 1.0\text{ METs} \times \text{weight (kg)} \times \text{total hours of activity/day}],
\]

where TEER\textsubscript{STAR-Q} was estimated from reported nightly hours of sleep from the STAR-Q multiplied by 0.9 METs (average energy expenditure during sleep (8, 9)) and added to total daily MET-hours during time spent awake multiplied by weight (in kilograms) measured at baseline. An additional 10% was added for energy expenditure attributed to the thermic effect of food. AEE\textsubscript{STAR-Q} was estimated from total daily MET-hours accumulated during waking time multiplied by weight (in kilograms) measured at baseline, from which energy expenditure attributed to resting metabolism (1.0 METs \times \text{weight} \times \text{activity hours/day}) was subtracted.

Finally, estimates of amounts of time spent in each activity domain and at various intensities—sedentary (≤1 METs), light (>1.5–≤3.0 METs), moderate (>3.0–≤6.0 METs), and vigorous (>6.0 METs)—were derived. We also determined AEE per kilogram of body weight (kcal/kg-day) for DLW- and STAR-Q-derived estimates in an effort to account for the known tendency of body mass or lean mass to inflate the association between questionnaire-derived AEE and DLW-derived AEE (10, 11).

**Validation study design**

An overview of the 21-day validation study protocol is provided in Figure 1. The DLW protocol was completed over a period of 14 days, with the STAR-Q1 being completed on day 14. Participants were instructed to begin prospectively recording daily activities in the 7-day diary on day 15. The relative timing of the STAR-Q1 and diary administrations was intended to minimize learning effects and altered awareness of behavior that might have occurred had the diaries been completed first.

**7-Day activity diary.** A 7-day activity diary was employed to compare domain- and intensity-specific AEE and duration estimates, respectively, with those of STAR-Q1. The 7-day diary was adapted from physical activity records described by Conway et al. (12) and was designed to ascertain sleep time as well as all activities engaged in (of ≥10 minutes duration and those requiring a change in physical effort) and posture while awake. Respondents recorded the time at which each activity began, described the activity, and selected one of 10 listed domains and subdomains (corresponding closely to the STAR-Q), posture (reclining, sitting, standing, walking, or “in motion” (e.g., swimming, cycling, dancing)), and the perceived intensity of the activity (effort level 1, 2, 3, or 4; definitions were provided). Participants also reported stair-climbing (number of flights up and down). Completed diaries were returned to the center by mail. Summary estimates of energy expenditure were derived for the 7-day diaries using reduction procedures as described for the STAR-Q.

**Anthropometric measures and resting metabolic rate.** Height and weight were measured at the test facility to the nearest 0.1 cm and 0.1 kg, respectively, on days 0 and 14 of the DLW protocol by a certified exercise physiologist using standard procedures (13). The TBF-310 Tanita Body Composition Analyzer and Scale (Tanita Corporation of America, Inc., Preston, Washington (TheCompetitiveEdge.com)) was used to measure body weight and to estimate percentage of body fat. Body mass index was determined as body weight (kg) divided by height (m) squared. Resting metabolic rate was estimated by means of the Schofield equation, which takes into account age, height, and weight (14).

**DLW protocol.** A 14-day DLW study was used as the primary criterion method against which AEE\textsubscript{STAR-Q} and TEER\textsubscript{STAR-Q} estimates were validated (Figure 1). Following an overnight fast, participants provided baseline urine and saliva samples for determination of background isotope levels (day 0). DLW doses were administered orally to provide each subject with 2.5 g 10 atom % of oxygen-18 (18O) per kg of total body water and 0.18 g 99 atom % of deuterium (2H) per kg of estimated total body water (Rotem Inc., Topsfield, Massachusetts). Each participant was asked to abstain from food and fluid intake for 4 hours after dosing. Postdose saliva samples were collected at 3 and 4 hours for the measurement of total body water from deuterium-isotope dilution. The Canadian Diet History Questionnaire (15), a food frequency questionnaire that ascertains past-year dietary intake, was completed on day 0 for food quotient estimation (proxy for respiratory quotient). Second-void urine samples were collected on days 1 and 8, with sampling time recorded on a time sheet and stored in a refrigerator in a 125-mL container until pick-up by study staff within the next day. On day 14, following an overnight fast, participants collected the final urine sample and returned to the test center for dosing with 0.18 g 99 atom % deuterium per kg of total body water, followed by postdose saliva sample collections as described for

