**Brief Report**

Self-Reported Versus Objectively Assessed Exercise Adherence

Hon K. Yuen, Ed Wang, Katy Holthaus, Laura K. Vogtle, David Sword, Hazel L. Breland, Diane L. Kamen

**OBJECTIVE.** We examined agreement of data between self-reported and objectively assessed exercise adherence among women with systemic lupus erythematosus.

**METHOD.** Eleven participants completed weekly exercise logs on date and duration of exercise during a 10-wk Wii Fit™ home-based program. Afterward, exercise data from the log were compared with those recorded in the Wii console.

**RESULTS.** Of the paired data, the mean duration of exercise recorded in the Wii was 29.5 min and that recorded in the log was 33.3 min. The composite intraclass correlation for exercise duration between exercise log and the Wii Fit was 0.4. The 95% limits of agreement indicated large between-subjects variability.

**CONCLUSION.** Exercise logs exhibit a marginally acceptable agreement with Wii estimation of exercise duration at a group level. However, caution should be applied when using the exercise log as a measure of a person's exercise behavior because of the tendency to overreport.


---

Hon K. Yuen, PhD, OTR/L, is Professor and Director of Research, Department of Occupational Therapy, School of Health Professions, University of Alabama at Birmingham, 1530 Third Avenue South, Birmingham, AL 35294; yuen@uab.edu

Ed Wang, PhD, is Professor, School of Health Professions, University of Alabama at Birmingham.

Katy Holthaus is Research Coordinator, Department of Health Sciences and Research, College of Health Professions, Medical University of South Carolina, Charleston.

Laura K. Vogtle, PhD, OTR/L, FAOTA, is Professor and Director, Postprofessional Master’s Program, Department of Occupational Therapy, School of Health Professions, University of Alabama at Birmingham.

David Sword, PT, DPT, CCS, is Associate Professor, Division of Physical Therapy, College of Health Professions, Medical University of South Carolina, Charleston.

Hazel L. Breland, PhD, OTR/L, is Assistant Professor, Division of Occupational Therapy, College of Health Professions, Medical University of South Carolina, Charleston.

Diane L. Kamen, MD, MSCR, is Associate Professor, Division of Rheumatology and Immunology, Department of Medicine, Medical University of South Carolina, Charleston.

---

Home-based exercise training has been shown to result in greater long-term adherence to exercise (Ashworth, Chad, Harrison, Reeder, & Marshall, 2005). Adherence to prescribed home-based exercise programs is typically assessed using a self-report exercise diary or log. For short-term research studies, weekly or biweekly telephone reminders are often used to prompt participants to complete the exercise log or to collect the exercise adherence data (Brubaker et al., 2000; King, Taylor, Haskell, & Debusk, 1988). For long-duration studies, such as 6 mo or longer, weekly or biweekly telephone monitoring or reminders may not be feasible (King et al., 1988).

Despite the common use of exercise logs, compliance with documentation of prescribed exercise remains a big challenge for researchers (Sallis & Saelens, 2000). Delayed recording in the exercise log may lead to inaccurate recall of details such as the date, frequency, duration of exercise or missed recording of the event altogether (Sallis & Saelens, 2000). In addition, some studies have suggested that participants tend to overreport their exercise levels in terms of frequency, duration, and intensity (Sallis & Saelens, 2000; Yore et al., 2007). However, validation of this concern is difficult because no gold standard or true criterion exists to measure these exercise parameters when participants exercise at home.

Wearable motion sensors, such as accelerometers and pedometers, are viewed as a more objective, direct, and accurate way to measure frequency, duration, and intensity of participation in home-based exercise than exercise logs (Trost, McIver, & Pate, 2005). However, these devices have several limitations that could prevent their use in longitudinal studies, including participants not wearing the monitor consistently during the allocated study periods and malfunction of the device, which together constitute 83.4% of noted problems (Perry et al., 2010).

A new-generation “exergame” system, Wii Fit™, has the capacity to keep a record of the date and duration of workouts performed and tracks the user’s progress over time (Anderson, Annett, & Bischof, 2010). These data can serve as a true criterion for adherence to the exergame program. The
purpose of this study was to evaluate the agreement of exercise data from self-report exercise logs with the built-in recorded data in the Wii console among African-American women with systemic lupus erythematosus (SLE).

