# Novel method for measurement of fatigue in multiple sclerosis: Real-Time Digital Fatigue Score

**Edward Kim, MD;[1–2]\* Jesus Lovera, MD, MsPH;[3] Laura Schaben, MD;[1–2] J. Melara, MPH;[3] Dennis Bourdette, MD;[1–2] Ruth Whitham, MD[1–2]**

[1]*Department of Neurology, School of Medicine, Oregon Health & Science University, Portland, OR;* [2]*Multiple Sclerosis Center of Excellence—West, Portland Department of Veterans Affairs Medical Center, Portland, OR;* [3]*Department of Neurology, School of Medicine, Louisiana State University Health Sciences Center—New Orleans, New Orleans, LA*

**Abstract**—The study's objective was to develop a real-time measurement for fatigue and to evaluate whether it is an effective clinical trial outcome measure compared with the Fatigue Severity Scale (FSS) and the Modified Fatigue Impact Scale (MFIS). Forty-nine subjects with MS and an FSS >4 recorded Real-Time Digital Fatigue Scores (RDFSs) on a wrist-worn device four times a day over 3 weeks. Scores were scaled 0–10, with 10 representing the worst possible fatigue. FSS and MFIS were evaluated and compared with RDFSs. Mean RDFSs significantly correlated with FSS ($r = 0.55$, $p < 0.001$) and MFIS ($r = 0.55$, $p < 0.001$). RDFS captured circadian variations in fatigue, with scores increasing from mean 3.4 at 9 a.m., to 4.0 at 1 p.m., 4.5 at 5 p.m., and 5.0 at 9 p.m. When all scores over all days were included in a mixed-model analysis of circadian variation, the differences in RDFS between times were more significant than in an analysis that included only single scores of data isolated from the first day of monitoring. RDFS is a promising measure. RDFS significantly correlated with FSS and MFIS, captured real-time daily and circadian variations in fatigue, and provided multiple measurements of fatigue that provided statistical advantages over FSS and MFIS.

## INTRODUCTION

Fatigue is a major cause of disability in individuals with multiple sclerosis (MS) and is associated with a reduced quality of life. Fatigue is reported in 75 to 95 percent of individuals with MS, and 50 to 60 percent of individuals with MS report that fatigue is their most disabling complaint [1–2]. Despite the prevalence and severity of fatigue in MS, treatment options remain limited. Better treatments for fatigue still need to be developed, and their efficacy must be demonstrated through clinical trials.

Current limitations in measuring fatigue remain an ongoing challenge for clinical trials of fatigue in MS. Assessment of fatigue is a difficult task, given differing definitions of fatigue, confounding factors such as loss of

motivation or sleepiness, and inherent subjectivity between individuals. Most fatigue studies in MS rely on questionnaires like the Fatigue Severity Scale (FSS) [3] and the Modified Fatigue Impact Scale (MFIS) [4–5]. Both FSS and MFIS are validated measures for fatigue, but both depend on retrospective recall of fatigue over the prior week and month, respectively, rather than real-time assessment. Schwid et al. suggested the potential efficacy of a regularly repeated fatigue measurement in real time with the demonstration of the Rochester Fatigue Diary [6]. We developed the Real-Time Digital Fatigue Score (RDFS) to more easily measure and digitally record the daily experience of fatigue in real time at regular time intervals (9 a.m., 1 p.m., 5 p.m., and 9 p.m.).

Development of better measurements of fatigue should further improve clinical research outcomes for future therapeutic trials in MS. Our main objectives of this study were to develop a novel method for the improved, easier measurement of MS-related fatigue in real time and to compare it with the FSS and MFIS as an outcome measure for clinical trials.

## METHODS

The data presented here were collected as part of a baseline pretreatment assessment for subjects entering a clinical trial of American ginseng for fatigue in MS. The study was accepted by the Oregon Health & Science University Institutional Review Board. Subjects signed an informed consent, and then we assessed them for inclusion/exclusion criteria to determine whether they qualified for entry into the study. Confidentiality was maintained by assignment of deidentified coding for each subject.

