



Published in final edited form as:

J Psychosom Res. 2014 September ; 77(3): 205–212. doi:10.1016/j.jpsychores.2014.07.006.

Distinguishing between frequency and intensity of health-related symptoms from diary assessments

Stefan Schneider, PhD¹ and Arthur A. Stone, PhD^{1,2}

¹University of Southern California

²Stony Brook University

Abstract

Objective—This study investigated the utility of distinguishing between the frequency and intensity of self-reported symptoms using diary-based assessments in a representative sample of U.S. residents.

Methods—Data from the 2010 American Time Use Survey were analyzed, in which 12,000 respondents provided a diary about the prior day and rated their pain, tiredness, stress, and sadness for three of the day's episodes. A "two-part" latent variable modeling strategy was applied to estimate the frequency (propensity of its presence) and intensity (mean level when present) of each symptom from the diary ratings. Regression analyses comparing differences in symptom frequency and intensity across demographic factors (gender, age, income, education) were conducted to evaluate the utility of the distinction.

Results—Frequency and intensity measures were reliably estimated from 3 daily episodes, were moderately intercorrelated for each symptom domain (r_s .39 to .60), and were differentially associated with demographic factors. Gender differences were evident only in symptom intensity, not frequency, with women reporting more intense symptoms. Comparisons by age showed pronounced declines in the frequency of tiredness and stress in older age, with no age-differences in the intensity of these symptoms. Higher socioeconomic status was associated with a lower intensity of pain, tiredness, stress, and sadness, but a higher frequency of tiredness and stress.

Conclusion—A useful distinction between symptom frequency and intensity may be made from diary-based assessments. It reveals demographic differences that are otherwise obscured and enables a more detailed characterization of health-related experiences in people's daily life.

© 2014 Elsevier Inc. All rights reserved.

Corresponding author: Stefan Schneider, Ph.D., Dornsife Center for Self-Report Science and Center for Social & Economic Research, University of Southern California, 12015 Waterfront Drive, Playa Vista, CA 90094-2536, Phone: 310-448-5354, schneids@usc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Competing Interest Statement

Arthur Stone has been a Senior Scientist with the Gallup Organization and a Senior Consultant with ERT, inc., which could be perceived to constitute a conflict of interest.

Keywords

Diary; frequency; intensity; two-part model; pain; tiredness; stress; sadness

Introduction

Interest in self-reported somatic and affective symptoms is high in research, clinical, and health policy settings. Knowledge about health-related symptoms is important for evaluating health care and treatment, for understanding health disparities, and for tracking population trends in health and wellbeing over time [1, 2]. To date, two characteristics of symptom experience – their frequency and their intensity – have often been overlooked or simply combined into single measures, in part, because there has been limited empirical study of the distinction [3]. The question addressed here is whether symptom frequency and intensity should be viewed and can be measured as distinctive health outcomes.

There are compelling conceptual arguments for separating the frequency and intensity of symptom experiences. A person could have symptoms of pain, fatigue, or emotional distress at *mild* levels yet very *often*, whereas another person could have symptoms at *high* levels but only *occasionally*, as in the case of symptom flares. The overall symptom severity (i.e., its average magnitude across time) could be very similar for both people, despite pronounced differences in the composition of symptom frequency and intensity. Discriminating these patterns could have implications for practice and research, perhaps suggesting different mechanisms and indicating different treatment strategies [3, 4].

Despite its theoretical appeal, the frequency-intensity distinction has received little empirical justification in past research on self-reported somatic symptoms. Chang et al. [3] compared retrospective self-report ratings of fatigue using a frequency (none of the time – all of the time) or intensity (not at all – very much) response format and found that the two produced largely corresponding scale scores (correlation of .86). Similarly, scales that ask participants to rate either the frequency (not at all -- almost always) or intensity (not at all -- extremely) of posttraumatic stress disorder symptoms have been found to yield highly overlapping information (correlation of .93) [5]. Based on these findings, it has been argued that the concepts are virtually redundant and that there is little use in querying frequency and intensity of somatic symptoms separately [3–5].

Importantly, however, these studies examined retrospective questionnaires, where respondents were asked to summarize their symptoms over several days (e.g., the past 7 days)[3]. Recall ratings can be impacted by memory biases [6], and contextual factors can influence how people use and interpret frequency and intensity response scales in retrospective self-reports [7]. Symptom diaries, such as ecological momentary assessment (EMA) and the Day Reconstruction method (DRM), mitigate or eliminate the effects of recall bias [6, 8]. In addition, by collecting experience ratings across multiple moments or episodes, outcome measures that summarize these experiences are created by the researcher and not implicitly by the respondents. In other words, measures of symptom frequency and intensity can be directly computed from the diary data instead of relying on the participant's

ability to meaningfully map their experiences onto a response scale that queries either frequency or intensity [9].