day 0. To measure the decline in isotope enrichment, samples were batch-analyzed in duplicate using the Isoprime Stable Isotope Ratio Mass Spectrometer (Isoprime Ltd., Cheadle Hulme, United Kingdom). A MultiFlow-Bio module for Isoprime (Isoprime Ltd.) equipped with a Gilson 222XL Autosampler (GV Instruments Ltd., Manchester, United Kingdom) was used for daily energy expenditure measurements. Data processing was performed using IonVantage software (2012) for Isoprime (Isoprime Ltd.). Stability tests were performed each day before testing, yielding standard deviations of 0.026% for deuterium and 0.004% for oxygen-18.

TEE was calculated according to the method of Racette et al. (16), using a modified Weir equation and an assumed respiratory quotient of 0.85. In addition, a food quotient (proxy for respiratory quotient) was estimated from the Diet History Questionnaire (15). AEE was subsequently derived from TEE as follows:

\[
\text{AEE}_{\text{DLW}} = \left[ \text{TEE}_{\text{DLW}} \times 0.90 \right] - \left\{ \left[ \left( \frac{\text{RMR}}{24} \right) \times (24 \text{ hours of sleep}) \right] + 0.9 \left( \frac{\text{RMR}}{24} \times \text{hours of sleep} \right) \right\},
\]

where RMR represents resting metabolic rate. As with the self-reported data, resting metabolic rate was estimated using the Schofield equation (14). The thermic effect of food (10%) was subtracted from TEE, and resting metabolic rate was subtracted after accounting for a 10% reduction in energy expenditure during the self-reported sleep hours from STAR-Q1.

Statistical analysis

Baseline characteristics were examined for participants who did not return a STAR-Q or had implausibly high STAR-Q-derived AEE levels. Their characteristics were compared with those of the rest of the group.

Reliability study. Descriptive statistics were calculated for all derived variables. Agreement across STAR-Q administrations was assessed using intraclass correlation coefficients (ICCs) based on 2-way mixed models without interaction (17), treating subjects as random effects and the activity assessment methods as fixed effects. Concordance correlation coefficients (CCCs) were estimated using variance components of a 2-way mixed-effects model (18). CCCs are practically and asymptotically equivalent to ICCs but do not require analysis-of-variance model assumptions, reflecting accuracy and precision, in a 2-way mixed-effects model with adjustment for age, sex, and season (18). Bland-Altman plots (24) were used to appraise group-level agreement and to identify trends in under- and overestimation of AEE and TEE by the STAR-Q relative to DLW. The 95% limits of agreement (25) around the mean difference showed with 95% certainty the range in which individual differences would occur for a similar population. Sample size was estimated for 2 mixed-effects analysis-of-variance models, assuming an ICC of at least 0.65 with a 95% confidence interval of ±0.10 (26).

RESULTS

Participation

Between July 2009 and July 2010, 106 participants were recruited; data collection was completed by January 2011. One participant was excluded based on illegible STAR-Q responses; 5 did not return STAR-Q2s, and 5 did not return STAR-Q3s. Others were excluded based on a priori criteria for implausible values (AEE ≥ 3,600 kcal/day). Final sample sizes were 102 (STAR-Q1), 96 (STAR-Q2), and 97 (STAR-Q3). Estimated CCCs comparing all 3 STAR-Qs for reliability included only those participants with data for all 3 STAR-Qs (n = 91). Nine participants who did not return at least 1 STAR-Q were similar to the remaining participants in terms of age, body mass index, sex, and percentage of body fat. Persons excluded because of implausibly high AEEs (n = 7) were mostly men (86% men, as opposed to 40% in the analyzed group).