### Subjects

Between January 2006 and October 2008, 49 subjects with clinically definite MS by "McDonald Criteria" [7] were enrolled in a prospective double-blinded, placebo-controlled crossover trial of American ginseng extract for MS fatigue [7]. They completed all baseline assessments. The sample of patients with MS was recruited primarily from an outpatient MS subspecialty clinic at Oregon Health & Science University. Some subjects were also recruited from outpatient community neurology clinics. Subjects aged 18 to 70 with clinically definite MS (McDonald Criteria [7]), as determined by

one of the evaluating study neurologists, were included if they had a complaint of fatigue for >2 months with FSS >4. Subjects with any MS disease subtype (relapsing-remitting MS, secondary progressive MS, and primary progressive MS) and any Expanded Disability Status Scale (EDSS) score were eligible for entry. Subjects were excluded from the study for the following seven reasons:

1. Use of ginseng or stimulants in the prior 6 weeks (modafinil, pemoline, methylphenidate).
2. Treatment with glucocorticoids in the prior 6 weeks.
3. Severe depression with Beck Depression Inventory (BDI) score [8] >31.
4. Exacerbation in prior month.
5. Pregnancy.
6. Other serious medical disease (including diabetes requiring insulin; anemia; uncontrolled hypothyroidism; liver, kidney, or heart failure; or chronic obstructive pulmonary disease).
7. Inability to complete questionnaires or study outcome measures (severe cognitive impairment, illiteracy, or severe visual impairment).

### Assessments

At the first visit, we assessed all subjects using the EDSS [9], BDI, FSS, and MFIS. EDSS was assessed by one of the evaluating study neurologists. Subjects were taught by a trained research assistant to use a wrist-worn device called Actiwatch Score (Mini Mitter, Respironics; Bend, Oregon) for recording RDFS. Actiwatch Score has a display on the front panel, contains a built-in beeper, and records scores of any parameter on a numerical scale (**Figure 1**). Watches were programmed to beep at 9 a.m., 1 p.m., 5 p.m., and 9 p.m. over 3 weeks. At each scheduled time, subjects were prompted to record a numerical score of their fatigue at that real-time moment following a scale provided on a laminated card that they carried with them (**Figure 2**). The numerical rating scale was from 0 to 10, with 0 representing none, 5 representing moderate, and 10 representing worst possible fatigue.

After baseline pretreatment monitoring for 3 weeks, subjects returned for evaluation with FSS and MFIS. FSS and MFIS data obtained at this visit were used for the comparison analyses to be described. RDFS data were downloaded onto a computer with Actiware software, version 5.5 (Mini Mitter Respironics; Bend, Oregon). From these recorded data, each subject's RDFS at different times of the day was captured. This process permitted calculation of mean RDFS for each subject over the

479

**KIM et al. Real-Time Digital Fatigue Scores in MS**

**Figure 1.**
Actiwatch Score wrist-worn device (Mini Mitter, Respironics; Bend, Oregon) for recording Real-Time Digital Fatigue Scores. Device has a display on front panel, contains a built-in beeper, and records scores of any parameter on a numerical scale.



**Figure 2.**
Numerical fatigue rating scale provided to subjects with multiple sclerosis on a separate form as a reference for entering their Real-Time Digital Fatigue Score of Actiwatch Score.

dian variations, all RDFSs from all subjects were analyzed with a mixed-model repeated-measures approach [10–11]; the model included random effects for each time of the day and for each day with an unstructured correlation matrix for times of the day and an autoregressive correlation matrix for days. We performed this analysis to explore how the RDFS could capture circadian variations in fatigue and to explore the potential statistical advantages of using the RDFS to apply a mixed-model analysis.

## RESULTS

Forty-nine subjects (45 women, 4 men) entering the American ginseng trial successfully completed all aspects of the 3-week baseline pretreatment monitoring. Mean age was 47 years (range 25–67), mean disease duration was 15.7 years (range 1.3–48), mean EDSS was 3.2 (range 0–7), and mean BDI was 8.9 (range 0–29). Thirty-nine had relapsing-remitting MS, seven had secondary progressive MS, and three had primary progressive MS. At the first visit, mean FSS was 6.1 (range 4–7) and mean MFIS was 50.0 (range 15–75). At the completion of the 3 weeks of pretreatment baseline monitoring, mean FSS was 5.7 (range 3.6–7), mean MFIS was 48.5 (range 15–74), and mean RDFS was 4.3 (range 0.7–7.3). RDFS was demonstrated to vary over the course of the day and from day to day (**Figure 3**).