The purpose of this study was to investigate whether symptom diaries allow for a reliable and useful distinction between the frequency and intensity of health-related outcomes. Data from a nationally representative sample of over 12,000 individuals collected by the 2010 American Time Use Survey (ATUS, <http://www.bls.gov/tus/>) were utilized. Similar to the DRM, respondents were interviewed about the prior day, provided a “chunking” of the day into distinct episodes, and rated their pain, tiredness, sadness, and stress for 3 selected episodes. We conceptualized the frequency of a symptom as the proportion of episodes in which it was endorsed as present, and its intensity as the average level of the symptom when it was present, consistent with prior related literature examining basic components of affect [10–12].

To evaluate the utility of the distinction, we examined the extent to which symptom frequency and intensity were differentially associated with demographic characteristics, notably, gender, age, income, and educational attainment. Insight into the prevalence of emotional and somatic symptoms across demographic groups is important for understanding of who is more likely to seek healthcare and to facilitate more cost-effective utilization of healthcare resources. However, it is important that the derived prevalence rates be as precise and informative as possible. Thus, the question we addressed here was whether separating frequency and intensity symptom components reveals demographic differences that are otherwise obscured.

Methods

Participants and procedure

Data collected as part of the US Bureau of Labor Statistics’ 2010 ATUS project with addition of the NIA-supported Wellbeing Module (WBM) were used for this study. The main purpose of ATUS is to develop nationally representative estimates of how people spend their time based on a subset of households who recently completed the Current Population Survey (CPS). Respondents are interviewed over the telephone to provide a detailed time diary of the previous day. In a series of questions, the interviewer asks: “What were you doing?”; “How long did you spend [ACTIVITY]?”; “What did you do next?”, starting at 4 AM of yesterday and ending at 4 AM on the interview day. Thus, episodes are defined based on the temporal sequence of yesterday’s activities. In the 2010 WBM completed after the time-use interview, 3 of the episodes were randomly selected for each respondent to ask about symptoms experienced.¹ Most episodes from the ATUS time diary were eligible for the WBM questions, with the exception of episodes that were shorter than 5 minutes and those that were coded as sleeping, grooming (e.g., personal hygiene), and personal/private activities (e.g., intimacy), given that people may not be able or inclined to report symptoms experienced during these episodes. In addition, due to a programming error in the computer-assisted interview, the last daily episode was excluded from selection for most participants (see <http://www.bls.gov/tus> for sampling and interviewing procedures).

¹The restriction of 3 randomly selected episodes was imposed to limit participant burden.

Measures

Symptoms for each of the selected episodes were rated on a unipolar 7-point scale. For pain, the question was "From 0 to 6, where a 0 means you did not feel any pain at all and a 6 means you were in severe pain, how much pain did you feel during this time?" For tiredness, sadness, and stress, the question was "From 0 to 6, where a 0 means you were not tired/sad/stressed at all and a 6 means you were very tired/sad/stressed, how tired/sad/stressed did you feel during this time?" The order of the symptom domains for each episode was assigned at random for each respondent.

Demographic characteristics gender and age were assessed during the ATUS interview. Socioeconomic status variables income and education are available from CPS interviews conducted 2–5 months prior to ATUS. There were no missing values for gender, age, and education. Family income was missing for 4% of the sample (and about 13% of nonresponses for income were allocated or imputed by ATUS); response categories for income approximated a logarithmic scale of the annual dollar amount, which is the preferred metric for analyses involving income and subjective wellbeing [13]. The sample was 44% female, 80% White, and 48% married, with a mean age of 46.7 ($SD=17.6$) years, and a median family income of \$40,000–\$49,999. One sixth (16%) had less than high school education and 59% at least some college. Weekdays and weekend days were sampled at a 1:1 ratio (see <http://www.bls.gov/tus> for detailed sampling characteristics).

Statistical methods

There were two phases of the analyses. The first was to estimate frequency and intensity from the 3 daily episodes, and results of these analyses were intended to address the reliability and overlap of the two concepts. The second phase assumes enough unique information to test differences in the association between demographic variables and the frequency and intensity measures of the outcome variables.