In the validity study, of 102 participants with valid STAR-Q1 data, 1 did not return a diary, 1 submitted a diary with insufficient detail for activity estimation, and 1 was excluded based on a diary-estimated AEE greater than 3,600 kcal/day. Two DLW measures were excluded because of incomplete equilibration and high baseline urine isotope levels. Thus, 99 and 100 subjects were included in the diary and DLW comparisons, respectively.

On average, study participants were middle-aged, Caucasian, employed, and well-educated (Table 1). Although participants represented a range of physical activity levels using Institute of Medicine (29) criteria and DLW estimates, TEE_{DLW} showed that the majority were “active” (physical activity level 1.60–<1.90) or “very active” (physical activity level ≥1.90).

The median self-reported time required for STAR-Q1 completion was 34 minutes. The proportions of STAR-Qs completed during the warmer months (May–October) were 45.7% (STAR-Q1), 53.0% (STAR-Q2), and 43.0% (STAR-Q3).

Estimated energy expenditure

Numbers of hours per day of self-reported activity (Table 2 and Web Table 1) were reasonable, on average. However,
median physical activity levels derived from the STAR-Qs and diary were relatively high (approximately 2.0, or "very active" based on Institute of Medicine criteria, in comparison with the DLW-derived average physical activity level of 1.8 ("active" using Institute of Medicine criteria)). TEEDLW estimation using a respiratory quotient of 0.85 or a Diet History Questionnaire-estimated food quotient were similar, with median TEEs differing by less than 20 kcal/day (data not shown); hence, we assumed a respiratory quotient of 0.85 in all subsequent TEEDLW analyses.

Reliability

ICCs and CCCs for kcal/day and MET-hours/day are presented in Table 3. Generally the ICCs for STAR-Q1 versus STAR-Q2 (completed 3 months apart) were higher than ICCs for STAR-Q1 versus STAR-Q3 (completed 6 months apart). When comparing STAR-Q1 with STAR-Q2, good-to-excellent ICCs (>0.70) were observed for TEE, AEE, sleep, overall occupational activity, occupational sedentary behavior, and television-watching. Remaining activities were associated with fair-to-moderate ICCs. Generally, adjusted CCCs (CCCadj) were lower than ICCs, but associations remained moderate to good (approximately 0.60–0.70) for TEE, AEE, sleep, overall occupational activity, occupational sedentary behavior, and television-watching. Notable differences between unadjusted and adjusted CCCs were found for TEE and AEE; covariate adjustment otherwise had little impact.

Validity

STAR-Q1 versus 7-day activity diary. On average, the diary numbers of hours per day closely approximated a
24-hour period, while the STAR-Q underestimated hours per day by just over 2 hours (Table 2). Reporting on sedentary time probably contributed to this difference (Figure 2). Additionally, total occupational activity, occupational light activity, total light leisure activity, and sedentary behavior during light leisure activity accounted for less time using the diary (Figure 3).

Results of the STAR-Q1-versus-diary comparison are presented in Table 4. For TEE and AEE, negligible median differences were observed at the group level; however, larger interquartile ranges indicated high between-subject variability. Agreement and consistency for TEE, AEE, and AEE/kg body weight were moderate to good depending on the test statistic; Spearman correlations were relatively high ($\rho_s = 0.74$, $P < 0.001$), whereas CCC_adj values were lower (0.51, 0.45, and 0.41). Results for Overall occupational activity, occupational sitting, and occupational sedentary time were consistently high, with CCC_adj values of approximately 0.70, and CCC_adj was 0.64 for overall strenuous activity (hours/day). In contrast, consistency between STAR-Q1- and diary-derived estimates for light activities were poor (CCC_adj < 0.30) overall and within occupation and exercise domains. Adjusting CCCs for covariates had the strongest effect for TEE and AEE but an otherwise negligible impact.

