# Two-Part Mixed-Effects Location Scale Models for Health Diary Data

*Shelley A. Blozis and Hayat Botan*



**Background:** The analysis of health diary data has long relied on inferential statistical methods focusing on sample means and ad hoc methods to calculate each individual's variation in health outcomes.

**Objectives:** In this paper, an advanced statistical model is applied to daily diary self-reported health outcomes to simultaneously study an individual's likeliness to report an outcome, daily mean intensity level, and variability in daily measures.

**Methods:** Using observational, secondary data from 782 adults, we analyzed self-report daily fatigue symptoms, distinguishing between whether an individual reported fatigue and its severity when reported. Self-reported depressed affect and participant characteristics were used as predictors of daily fatigue symptoms.

**Results:** A higher likeliness to report fatigue correlated with higher mean fatigue severity and greater stability in severity ratings. Higher mean severity correlated with greater stability in severity ratings. Females and those with high depressed affect were more likely to report fatigue. Females and those with high depressed affect reported greater mean severity.

**Discussion:** The model applied to daily measures allowed for the simultaneous study of an individual's likeliness to report a symptom, daily mean symptom severity, and variability in severity across days. An individual's daily variation in symptom severity was represented as a model parameter that did not contain measurement error that is present in ad hoc methods.

**Key Words:** individual differences, intensive longitudinal data, interindividual variation, intraindividual variation, patient-reported outcomes, self-report data, semicontinuous variables

(*Nurs Res* 2025;74: 225–232)

Intensive monitoring of patient-reported outcomes has been a nursing practice for many years (Hale, 2016). Collecting data on such outcomes facilitates a structured approach to nursing that is optimally implemented using health diaries (Verbrugge, 1980). Historically, health diaries have been instrumental in identifying important trends in epidemiological research. In a more contemporary sense, diaries are used as health event and symptom identification measures and help bridge a gap between acute patient-reported outcomes and patient-centered care. Health diaries serve various methodological purposes, but above all, the facilitation of experience from patient to provider contributes to the now so-called primary source of data in health research (Burman, 1995). Daily monitoring of health events and symptoms—coupled with medical insights—has become central to providing quality care and, thus, has contributed to nursing research and best practices (Lischetzke & Könen, 2023). Indeed, documenting patterns in acute health events and symptoms has contributed to a greater understanding of the links between acute reports and future health outcomes (Bartley et al., 2018; Pringle et al., 2003; Rothwell et al., 2010; Shimbo et al., 2012; Topaz et al., 2021).

Given their nature, diary health data can present challenges for analysis, including the need to address within-subject dependencies of repeated assessments. Diary measures are likely to result in very different patterns of data collection between individuals, such as individuals having a different number of assessments or missing data for planned assessments. Diary data are also expected to show individual differences in responses, such as differences in the degree of within-person variability in reports over time, in addition to differences in response levels. Symptom reports, for instance, can exhibit differences between individuals, such that some individuals may not experience the symptom during the study period, others may consistently experience the symptom, and others may report symptoms on some days but not others.

Depending on the aims of an investigation, data analysis options for diary health data most generally include calculations of descriptive statistics, such as proportions and sample means, combined with data visualizations that can provide a basis for understanding trends in patient-reported outcomes. Inferential statistical methods add to the characterization by permitting inferences about a population of interest, such as to study relationships between health outcomes, concomitant variables, and patient characteristics in a patient population or draw comparisons between different populations. Due to the serial nature of diary data, methods must address the dependencies of observations within individuals, a challenge that has been met by methods, such as repeated measures analysis of variance (RM-ANOVA), that summarize population-level trends (Aroian & Vander Wal, 2007; Berger et al., 2003; Ito & Tadaka, 2022).

Mixed-effects models have become a standard for the analysis of repeated measures data. Mixed-effects models combine population-level data summaries with summaries of individual differences in select features of the repeated measures. The fixed effects of a model describe aspects of a variable that all individuals share in a population, so they characterize population-level trends, such as the mean response level of an outcome over time. The random subject effects are specific to the individual and allow features that are used to describe population-level trends to vary between individuals, such as allowing for the response level to vary between individuals. For instance, Bartley et al. (2018) applied a mixed-effects model to repeated measures of pain and fatigue symptoms to characterize population-level trends and individual differences in symptoms of fibromyalgia patients assessed over time. In their study, individuals differed in their pain and fatigue symptoms, so the random effects of the model allowed for these differences. In

Shelley A. Blozis, PhD, is Professor, University of California, Davis, Davis, California. Hayat Botan, is Research Assistant, University of California, Davis, California.

Accepted for publication September 28, 2024.

The corresponding author states that there are no conflicts of interest and that no funding was received for the work submitted.

Ethical Conduct of Research: This study involves a secondary analysis of publicly available data from the Midlife in the United States study series for which individuals are not identifiable. Consequently, the authors' institution does not require ethics review of studies using this data set.

Corresponding author: Shelley A. Blozis, PhD, University of California, Davis, Psychology Department, One Shields Ave, Davis, CA 95616 (e-mail: sablozis@ucdavis.edu).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.nursingresearchonline.com).



ISSN: 1538-9847/25/7403–0225
DOI: 10.1097/NNR.0000000000000810

addition to accounting for individual differences in symptom patterns, mixed-effects models allowed for patients to have different patterns of data collection and be assessed a different number of times because these models do not require complete data or for respondents to be measured according to identical time points.

A significant purpose of collecting health data intensively over time is due to an expectation that patients will show differences in the variability of measures over time, and mixed-effects models are ideally suited to capturing this aspect of diary data. A special version of a mixed-effects model—known as a mixed-effects location scale (MELS) model—was specially developed for data collections that exhibit individual differences in the variability of scores within individuals (Hedeker et al., 2012). That is, a MELS model includes a random scale to permit between-subject heterogeneity of the within-subject variance of a variable measured over time. In other words, the degree of variability in a measured outcome over time can differ from one person to the next. For health events and symptom data in particular, not all individuals may report the outcome at any time (such as when a health screening tool is being used), whereas others may experience the outcome intermittently, and as a result, a data set can include a high number of zeros (or other value used to denote the absence of the target outcome). Further, this is often in conjunction with individual differences in reports of the severity of the outcome when it is reported. In such cases, a special version of a MELS model can be a useful analytic tool in which each of these aspects of a symptom is considered a unique feature to study. A two-part MELS model provides an analytic framework to simultaneously model a binary variable that indicates the presence or absence of a targeted outcome and a continuous variable that reflects the magnitude of the outcome when it is present (Blozis, 2024b; Blozis et al., 2020).