We targeted a population with SLE because fatigue and physical deconditioning are common among people with SLE. For example, a study reported that more than 80% of study participants experienced clinically significant levels of fatigue (Tench, McCurdie, White, & D’Cruz, 2000). The prevalence of SLE in women is approximately 10 times higher than in men, with a disproportionately higher prevalence of SLE reported in African Americans (Jacobson, Gange, Rose, & Graham, 1997).

Physical exercise has been shown to have multiple benefits for people with SLE, including significant reduction in fatigue levels, weight loss, and improvement in physical fitness (Ayán & Martin, 2007; Strömbeck & Jacobsson, 2007). However, only about one-third of those with SLE meet physical activity goals of exercising ≥150 min/wk or having energy expenditure of ≥1,400 kcal/wk (Mancuso, Perna, Sargent, & Salmon, 2011). The incorporation of exergames such as Wii Fit into home-based exercise programs for people with SLE is believed to help reduce or eliminate the exercise barriers cited in the literature (Mancuso et al., 2011) and sustain motivation and adherence to exercise (Lieberman et al., 2011).

Method

Research Design

This was a methodological study embedded in a one-group pretest–posttest clinical trial. The study was approved by the institutional review board of the Medical University of South Carolina, where the study was conducted. Signed informed consent was obtained from all participants before the trial started.

Participants

We recruited a convenience sample of 15 sedentary African-American women with SLE through the Medical University of South Carolina Lupus Clinic. Recruitment was facilitated through the longitudinal observational SLE Clinic Database and SLE in Gullah Health Database (Kamen et al., 2008) and conducted by a research coordinator. To be eligible for participation in the study, participants had to meet the following inclusion criteria: (1) female, age 18 yr or older; (2) self-identification as African-American; (3) diagnosed with SLE and fulfilling at least four of the revised American College of Rheumatology classification criteria for SLE (Hochberg, 1997); (4) ambulatory; (5) experiencing fatigue for the past 3 mo or longer as indicated by a rating of at least 4 (moderate fatigue) on a Fatigue Visual Analog Scale ranging from 0 to 10, with 10 being the most fatigue (Piper et al., 2008); (6) sedentary (exercise <3 times per week for 30 min in the past 6 mo); (7) functionally literate (i.e., able to read and follow exercise directions in English); and (8) having permission from their physician to participate in the study.

The exclusion criteria were as follows: (1) cognitive dysfunction, as indicated by a score <24 on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975); (2) anemia (with hemoglobin <8 g/dL); (3) poor control of metabolic diseases or other concurrent systemic health problems (e.g., infections, malnutrition), which are known to contribute to increased fatigue levels; (4) known electrolyte abnormalities; (5) documented psychiatric diagnosis of any major Axis I psychiatric disorder, such as melancholia; (6) severe visual or hearing impairment that cannot be corrected using assistive devices; (7) significant functional impairments resulting from heart disease, arrhythmias, chronic pulmonary disease, or conditions such as avascular necrosis of the hip or knee or severe arthritis of three or more weight-bearing joints that prevent exercising; or (8) systolic blood pressure >200 mm Hg or diastolic blood pressure >115 mm Hg.

Procedures

A detailed description of the procedures for this single-group pretest–posttest clinical trial to evaluate the effectiveness of a 10-wk home-based Wii Fit exercise program for reducing severity of fatigue was previously reported (Yuen, Holthaus, Kamen, Sword, & Brelend, 2011). Briefly, study participants were asked to perform Wii Fit exercise at home for 30 min, 3 times/wk for 10 wk. However, participants were allowed to increase their frequency or duration of Wii Fit exercise if they wanted to do so. After completion of the baseline evaluation and receiving instructions on and training in the Wii Fit exercise protocol, participants were issued 10 weekly exercise logs together with instructions for completion in a folder. Participants were instructed to record in the log the mode of Wii Fit exercise, date (i.e., frequency), duration of engaging in Wii Fit exercise, and their perceived exertion rating after each exercise session. They were allowed to play games other than those in the Wii Fit protocol, such as Wii Sports, but were told not to record this information (i.e., mode of game and its duration) in the log. Exercise information from Wii Sports does not register in the Wii console. Participants were also reminded that no one besides themselves should play their Wii avatar. After the 10-wk exercise program, the research coordinator collected the exercise logs and downloaded the recorded exercise data from the participants’ Wii consoles.