Among the 49 subjects included in the analysis, 2,246 scheduled observations of timed RDFS were recorded. Each subject was prompted to record a scheduled RDFS four times a day over 3 weeks. The mean number of scheduled RDFS entered was 45.8 (median =

entire 3-week period, as well as mean RDFS for each time of day.

### Statistical Analyses

We completed statistical analyses of the intraclass correlation coefficients (ICCs) for the RDFS, FSS, and MFIS using SPSS (version 16.0.1, 2008; SPSS Inc; Chicago, Illinois). A two-way random effects model was applied in which subjects and days were treated as random effects.

We completed all other statistical analyses using SAS 9.1 (SAS Institute Inc; Cary, North Carolina). Mean RDFS for each subject was correlated with their FSS, MFIS, and other continuous variables with use of Pearson correlation coefficient. We performed this analysis to explore how the newly developed RDFS correlated with previously established outcome measures of MS-related fatigue, as well as other factors. In an analysis of circa-
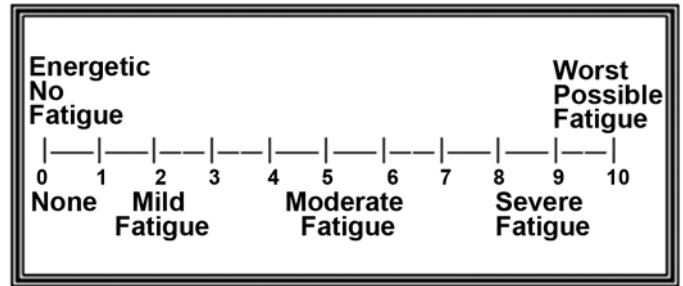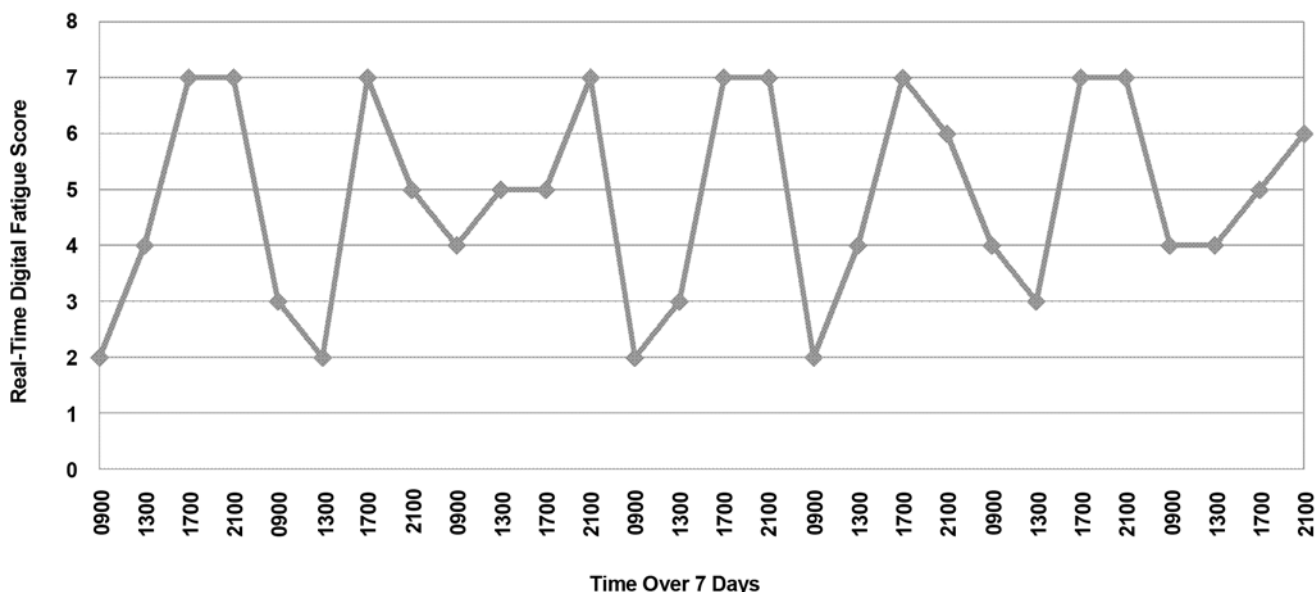
**Figure 3.**
Real-Time Digital Fatigue Scores over 7 days for individual subject with multiple sclerosis.