In this report, we illustrate a novel application of a two-part MELS model for daily symptom reports to jointly analyze binary daily symptom indicators (symptom presence vs. absence) and daily symptom severity when a symptom is reported to understand the health course of a patient. The simultaneous examination of these different aspects of patient-reported outcomes forms the methodological foundation of this paper to offer an analytic framework for understanding acute outcomes. While this paper provides a thorough and applied example of this methodology, the focus is to advance MELS models in areas where traditional methods that rely on aggregated data, such as patient reports that have been averaged across the study period, are used to guide treatment plans. It is well understood that traditional methods can be inadequate, as they are not set up to capture the varied progression and nature of acute outcomes over an extended period of time (Kent & Hayward, 2007). In contrast, a two-part MELS model provides a framework to capture intra- and interpatient variability; it provides nurses with the necessary tools to approach clinical and nonclinical studies where the presence (vs. absence) of an acute health outcome and its severity are informative about a patient's state. To meet these aims, the remainder of the paper is organized as follows. A data set from a publicly available daily diary study is described. Selected variables that change with time (time-varying) and some that do not change with time (time-invariant) are used to study daily symptom reports.

## METHODS

Data from the Midlife in the United States (MIDUS) Project are used to illustrate the methods discussed. This involves a secondary analysis of publicly available data for which individuals are not identifiable. Consequently, the authors' institution does not require ethics review of studies using the data. The data presented are from the MIDUS Refresher 1 (Ryff et al., 2011–2014) and the Daily Diary Project (Ryff & Almeida, 2012–2014). For the daily diary study, $n$ = 782 adults (55.6% female) between the ages of 25 and 75 (M = 47.91, SD = 12.67) were surveyed by telephone over eight consecutive days.

We illustrate how to apply a two-part MELS model and incorporate covariates into the model. A measure of depressed affect was taken from the MIDUS Refresher 1 project, and selected covariates were used to study their relationships with daily fatigue reports. Fatigue is a common complaint linked to a wide range of medical diagnoses and medical treatments (Tiesinga et al., 1996) and its definitions vary considerably (Finsterer & Mahjoub, 2014). Characterizing patterns in perceptions of daily fatigue, including its occurrence and severity, has been helpful in understanding daily experiences of patient populations (Kratz et al., 2017; Powell et al., 2017) and responses to therapeutic treatments in certain populations (Parrish et al., 2008; Schwartz, 2000). As a symptom, fatigue has been linked to depressed affect and indicated to be a symptom following treatment for clinical depression (Barkham et al., 1996; Nierenberg et al., 2010). The aim of the analyses here is to show how the association between depressed affect and daily measures of fatigue can be studied using a two-part MELS model.

### Measures

Daily fatigue was assessed using two questions from the daily diary survey. Participants were asked a series of questions about symptoms they may have experienced, all in reference to the day of the interview. The first question asked about fatigue. If the respondent responded "yes" to fatigue, then a follow-up question asked the participant to rate the severity on a 1 (very mild) to 10 (very severe) scale. Previous evidence supports a similar single question about current fatigue as a valid nonspecific measure of fatigue (van Hooff et al., 2007).

The level of depressed affect was assessed by a scale included in the survey battery of MIDUS Refresher 1. Prior to being asked questions from this scale, participants were asked if during the previous 12 months, there was a period of at least 2 weeks when they felt sad, blue, or depressed. If a positive response was given, they were asked seven questions about those 2 weeks; if these feelings were experienced for more than 2 weeks during the previous 12 months, then they were asked to think about the worst 2-week experience. The first two of the seven questions, for example, asked if (yes or no) during that 2-week period, "did you lose interest in most things" and "feel more tired out or low on energy than is usual?" An individual's score from the scale, hereafter called $DA_i$, was created by summing responses to the seven items. Individuals not answering positively about having experienced a 2-week period or longer of feeling sad, blue, or depressed were assigned a score of 0.

### Covariates

Biological sex was coded as $female_i$ = 1 if female and 0 if male. Age was measured at the time of the daily survey. The subscript $i$ indicates that the variable is subject-specific and time-invariant. From previous research (Shrout et al., 2018) and research using data from the MIDUS project series that included an analysis of fatigue measures (Blozis, 2024b), self-report measures at the first assessment are known to exhibit an "initial elevation" (IE) bias in which responses at the first assessment tend to be more extreme relative to the responses that follow. To account for IE effects, an indicator of the first interview day was included in the model, where $IE_{ti}$ = 1 if the response was from the first interview and $IE_{ti}$ = 0 if not.

### Missing Data

Data for covariates and depression scores were complete. About 74.2% of the participants had eight daily measures of fatigue, and 14.3% were missing data for 1 day. The remaining 11.5% were

missing fatigue measures for 2 to 6 days. Daily fatigue was assumed to be missing at random based on results from fitting a random pattern-mixture model that assumed data were not missing at random (Blozis, 2024a; see Supplemental Digital Content [http://links.lww.com/NRES/A552]). Missing at random means that whether or not data are missing is independent of the missing data.



## Mixed-Effects Models

Statistical analysis of repeated measures has historically relied on classic methods (e.g., RM-ANOVA) that eventually gave way to mixed-effects models that simultaneously describe the mean response and individual differences in characteristics used to describe responses. Shin (2009) demonstrates applications of mixed-effects models (a.k.a. hierarchical linear models) in nursing research. Mixed-effects models can be estimated using maximum likelihood (ML), and consequently, missing data and data observed at different time points for different subjects are readily handled (Schafer & Graham, 2002).

In a mixed-effects model, there is at least one random subject effect. The fixed effects are constant across all population members, whereas a random subject effect varies between subjects. For example, a model that assumes a random intercept and a fixed effect for a time-varying covariate would permit the intercept to differ between individuals, and the effect of the covariate would be constant across individuals. In this case, the individual response levels would deviate above and below the mean response level (when the covariate is equal to 0); however, the effect of the covariate would be held constant for all individuals. Conversely, a model that includes a random intercept and a random effect for a time-varying covariate would permit the intercept and the effect of the covariate to differ between individuals. In this case, the individual intercepts would deviate above and below the mean intercept, and the individual effects of the covariate would deviate above and below the mean effect of the covariate. Whereas a random subject effect varies between individuals, the residual at the occasion level varies between both occasions and subjects. The residual reflects the difference between an observed score and the fitted value for a given model and is often used to assess how well a model fits a set of data.

