The Doubly Labeled Water Method Produces Highly Reproducible Longitudinal Results in Nutrition Studies


USDA/Agricultural Research Service Children’s Nutrition Research Center, Baylor College of Medicine, Houston, TX; Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA; Washington University School of Medicine, St. Louis, MO; Pennington Biomedical Research Center, Baton Rouge, LA; and Duke Clinical Research Institute, Duke University, Durham, NC

Abstract

The doubly labeled water (DLW) method is considered the reference method for the measurement of energy expenditure under free-living conditions. However, the reproducibility of the DLW method in longitudinal studies is not well documented. This study was designed to evaluate the longitudinal reproducibility of the DLW method using 2 protocols developed and implemented in a multicenter clinical trial—the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE). To document the longitudinal reproducibility of the DLW method, 2 protocols, 1 based on repeated analysis of dose dilutions over the course of the clinical trial (dose-dilution protocol) and 1 based on repeated but blinded analysis of randomly selected DLW studies (test-retest protocol), were carried out. The dose-dilution protocol showed that the theoretical fractional turnover rates for $^2H$ and $^{18}O$ and the difference between the 2 fractional turnover rates were reproducible to within 1% and 5%, respectively, over 4.5 y. The Bland-Altman pair-wise comparisons of the results generated from 50 test-retest DLW studies showed that the fractional turnover rates and isotope dilution spaces were highly reproducible over 2.4 y. Our results show that the DLW method is reproducible in longitudinal studies and confirm the validity of this method to measure energy expenditure, define energy intake prescriptions, and monitor adherence and body composition changes over the period of 2.5–4.4 y.

The 2 protocols can be adopted by other laboratories to document the longitudinal reproducibility of their measurements to ensure the long-term outcomes of interest are meaningful biologically. This trial was registered at clinicaltrials.gov as NCT00427193. J. Nutr. 144: 777–783, 2014.

Introduction

The doubly labeled water (DLW) method was originally developed and validated for measuring energy utilization in small mammals (1–9). Following the validation of the DLW method for measuring energy expenditure (EE) against indirect calorimetry in humans (10–16), the method quickly became the reference method for measuring habitual EE in humans, including premature infants (17,18), newborns (19,20), children (21), adolescents (22,23), pregnant women (24,25), lactating women (26,27), and adults (28,29), as well as individuals with various diseases (30–33). The DLW method is noninvasive and does not require blood sampling. The method also has minimal participant burden and can be used anywhere. Briefly, the DLW method is based on the principle that the disappearance rate of the heavier stable isotope of hydrogen ($^2H$) reflects water turnover rate, whereas the disappearance rate of the heavier stable isotope of oxygen ($^{18}O$) reflects both water and CO$_2$ turnover rates. Therefore, with time, the difference between the 2 fractional turnover rates were reproducible to within 1% and 5%, respectively, over 4.5 y. The Bland-Altman pair-wise comparisons of the results generated from 50 test-retest DLW studies showed that the fractional turnover rates and isotope dilution spaces were highly reproducible over 2.4 y. Our results show that the DLW method is reproducible in longitudinal studies and confirm the validity of this method to measure energy expenditure, define energy intake prescriptions, and monitor adherence and body composition changes over the period of 2.5–4.4 y. The 2 protocols can be adopted by other laboratories to document the longitudinal reproducibility of their measurements to ensure the long-term outcomes of interest are meaningful biologically. This trial was registered at clinicaltrials.gov as NCT00427193.
reproducibility of the DLW method, which is critical for longitudinal studies to monitor changes in EE, energy intake, and body composition, has not been documented.

The objective of this study was to evaluate the longitudinal reproducibility of the DLW method based on 2 protocols: 1 for the study dose dilutions and 1 for the test-retest reliability, which were developed and implemented in the National Institute on Aging’s multicenter clinical trial, the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE).