50, range 2–76). The average compliance for entering a RDFS at each scheduled time was 64 percent.

**Correlation of Fatigue Scores**

RDFSs were compared with the FSS and MFIS measurements obtained at the visit after the 3-week period of pretreatment baseline monitoring with the RDFS complete. Mean RDFS significantly correlated with FSS ($r = 0.55$, $p < 0.001$) and MFIS ($r = 0.55$, $p < 0.001$) (**Figure 4**). Mean RDFS did not correlate significantly with BDI ($r = 0.20$, $p = 0.2$) or EDSS ($r = 0.20$, $p = 0.30$). Other covariate measures, including age, disease duration, and physical disability (25-foot timed walk and 9-hole peg test), failed to correlate with mean RDFS or other fatigue measures.

**Intraclass Correlation Coefficients of Fatigue Scores**

To demonstrate statistical reliability of the RDFS, we assessed and compared ICCs between the FSS, MFIS, and RDFS. The ICC is the ratio of between-subjects variance to the total variance. A higher ICC corresponds to better statistical reliability of a measurement scale. The ICC for the RDFS was calculated with a sample data set of the first 10 RDFS entries from each subject. Subjects with ≤10 RDFS entries were excluded. The ICC for the RDFS (0.968, 95% confidence interval (CI): 0.953–

0.980) was significantly higher than the ICCs for the FSS (0.833, 95% CI: 0.704–0.906) and MFIS (0.883, 95% CI: 0.791–0.934). No formal statistical test exists for comparing ICCs. However, the differences between the ICCs of the RDFS with the FSS and MFIS are significant because the CIs for the coefficients do not overlap.

**Circadian Variations in Fatigue Scores**

RDFS also captured significant variability in fatigue throughout the day. Using a mixed-model repeated-measures analysis that included all RDFSs assessed over all days, we observed a significant incremental increase in circadian fatigue over each day. At 9 a.m., mean RDFS was 3.4, increasing to 4.1 at 1 p.m., 4.5 at 5 p.m., and 5.0 at 9 p.m. (**Figure 5**). In this mixed-model analysis, the circadian differences in RDFS between times of day were all significant.

Furthermore, the differences in the mixed-model analysis were more statistically significant than in an analysis of variability over time that only used subjects' isolated fatigue scores from the first day of monitoring (**Table**). In the mixed-model analysis, each difference between fatigue scores at different times of day demonstrated significance of $p < 0.005$, with the majority of values as $p < 0.001$ (range <0.001–0.004). In contrast, the paired $t$-test analysis of variability comparing only isolated scores

481

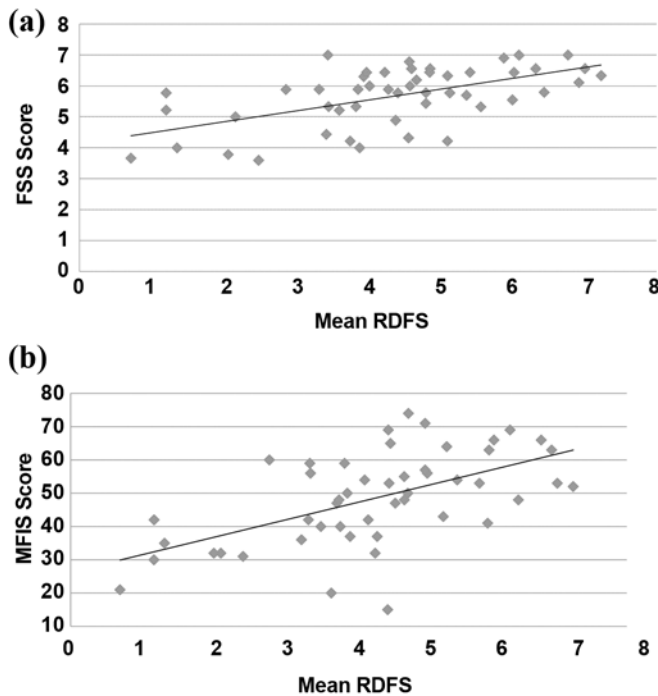KIM et al. Real-Time Digital Fatigue Scores in MS



**Figure 4.**
(a) Fatigue Severity Scale (FSS) scores vs mean Real-Time Digital Fatigue Scores (RDFSs) and (b) Modified Fatigue Impact Scale (MFIS) scores vs mean RDFSs in 49 subjects with multiple sclerosis.