## Two-Part MELS Models

The focal point of this paper is to illustrate a two-part MELS model to jointly analyze a binary variable that denotes the presence versus the absence of a symptom and a continuous variable that denotes the intensity of the symptom when reported. From the daily diary data, we use the binary indicators and severity measures of fatigue to show how to describe the daily variability in each aspect of fatigue, how to examine the relationship between the two aspects and, finally, how to incorporate time-varying and time-invariant covariates into a model.

For the 782 participants of the daily survey, there were 5,756 daily binary indicators of fatigue, and on those days when fatigue was reported, there were 1,411 severity ratings. For notation, let $u_{ti}$ be the binary indicator of fatigue on day $t$ for individual $i$, where $i = 1,\ldots, 782$, and $t = 1,\ldots, n_i$, where $n_i$ is the number of daily indicators of fatigue for the individual. That is, $u_{ti} = 1$ if fatigue was reported and $u_{ti} = 0$ if fatigue was not reported. Daily indicators for nine individuals are displayed in Figure 1. As shown, five participants (IDs 30015, 30052, 30092, 30113, 30140) did not report fatigue, whereas four others (IDs 30126, 30144, 30151, 30154) showed intermittent patterns of fatigue. Next, let $v_{ti}$ be the fatigue

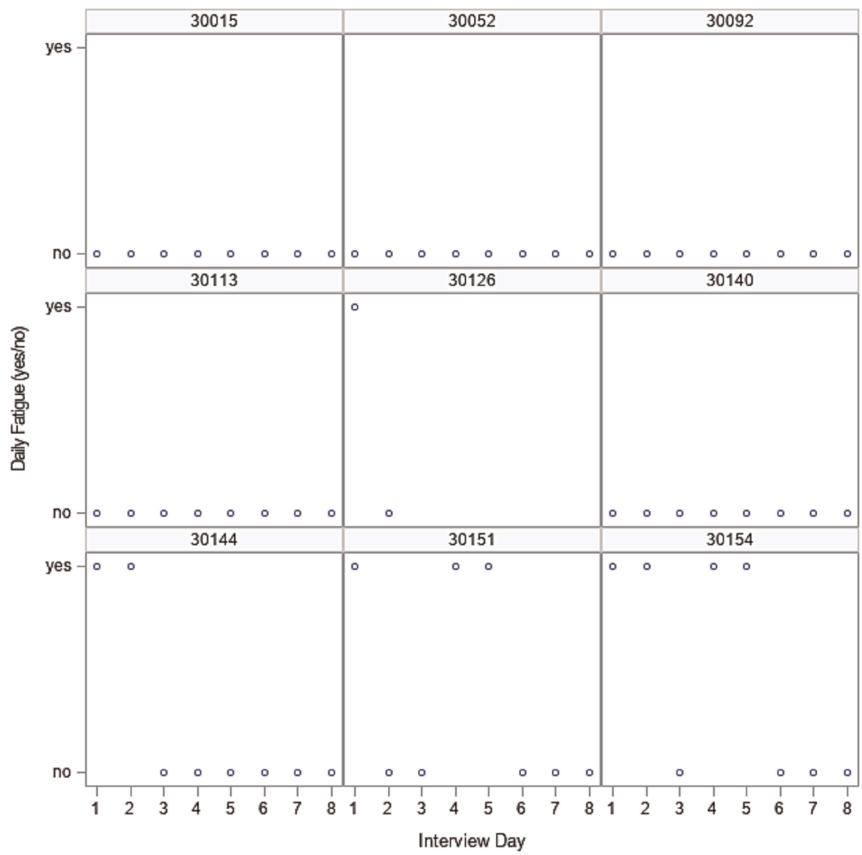


**FIGURE 1.** Binary indicators of fatigue by interview day for a select set of nine subjects.

severity on day $t$ for individual $i$ conditional that fatigue was reported, where $t = 1,\ldots, m_i$ and $m_i$ is the number of days when fatigue was reported. Daily fatigue severity ratings (conditional that fatigue was reported on a given day) for nine individuals are displayed in Figure 2. Participant ID 30151, for example—whose data appear in both Figures 1 and 2—reported fatigue symptoms on interview days 1, 4, and 5, so severity ratings for this participant appear in Figure 2 for only those particular days. Ratings for all who had at least one positive report of fatigue ($n = 458$, 58.6% of the total sample of 782) are displayed in Figure 3.

The daily survey was an observational study about the everyday lives of adults living in the United States, so fatigue responses were not expected to follow a particular trend, on average, such as increasing or decreasing across days. Given this, the (log odds) of the binary indicators of fatigue and the severity ratings were hypothesized to not change according to the interview day. For both aspects of fatigue, we begin with unconditional models for which no predictors are included. For the binary indicator, a logistic mixed-effects model was applied. For this measure, the logit of the probability that individual $i$ reported fatigue on day $t$, denoted by $\eta_{ti}$, was

$$\eta_{ti} = \log\left[P(u_{ti} = 1)/(1-P(u_{ti} = 1))\right],$$

that was then assumed to follow a two-level model:

$$\eta_{ti} = \alpha_0 + a_i, \quad (1)$$

where $\alpha_0$ is the population log odds across days and subjects of reporting fatigue, and $a_i$ is a random subject effect that reflects the extent to which the individual's log odds differ from the population log odds.

Severity, conditional that fatigue was reported, followed a linear mixed-effects model:

$$v_{ti} = \beta_0 + b_i + e_{ti}, \quad (2)$$

where $\beta_0$ is the mean severity across days and subjects; $b_i$ is a random subject effect that reflects the difference between an individual's daily mean severity and the population's mean severity. The residual $e_{ti}$ is the difference between an individual's observed and fitted values. This second model part represents a common formulation of a mixed-effects model (see Shin, 2009).

In Equations (1) and (2), the random subject effects $a_i$ and $b_i$, respectively, are assumed to be independently and normally distributed between subjects with means equal to 0 and variances $\phi_a^2$ and $\phi_b^2$, respectively. The variances indicate how much individuals vary from each other in terms of their respective population values. Thus, the extent to which $\phi_a^2$ differs from 0 is reflective of how much individuals vary from each other in their log odds of reporting fatigue across days, Similarly, the extent to which $\phi_b^2$ differs from 0 is reflective of how much individuals vary in their reported daily mean severity ratings across days.