Two-part modeling of symptom frequency and intensity—We operationally defined the frequency of a symptom as the proportion of episodes in which it was present (i.e., a rating greater than zero) and its intensity as the average level during episodes when it was present. A "two-part" structural equation modeling [14, 15] strategy, was applied for this purpose. As outlined by Olsen and Schafer [14], two-part models have been developed for variables that have a proportion of responses at a single value (often zero) and a continuous distribution among the remaining responses. Zeros are assumed to be *bona fide* valid data values indicating the absence of a symptom, not proxies for negative responses in a truncated distribution. As shown in Figure 1 (left side), the responses are recoded into two indicator variables. A dichotomous indicator distinguishes the absence of a symptom (0=not at all) from its presence at any level (greater than 0) for each episode. A continuous indicator represents the symptom level (1 to 6) if the symptom was present, and is coded missing if the symptom was absent during the episode.

It should be noted that this coding results in unbalanced (i.e., missing) data, with respondents contributing between 0 and 3 observations to the symptom intensity part. However, the missingness is determined by the value on the frequency part, which is always

observed [14]. The two-part model explicitly accounts for the missing data by estimating frequency and intensity as correlated latent factors. As shown in Figure 1 (right side), the model was specified so that the frequency factor represented the latent average of the dichotomous (presence—absence) indicators, and the intensity factor represented the latent average of the continuous (level when present) indicators.² With correlated factors, the nonresponse mechanism for missing values on the intensity part becomes ignorable (missing at random, MAR) in the sense defined by Little and Rubin [16], and full-information maximum likelihood estimation provides unbiased estimates of symptom intensity under MAR [14].

For comparison, we also examined models in which the frequency and intensity symptom components were not distinguished. For this purpose, we estimated one-factor models for continuous variables representing “simple” latent averages of the ratings across the 3 episodes (referred to as “traditional average scoring”). Because the one-factor and two-part models are not nested in each other, we examined Bayesian Information Criterion [17] and log likelihood values to evaluate whether the two-part models had improved fit compared with a one-factor model.

Reliabilities of “scale” scores obtained from 3 episodes were estimated from the factor-analysis (one-factor and two-part) models. For continuous indicators (“traditional” average and intensity factors), we applied the Spearman-Brown formula [18, 19] for reliability $\rho = \sigma^2_{\eta} / [\sigma^2_{\eta} + \sigma^2_e / 3]$, where σ^2_{η} is the latent factor (true score) variance and σ^2_e denotes common error variance of the indicators. For dichotomous indicators (frequency factor), we used the latent-variable method of Raykov, Dimitrov, and Asparuhov [20], which assumes a linear model for latent response variables underlying the probabilities for dichotomous responses to derive the true score and error variances. Both methods provide unbiased reliability estimates under MAR in the presence of missing values.

Demographic differences in symptom frequency and intensity—To evaluate and compare the relationships of demographic characteristics with symptom frequency and intensity, the latent frequency and intensity variables for each symptom domain were regressed on gender, age, income, and education. Separate models were estimated for each demographic variable and symptom domain. Review of prior literature on demographic associations with the outcome measures informed our predictions in certain instances. For relationships with age, piecewise linear models with two segments (18 to 54 years and 54+ years) were used to estimate age-related changes during younger to middle age (segment 1) and during older age (segment 2), consistent with prior research [21]. For relationships with income and education, linear and quadratic effects were initially tested given prior evidence of curvilinear associations [13]; however, quadratic effects were not significant for any symptom domain and, therefore, only linear effects are shown. Given that unstandardized regression coefficients for the frequency and intensity components are not directly comparable because they are on different (cumulative normal probability versus

²We note that the two-part model originally suggested by Olsen and Schafer [14] involves an additional latent growth curve component capturing temporal change in each of the two model parts. For the present purposes of solely estimating symptom intensity and frequency as latent averages from multiple diary episodes, the model reduces to a two-part factor model without growth component, loadings fixed at unity, and homogeneous residual variances (a “random intercepts” model).

linear) metrics, the regression coefficients were standardized relative to the variance of the latent frequency and intensity factors for statistical comparison.

All models were estimated using a maximum likelihood estimator with standard errors robust for non-normality [22]. A probit link using numerical integration was employed for estimation of latent symptom frequency from dichotomous indicator variables (a linear link was used for latent intensity). All analyses were conducted at the individual (i.e., respondent) level, with parameter estimates adjusted by statistical weights provided by ATUS that correct for unequal sampling and response rates across demographic subgroups and days of the week (respondent weight WUFINLWGT). Because a series of comparisons were conducted, the statistical significance threshold was set at $p < .001$. *Mplus* version 7.11 [23] was used for all analyses.