**STAR-Q1 versus DLW.** On average, the STAR-Q1 significantly overestimated energy expenditure relative to DLW ($P<0.001$; Wilcoxon signed-rank test), with a median difference (STAR-Q1 minus DLW) of 367 kcal/day (interquartile range, 800 kcal/day) for TEE and a median difference of 293 kcal/day (interquartile range, 769 kcal/day) for AEE. Bland-Altman plots (Figure 4) revealed significant inverse trends in which overestimation of energy expenditure occurred for persons with the lowest $\text{TEE}_{DLW}$ (Figure 4A) or $\text{AEE}_{DLW}$ (Figure 4B), while for persons with the highest $\text{TEE}_{DLW}$ or $\text{AEE}_{DLW}$, energy expenditure was underestimated by the STAR-Q. This negative trend was most pronounced for AEE/kg body weight ($\rho_s = -0.61$, $P < 0.001$; Figure 4C).

Compared with DLW estimates, $\text{TEE}_{\text{STAR-Q1}}$ was moderately correlated ($\rho_s = 0.53$, $P < 0.001$), as was $\text{AEE}_{\text{STAR-Q1}}$ ($\rho_s = 0.40$, $P < 0.001$), while the correlation for body-weight-adjusted $\text{AEE}_{\text{STAR-Q1}}$ was slightly lower ($\rho_s = 0.34$, $P = 0.001$). ICCs using log-transformed variables were also fair to moderate, with the ICC for TEE being 0.41 (95% CI: 0.23, 0.56), that for AEE being 0.20 (95% CI: 0.01, 0.38), and that for AEE/kg body weight being 0.13 (95% CI: −0.07, 0.32). CCCs with log-transformed TEE, AEE, and AEE/kg body weight and adjusted for covariates were 0.24 (95% CI: 0.07, 0.36), 0.21 (95% CI: 0.11, 0.32), and 0.19 (95% CI: 0.08, 0.27), respectively.

**DISCUSSION**

The STAR-Q has reasonable reliability for estimating past-month AEE and, on average, fair validity for ranking individuals by AEE based on Spearman correlation comparisons with DLW. However, DLW comparisons also revealed proportional measurement error, with trends of AEE overreporting by the least active persons and AEE underreporting by the most active persons; hence, validity estimates based on adjusted CCCs were lower. Compared with prospectively completed 7-day activity diaries, the STAR-Q performed particularly well in estimating energy expenditure related to overall occupational activity, occupational sedentary and sitting time, and strenuous activity, but agreement was weaker for light activity.

To our knowledge, only 1 other questionnaire was designed with an objective similar to our own. Besson et al. (30) reported results for validation of the Recent Physical Activity Questionnaire against accelerometry and DLW. Similar to the STAR-Q, this questionnaire ascertains past-month activity and demonstrates substantial validity for ranking individuals on vigorous-intensity activity. Relative to the STAR-Q, Besson et al. described somewhat better validity for ranking individuals on TEE ($\rho_s = 0.67$ vs. $\rho_s = 0.53$) and AEE/kg body weight ($\rho_s = 0.39$ vs. $\rho_s = 0.34$) but lower validity for ranking them on sedentary time ($\rho_s = 0.27$ vs. $\rho_s = 0.40$).
In contrast to the STAR-Q, the Recent Physical Activity Questionnaire did not ascertain the number of hours slept per night, the duration of the workweek, or details on household activities. When comparing reported behaviors with 7-day diaries, the STAR-Q performed well, on average, despite nonoverlapping reference periods. True differences in energy expenditure might have occurred between the two assessment periods, perhaps accounting for some discrepancy between the STAR-Q and the diary. However, areas of substantial agreement between the STAR-Q and diary are informative for guiding the optimal use of the STAR-Q in future studies (e.g., occupational sedentary behavior, strenuous activity). These stronger results were not unexpected, since occupational and strenuous activities have been reliably reported in other studies (30, 31).