The residual $e_{ti}$ in Equation (2) is assumed to be independently and normally distributed between days and subjects with mean = 0 and variance $\sigma_e^2$. If individuals were not expected to differ from each other in their day-to-day variability in severity ratings, then it would be reasonable to assume homogeneity of the residual variance across days. For daily severity ratings, however, this was

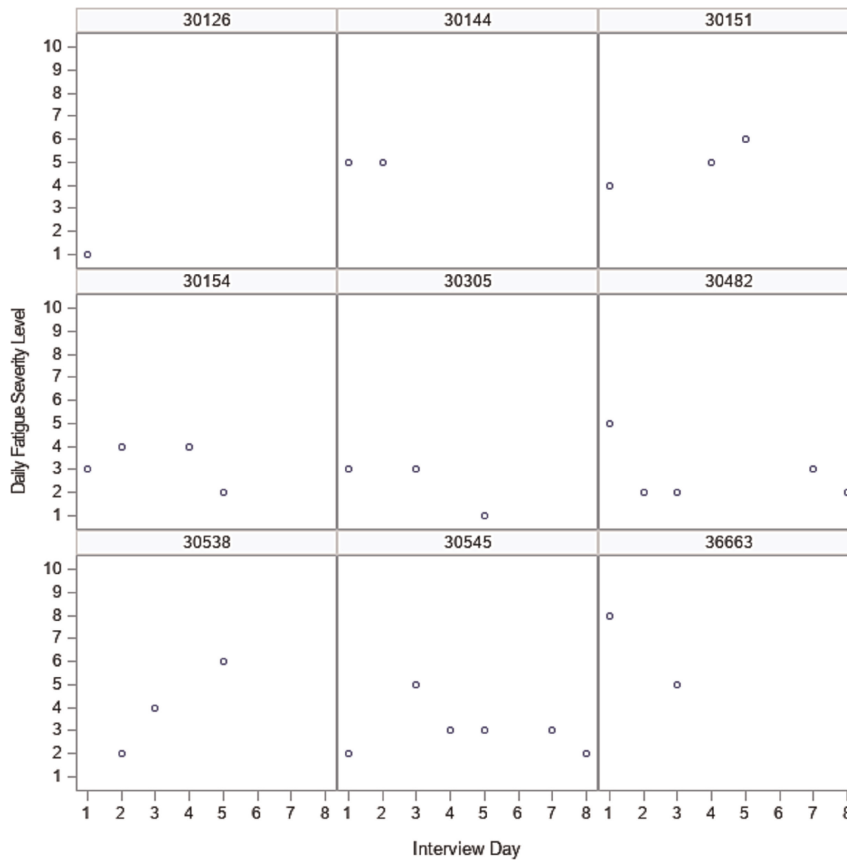


**FIGURE 2.** Fatigue severity ratings by interview day for a select set of nine subjects.

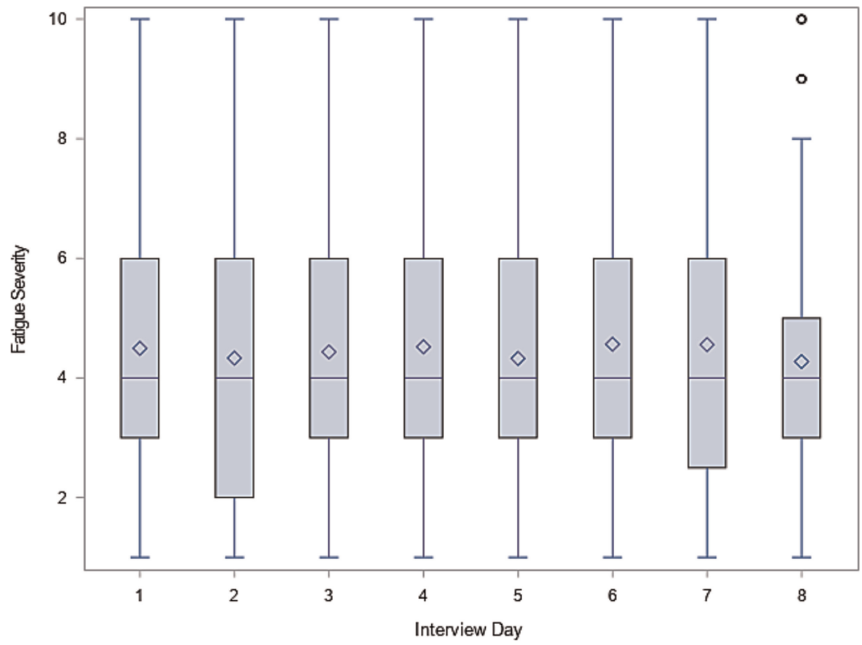


**FIGURE 3.** Boxplots of daily fatigue severity ratings by interview day ($n$ = 458 subjects). *Note.* Fatigue severity ratings are conditional on a positive fatigue report. Of the sample of 782 subjects, 458 had at least one severity score. In the boxplot displays, the solid line inside the box is the sample median and the diamond is the sample mean. Outliers are identified for Interview Day 8 by the open circles, indicating that two cases have scores that are greater than 1.5 times the interquartile range.



not expected. So, a model for the residual variance included a random scale effect to allow for between-person differences in the day-to-day variability:

$$\sigma_e^2 = \exp(\tau_0 + c_i), \tag{3}$$

where the exponentiated value of $\tau_0$ (i.e., $e^{\tau_0}$)) is the residual variance for an individual whose random scale effect $c_i$ is equal to 0, where $c_i$ is assumed to be independently and normally distributed with a mean of 0 and variance $\phi_c^2$. The extent to which $\phi_c^2$ differs from 0 is reflective of how much individuals vary in their day-to-day (log) variance of severity ratings.

The models for $u_{ti}$ and $v_{ti}$ are joined at the subject level by the covariances between their respective random subject effects. That is, the subject-level log odds of reporting fatigue, the daily mean severity rating, and the within-subject daily variability in severity ratings are allowed to covary and together are represented in the covariance matrix $\mathbf{\Phi}$:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_a^2 & \phi_{ab} & \phi_{ac} \\ \phi_{ba} & \phi_b^2 & \phi_{bc} \\ \phi_{ca} & \phi_{cb} & \phi_c^2 \end{bmatrix},$$

where the variances of the random subject effects are on the diagonal, and their covariances are on the off diagonal. The correlations between the random subject effects can be obtained by standardizing the covariance matrix. Assuming positive correlations between the random effects, for example, interpretation of the correlations would be as follows: A positive correlation between the random subject effect relating to the binary indicators of fatigue and the random subject effect relating to the severity ratings would indicate that for those who have a greater tendency to report fatigue to have relatively high daily mean severity ratings. A positive correlation between the random subject effect relating to binary indicators of fatigue and the random scale effect relating to daily variability in severity ratings would indicate that for those who have a greater tendency to report fatigue to also show greater day-to-day variability in severity ratings. Finally, a positive correlation between the random subject effect relating to daily mean severity ratings and the random scale effect would indicate that those with relatively high daily mean severity ratings also show greater day-to-day variability in severity.