**Figure 5.**
Fluctuation in fatigue throughout day of mean Real-Time Digital Fatigue Scores in 49 subjects with multiple sclerosis.

**Table.**
Differences in Real-Time Digital Fatigue Scores (RDFSs) between times of day in 49 subjects with multiple sclerosis.

| Scores from All Days | | Scores from Single Day | |
|---|---|---|---|
| Time vs Time | $p$-Value[*] | Time vs Time | $p$-Value[†] |
| 9 a.m. vs 1 p.m. | 0.008 | 9 a.m. vs 1 p.m. | 0.57 |
| 9 a.m. vs 5 p.m. | <0.001 | 9 a.m. vs 5 p.m. | 0.12 |
| 9 a.m. vs 9 p.m. | <0.001 | 9 a.m. vs 9 p.m. | 0.42 |
| 1 p.m. vs 5 p.m. | 0.004 | 1 p.m. vs 5 p.m. | 0.003 |
| 1 p.m. vs 9 p.m. | <0.001 | 1 p.m. vs 9 p.m. | 0.02 |
| 5 p.m. vs 9 p.m. | <0.001 | 5 p.m. vs 9 p.m. | 0.38 |

[*]Significance of differences in RDFSs between times using mixed-model repeated-measures analysis of all scores from all days.
[†]Significance of differences in RDFSs between times using paired $t$-test analysis of isolated scores from single day.

from a single day of observation was much less robust and demonstrated only two significant differences of $p < 0.05$, with the majority of values as $p > 0.10$ (range 0.003–0.42).

## DISCUSSION

This article is the first to demonstrate a wrist-worn device to digitally record fatigue in real time in MS. For our development of the RDFS, we used a numerical rating scale similar to a visual analog scale (VAS) developed by Krupp et al [12]. Krupp et al. previously evaluated the usefulness of a visual analog fatigue scale, comparing it with the FSS and the original Fatigue Impact Scale. Their VAS demonstrated adequate test-retest reliability and better responsiveness to changes in fatigue medications than the FSS and the original Fatigue Impact Scale [12]. Schwid et al. enhanced the VAS and further suggested the potential efficacy of a fatigue measurement in real time by demonstrating the Rochester Fatigue Diary [6]. It consisted of a diary of VASs to measure fatigue every hour
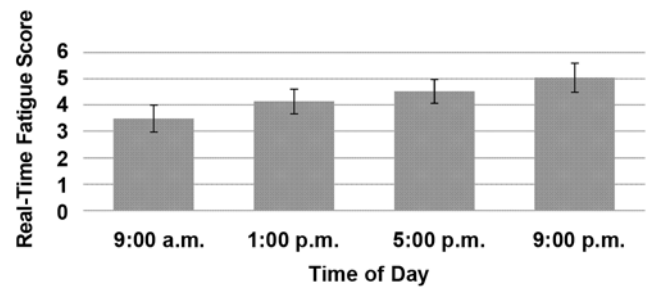
for 1 day at a time. In Schwid et al.'s study of the diary for 7 consecutive days, scoring moderately correlated with FSS ($r = 0.40$, $p = 0.05$) and a circadian pattern of fatigue increasing through the afternoon was demonstrated. While promising as a record of fatigue scores in real time, the diary was cumbersome to use, requiring handwritten hourly notations in a notebook. In comparison, RDFS is easy to use. The watch automatically prompts subjects to record fatigue scores at preprogrammed time points, and the data are recorded electronically. The digital electronic data format then permits easy downloading and data analysis.