Results

Descriptive results

Health-related symptom data were available from 12,829 ATUS Wellbeing Module respondents. Participants provided symptom ratings for 38,059 diary episodes or an average of 2.97 episodes per person, slightly less than the targeted 3 episodes per person (1.1% responses were missing). The mean episode length was 67 minutes ($SD=96$, range 5 to 1107 minutes). Figure 2 show the response distribution for pain, tired, sad, and stress. As can be seen, the distributions were heavily skewed with a large proportion of “not at all” (zero) responses for all symptom domains, indicating that symptoms were present during 23% (sad) to 67% (tired) of the episodes, and absent during the remaining episodes. Two-part models are especially useful for distributions with a preponderance of zeros, given that separating the zeros (captured in the frequency part) from the remaining responses (captured in the intensity part) normalizes the distributions [14]. Compared with one-factor models for traditional symptom averages, the twopart models yielded substantially lower Bayesian Information Criterion and log likelihood values (see Table 1), suggesting that separating frequency and intensity components in the two-part models considerably improved model fit.

Table 2 shows descriptive statistics for each symptom domain using traditional average scoring and using two-part models. The reliabilities of “scale” scores derived from 3 episodes ranged from .77 (tired/stress) to .89 (pain) using traditional averages. When symptom components were separated into frequency and intensity, resulting reliabilities were similar, ranging from .73 (tired) to .85 (pain) for frequency and from .76 (tired) to .91 (pain) for intensity. Correlations between different symptom domains ranged from .41 to .73 (for traditional averages), from .48 to .78 (for frequency) and from .50 to .79 (for intensity). Frequency and intensity measures for the same symptom domain were moderately positively correlated with r s ranging from .39 (sad) to .60 (pain).

Relationships of demographic variables with symptom frequency and intensity

Gender differences—Using traditional average scoring, women were found to report significantly more pain, tiredness, and stress than men, with no significant gender difference

in sadness (effect sizes ranging from $d = +.07$ to $+.25$, see Table 3). Two-part modeling results showed that men and women did not differ in symptom *frequency* on any of the symptom domains (ds ranging from $-.02$ to $+.10$), whereas women showed significantly greater symptom *intensity* than men on all four domains (ds ranging from $+.22$ to $+.36$, see Table 3). Gender differences were consistently more pronounced for symptom intensity than for symptom frequency (difference in ds ranging from $+.22$ to $.30$).

Age differences—Figure 2 shows the patterns of age-differences from traditional and two-part models for each symptom domain. Traditional average scoring suggested significant increases in pain ($\beta = +.17$) and sadness ($\beta = +.12$) over the course of young/middle age, with no further significant increase during older age. For tiredness and stress, traditional scoring showed no significant age-trends during young/middle age but significant declines in tiredness ($\beta = -.13$) and stress ($\beta = -.15$) in older age. When two-part modeling was applied to pain scores, frequency and intensity scores followed corresponding age-patterns, with significant increases in pain frequency ($\beta = +.20$) and intensity ($\beta = +.16$) during young/middle age and no further significant increases in older age. However, two-part modeling revealed divergent age-patterns in the frequency and intensity of tiredness and stress (see Figure 2). Specifically, only the *frequency* of tiredness ($\beta = -.15$) and stress ($\beta = -.21$) showed significant decreases in older age: at an age of 55 years, people reported being tired during 65% and stressed during 48% of daily episodes, whereas the rates declined to 45% for tiredness 21% for stress at 80 years of age. In contrast, no significant age-differences were evident in the *intensity* of tiredness or stress. Finally, sadness frequency and intensity showed trends in opposite directions in older age, with a decrease in frequency ($\beta = -.06$) and a nonsignificant increase in intensity ($\beta = +.04$; trends differed significantly from each other, $p < .001$).

Differences by income—Results for income are shown in Figure 4. When traditional average scoring was applied, higher income was associated with lower pain ($\beta = -.18$) and sadness ($\beta = -.15$), but income was not significantly related to tiredness ($\beta = .01$) and stress ($\beta = -.04$). As shown in Figure 4, two-part modeling revealed different patterns of association. Even though income was negatively related to the frequency ($\beta = -.13$) and intensity ($\beta = -.29$) of pain, as well as the frequency ($\beta = -.10$) and intensity ($\beta = -.25$) of sadness, the effects were significantly ($p < .001$) more pronounced for intensity than for frequency components. Furthermore, tiredness and stress yielded significantly different effects in opposite directions: whereas the *frequency* of tiredness ($\beta = +.10$) and stress ($\beta = +.06$) increased with higher income, the *intensity* of tiredness ($\beta = -.11$) and stress ($\beta = -.18$) decreased with higher income.