## RESULTS

ML estimation of a two-part MELS model was carried out using SAS (version 9.4) with PROC NLMIXED. Statistical software for estimation of nonlinear mixed-effects models is needed due to a nonlinear function that is used to model a variance, such as Equation (3). SAS scripts are in the SDC.

Fatigue severity ratings by interview day are positively skewed (see Figure 2). The fit of the model that assumed that severity ratings were normally distributed was compared to the fit of a model that assumed ratings were log-normally distributed, with the latter being an improvement in fit. The model for the binary fatigue indicator remained the same in both models. Going forward, the fatigue severity ratings were assumed to be log-normally distributed.

Table 1 includes estimates of the fixed effects, and Table 2 gives estimates relating to the variances of the random subject effects and those relating to the variance of the occasion-level residuals. In Table 1, the estimated population logit was $\hat{\alpha}_0 = -1.89$ (95% CI: −2.09, −1.69), corresponding to an overall probability of .13 that an individual reported fatigue. The estimated mean log severity rating was $\hat{\beta}_0 = 1.16$ (95% CI: 1.10, 1.22), corresponding to a rating of $e^{1.16} = 3.18$.

Table 2 reports the estimated standard deviations of the random subject effects and their correlations. Deviance tests evaluated the need for each of the three random effects; they supported their inclusion, suggesting individual differences in the likeliness to report fatigue, the daily mean log severity rating, and the day-to-day variability in log severity ratings. Regarding the correlations between the random subject effects, the correlation between the individual's

**TABLE 1.** ML Estimates of Fixed Effects of the Mean Structure ($n$ = 782)

| | Unconditional model | Conditional model |
|---|---|---|
| Outcome: fatigue (yes/no) | | |
| Logistic model parameters | Estimate (95% CI) | Estimate (95% CI) |
| Log odds, $\alpha_0$ | −1.89 (−2.09 to −1.69) | −3.02 (−3.39 to −2.64) |
| IE, $\alpha_1$ | | 1.51 (1.27 to 1.76) |
| female$_i$, $\alpha_2$ | | 1.08 (0.66 to 1.49) |
| age$_i$, $\alpha_3$ | | −0.00 (−0.02 to 0.01) |
| DA$_i$, $\alpha_4$ | | 0.28 (0.18 to 0.38) |
| Outcome: fatigue severity | | |
| Generalized linear model parameters | Estimate (95% CI) | Estimate (95% CI) |
| Daily mean, $\beta_0$ | 1.16 (1.10 to 1.22) | 1.03 (0.94 to 1.11) |
| IE, $\beta_1$ | | 0.15 (0.08 to 0.21) |
| female$_i$, $\beta_2$ | | 0.09 (0.01 to 0.18) |
| age$_i$, $\beta_3$ | | 0.00 (−0.00 to 0.00) |
| DA$_i$, $\beta_4$ | | 0.05 (0.03 to 0.07) |
| −2lnL | 10724 | 10444 |
| AIC | 10742 | 10502 |
| BIC | 10784 | 10637 |

*Note*. ML = maximum likelihood; Est(95% CI) = ML estimate with the 95% confidence interval in parentheses; IE = initial elevation; DA = depressed affect; −2lnL = −2∗loglikelihood; AIC = Akaike information criterion: AIC = $-2lnL + 2k$, where k is the number of parameters estimated; BIC = Bayesian information criterion: BIC = $-2lnL + k *$ log ($n$), where $n$ is the number of subjects.

likeliness to report fatigue and daily mean log severity was $r$ = .46, indicating a moderate tendency for those with a relatively high likeliness to report fatigue to have a relatively high daily mean log severity rating. The correlation between the likeliness to report fatigue and the day-to-day variability in log severity ratings was $r = -.32$, indicating a moderate tendency for those with a relatively high likeliness to report fatigue to have relatively less day-to-day variability in log severity ratings. Thus, individuals who were more likely to report fatigue had greater stability in their log severity ratings across days. The correlation between the daily mean log severity rating and the day-to-day variability in log severity ratings was $r = -.73$, indicating a moderately strong tendency for those with a higher daily mean log severity rating to have less day-to-day variability (i.e., greater stability) in severity ratings.

## Covariates for the Mean and Covariance Structures

The two-part MELS model was expanded to include covariates and the DA score. The logit and the fatigue severity rating were assumed to be functions of the covariates and DA:

$$\eta_{ti} = \alpha_0 + \alpha_1 \mathrm{IE}_{ti} + \alpha_2 \mathrm{female}_i + \alpha_3 \mathrm{age}_i + \alpha_4 \mathrm{DA}_i + a_i, \quad (4)$$

$$v_{ti} = \beta_0 + \beta_1 \mathrm{IE}_{ti} + \beta_2 \mathrm{female}_i + \beta_3 \mathrm{age}_i + \beta_4 \mathrm{DA}_i + b_i. \quad (5)$$

The intercept $\alpha_0$ of the logit model is the population-level logit on an interview day other than the first day for a male whose age was equal to the sample mean age and whose depressed affect scale score and random effect $a_i$ were equal to 0. The coefficient $\alpha_1$ is the difference in the logit between the first interview day and the following days, $\alpha_2$ is the difference in the logit between females and males, $\alpha_3$ is the effect of age on the logit, and $\alpha_4$ is the effect of DA on the logit. For the fatigue severity rating in Equation (5), the coefficients have similar interpretations.

Next, the random subject effects variances were specified as functions of the covariates and DA. It is important to note that the random subject effects $a_i$ and $b_i$ in Equations (4) and (5) are conditional on the variables that predict the daily logit and severity rating, respectively. As a result, the variances of $a_i$ and $b_i$, $\phi_a^2$ and $\phi_b^2$, are