Because the RDFS is based on a 0- to 10-point numerical rating scale that queries each subject to longitudinally quantify their absolute level of fatigue, RDFS is constructed to have good face validity as a direct measure of fatigue severity. RDFS correlated strongly with validated traditional measures of MS-related fatigue, the FSS and MFIS, and was independent of the BDI and EDSS.

Construct validity for the RDFS was supported by this demonstration of strong associations between the RDFS and the FSS and MFIS. Validity was further supported by the demonstration that RDFS was not related to depression and disability. These observations support the reliability of the RDFS as a measure of fatigue rather than depression or neurological impairment.

To demonstrate more clearly the statistical reliability of the RDFS compared with the reliability of FSS and MFIS, we compared the ICCs for each of the fatigue measurement scales. A higher ICC corresponds to better reliability of a measurement scale. When multiple measurements of RDFS for each subject were collected over time and compared, RDFS was demonstrated to have a significantly higher ICC than the ICCs for the FSS and MFIS. This comparison analysis of ICCs demonstrates that the RDFS is more reliable and that the differences in reliability between the RDFS and the FSS and MFIS are significant. These data support the internal reliability of scoring of RDFS for each individual subject and suggest the potential efficacy of mixed-modeling analysis.

Unlike the single retrospective measurements of fatigue obtained by the FSS and MFIS, the RDFS prospectively captures real-time fatigue over multiple time points. Therefore, it also likely provides a more responsive, accurate, and up-to-the-moment assessment of self-reported fatigue than the FSS and MFIS questionnaires in which subjects are asked to make difficult retrospective reports of their fatigue level from the preceding 1 week and 1 month for the FSS and MFIS, respectively. Consequently, the RDFS is less subject to recall bias than the FSS and MFIS.

Because of the multiple observations, the RDFS data set provides statistical advantages for analysis over the traditional FSS and MFIS scores. As one example of the utility of the RDFS, we measured and demonstrated a significant circadian variation in fatigue that cannot be observed with the FSS or MFIS. This capability of the RDFS to capture significant circadian fatigue changes over time demonstrates its responsiveness as a measure of fatigue. In our analysis, we analyzed the data set of all subjects' RDFSs over all days using a mixed-model repeated-measures approach, which takes advantage of the multiple observations of fatigue for each subject. We used this statistical approach to include all scores from each subject, take advantage of the correlation between daily scores, and decrease within-subject error. This statistical approach also naturally adjusts for missing values, minimizing the impact of any missed score entries. The

mixed-model repeated-measures analysis demonstrated more statistically significant differences than an analysis comparing subjects' isolated fatigue scores from 1 day. This comparison demonstrates the potential statistical power of multiple real-time observations of fatigue over time compared with the single retrospective assessment of an isolated fatigue rating score.

With mixed-model analysis, we minimized the impact of missed scheduled fatigue score entries while we emphasized the strength of collecting multiple fatigue scores over time. The strength of the RDFS is the high volume of fatigue data collected. Even with many missed score entries, a very high volume of fatigue data is collected. Statistical power remains significant even with missed entries because of the high frequency of regular fatigue observations for each individual subject. The average subject in this study entered over 45 observations of fatigue. For the entire sample of 49 subjects, 2,246 scores of real-time fatigue were collected for analysis. Even if many scheduled fatigue scores are not entered, the collective volume of RDFS data still contrasts significantly with the isolated scores of FSS and MFIS assessed for each individual during the study. Furthermore, because a mixed-model repeated-measures approach can be applied to the analysis of RDFS data, an additional statistical adjustment is built-in for any missed fatigue score entries of each individual. However, compliance for score entries remains a potential problem for application of the RDFS. Future studies may need to explore different strategies to improve compliance.

Our study population had a higher preponderance of women compared with men (45:4) than is observed in the general population with MS or in previous fatigue studies. Although MS-related fatigue may behave differently between sexes, other fatigue studies have not demonstrated a significant difference [13]. Furthermore, both FSS and MFIS have been validated for use in men and women with MS, and RDFS correlates well with both. Even though our data overrepresent women, we anticipate the RDFS to remain applicable to the general population with MS. Most likely, differences in sex would not alter this novel approach to studying fatigue in men and women with MS. The data presented currently represent an exploratory analysis of this new fatigue outcome measure. Future studies may need to be designed to explore any potential differences between the application of the RDFS in men and women.