Data Analysis

We used a paired t test to compare the exercise duration recorded on the Wii console and that recorded on the log. Statistical significance was set at p < .05. Reliability of recorded exercise duration between the Wii Fit and the log was assessed using the intraclass correlation (ICC). Because the ICC does not measure agreement that leads to ignoring any systematic bias between the two methods of measurement (Wii console and exercise log: Bland & Altman, 2003), we used the Bland–Altman method of limits of agreement (LoA) to provide an indication of the systematic random error of the data (Bland & Altman, 1986). Agreement between the two measurement methods of exercise duration was assessed using the Bland–Altman plot, which plots the differences between the exercise duration recorded in the exercise log and that in the Wii console against the mean duration of these two methods of data collection. We used the 95% LoA to describe the total error between the two methods of measurement, calculated as mean
difference ± 1.96 \times \text{standard deviation of the difference scores} \quad (\text{Bland \& Altman, 1986}). \text{Values closer to 0 indicate a greater level of agreement.}

This study involved each participant repeatedly recording exercise duration across 10 wk (expected to be 3 times/wk). The data structure of this study therefore consisted of both between- and within-subject information on the difference in exercise duration between the two methods of measurement. Treating each pair of data points (recorded exercise duration from the Wii console and exercise log) as though they were from a different participant would likely bias the estimation of the interval between the LoA (i.e., make it narrower). Alternatively, if an average were taken of all the exercise records within each participant (i.e., ignoring the within-subject variation), the interval between the LoA would also be biased (i.e., too narrow).

To obtain a correct 95% LoA, Bland and Altman (2007) proposed a new statistical method to analyze such clustered data by deriving the estimated standard deviation for individual differences from the variance of the subject mean differences in exercise duration and an extra term derived from the separate measurement error of each measurement method. Bland and Altman proposed two situations in which this new statistical method is used to compute the LoA for repeated measurements—one is when the quantity of the measure varies within a series of observations for each participant does not vary much (i.e., the true value is constant), and the other is when the quantity of the measure varies (i.e., the true value varies). Because the recorded exercise duration varied from session to session over the 10-wk period, we chose to use the formula for when the true value varies to compute the 95% LoA.

### Results

Of the 15 participants recruited and enrolled, 1 lost her exercise log folder and another lost her data because it was accidentally deleted from the Wii console. In addition, inspection of the remaining participants’ paired data from the Wii Fit and exercise logs revealed that >90% of the data pairs for 2 participants were exactly the same, which was highly unlikely unless the participants copied the data from the Wii console to the exercise log. Because the purpose of this study was to evaluate the agreement between exercise frequency and duration recorded in the log and recorded in the Wii console, we eliminated these 4 participants from the analysis. Therefore, the analytic sample consisted of 11 participants with usable data.

The mean age of the 11 participants was 48.8 yr (standard deviation [SD] = 14), and the mean body mass index was 28.6 (SD = 5.6). Of the 261 data points recorded, 187 (72%) included both Wii and exercise log data. Of the 74 nonpaired data points, 59 (80%) entries were recorded in the Wii console only and 15 (20%) entries were recorded in the exercise log only. The mean exercise frequency adherence reported for the 11 participants was 75% for the Wii Fit, 61% for the exercise log, and 79% for both combined.

Of the 187 paired data points, the mean duration of exercise recorded in the Wii was 29.5 min (SD = 13.4) and in the log, 33.3 min (SD = 13.8). The mean difference between the two methods in recorded exercise duration was 3.8 min, which is statistically significant ($p < .0001$). However, when using the mean exercise duration for each participant (i.e., 11 data points) for comparison, we observed no statistically significant difference in exercise duration between the two methods of measurement ($p = .30$). The bias (i.e., percentage of mean difference) on exercise duration between the two methods for each participant is shown in Table 1.