**TABLE 2.** ML Estimates of the Between- and Within-Person Covariance Structures ($n$ = 782)

| | Unconditional model | Conditional model |
|---|---|---|
| Between-subject | | |
| Outcome: fatigue (yes/no) | | |
| Logistic model parameters | Estimate (95% CI) | Estimate (95% CI) |
| Log odds, $\gamma_{a0}$ | 1.57 (1.38 to 1.76) | 1.87 (1.58 to 2.17) |
| IE, $\gamma_{a1}$ | | −0.58 (−0.99 to −0.16) |
| female$_i$, $\gamma_{a2}$ | | −0.55 (−0.93 to −0.18) |
| age$_i$, $\gamma_{a3}$ | | 0.01 (−0.01 to 0.02) |
| DA$_i$, $\gamma_{a4}$ | | 0.12 (0.03 to 0.21) |
| Outcome: fatigue severity | | |
| Generalized linear model parameters | Estimate (95% CI) | Estimate (95% CI) |
| Daily mean, $\gamma_{b0}$ | −1.76 (−1.88 to −1.63) | −2.05 (−2.42 to −1.69) |
| IE, $\gamma_{b1}$ | | −0.53 (−0.81 to −0.25) |
| female$_i$, $\gamma_{b2}$ | | 0.40 (0.04 to 0.76) |
| age$_i$, $\gamma_{b3}$ | | 0.01 (−0.01 to 0.02) |
| DA$_i$, $\gamma_{b4}$ | | 0.05 (−0.02 to 0.12) |
| Additional estimates | | |
| SD of subject log odds, $\phi_a$ | 2.19 | 2.55 |
| SD of subject daily mean, $\phi_b$ | 0.42 | 0.36 |
| SD of within-subject scale, $\phi_c$ | 0.86 | 0.87 |
| Corr($b_i$,$a_i$) | .46 | .47 |
| Corr($c_i$,$a_i$) | −.32 | −.34 |
| Corr($c_i$,$b_i$) | −.73 | −.70 |
| Within-subject | Estimate (95% CI) | Estimate (95% CI) |
| $\tau_0$ | −1.89 (−2.15 to −1.64) | −1.76 (−2.11 to −1.41) |
| IE, $\tau_1$ | | −0.16 (−0.42 to 0.11) |
| female$_i$, $\tau_2$ | | 0.02 (−0.26 to 0.30) |
| age$_i$, $\tau_3$ | | −0.00 (−0.02 to 0.00) |
| DA$_i$, $\tau_4$ | | −0.11 (−0.06 to 0.00) |
| Additional estimate | | |
| Variance of the within-subject residual when $c_i$ = 0, $\sigma_e^2$ | .15 | .17 |
| −2lnL | 10724 | 10444 |
| AIC | 10742 | 10502 |
| BIC | 10784 | 10637 |

*Note.* ML = maximum likelihood; Est(95% CI) = maximum likelihood estimate with the 95% confidence interval in parentheses. Exponentiating $\gamma_{a0}$ yields the variance of the subject-specific log odds: exp($\gamma_{a0}$) = $\phi_a^2$. Exponentiating $\gamma_{b0}$ yields the variance of the subject-specific daily mean log severity: exp($\gamma_{b0}$) = ($\phi_b^2$). The standard deviation of the random scale effect, $\phi_c$, = component of the generalized linear model part relating to fatigue severity ratings. Estimation of correlations between the random subject effects were based on Fisher's transformation of the correlation coefficient and then using the inverse transformation to define the correlation in the correlation matrix of the random. Exponentiating $\tau_0$ yields the variance of the within-subject residual of the generalized linear model: exp($\tau_0$) = $\sigma_e^2$ for a subject with a random scale effect $c_i$ equal to 0. IE = initial elevation; DA = depressed affect; −2lnL = −2∗loglikelihood; AIC = the Akaike information criterion: AIC = −2lnL + 2 k, where $k$ is the number of parameters estimated; BIC = Bayesian information criterion: BIC = −2lnL + k*log($n$), where $n$ is the number of subjects.

the variances of the conditional random effects. To do this, an exponential function—similar to that in Equation (3)—was used to express each conditional variance:

$$\phi_a^2 = \exp(\gamma_{a0} + \gamma_{a1}\mathrm{IE}_{ti} + \gamma_{a2}\mathrm{female}_i + \gamma_{a3}\mathrm{age}_i + \gamma_{a4}\mathrm{DA}_i),$$

$$\phi_b^2 = \exp(\gamma_{b0} + \gamma_{b1}\mathrm{IE}_{ti} + \gamma_{b2}\mathrm{female}_i + \gamma_{b3}\mathrm{age}_i + \gamma_{b4}\mathrm{DA}_i).$$

The intercept $\gamma_{a0}$ of the model for the variance of the random effect $a_i$ from the logit model, after its value is exponentiated, is the between-subject variance of the logit on an interview day other than the first day for a male at the sample mean age and whose DA score was equal to 0. The coefficient $\gamma_{a1}$ is the difference in the variance between the first interview day and the following days, $\gamma_{a2}$ is the difference between females and males, $\gamma_{a3}$ is the effect of age, and $\gamma_{a4}$ is the effect of DA. The model's coefficients for the variance of the random effect $b_i$ from the linear model have similar interpretations. A positive effect of an explanatory variable indicates greater between-subject variation of the random effect given a higher level of the explanatory variable. Conversely, a negative effect indicates greater between-subject variation in the random effect, given a lower level of the explanatory variable.

Finally, the within-subject residual variance $\sigma_e^2$ was a function of the covariates and DA, in addition to the random scale effect that was previously included in Equation (3):

$$\sigma_e^2 = \exp(\tau_0 + \tau_1 IE_{ti} + \tau_2 female_i + \tau_3 age_i + \tau_4 DA_i + c_i).$$

The intercept $\tau_0$, after it is exponentiated, is the within-subject variance on a day other than the first interview for a male at the sample mean age and whose DA scale score and random scale effect were equal to 0. The coefficient $\tau_1$ is the difference in the variance between the first interview and the following days, $\tau_2$ is the difference between females and males, $\tau_3$ is the effect of age, and $\tau_4$ is the effect of DA. The random scale effect $c_i$ accounts for heterogeneity of the residual variance that is not due to the measured covariates and DA. Interpretation of the effects is similar to those described earlier with regard to the variances of the random subject effects.

ML estimates are given separately in the last columns of estimates of Tables 1 and 2. From Table 1, fatigue was more likely reported at the first interview relative to other days ($\hat{\alpha}_1 = 1.51$, 95% CI: 1.27, 1.76). Females were more likely to report fatigue than were males ($\hat{\alpha}_2 = 1.08$, 95% CI: 0.66, 1.49). The estimated effect of age on the likeliness to report was about 0 ($\hat{\alpha}_3 = -0.00$, 95% CI: −0.02, 0.01). Relatively high DA scores were positively related to a higher likeliness to report ($\hat{\alpha}_4 = 0.28$, 95% CI: 0.18, 0.38). Daily mean log severity ratings were higher at the first interview relative to other days ($\hat{\beta}_1 = 0.15$, 95% CI: 0.08, 0.21). Females had higher daily mean log severity ratings relative to males ($\hat{\beta}_2 = 0.09$, 95% CI: 0.01, 0.18). The effect of age was close to 0 ($\hat{\beta}_3 = 0.00$, 95% CI: −0.00, 0.00), and relatively high DA scores were positively related to higher daily mean log severity ratings ($\hat{\beta}_4 = 0.05$, 95% CI: 0.03, 0.07).