Participants and Methods

Study design

CALERIE was a multicenter, parallel-group, randomized controlled clinical trial conducted between 2005 and 2012 at the Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Washington University School of Medicine, and Pennington Biomedical Research Center. Duke Clinical Research Institute served as the coordinating center and the gas-isotope-ratio MS laboratory at Baylor College of Medicine’s USDA/Agricultural Research Service Children’s Nutrition Research Center served as the central DLW laboratory. The design and conduct of the clinical trial were published elsewhere (39). Study participants provided written informed consent. The institutional review board for human studies at each participating institution approved the study’s protocol.

Healthy adults were recruited for the 2-y clinical trial. Eligible participants were between 20 and 50 y of age (men) or between 20 and 47 y of age (women), either of normal body weight or slightly overweight (BMI of ≥22.0 and <28.0 kg·m⁻²), nonsmoking, nondiabetic, and not on any medications. Female participants had to use an acceptable form of contraception during the clinical trial. A total of 238 participants began the baseline evaluations, of which 218 were randomly assigned, with a 2:1 allocation to the caloric restriction (CR) and control groups, respectively, and were provided at least 1 follow-up evaluation (40). A total of 191 participants provided complete follow-up data.

DLW method

The DLW method was used to determine the habitual energy intake of each study participant. Two consecutive 14-d DLW protocols were developed and implemented in the CALERIE multicenter clinical trial. The DLW method was used to determine the habitual energy intake of healthy adults recruited for the 2-y clinical trial. Eligible participants were between 20 and 50 y of age (men) or between 20 and 47 y of age (women), either of normal body weight or slightly overweight (BMI of ≥22.0 and <28.0 kg·m⁻²), nonsmoking, nondiabetic, and not on any medications. Female participants had to use an acceptable form of contraception during the clinical trial. A total of 238 participants began the baseline evaluations, of which 218 were randomly assigned, with a 2:1 allocation to the caloric restriction (CR) and control groups, respectively, and were provided at least 1 follow-up evaluation (40). A total of 191 participants provided complete follow-up data.

For each DLW period, 2 baseline urine samples were collected. The participant was then administered by mouth a mixed cocktail containing 0.1 g of H₂O₂ at 99.98 atom percent H and 0.16 g of 100% ¹⁸O per kg of body weight. The DLW dose was designed to minimize potential errors introduced by the anticipated fluctuation in natural abundances of the 2 isotopes during the CR intervention, to reduce the effect of analytical errors on the precision of the DLW method, and to ensure there were sufficient isotopes at the end of each 14-d DLW study period for accurate and precise isotope ratio measurements (41–43). Six postdose urine samples were collected: 2 at 5–6 h postdose, 2 on day 7, and 2 on day 14. Study participants were instructed to void at home in the morning on days 7 and 14 before the postdose urine samples were collected in the clinic. The exact time of dosing and sample collection times were recorded.

Urine samples were transferred to 3 sets of o-ring cryovials. Encrypted ID labels, created and printed by the coordinating center, were affixed by site personnel to the cryovials in all follow-up DLW studies to ensure the DLW Laboratory was unaware of the treatment assignment and participant ID. One set of cryovials was shipped on dry ice to the DLW Laboratory for isotope ratio measurements.

For ¹H assays, 10 µL of urine without further treatment was converted to H₂ using the zinc reduction method (36,37). The H₂ was introduced via the automated sample inlet system directly into a Finnigan instrument for hydrogen isotope ratio measurement. For ¹⁸O assays, ISOPREP-18 H₂O-CO₂ equilibrium chambers were used, in which 100 µL of urine was equilibrated with 300 mbar of CO₂ of known ¹³C/O content for 10 h prior to admission to the ion source of a VG instrument for oxygen isotope ratio measurement (36). The isotope ratio measurements were expressed in delta (δ) per mil (parts per 1,000 or ‰) as follows:

\[ \delta^{18}O = \frac{R_{\text{sample}} - 1}{R_{\text{standard}} - 1} \times 10^3, \]