On the other hand, activities with poor agreement highlight problematic domains and intensities—for example, light-intensity activity that may require alternate assessment methods combining objective and self-reported data.

The strengths of our study relative to past DLW studies (3) include a large sample size, appropriate use of DLW as a criterion method, and the use of multiple test statistics, including

### Table 3. Intraclass and Concordance Correlation Coefficients (MET-hours/day) for Reliability Assessments in the STAR-Q Validation Study, Alberta, Canada, 2009–2011

<table>
<thead>
<tr>
<th></th>
<th>STAR-Q1 vs. STAR-Q2 (n=95)</th>
<th>STAR-Q1 vs. STAR-Q3 (n=96)</th>
<th>STAR-Q1 vs. STAR-Q2 vs. STAR-Q3 (n=91)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>95% CI</td>
<td>ICC</td>
</tr>
<tr>
<td>TEE, kcal/day</td>
<td>0.84</td>
<td>0.77, 0.89</td>
<td>0.74</td>
</tr>
<tr>
<td>AEE, kcal/day</td>
<td>0.73</td>
<td>0.62, 0.81</td>
<td>0.60</td>
</tr>
<tr>
<td>Sleeping</td>
<td>0.79</td>
<td>0.70, 0.85</td>
<td>0.69</td>
</tr>
<tr>
<td>Stair-climbing, flights/day</td>
<td>0.45</td>
<td>0.28, 0.60</td>
<td>0.44</td>
</tr>
<tr>
<td>Active sitting</td>
<td>0.45</td>
<td>0.28, 0.60</td>
<td>0.46</td>
</tr>
</tbody>
</table>

#### Overall activity

<table>
<thead>
<tr>
<th></th>
<th>Sedentary</th>
<th>Light</th>
<th>Moderate</th>
<th>Strenuous</th>
<th>Exercise</th>
<th>Occupational activity</th>
<th>Sitting</th>
<th>Sedentary</th>
<th>Light</th>
<th>Moderate</th>
<th>Strenuous</th>
<th>Light leisure activity</th>
<th>Television-viewing</th>
<th>Computer use</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.53</td>
<td>0.60</td>
<td>0.42</td>
<td>0.59</td>
<td>0.76</td>
<td>0.69</td>
<td>0.71</td>
<td>0.65</td>
<td>0.44</td>
<td>0.65</td>
<td>0.33</td>
<td>0.56</td>
<td>0.72</td>
<td>0.60</td>
<td>0.56</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.37, 0.66</td>
<td>0.46, 0.71</td>
<td>0.26, 0.57</td>
<td>0.44, 0.71</td>
<td>0.66, 0.83</td>
<td>0.57, 0.78</td>
<td>0.60, 0.80</td>
<td>0.52, 0.75</td>
<td>0.46, 0.59</td>
<td>0.52, 0.75</td>
<td>0.33, 0.49</td>
<td>0.49, 0.67</td>
<td>0.61, 0.70</td>
<td>0.52, 0.61</td>
<td>0.41, 0.58</td>
</tr>
</tbody>
</table>

#### Abbreviations: AEE, activity energy expenditure; CCC, concordance correlation coefficient; CI, confidence interval; DLW, doubly labeled water; ICC, intraclass correlation coefficient; MET, metabolic equivalent of task; STAR-Q, Sedentary Time and Activity Reporting Questionnaire; TEE, total energy expenditure.

a STAR-Q1, STAR-Q2, and STAR-Q3 represent the first, second, and third administrations of the STAR-Q.

b Not adjusted for any variable.
c 95% CIs were estimated by means of bootstrapping for both adjusted and unadjusted CCCs.
d Adjusted for sex, age, and body mass index at baseline, as well as season of STAR-Q completion.
e Log-transformed data.
f Excludes napping.
g Results are shown for climbing up stairs (self-reported number of flights per day); results for going down stairs were very similar.
h Represents the sum of all activities performed while sitting and expending energy at >1.5–≤2.5 METs.
i Intensity categories were defined as follows: sedentary (excluding sleep), ≤1.5 METs; light, >1.5–≤3 METs; moderate, >3–≤6 METs; strenuous, >6 METs. Strenuous occupational activity was not analyzable because of insufficient data.