Table 2 contains the estimated effects of the covariates and DA on the variances of the conditional random subject effects. There was less variation in the conditional logits on the first interview day relative to subsequent days ($\hat{\gamma}_{a1} = -0.58$, 95% CI: −0.99, −0.16). Females were more similar than were males ($\hat{\gamma}_{a2} = -0.55$, 95% CI: −0.93, −0.18). The estimated effect of age on the variance was close to 0 ($\hat{\gamma}_{a3} = 0.01$, 95% CI: −0.01, 0.02). Respondents with high DA scores varied more in their conditional logits than those with lower DA scores ($\hat{\gamma}_{a4} = 0.12$, 95% CI: 0.03, 0.21). Next, we describe how predictors were related to the conditional variance of the random effect relating to fatigue severity. This variance reflects the extent to which respondents were dissimilar in their estimated daily mean log severity ratings across days after the daily means were regressed on the covariates and DA scale scores. There was less variation in the conditional means on the first interview day relative to subsequent days ($\hat{\gamma}_{b1} = -0.53$, 95% CI: −0.81, −0.25). Males were more similar to each other than were females ($\hat{\gamma}_{b2} = 0.40$, 95% CI: 0.04, 0.76). The estimated effect of age was close to 0 ($\hat{\gamma}_{b3} = 0.01$, 95% CI: −0.01, 0.02). The direction of the effect of DA, $\hat{\gamma}_{b4} = 0.05$, was not clear (95% CI: −0.02 to 0.12).

Table 2 additionally contains the estimated effects of the covariates and DA scores on the within-subject residual variance. This variance reflects the extent to which respondents varied from day-to-day in the logs of their severity ratings after ratings had been regressed on the covariates and DA. Similar to the preceding results, the effect of age was close to 0 ($\hat{\tau}_3 = -0.00$, 95% CI: −0.02, 0.00). The estimates relating to the effects of the other variables were not clear because all of the estimated 95% CIs spanned—sometimes widely—from negative to positive values: For the effect of the first interview day, $\hat{\tau}_1 = -0.16$, 95% CI: −0.42, 0.11. For biological sex, $\hat{\tau}_2 = 0.02$, 95% CI: −0.26, 0.30. For DA, $\hat{\tau}_4 = -0.11$, 95% CI: −0.06, 0.00. Given its relevance to the within-subject residual, it is notable that the variance of the random scale was estimated to be 0.87 (see Table 2), suggesting that respondents differed in the degree of day-to-day variability in severity ratings after accounting for the DA score and covariates.

## DISCUSSION

Health diaries offer insights into the individual patients' experiences and improve data quality by reducing response biases and errors in recall that can occur with retrospective reports. A daily inquiry helps measure variation in health reports within patients, giving insights into the state of a patient's health status, daily functionality, and quality of life, ultimately providing a more engaged patient care experience (Veloso Costa et al., 2021). Models like the MELS model offer increased insights into how quality of care can be improved. That is, MELS models can be used to assess the nuances in health outcome variability both within and between patients. While this report illustrated these qualities of the method using data from the general population, a pressing challenge in quality nursing care is the management of illnesses in particular populations. In these areas, a MELS model could be used to assess how factors, including medication adherence, lifestyle changes, and patient demographics, influence health outcomes over time. Considering the mean response (location) and the variability in responses (scale), the MELS model provides a comprehensive understanding of how different patients manage their conditions, which then allows nurses to create detailed and analytic-backed plans that incorporate these patterns to help predict clinical outcomes.

Other approaches aimed at measuring individual differences in within-subject variation have relied on ad hoc methods by calculating each individual's SD or coefficient of variation (i.e., SD/mean) of repeated measures about each individual's mean (Bartley et al., 2018; Pringle et al., 2003; Rothwell et al., 2010; Shimbo et al., 2012) and used these measures of variation to characterize individual differences of daily reports. A strength of a MELS model is that the day-to-day variation is represented as a model parameter that does not contain the measurement error that is present in these ad hoc methods.

## CONCLUSION

Understanding health changes over time is crucial to assessing the progression of illness and treatment effects. The MELS model provides a way to quantitatively assess individual differences in

variability in health reports and a means for testing the associations between this variability and other variables.

## REFERENCES


Aroian, K. J., & Vander Wal, J. S. (2007). Measuring elders' symptoms with daily diaries and retrospective reports. *Western Journal of Nursing Research*, 29, 322–343. https://doi.org/10.1177/0193945906293814

Barkham, M., Rees, A., Stiles, W. B., Shapiro, D. A., Hardy, G. E., & Reynolds, S. (1996). Dose–effect relations in time-limited psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, 64, 927–935. https://doi.org/10.1037/0022-006X.64.5.927

Bartley, E. J., Robinson, M. E., & Staud, R. (2018). Pain and fatigue variability patterns distinguish subgroups of fibromyalgia patients. *Journal of Pain*, 19, 372–381. https://doi.org/10.1016/j.jpain.2017.11.014

Berger, A. M., VonEssen, S., Kuhn, B. R., Piper, B. F., Agrawal, S., Lynch, J. C., & Higginbotham, P. (2003). Adherence, sleep, and fatigue outcomes after adjuvant breast cancer chemotherapy: Results of a feasibility intervention study. *Oncology Nursing Forum*, 30, 513–522. https://doi.org/10.1188/03.ONF.513-522

Blozis, S. A. (2024a). Bayesian pattern-mixture models for dropout and intermittently missing data in longitudinal data analysis. *Behavior Research Methods*, 56, 1953–1967. https://doi.org/10.3758/s13428-023-02128-y

Blozis, S. A. (2024b). First-interview response patterns of intensive longitudinal psychological and health data. *Journal of Health Psychology*, 30, 13591053241235751. https://doi.org/10.1177/13591053241235751

Blozis, S. A., McTernan, M., Harring, J. R., & Zheng, Q. (2020). Two-part mixed-effects location scale models. *Behavior Research Methods*, 52, 1836–1847. https://doi.org/10.3758/s13428-020-01359-7

Burman, M. E. (1995). Health diaries in nursing research and practice. *Journal of Nursing Scholarship*, 27, 147–152. https://doi.org/10.1111/j.1547-5069.1995.tb00839.x