where \( R_{\text{Sample}} \) and \( R_{\text{Standard}} \) were the \(^{2}H/H\) or \(^{18}O/^{16}O\) isotope ratios of the sample or the laboratory working standard, respectively. The isotope ratios were then normalized against 2 international water standards: Vienna-Standard Mean Ocean Water and Standard Light Antarctic Precipitation (44). The precision (SD) for the ²H assay was 1.0‰ for samples with natural abundance of ²H and 1.8‰ for samples with enriched amounts of ²H (37). For ¹⁸O assays, the precision was 0.21‰ for samples with natural abundance of ¹⁸O and 0.97‰ for samples with enriched amounts of ¹⁸O (36).

The isotope dilution space of ²H (N₉₁) and ¹⁸O (N₃₀) was calculated as follows:

\[ N_{H} \text{ or } N_{O} \text{ (mol)} = \frac{\delta^{2}H \times A \times E_{a}}{E_{d} \times 18.02}. \]

where \( d \) was the dose of ²H₂O or ²H₂¹⁸O in grams, \( A \) was the amount of laboratory water in grams used in the dose dilution, \( \alpha \) was the amount of ²H₂O or ²H₂¹⁸O in grams added to the laboratory water in the dose dilution, \( E_{a} \) was the rise in δ²H or δ¹⁸O values in the laboratory water after the addition of the isotopic water, and \( E_{d} \) was the rise in δ²H or δ¹⁸O values in the urine samples at time zero obtained from the zero-time intercepts of the ²H and ¹⁸O decay curves in the urine samples. The use of dose dilution in the calculation of isotope dilution spaces was recommended by the International Dietary Energy Consultancy Groups to ensure accuracy of the isotope dilution calculations (45). Carbon dioxide production rate (VCO₂) was calculated from the fractional turnover rate of ²H (k₁) and ¹⁸O (k₃) as follows (46):

\[ \dot{V}CO_{2} \text{ (mol d}^{-1}) = 0.4812 \times \left[ \frac{(k_{3} \times N_{O}) - (k_{1} \times N_{H})}{k_{3} - k_{1}} \right] - 0.0246 \times r_{g}, \]

where \( r_{g} \) was the fractionated water loss, which was calculated as 1.05 \( \times \) \( N_{O} \times k_{O} - N_{H} \times k_{H} \). The VCO₂ was converted to EE based on an energy equivalent of 1 L of CO₂ to be 3.815/RQ + 1.2321 (16), where RQ was the respiratory quotient provisionally estimated to be 0.86 for all DLW measurements in this study (47).

Longitudinal reproducibility of the DLW method

To assess the longitudinal reproducibility of the DLW method, 2 protocols were developed and implemented in the CALERIE multicenter clinical trial.

Dose-dilution protocol. Two dose dilutions that spanned the range of isotopic enrichments anticipated at 5–6 h postdose (at 1:400 dilution) and at 14-d postdose (at 1:1300 dilution) were prepared from the DLW dose mixture used in the CALERIE clinical trial. Sufficient quantities of the 2 dose dilutions, along with the laboratory water that was used to prepare the dose dilutions, were stored in leak-proof containers at 5°C for the duration of the clinical trial. Initially, the 2 dose dilutions and the laboratory water were analyzed 10 times each for ²H and ¹⁸O content each day for 10 d. The mean values were used to generate the conversion constants to convert the monthly ²H and ¹⁸O measurements of the dose dilutions and the laboratory water into the theoretical fractional turnover rates of 0.1 for ²H and 0.13 for ¹⁸O. The conversion constants were calculated as follows:

\[ C_{\delta_{H}} \text{ or } C_{\delta_{O}} = \frac{\ln(\delta_{E_{1:400}}) - \ln(\delta_{E_{1:1300}})}{0.1 \text{ or } 0.13}, \]