j MET level ≤2.5.
the novel use of the CCC. The CCC is a rigorous estimate of agreement that allows for the adjustment of covariates that might contribute to between-subject variability. Adjusted coefficients may be lower than unadjusted estimates; however, the remaining explained variance more clearly reflects the measure of interest (18). For example, a $\text{CCC}_{\text{adj}}$ of 0.21 for AEE indicates that STAR-Q-derived AEE explained 21% of DLW-derived AEE. To our knowledge, only 2 other DLW validation studies have conducted analyses that might be considered analogous to this. Masse et al. (32)

![Figure 2. Self-reported mean duration (hours/day) of active and inactive behavior by intensity in the Sedentary Time and Activity Reporting Questionnaire (STAR-Q) validation study, Alberta, Canada, 2009-2011. STAR-Q1, STAR-Q2, and STAR-Q3 represent the first, second, and third administrations of the STAR-Q.](image)

![Figure 3. Self-reported mean duration (hours/day) of active and inactive behavior by domain in the Sedentary Time and Activity Reporting Questionnaire (STAR-Q) validation study, Alberta, Canada, 2009-2011. STAR-Q1, STAR-Q2, and STAR-Q3 represent the first, second, and third administrations of the STAR-Q.](image)
used regression analyses to show that 2 different types of questionnaires (checklist and global), when administered to 260 women, explained 6.4% and 5.0% of DLW-derived AEE, respectively. Neuhouser et al. (20) reported that the Women’s Health Initiative Personal Habits Questionnaire, the Arizona Activity Frequency Questionnaire, and the 7-Day Physical Activity Recall, administered to 450 women, explained 7.6%, 4.8%, and 3.4% of DLW-derived AEE, respectively.

While our body mass index-adjusted analysis controlled for variance due to misreporting that may be associated with body mass index, it may not have sufficiently accounted for the known tendency for body mass or lean body mass to inflate the association between questionnaire- and DLW-derived AEE (33). We therefore also divided AEE by kilograms of body weight, but we observed only a negligible decrease in the $\text{CCC}_{\text{adj}}$ (from 0.21 to 0.19), whereas larger decreases occurred with Spearman correlations (from

### Table 4. Agreement and Correlations (MET-hours/day) Between STAR-Q1a and 7-Day Diary Activity, by Intensityb and Domain, in a Validity Assessment of the STAR-Q (n = 99), Alberta, Canada, 2009–2011