Differences by education—Results for educational attainment largely paralleled those obtained for income. Traditional scoring showed negative linear effects of education on pain ($\beta = -.10$) and sadness ($\beta = -.08$), but not on tiredness ($\beta = -.03$) and stress ($\beta = +.02$). In two-part models, even though education was negatively related to pain frequency ($\beta = -.06$) and sadness frequency ($\beta = -.04$), the effects were significantly more pronounced for pain intensity ($\beta = -.22$) and sadness intensity ($\beta = -.23$). Furthermore, higher education was associated with a greater frequency of tiredness ($\beta = +.05$) and stress ($\beta = +.12$) but lower intensity of tiredness ($\beta = -.12$) and stress ($\beta = -.12$).

Discussion

The results of this study support the distinction between the frequency and intensity of health-related symptoms in the context of a multi-episode, diary-based assessment method. Both symptom components could be captured with reasonably good reliability of $>.70$ based on no more than 3 diary episodes collected in the ATUS wellbeing module. Notably, the reliabilities were similar to those obtained when applying the traditional strategy of averaging the ratings, suggesting that decomposing symptom ratings into frequency and intensity parts does not yield a reduction in measurement reliability. Symptom frequencies and intensities were positively intercorrelated for all domains; people who experienced a given symptom more often also tended to experience it at higher levels. However, the magnitude of these correlations was moderate even with correction for unreliability, suggesting sufficiently distinct constructs.

Compelling evidence for the distinction was that demographic characteristics were differentially associated with symptom frequencies and intensities. As pointed out by Olsen and Schafer [14], for survey responses that are heavily skewed and piled up at zero “it is natural to view a semi-continuous response as a result of two processes, one determining whether the response is zero and the other determining the actual level if it is non-zero. The two processes are distinct and may be influenced by covariates in different ways” (p. 730). Gender differences were evident only in the intensity, not the frequency, components of pain, tiredness, stress, and sadness, with women reporting more intense symptoms than men. Differences by age were more complex; however, except for pain, they were more pronounced for the frequency than the intensity of symptoms. Over the course of older age, steep declines were evident especially in how often (but not: how much) people reported feeling tired or stressed. Relationships with socioeconomic status variables income and education yielded yet a different pattern involving a reversal in the direction of effects: whereas higher socioeconomic status was associated with a lower intensity of pain, tiredness, stress, and sadness, it was associated with a higher frequency of tiredness and stress (even though effect sizes were very small in several instances). Notably, in all cases, the traditional average scoring approach yielded patterns of association with demographic characteristics in between those obtained for frequency and intensity components. Thus, this suggests that when frequency and intensity are not decomposed, a mixing of the effects occurs that can blur demographic differences and in some cases completely conceal them.

Our findings can be interpreted within the context of prior related literature. Studies on gender differences in emotion have consistently shown that women tend to experience both positive and negative affect more intensely (but not necessarily more often) than men [24, 25]. Cultural gender expectations include greater caregiving responsibilities, which may encourage women to be more sensitive to their inner sensations and respond more strongly to them [24]. The present results suggest that these processes may not be limited to emotionality but expand to gender differences the area of somatic experiences.

The pattern of relationships between age and pain closely replicate previous results from a diary-based national survey [26]. For the other symptom domains, our results are in line with emerging evidence suggesting that while aging is associated with declines in some

areas of functioning, several aspects of wellbeing are maintained and even improve with age [11, 27]. In fact, our finding of pronounced declines in the frequency (but not the intensity) of stress and tiredness in older age may be due to a decrease in the number (but not the magnitude) of stressful and taxing daily life experiences in later life, perhaps partly reflecting effects of retirement.

With respect to socioeconomic status, prior research has documented weak and inconsistent relationships between income and self-reports of stress or emotional wellbeing [13, 28]. It has been argued that the types of activities that wealthier people are engaged in on a daily basis are not necessarily associated with more positive wellbeing experiences [28]. However, the present study findings suggest that significant effects of socioeconomic status can be obscured by combining the frequency and intensity of experiences into a single metric. Whereas no effects were evident when examining traditional averages of tiredness and stress, higher income and education were consistently associated with lower symptom intensities, but with more frequent of tiredness and stress.