where \( C_{\delta_{H}} \) and \( C_{\delta_{O}} \) were the conversion constants for ²H and ¹⁸O, respectively; and \( \delta_{E_{1:400}} \) and \( \delta_{E_{1:1300}} \) were the ²H and ¹⁸O content of the dose dilution above the isotopic content of the laboratory water.
anticipated at 5–6 h and at 14-d postdose, respectively. To monitor the reproducibility of the $^3$H and $^{18}$O measurements over the course of the clinical trial, monthly measurements of the 2 dose dilutions and the laboratory water were performed and the values were converted to $k_\text{H}$ and $k_\text{O}$ values using the respective conversion constant ($C_\text{H}$ or $C_\text{O}$). The percentage difference of the $k_\text{H}$, $k_\text{O}$, and $k_\text{O}-k_\text{H}$ values generated from the monthly measurements of the 2 dose dilutions and the laboratory water with respect to the theoretical values of 0.10, 0.13, and 0.03, respectively, was plotted against the date of analysis to monitor the long-term reproducibility of these measurements.

**Test-retest protocol.** All DLW studies that were performed postrandomization from participants in both study arms were eligible for the test-retest protocol. Baseline studies were excluded because of the requirement to provide baseline total energy expenditure values quickly to the clinical sites so that the correct energy prescription could be determined. Sample size calculations (48) indicated that a sample of at least 46 duplicate DLW studies would be required to demonstrate that the intra-class correlation was $>0.8$ with a type-I error of $\alpha = 0.05$ and type-II error of $\beta = 0.2$. Thus, 50 DLW studies, or $\sim 8\%$ of postrandomization DLW studies, were included.

At periodic intervals in calendar time, DLW studies were selected for the study. The goal was to select them when $\sim 120$ new postrandomization DLW studies had been performed since the previous calendar point. However, because of administrative issues, the samples were actually selected at 4 time points when 427, 84, 116, and 10 additional samples had accumulated. The first sampling was delayed to allow the DLW laboratory to focus on the baseline studies. Moreover, because more samples were found to be ineligible than expected, the sample rate was increased toward 15% by the end of the study to meet the required study size. Samples were selected from all new postrandomization DLW studies using simple random sampling by the statistician at the coordinating center. The selection was stratified by site but not by treatment assignment. DLW studies were not performed in the control group at months 6 and 18, because this might provide the DLW laboratory with knowledge of treatment groups, participant ID, and protocol time point so that only the sample collection sequence was identified. The clinical sites retrieved the duplicate urine sample sets from their freezers, affixed the blinded labels, and forwarded them to the DLW laboratory for analysis. When the mass spectrometric measurements were completed, the isotopic data were submitted to the coordinating center, and the study site then forwarded the study information to the DLW laboratory to generate the DLW outcome variables $k_\text{H}$, $k_\text{O}$, $N_\text{H}$, $N_\text{O}$, and EE.

**Statistical methods**

Descriptive statistics were used to calculate the mean, SD, and range of the participants’ physical characteristics, the percentage difference from the theoretic $k_\text{H}$, $k_\text{O}$, and $k_\text{O}-k_\text{H}$ values under the dose-dilution protocol, and the mean, mean difference, and the corresponding SD and range of the DLW outcome variables under the test-retest protocol. Independent samples $t$ test and chi-square test were used to compare the continuous variables and categorical variables, respectively, between the retest participants and the nonretest participants. Paired samples $t$ test was used to compare the test-retest outcome variables. The Bland-Altman pair-wise comparison (49,50) was used to evaluate the reproducibility of the test-retest results. Statistical analyses were performed with SPSS software (SPSS).