Table 1 lists the ICCs for concordance between Wii and exercise log on exercise duration reported by each participant and in total. The ICCs (3,1) between the Wii Fit and exercise log for exercise duration ranged from .13 to .90, with a composite ICC coefficient of .40 (95% confidence interval [.27–.51]). As expected, the ICC was notably lower (ICC = .29) when using the aggregate mean for each participant (i.e., 11 data points).

Figure 1 displays the Bland–Altman plot depicting the difference in exercise duration between the exercise logs and the objectively measured (i.e., Wii consoles) plotted against the average exercise duration of the two methods. A systematic bias was noted, with the exercise log reporting lower values than the Wii Fit.

### Table 1. Bias Between Wii and Log Data (for Paired Data Only)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Count</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>% Diff</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>32.4</td>
<td>9.1</td>
<td>42.7</td>
<td>11.1</td>
<td>31.6</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>31.7</td>
<td>1.3</td>
<td>36.6</td>
<td>13.8</td>
<td>15.4</td>
<td>.0945</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>37.3</td>
<td>15.9</td>
<td>37.0</td>
<td>13.6</td>
<td>-0.8</td>
<td>.8754</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>9.9</td>
<td>5.5</td>
<td>26.8</td>
<td>12.0</td>
<td>17.9</td>
<td>.0052</td>
</tr>
<tr>
<td>7</td>
<td>34</td>
<td>34.3</td>
<td>8.1</td>
<td>33.1</td>
<td>5.7</td>
<td>-3.7</td>
<td>.2193</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>31.5</td>
<td>12.5</td>
<td>32.6</td>
<td>11.6</td>
<td>3.6</td>
<td>.5158</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>15.3</td>
<td>6.7</td>
<td>9.8</td>
<td>8.7</td>
<td>-36.1</td>
<td>.3217</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>18.8</td>
<td>4.9</td>
<td>33.0</td>
<td>8.5</td>
<td>75.2</td>
<td>.0008</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>37.2</td>
<td>13.6</td>
<td>13.2</td>
<td>2.6</td>
<td>-64.6</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>32.5</td>
<td>26.2</td>
<td>4.5</td>
<td>14.8</td>
<td>24.6</td>
<td>.5000</td>
</tr>
<tr>
<td>15</td>
<td>23</td>
<td>14.5</td>
<td>6.1</td>
<td>34.2</td>
<td>18.1</td>
<td>135.6</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Total</td>
<td>187</td>
<td>29.5</td>
<td>13.4</td>
<td>33.3</td>
<td>13.8</td>
<td>12.7</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note. SD = standard deviation; % Diff = percentage of difference, calculated as [(Log data – Wii data)/Wii data] x 100%

### Table 2. ICCs for Concordance Between Wii and Log Data

<table>
<thead>
<tr>
<th>Participant</th>
<th>Count</th>
<th>ICC (3,1)</th>
<th>95% CI for ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>.51</td>
<td>[.17, .74]</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>.60</td>
<td>[.15, .84]</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>.90</td>
<td>[.77, .96]</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>.18</td>
<td>[.60, .78]</td>
</tr>
<tr>
<td>7</td>
<td>34</td>
<td>.65</td>
<td>[.40, .81]</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>.69</td>
<td>[.45, .84]</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>.29</td>
<td>[.93, .98]</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>.76</td>
<td>[.14, .97]</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>.13</td>
<td>[.46, .63]</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>.86</td>
<td>[.59, 1.00]</td>
</tr>
<tr>
<td>15</td>
<td>23</td>
<td>.17</td>
<td>[.26, .54]</td>
</tr>
<tr>
<td>Total</td>
<td>187</td>
<td>.40</td>
<td>[.27, .51]</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; ICC = intraclass correlation coefficient.
was evident from the Bland–Altman plot, which showed that the level of agreement was poor. In general, participants tended to over-report the duration of their exercise in the exercise log, with the trend indicating that the overestimation became larger as the magnitude of reported Wii duration decreased. In fact, we observed a significant negative correlation between the difference in exercise duration recorded in the exercise logs and in the Wii and the exercise duration recorded in the Wii ($r = -0.53, p < .0001$). The 95% LoA used to describe the total measurement error between the two methods was 35 to −27, which means the amount of measurement error varied across participants from 0 to as much as 35 and −27 min, a range of 62 min, suggesting a large between-subjects variability.