483

KIM et al. Real-Time Digital Fatigue Scores in MS

## CONCLUSIONS

Because the wrist-worn Actiwatch Score device can easily record extensive digital fatigue data that can then be downloaded and analyzed for an objective measurement of treatment response over any defined time period, RDFS may prove to be very effective as an outcome measure to monitor the effects of a new fatigue therapy in a clinical trial. As discussed, the RDFS demonstrates properties of good face validity, construct validity, reliability, and responsiveness. Demonstration of its responsiveness to time was further strengthened by the application of a mixed-model repeated-measures analysis. With the potential applications of a mixed-model analysis, the RDFS may be very effective as an outcome measure for clinical trials. A significant interest will be to observe how the analysis of an RDFS data set by a mixed-model repeated-measures approach responds to change compared with the analysis of FSS or MFIS in a clinical trial. As an example, we have completed a double-blinded, placebo-controlled crossover trial of American ginseng for MS-related fatigue using the FSS, MFIS, and RDFS as outcome measures (manuscript in preparation) [14]. We hope that this novel method of assessing real-time fatigue will prove to be a valuable outcome measure in clinical trials of drugs and other interventions for treating fatigue in MS.

## ACKNOWLEDGMENTS

## REFERENCES

1. Freal JE, Kraft GH, Coryell JK. Symptomatic fatigue in multiple sclerosis. Arch Phys Med Rehabil. 1984;65(3): 135–38. [PMID: 6703889]

2. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. Can J Neurol Sci. 1994;21(1):9–14. [PMID: 8180914]

3. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol. 1989;46(10):1121–23. [PMID: 2803071]

4. Paralyzed Veterans of America. Fatigue and multiple sclerosis: Evidence-based management strategies for fatigue in multiple sclerosis. Washington (DC): Multiple Sclerosis Council for Clinical Practice Guidelines; 1998.

5. Kos D, Kerckhofs E, Carrea I, Verza R, Ramos M, Jansa J. Evaluation of the Modified Fatigue Impact Scale in four different European countries. Mult Scler. 2005:11(1):76–80. [PMID: 15732270]
DOI:10.1191/1352458505ms1117oa

6. Schwid SR, Covington M, Segal BM, Goodman AD. Fatigue in multiple sclerosis: Current understanding and future directions. J Rehabil Res Dev. 2002;39(2):211–24. [PMID: 12051465]

7. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, Lublin FD, Metz LM, McFarland HF, O'Connor PW, Sandberg-Wollheim M, Thompson AJ, Weinshenker BG, Wolinsky JS. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria." Ann Neurol. 2005;58(6):840–46. [PMID: 16283615]
DOI:10.1002/ana.20703

8. Beck AT, Beamesderfer A. Assessment of depression: The depression inventory. Mod Probl Pharmacopsychiatry. 1974; 7(0):151–69. [PMID: 4412100]

9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology. 1983;33:1444–52. [PMID: 6685237]

10. Brown H, Prescott R. Applied mixed models in medicine. Chichester (United Kingdom): John Wiley & Sons; 2003.

11. McCulloch CE, Searle SR. Generalized, linear, and mixed models. New York (NY): John Wiley & Sons; 2001.

12. Krupp LB, Soefer MH, Pollina DA, Smiroldo J, Coyle PK. Fatigue measures for clinical trials in multiple sclerosis. Neurology. 1998;5:S38.

13. Bakshi R, Shaikh ZA, Miletich RS, Czarnecki D, Dmochowski J, Henschel K, Janardhan V, Dubey N, Kinkel PR. Fatigue in multiple sclerosis and its relationship to depression and neurologic disability. Mult Scler. 2000;6(3):181–85. [PMID: 10871830]

14. Kim E, Lovera J, Schaben L, Bourdette D, Whitham R. A single center, randomized, double-blinded placebo-controlled crossover pilot study of the effects of American ginseng on multiple sclerosis fatigue. 2009 Fall Conference: Turning the Spotlight of Neurology; 2009 Nov 6–8; Las Vegas, NV. St. Paul (MN): American Academia of Neurology; 2009. A255 p.