**Results**

**Longitudinal reproducibility of the DLW method**

**Dose-dilution protocol.** The longitudinal reproducibility of the DLW method based on the dose-dilution protocol is summarized in Figure 1. Fig. 1A illustrates the reproducibility of the $k_\text{H}$ values over a period of 4.4 y and, as shown, the $k_\text{H}$ values generated from the dose-dilution measurements over 4.4 y were within 1% of the theoretic value of 0.10 for $k_\text{H}$, with a mean difference of 0.11 $\pm$ 0.25% (mean $\pm$ SD; range: $-0.47$–$0.81\%$). Fig. 1B illustrates the reproducibility of the $k_\text{O}$ values over 4.4 y and, as shown, the $k_\text{O}$ values generated from the dose-dilution measurements were within 1% of the theoretic value of 0.13 for $k_\text{O}$ with a mean difference of 0.06 $\pm$ 0.33% (range: $-0.76$–$0.76\%$). The reproducibility of the difference between $k_\text{O}$ and $k_\text{H}$ is summarized in Fig. 1C. As shown in the figure, the $k_\text{O}$–$k_\text{H}$ values were reproducible within 5% of the theoretic value of 0.03 with a mean difference of $-0.06 \pm 0.19\%$ (range: $-3.63$–$4.12\%$).

**Test-retest protocol.** The demographic and baseline physical characteristics of the participants who were randomly selected under the test-retest protocol and the nonretest participants are summarized in Table 1. The follow-up DLW studies ($n = 50$)
used in the test-retest protocol were obtained from 46 randomly selected participants, because 4 participants had 2 DLW studies that were carried out at different time points of the clinical trial. The demographic and baseline characteristics of the 46 participants selected for the test-retest protocol were not different from the 172 participants who were not selected for the test-retest protocol.

Table 2 provides the descriptive statistics for the DLW outcome variables obtained from the 50 DLW studies randomly selected for the test-retest protocol. The original fractional turnover rates for $^2$H and $^{18}$O ranged from $-0.176$ to $-0.058$ d$^{-1}$ and from $-0.201$ to $-0.081$ d$^{-1}$, respectively. The original isotope dilution spaces for $^2$H and $^{18}$O also ranged from 27.7 to 49.7 kg and from 26.8 to 47.9 kg, respectively. With respect to EE, the original values ranged from 1561 to 3675 kcal d$^{-1}$. Therefore, the 50 DLW studies randomly selected for the test-retest protocol provided a wide range of these measurements to fully evaluate the longitudinal reproducibility of the DLW method. A paired-samples t test showed that the differences between the original tested and retested fractional turnover rates ($k_{_{2H}}$ and $k_{^{18}O}$) were significant ($P = 0.02$). However, none of the differences between the original tested and retested values for $N_{^{2H}}$, $N_{^{18}O}$, and EE was found to be significant ($P \geq 0.3$). The small differences observed between the original tested and retested values for $k_{_{2H}}$ and $k_{^{18}O}$ are considered physiologically irrelevant because no significant difference was observed among the major DLW outcomes (isotope dilution spaces and EE), which were derived from these fractional turnover rates.

Figure 2 summarizes the Bland-Altman pair-wise comparisons between the retested DLW outcome variables and the original tested values. Fig. 2A shows that the retested $k_{_{2H}}$ values, when compared with the original values, had a bias of 0.0004 d$^{-1}$ with a lower and upper limit of agreement between $-0.002$ and 0.003 d$^{-1}$, respectively. With the exception of 1 data point, the rest of the differences were within the lower and upper limit of agreement. The comparison between the retested $k_{^{18}O}$ values and the original values showed a bias of 0.0005 d$^{-1}$, with a lower and upper limit of agreement between $-0.003$ and 0.004 d$^{-1}$, respectively (Fig. 2B). Again, with the exception of 1 data point, the rest of the differences all fall within the limit of agreement. When compared with the original values, the retested $N_{^{2H}}$ values (Fig. 2C) had a bias of $-0.1$ kg with a lower and upper limit of agreement between $-2.1$ and 2.0 kg, respectively. The individual differences again fall within the limit of agreement, with the exception of 1 data point. Similar results were obtained for the retested $N_{^{18}O}$ values (Fig. 2D), with a bias of $-0.1$ kg and a lower and upper limit of agreement between $-1.8$ and 1.7 kg, respectively. For the retested EE values (Fig. 2E), a bias of $-5$ kcal d$^{-1}$, with a lower and upper limit of agreement between $-148$ and 137 kcal d$^{-1}$, respectively, was observed. With the exception of 1 data point, the other differences were all within the limit of agreement. Regression analyses indicated that the differences in $k_{_{2H}}$ and $k_{^{18}O}$