Finsterer, J., & Mahjoub, S. Z. (2014). Fatigue in healthy and diseased individuals. *The American Journal of Hospice & Palliative Care*, 31, 562–575. https://doi.org/10.1177/1049909113494748

Hale, C. (2016). Review: Fostering self-care behaviours through symptom diary use? An exploratory, cross-sectional study about the use of and attitude towards a symptom diary of patients with heart failure. *Journal of Research in Nursing*, 21, 542–543. https://doi.org/10.1177/1744987116674404

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, 31, 3328–3336. https://doi.org/10.1002/sim.5338

Ito, E., & Tadaka, E. (2022). Effectiveness of the Online Daily Diary (ONDIARY) program on family caregivers of advanced cancer patients: A home-based palliative care trial. *Complementary Therapies in Clinical Practice*, 46, 101508. https://doi.org/10.1016/j.ctcp.2021.101508

Kent, D. M., & Hayward, R. A. (2007). Limitations of applying summary results of clinical trials to individual patients: The need for risk stratification. *JAMA*, 298, 1209–1212. https://doi.org/10.1001/jama.298.10.1209

Kratz, A. L., Braley, T. J., Foxen-Craft, E., Scott, E., Murphy, J. F., 3rd, & Murphy, S. L. (2017). How do pain, fatigue, depressive, and cognitive symptoms relate to well-being and social and physical functioning in the daily lives of individuals with multiple sclerosis? *Archives of Physical Medicine and Rehabilitation*, 98, 2160–2166. https://doi.org/10.1016/j.apmr.2017.07.004

Lischetzke, T., & Könen, T. (2023). Daily diary methodology. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research*. Springer. https://doi.org/10.1007/978-3-031-17299-1_657

Nierenberg, A. A., Husain, M. M., Trivedi, M. H., Fava, M., Warden, D., Wisniewski, S. R., Miyahara, S., & Rush, A. J. (2010). Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: A STAR*D report. *Psychological Medicine*, 40, 41–50. https://doi.org/10.1017/S0033291709006011

Parrish, B. P., Zautra, A. J., & Davis, M. C. (2008). The role of positive and negative interpersonal events on daily fatigue in women with fibromyalgia, rheumatoid arthritis, and osteoarthritis. *Health Psychology*, 27, 694–702. https://doi.org/10.1037/0278-6133.27.6.694

Powell, D. J. H., Liossi, C., Schlotz, W., & Moss-Morris, R. (2017). Tracking daily fatigue fluctuations in multiple sclerosis: Ecological momentary assessment provides unique insights. *Journal of Behavioral Medicine*, 40, 772–783. https://doi.org/10.1007/s10865-017-9840-4

Pringle, E., Phillips, C., Thijs, L., Davidson, C., Staessen, J. A., de Leeuw, P. W., Jaaskivi, M., Nachev, C., Parati, G., O'Brien, E. T., Tuomilehto, J., Webster, J., Bulpitt, C. J., & Fagard, R. H., Syst-Eur investigators. (2003). Systolic blood pressure variability as a risk factor for stroke and cardiovascular mortality in the elderly hypertensive population. *Journal of Hypertension*, 21, 2251–2257. https://doi.org/10.1097/00004872-200312000-00012

Rothwell, P. M., Howard, S. C., Dolan, E., O'Brien, E., Dobson, J. E., Dahlöf, B., Sever, P. S., & Poulter, N. R. (2010). Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. *Lancet*, 375, 895–905. https://doi.org/10.1016/S0140-6736(10)60308-X

Ryff, C. D., & Almeida, D. M. (2012–2014). Midlife in the United States (MIDUS refresher 1): Daily Diary Project. Inter-university Consortium for Political and Social Research [distributor], 2020-12-14. https://doi.org/10.3886/ICPSR37083.v2

Ryff, C., Almeida, D. M., Ayanian, J. Z., Binkley, N., Carr, D. S., Coe, C., Davidson, R., Grzywacz, J., Karlamangla, A., Krueger, Lachman, M. E., Love, G., Mailick, M., Mroczek, D. K., Radler, B., Seeman, T. E., Sloan, R., Thomas, D., Weinstein, M., & Williams, D. R. (2011–2014). Midlife in the United States (MIDUS Refresher 1), Inter-university Consortium for Political and Social Research [distributor], 2017-11-20. https://doi.org/10.3886/ICPSR36532.v3

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schwartz, A. L. (2000). Daily fatigue patterns and effect of exercise in women with breast cancer. *Cancer Practice*, 8, 16–24. https://doi.org/10.1046/j.1523-5394.2000.81003.x

Shimbo, D., Newman, J. D., Aragaki, A. K., LaMonte, M. J., Bavry, A. A., Allison, M., Manson, J. E., & Wassertheil-Smoller, S. (2012). Association between annual visit-to-visit blood pressure variability and stroke in postmenopausal women: Data from the Women's Health Initiative. *Hypertension*, 60, 625–630. https://doi.org/10.1161/hypertensionaha.112.193094

Shin, J. H. (2009). Application of repeated-measures analysis of variance and hierarchical linear model in nursing research. *Nursing Research*, 58, 211–217. https://doi.org/10.1097/NNR.0b013e318199b5ae

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences*, 115, E15–E23. https://doi.org/10.1073/pnas.1712277115

Tiesinga, L. J., Dassen, T. W., & Halfens, R. J. (1996). Fatigue: A summary of the definitions, dimensions, and indicators. *International Journal of Nursing Terminologies and Classifications*, 7, 51–62. https://doi.org/10.1111/j.1744-618x.1996.tb00293.x

Topaz, M., Koleck, T. A., Onorato, N., Smaldone, A., & Bakken, S. (2021). Nursing documentation of symptoms is associated with higher risk of emergency department visits and hospitalizations in homecare patients. *Nursing Outlook*, 69, 435–446. https://doi.org/10.1016/j.outlook.2020.12.007

van Hooff, M. L. M., Geurts, S. A. E., Kompier, M. A. J., & Taris, T. W. (2007). "How fatigued do you currently feel?" Convergent and discriminant validity of a single-item fatigue measure. *Journal of Occupational Health*, 49, 224–234. https://doi.org/10.1539/joh.49.224

Veloso Costa, A., Padfield, O., Elliott, S., & Hayden, P. (2021). Improving patient diary use in intensive care: A quality improvement report. *Journal of the Intensive Care Society*, 22, 27–33. https://doi.org/10.1177/1751143719885295

Verbrugge, L. M. (1980). Health diaries. *Medical Care*, 18, 73–95.