There are several limitations to these results that we acknowledge. Foremost among them is that this is essentially a conceptual demonstration concerning the utility of examining frequency and intensity as separate concepts over a single metric. The presented data were representative of the US population and can inform and improve understanding of health-related symptoms in national surveys. However, only a small sample of variables – both predictor and outcomes – were examined, and we do not know the degree to which these findings generalize to other health outcomes or to specific groups of individuals with a chronic illness or medical condition. Moreover, this study was limited to demographic differences and future research will need to address the relative utility of symptom frequencies and intensities for predicting health behaviors or changes in medical status.

Additional limitations with respect to the symptom diaries analyzed in this study should also be noted. ATUS collects diary data about the prior day, which may introduce recall bias to some extent. However, the data collection method is consistent with the Day Reconstruction Method, which has shown to provides ratings comparable with experience sampling data [8]. Respondents were not informed about the specific questions prior to the interview (they received a generic invitation letter); this may facilitate genuine responses, but could also increase the risk of unreliable information if some participants are not attuned to the areas of their lives that they are asked to report on. Furthermore, each symptom domain was measured with a single item to limit participant burden, but the use of multi-item scales would be preferable to capture the full breadth and complexity of the constructs. In addition, symptom ratings were collected only for 3 diary episodes; while this provided satisfactory reliability, data on more episodes per person would yield a more precise characterization of symptom frequency and intensity, and using less than 3 episodes would not be recommended for future research.

These limitations notwithstanding, the results of this study demonstrate the principles of a data collection and data analysis strategy for decomposing symptom ratings into frequency and intensity components from diary ratings. They challenge the notion that frequency and intensity of health-related symptoms are redundant and interchangeable constructs and

instead suggest that separating the two components enables a more detailed characterization of symptom experiences in people's daily life. The diary measurement method used in ATUS, while applicable to large-scale data collection in the general population, can easily be implemented in smaller studies examining samples with a specific medical condition. Future studies applying the proposed methods for studying frequency and intensity symptom components in medical samples could use the present findings as a benchmark to evaluate the generalizability of their results. Important directions for future research will be to evaluate the clinical utility of the distinction between symptom frequency and intensity for characterizing the symptomatology and etiology of clinical diagnoses, and for gaining a more differentiated understanding of treatment effects in clinical trials.

Acknowledgments

This research was supported by a grant from the National Institute on Aging (P30 AG024928; Stone, PI).

References

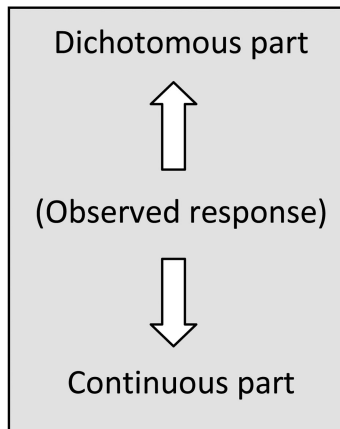
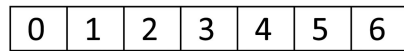
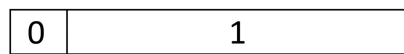
1. Guyatt GH, Feeny DH, Patrick DL. Measuring Health-Related Quality-of-Life. *Ann Intern Med.* 1993; 118(8):622–629. [PubMed: 8452328]
2. Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. *Qual Life Res.* 2008; 17(9):1125–1135. [PubMed: 18836850]
3. Chang C, Cella D, Clarke S, Heinemann A, Von Roenn J, Harvey R. Should symptoms be scaled for intensity, frequency, or both? *Palliative & supportive care.* 2003; 1(1):51. [PubMed: 16594288]
4. Mannion AF, Balague F, Pellise F, Cedraschi C. Pain measurement in patients with low back pain. *Nat Clin Pract Rheum.* 2007; 3(11):610–618.
5. Elhai JD, Lindsay BM, Gray MJ, Grubaugh AL, North TC, Frueh BC. Examining the uniqueness of frequency and intensity symptom ratings in posttraumatic stress disorder assessment. *J Nerv Ment Dis.* 2006; 194(12):940–944. [PubMed: 17164633]
6. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain.* 2004; 107(1–2):61–69. [PubMed: 14715390]
7. Schwarz, N. Retrospective and concurrent self-reports: The rationale for real-time data capture. In: Stone, AA.; Shiffman, SS.; Atienza, A.; Nebeling, L., editors. *The science of real-time data capture: Self-reports in health research.* New York: Oxford University Press; 2007. p. 11–26.
8. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: The day reconstruction method. *Science.* 2004; 306(5702):1776–1780. [PubMed: 15576620]
9. Stone AA, Broderick JE, Schneider S, Schwartz JE. Expanding Options for Developing Outcome Measures From Momentary Assessment Data. *Psychosom Med.* 2012; 74(4):387–397. [PubMed: 22582336]
10. Schimmack U, Diener E. Affect intensity: Separating intensity and frequency in repeatedly measured affect. *J Pers Soc Psychol.* 1997; 73(6):1313–1329.
11. Carstensen LL, Pasupathi M, Mayr U, Nesselroade JR. Emotional experience in everyday life across the adult life span. *J Pers Soc Psychol.* 2000; 79(4):644–655. [PubMed: 11045744]
12. Diener E, Larsen RJ, Levine S, Emmons RA. Intensity and Frequency - Dimensions Underlying Positive and Negative Affect. *J Pers Soc Psychol.* 1985; 48(5):1253–1265. [PubMed: 3998989]
13. Kahneman D, Deaton A. High income improves evaluation of life but not emotional well-being. *P Natl Acad Sci USA.* 2010; 107(38):16489–16493.
14. Olsen MK, Schafer JL. A two-part random-effects model for semicontinuous longitudinal data. *J Am Stat Assoc.* 2001; 96(454):730–745.