### Table 1 Demographic and baseline physical characteristics of participants in the CR clinical trial

<table>
<thead>
<tr>
<th>Variable</th>
<th>Test-retest participants (n = 46)</th>
<th>Nonretest participants (n = 172)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>39.7 ± 6.6 (22.9–50.6)</td>
<td>37.5 ± 7.3 (20.7–50.8)</td>
<td>0.06</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td>Male</td>
<td>18 (39.1)</td>
<td>48 (27.9)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>28 (60.9)</td>
<td>124 (72.1)</td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td>White</td>
<td>34 (73.9)</td>
<td>134 (77.9)</td>
<td></td>
</tr>
<tr>
<td>Nonwhite</td>
<td>12 (26.1)</td>
<td>38 (22.1)</td>
<td></td>
</tr>
<tr>
<td>Ethnic group, n (%)</td>
<td></td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>2 (4.3)</td>
<td>5 (2.9)</td>
<td></td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>44 (95.7)</td>
<td>165 (95.9)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0.0)</td>
<td>2 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Weight, kg</td>
<td>73.5 ± 9.5 (52.8–97.7)</td>
<td>71.3 ± 9.1 (51.8–97.5)</td>
<td>0.14</td>
</tr>
<tr>
<td>Height, cm</td>
<td>167.9 ± 9.0 (153.2–191.4)</td>
<td>168.4 ± 8.4 (147.7–195.5)</td>
<td>0.38</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>25.5 ± 1.7 (22.0–28.5)</td>
<td>25.1 ± 1.7 (21.3–29.0)</td>
<td>0.14</td>
</tr>
<tr>
<td>BMI status, n (%)</td>
<td></td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>Normal weight</td>
<td>21 (45.7)</td>
<td>84 (48.8)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>25 (54.3)</td>
<td>88 (51.2)</td>
<td></td>
</tr>
</tbody>
</table>

1 Values are means ± SDs (ranges) or n (%). P values by independent samples t test for continuous variables and chi-square test for categorical variables. CR, caloric restriction.

### Table 2 DLW outcome variables calculated from samples collected from human studies randomly selected from the CR clinical trial under the test-retest protocol

<table>
<thead>
<tr>
<th>Variables</th>
<th>Original</th>
<th>Retest</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{_{2H}}$ d$^{-1}$</td>
<td>$-0.098$ ± $0.024$ (−0.176 to −0.058)</td>
<td>$-0.098$ ± $0.024$ (−0.178 to −0.058)</td>
<td>$0.000$ ± $0.001$ (−0.007–0.001)</td>
</tr>
<tr>
<td>$k_{^{18}O}$ d$^{-1}$</td>
<td>$-0.122$ ± $0.025$ (−0.201 to −0.081)</td>
<td>$-0.122$ ± $0.026$ (−0.202 to −0.081)</td>
<td>$0.001$ ± $0.002$ (−0.011–0.001)</td>
</tr>
<tr>
<td>$N_{^{2H}}$ kg</td>
<td>35.7 ± 6.1 (27.7–49.7)</td>
<td>35.8 ± 6.4 (27.7–51.8)</td>
<td>$-0.1$ ± $1.0$ (−0.9–7.0)</td>
</tr>
<tr>
<td>$N_{^{18}O}$ kg</td>
<td>34.4 ± 5.9 (26.8–47.9)</td>
<td>34.5 ± 6.1 (26.8–49.1)</td>
<td>$-0.1$ ± $0.9$ (−1.2–5.9)</td>
</tr>
<tr>
<td>EE, kcal d$^{-1}$</td>
<td>2242 ± 407 (1561–3675)</td>
<td>2247 ± 397 (1584–3622)</td>
<td>$-5$ ± $73$ (−118–161)</td>
</tr>
</tbody>
</table>