Discussion

About 72% (187 of 261) of the records had matched exercise dates from both the Wii console and the exercise log, indicating that the exercise log seems an appropriate way to document the frequency of exercise (Krippendorff, 1980). The adherence to exercise frequency tended to be underreported in the exercise log. The majority (80%) of the unpaired data points between the two methods were attributed to missing documentation from the exercise logs. One of the obvious explanations was that participants occasionally forgot to record their exercise duration. Other, less likely explanations included someone other than the participants playing the Wii Fit. However, participants were told not to let other people play their Wii avatar. The research coordinators noticed that most participants who had other family members in the household created additional Wii avatars for the Wii Fit. Therefore, it was not likely that other people in the household affected the accuracy of the data. Incorporating appropriate mnemonic strategies after exercise may help participants remember to complete the exercise log.

Some data, although small amounts (20%), were recorded in exercise logs but not in the Wii console. One of the possible explanations for this phenomenon was participants retrospectively recording exercise information on the basis of recall. As a result, they may have thought they did the exercise on that particular day but in fact did not. Alternatively, they may have mistakenly recorded other non–Wii Fit exercise such as Wii Sports in their exercise log.

Although the mean difference (3.8 min) in recorded exercise duration between the two methods was statistically significant, on the basis of the mean recorded exercise duration of 30 min, this difference (about 13%) seems relatively small and is certainly within acceptable clinical error range. However, the amount of error between the two methods across participants—62 min—was huge, indicating considerable between-subjects variation in the level of agreement in recorded exercise duration between these two methods of measurement.

We intended to assess the absolute validity of the exercise log using the built-in recording system for Wii Fit data, which served as the best estimate of the true (criterion) value. Although exercise logs have a marginally acceptable agreement with Wii estimation of exercise duration at a group level (based on ICC = .4; Fleiss, 1986), caution should be applied when using the exercise log as a measure of a person’s exercise behavior because of the tendency to overreport. To our knowledge, this study is the first to truly validate that participants tend to overreport their exercise duration but underrecord exercise frequency in a 10-wk home-based exercise program.

Limitations

One of the drawbacks of Wii Fit compared with an exercise log in documenting exercise information is that it does not record exercise intensity. In addition, few exercise or commercially available exergames can record the date and duration of exercise that the users perform over a period of time. As a result, similar validation for non–Wii Fit exercise may be less feasible. Other limitations related to this study include small sample size and a specific disease group whose results might not be generalizable to populations with other disabilities.

Future Research

Future research aiming to explore the psychosocial aspects of using an external monitoring system such as the Wii Fit and iPhone to track exercise patterns may provide valuable insight for the development of a better system or strategy to improve patients’ adherence to exercise as well as documentation of their exercise performance.

Implications for Occupational Therapy Practice

Occupational therapy clinicians often ask patients to document home-based exercise to track patients’ adherence to prescribed exercise programs. This information may be used to identify barriers that affect adherence to prescribed exercise.

- On the basis of this study’s findings, frequency of exercise documented in the
log seems reasonably accurate (although underrecorded), but the main concern is overreporting of exercise duration.

- If the goal is to check whether patients adhere to frequency of exercise, exercise logs seem an appropriate way to track this exercise parameter and obtain a rough estimation (an underestimation) of their adherence. However, long-term tracking without monitoring is required to confirm this study’s findings.
- If the goal is to track exercise duration, the group mean exercise duration is acceptable, but individual recording of sessions tends to be overreported.
- Clinicians and researchers should be mindful of overreporting when selecting exercise logs to document duration of exercise. Incorporating appropriate strategies to assist patients in keeping track of the duration of exercise is recommended when using exercise logs.

Acknowledgments

We thank G. Y. Zou, who provided the SAS macro program to compute the LoA analysis for repeated measurements. This project was supported by the South Carolina Clinical and Translational Research Institute, the Medical University of South Carolina’s Clinical and Translational Science Award, and the National Institutes of Health/National Center for Research Resources (Grant UL1RR029882). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or the National Center for Research Resources.

References