15. Duan N, Manning WG Jr, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business & Economic Statistics*. 1983; 1(2):115–126.
16. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. New York: Wiley; 1987.
17. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6:461–464.
18. Raykov T, Marcoulides GA. On multilevel model reliability estimation from the perspective of structural equation modeling. *Structural Equation Modeling*. 2006; 13(1):130–141.
19. Raykov T, Penev S. Evaluation of Reliability Coefficients for Two-Level Models via Latent Variable Analysis. *Struct Equ Modeling*. 2010; 17(4):629–641.
20. Raykov T, Dimitrov DM, Asparouhov T. Evaluation of Scale Reliability With Binary Measures Using Latent Variable Modeling. *Struct Equ Modeling*. 2010; 17(2):265–279.
21. Stone AA, Schwartz JE, Broderick JE, Deaton A. A snapshot of the age distribution of psychological well-being in the United States. *P Natl Acad Sci USA*. 2010; 107(22):9985–9990.
22. White H. A Heteroskedasticity-Consistent Covariance-Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*. 1980; 48(4):817–838.
23. Muthén, LK.; Muthén, BO. *Mplus user's guide*. 7th ed.. Los Angeles, CA: Muthén & Muthén; 1998–2012.
24. Grossman M, Wood W. Sex-Differences in Intensity of Emotional Experience - a Social-Role Interpretation. *J Pers Soc Psychol*. 1993; 65(5):1010–1022. [PubMed: 8246109]
25. Fujita F, Diener E, Sandvik E. Gender differences in negative affect and well-being: the case for emotional intensity. *J Pers Soc Psychol*. 1991; 61(3):427–434. [PubMed: 1941513]
26. Krueger AB, Stone AA. Assessment of pain: a community-based diary survey in the USA. *Lancet*. 2008; 371(9623):1519–1525. [PubMed: 18456101]
27. Carstensen LL, Turan B, Scheibe S, Ram N, Ersner-Hershfield H, Samanez-Larkin GR, et al. Emotional Experience Improves With Age: Evidence Based on Over 10 Years of Experience Sampling. *Psychol Aging*. 2011; 26(1):21–33. [PubMed: 20973600]
28. Kahneman D, Krueger AB, Schkade D, Schwarz N, Stone AA. Would you be happier if you were richer? A focusing illusion. *Science*. 2006; 312(5782):1908–1910. [PubMed: 16809528]

Highlights

- We ask if symptom frequency and intensity should be viewed as distinctive outcomes.
- A statistical "two-part" model is applied to multiple diary ratings.
- Symptom frequency and intensity can be reliably distinguished using this strategy.
- Differential associations with demographic factors suggest utility of the approach.