<table>
<thead>
<tr>
<th>Median Difference, STARQ1 – Diary (IQR)</th>
<th>Spearman Correlation Coefficient ($r_s$)</th>
<th>ICC</th>
<th>95% CI</th>
<th>Unadjusted CCCd</th>
<th>95% CI</th>
<th>Adjusted CCCf</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEE, kcal/dayg</td>
<td>–98.2 (615.2)</td>
<td>0.74</td>
<td>0.73</td>
<td>0.62, 0.81</td>
<td>0.73</td>
<td>0.61, 0.81</td>
<td>0.51</td>
</tr>
<tr>
<td>AEE, kcal/dayfh</td>
<td>–81.7 (540.5)</td>
<td>0.61</td>
<td>0.55</td>
<td>0.40, 0.67</td>
<td>0.55</td>
<td>0.36, 0.73</td>
<td>0.45</td>
</tr>
<tr>
<td>AEE, kcal/kg-dayfh</td>
<td>–1.3 (7.8)</td>
<td>0.45</td>
<td>0.42</td>
<td>0.25, 0.57</td>
<td>0.42</td>
<td>0.24, 0.60</td>
<td>0.41</td>
</tr>
<tr>
<td>Sleepingf</td>
<td>–0.8 (0.8)</td>
<td>0.62</td>
<td>0.18</td>
<td>–0.02, 0.36</td>
<td>0.37</td>
<td>0.24, 0.49</td>
<td>0.37</td>
</tr>
<tr>
<td>Stair-climbing, flights/dayi</td>
<td>0.1 (8.5)</td>
<td>0.41</td>
<td>0.37</td>
<td>0.19, 0.53</td>
<td>0.36</td>
<td>0.16, 0.52</td>
<td>0.36</td>
</tr>
<tr>
<td>Active sittinggh</td>
<td>0.0 (0.4)</td>
<td>0.30</td>
<td>0.30</td>
<td>0.11, 0.47</td>
<td>0.32</td>
<td>0.14, 0.47</td>
<td>0.33</td>
</tr>
<tr>
<td>Overall activityi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedentary</td>
<td>–3.9 (4.9)</td>
<td>0.40</td>
<td>0.12</td>
<td>–0.08, 0.31</td>
<td>0.26</td>
<td>0.11, 0.38</td>
<td>0.27</td>
</tr>
<tr>
<td>Lightg</td>
<td>0.6 (7.3)</td>
<td>0.26</td>
<td>0.26</td>
<td>0.07, 0.43</td>
<td>0.29</td>
<td>0.07, 0.47</td>
<td>0.22</td>
</tr>
<tr>
<td>Moderateg</td>
<td>0.2 (5.0)</td>
<td>0.57</td>
<td>0.49</td>
<td>0.33, 0.63</td>
<td>0.50</td>
<td>0.32, 0.64</td>
<td>0.50</td>
</tr>
<tr>
<td>Strenuousg</td>
<td>0.0 (3.3)</td>
<td>0.68</td>
<td>0.65</td>
<td>0.52, 0.75</td>
<td>0.66</td>
<td>0.54, 0.79</td>
<td>0.64</td>
</tr>
<tr>
<td>Exercise, sports, and leisure activityjh</td>
<td>–0.2 (4.2)</td>
<td>0.47</td>
<td>0.45</td>
<td>0.28, 0.59</td>
<td>0.43</td>
<td>0.24, 0.59</td>
<td>0.42</td>
</tr>
<tr>
<td>Lightg</td>
<td>0.0 (0.3)</td>
<td>0.21</td>
<td>0.27</td>
<td>0.08, 0.44</td>
<td>0.28</td>
<td>0.06, 0.49</td>
<td>0.29</td>
</tr>
<tr>
<td>Moderateg</td>
<td>0.0 (0.8)</td>
<td>0.35</td>
<td>0.36</td>
<td>0.18, 0.52</td>
<td>0.35</td>
<td>0.13, 0.53</td>
<td>0.35</td>
</tr>
<tr>
<td>Strenuousg</td>
<td>0.0 (3.4)</td>
<td>0.49</td>
<td>0.48</td>
<td>0.31, 0.62</td>
<td>0.48</td>
<td>0.33, 0.62</td>
<td>0.48</td>
</tr>
<tr>
<td>Occupational activityph</td>
<td>1.4 (3.6)</td>
<td>0.71</td>
<td>0.70</td>
<td>0.58, 0.79</td>
<td>0.71</td>
<td>0.53, 0.85</td>
<td>0.69</td>
</tr>
<tr>
<td>Sittingp</td>
<td>0.3 (2.2)</td>
<td>0.75</td>
<td>0.70</td>
<td>0.58, 0.79</td>
<td>0.71</td>
<td>0.59, 0.82</td>
<td>0.69</td>
</tr>
<tr>
<td>Sedentary</td>
<td>0.1 (1.8)</td>
<td>0.76</td>
<td>0.75</td>
<td>0.65, 0.82</td>
<td>0.75</td>
<td>0.63, 0.85</td>
<td>0.76</td>
</tr>
<tr>
<td>Lightp</td>
<td>0.6 (3.0)</td>
<td>0.24</td>
<td>0.08</td>
<td>–0.12, 0.27</td>
<td>0.19</td>
<td>0.04, 0.36</td>
<td>0.16</td>
</tr>
<tr>
<td>Moderatep</td>
<td>0.0 (0.3)</td>
<td>0.30</td>
<td>0.50</td>
<td>0.34, 0.63</td>
<td>0.46</td>
<td>0.19, 0.65</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Abbreviations: AEE, activity energy expenditure; CCC, concordance correlation coefficient; CI, confidence interval; ICC, intraclass correlation coefficient; IQR, interquartile range; MET, metabolic equivalents of task; STAR-Q, Sedentary Time and Activity Reporting Questionnaire; TEE, total energy expenditure.