1 Values are means ± SDs (ranges), n = 50. CR, caloric restriction; DLW, doubly labeled water; EE, energy expenditure; $k_{_{2H}}$, fractional turnover rate of $^2$H; $k_{^{18}O}$, fractional turnover rate of $^{18}$O; $N_{^{2H}}$, isotope dilution space of $^2$H; $N_{^{18}O}$, isotope dilution space of $^{18}$O.
Bland-Altman pair-wise comparison between the retested $\text{N}_2$ values and the originally calculated $\text{N}_2$ values (A); Bland-Altman pair-wise comparison between the retested EE values and the originally calculated EE values (B). The solid line within each panel represents zero difference. The dotted line within each panel represents the bias or mean difference between the retest and original values. The 2 dashed lines within each panel represent the 95% CIs of the bias. The symbols within each panel represent the individual difference between the retest and difference. The dotted line within each panel represents the bias or mean difference between the retest and original values. The 2 dashed lines isotope dilution space of $^2\text{H}$; $\text{N}_2$, isotope dilution space of $^{18}\text{O}$.

were inversely related to the mean $k_H$ and $k_O$ values ($r^2 = 0.09$, $P \leq 0.04$). No relation was detected between the differences and the mean values among the comparisons of the $N_H$ ($P = 0.07$), $N_O$ ($P = 0.17$), and EE ($P = 0.32$) measurements.

Among the 50 DLW studies chosen for the retest protocol, 2 studies were found to provide outcomes significantly different from the original values. Repeated MS analyses on the samples yielded the original values. Repeated MS analyses on the samples yielded the original values. Referring to the CALERIE clinical trial among study participants who were randomly assigned to the CR intervention.

### Adherence monitoring

Adherence measures were used to determine the degree of CR actually achieved. Adherence was characterized as the percentage of CR achieved and was calculated as follows:

$$%\text{CR} = 100[1 - (\text{EI}/\text{EI}_{\text{AL}})]$$

where $\text{EI}_{\text{AL}}$ represents mean daily energy intake over the period of interest and $\text{EI}_{\text{AL}}$ represents the ad libitum daily energy intake before the start of the intervention. Ad libitum energy intake was characterized by the mean of 2 consecutive measures of EE performed at baseline using the DLW method. Based on the relation, $\text{EI} = \text{EE} + \Delta\text{ES}$, where $\text{EE}$ was the mean daily energy expenditure during the period of interest and $\Delta\text{ES}$ was the change

### Applications of the DLW method in the CALERIE clinical trial

**CR prescription.** In the CALERIE clinical trial, 2 consecutive 14-d DLW studies were carried out at baseline to determine the EE of each study participant and to establish the CR prescription for those randomly assigned to the CR intervention. Because these participants were healthy and were not taking part in any dietary or physical activity programs to lose weight, the mean EE measurements derived from these 2 consecutive DLW studies were assumed to equal their ad libitum energy intakes. The pre-intervention energy intakes of 10 participants, 5 males and 5 females, who were assigned to the CR intervention in the CALERIE clinical trial with best adherence are summarized in Table 3. The 25% CR prescriptions were calculated as 75% of ad libitum energy intake as determined by the DLW method.
in body energy stores during the period of interest. For intervals between 2 DLW measures, EE was computed as the mean of the EE estimates across the 2 time points. For intervals spanning more than 2 DLW measures, the mean of the estimates for each interval, weighted by the duration of the interval, was applied. \( \Delta ES \) was estimated by calculating the change in energy stores (measured by dual energy x-ray absorptiometry) from the beginning to the end of the interval. \( \Delta ES \) was calculated using standard coefficients for changes in fat mass (FM) (FM: 9300 kcal/kg) and fat-free mass (FFM) (FFM: 1100 kcal/kg). The EE, \( E_{IS} \), \( \Delta ES \), and %CR for 10 participants who were assigned to the CR intervention over a 6-mo period are summarized in Table 3.