a The first STAR-Q (STAR-Q1) was completed 14 days after doubly labeled water dosing (day 0).
b Intensity categories were defined as follows: sedentary (excluding sleep), ≤1.5 METs; light, >1.5–≤3 METs; moderate, >3–≤6 METs; strenuous, >6 METs. Strenuous occupational activity was not analyzable because of insufficient data.
c $P < 0.001$ (2-sided t-test) for Spearman correlations unless otherwise indicated.
d Not adjusted for any variable.
e 95% CIs were estimated by means of bootstrapping for both adjusted and unadjusted CCCs.
f Adjusted for sex, age, and body mass index at baseline, as well as season of STAR-Q completion.
g Data were log-transformed for ICC and CCC analyses.
h Excludes sleep.
i Excludes napping.
j Results are shown for climbing up stairs (self-reported number of flights per day); results for going down stairs were very similar.
k Represents the sum of all activities performed while sitting and expending energy at >1.5–≤2.5 METs.
l $P = 0.003$.
m $P = 0.009$.
n $P = 0.04$.
o MET level ≤ 2.5.
p $P = 0.02$.  

Our results from the reliability study indicate that at the group level, past-month assessments of AEE can provide reasonable estimates of "usual" AEE. Although we adjusted for the effects of season in our CCC analyses, its impact was found to be negligible; yet we do not know whether true AEE was stable across assessment periods. Fluctuations in physical activity behavior have been associated with variations in temperature and weather conditions (22, 23, 34). It is possible that persons who are active in one season may remain active year-round by participating in season-appropriate activities.

Limitations of our study need to be acknowledged. First, we were unable to measure resting metabolic rate directly and instead used a standard equation to estimate it. While we accounted for body weight, age, and sex, other factors that predict resting metabolic rate (e.g., body composition) could have affected the validity of DLW-derived AEE (since $AEE_{DLW} = \frac{TEE_{DLW} \times 0.90}{(RMR/24) \times (24 - \text{hours of sleep}) + [0.9(RMR/24) \times \text{hours of sleep}]}$). Second, we used 7-day diaries as a criterion measure for validating domain- and intensity-specific estimates of AEE and activity, but we acknowledge that the diary method is an imperfect gold standard. Third, these results can only be generalized to middle-aged Canadian adults who are generally educated, employed, and physically active. Fourth, since ICCs are lower when between-subject variability is low relative to measurement error (whether random or systematic) (35), it is possible that the validity and reliability of the STAR-Q were underestimated for some domains and activities with restricted variability (e.g., sleeping, sedentary behavior) (Table 4). Fifth, our analyses did not quantify subject-related sources of measurement error, as has been reported by others (19, 20), since we felt it was beyond the scope of the present study.

In summary, using rigorous methods, we showed that the STAR-Q on average had fair validity for ranking individuals by overall AEE and was reliable over a period of several months. Our study also highlighted systematic measurement error that was proportional to AEE, which should not be ignored in the design and analysis of future etiological studies that use the STAR-Q. The comprehensive design of the STAR-Q allows for the assessment of all domains of activity at various intensities, including sedentary behavior. The STAR-Q is designed to meet the needs of these evolving, hybrid research methods.

**ACKNOWLEDGMENTS**

Author affiliations: Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, Alberta, Canada (Ilona Csizmadi; 433)
REFERENCES