**Body compositional changes.** Isotope dilution has been long considered one of the reference methods for the measurements of body composition. It has been well documented that FFM is one of the reference methods for the measurements of body composition. It has been well documented that FFM in healthy adults has a hydration of 75% (51). Knowing the \( N_0 \) from the DLW protocol, total body water (TBW) can be calculated using the equation \( TBW = N_0/1.01 \). Therefore, FFM can be calculated from TBW using the equation \( FFM = TBW/0.73 \). FFM is simply the difference between body weight and FFM. The changes in body composition (body weight, FFM, FM) among 10 participants who were assigned to the CR group over a 6-mo period are summarized in Table 3.

**Discussion**

Our results represent the first study to document the longitudinal reproducibility of the DLW method.

The DLW method is considered the reference method for EE measurements under free-living conditions because it is noninvasive, nonrestrictive with minimal participant burden, and has no known adverse effects. The other advantage of the DLW method is that it can be implemented almost anywhere and the samples can be shipped back to the analytical laboratory. Because both \(^2\text{H}\) and \(^18\text{O}\) are nonradioactive stable isotopes, they do not decay or emit harmful radiation and therefore can be kept for a long time under proper conditions to support longitudinal studies. As shown in Fig. 1, the DLW method was highly reproducible over a period of 4.4 y. The results also demonstrated that the isotope ratio measurements by gas-isotope-ratio MS were highly reproducible. The longitudinal reproducibility of the DLW method was further supported by the results obtained from the blinded test-retest protocol (Fig. 2), showing that the results were highly reproducible up to 2.5 y.

One previous study examined the reliability of the DLW method in 5 participants (52). However, that study was not blinded and the DLW protocol was repeated on the same participants after a 3-d break. Therefore, although that study could be used to evaluate the reliability of the DLW method within participants, it could not be used to evaluate the longitudinal reproducibility of the DLW method.

Unfortunately, the DLW method is not widely used in cross-sectional or longitudinal studies because the method is expensive and requires specialized instrumentation such as isotope ratio MS to measure the stable isotopes. Therefore, other dietary assessment methods such as 24-h dietary recalls and FFQs often are employed in surveys and longitudinal studies. However, these less-expensive methods are known to have large measurement errors, particularly among children, different ethnic groups, and overweight or obese participants (53–56).

Because the reproducibility results were obtained using isotope ratio MS, the results might not be applicable to DLW studies carried out using other instrumentation such as cavity ring-down spectroscopy (57–60) or Fourier transform infrared spectroscopy (61,62). The long-term reproducibility of the DLW method using these other instruments will need to be documented.

Our results demonstrate that the DLW outcome variables are highly reproducible longitudinally. Therefore, other laboratories can use these 2 protocols to document the longitudinal reproducibility of their measurements to ensure the biologic significance of the long-term outcomes of interest.

**Acknowledgments**

All authors were involved in the research design to evaluate the longitudinal reproducibility of the DLW method. S.B. Roberts and S.K.D. were the principal investigators at the Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University. S.B. Racette was the principal co-investigator at Washington University School of Medicine. L.M.R. was the principal coinvestigator at the Pennington Biomedical Research Center. J.R. and W.E.K. were the principal investigators at the Duke Clinical Research Institute, Duke University. M.V.B. was the study database manager at the coordinating center. W.W.W. was the principal investigator of the DLW Laboratory at the USDA/Agricultural Research Service Children’s Nutrition Research Center, Baylor College of Medicine. L.L.C. performed all the MS measurements in the DLW Laboratory. W.W.W. and M.V.B. performed the statistical analysis. W.W.W. wrote the initial draft of the manuscript and had primary responsibility for the final content. All authors read and approved the final manuscript.

**Literature Cited**

15. Howard WA, Prentice AM, Murgatroyd PR, Davies HL, Cole TJ, Saylor KE, Glodberg GR, Halliday D, Macnamara JP. Measurement of \( \text{CO}_2 \) and