Coding of episode-level ratings



Two-part model for 3 episodes

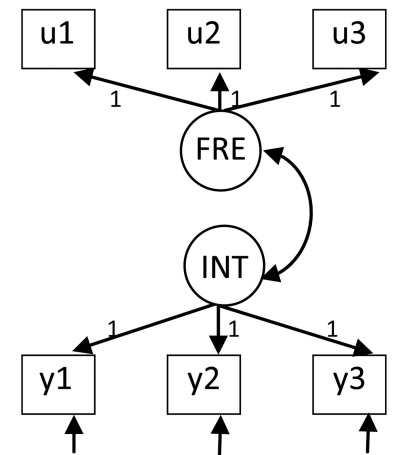


Figure 1.

Coding of episode-level ratings (left) for use in two-part model (right). In the two-part model, u_1 - u_3 represent binary indicators of a latent frequency factor FRE, y_1 - y_3 represent continuous indicators of a latent intensity factor INT. To ensure that all 3 diary episodes contributed equally to the latent factors, loadings were fixed to unity and residual variances for the 3 continuous indicators were held equal.

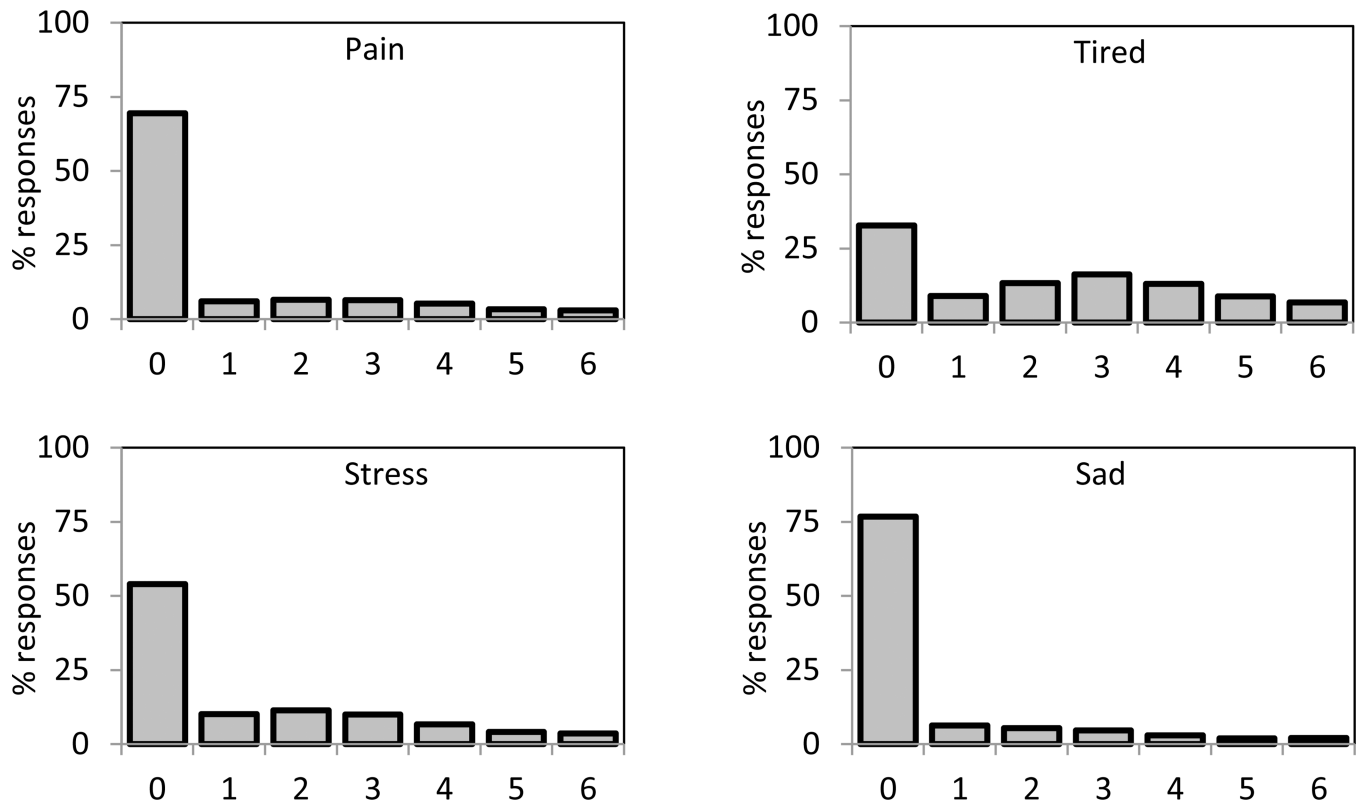


Figure 2.
Observed frequency distribution of episode-level ratings. Ratings are on a unipolar scale where 0 = not at all and 6 = very/severe.

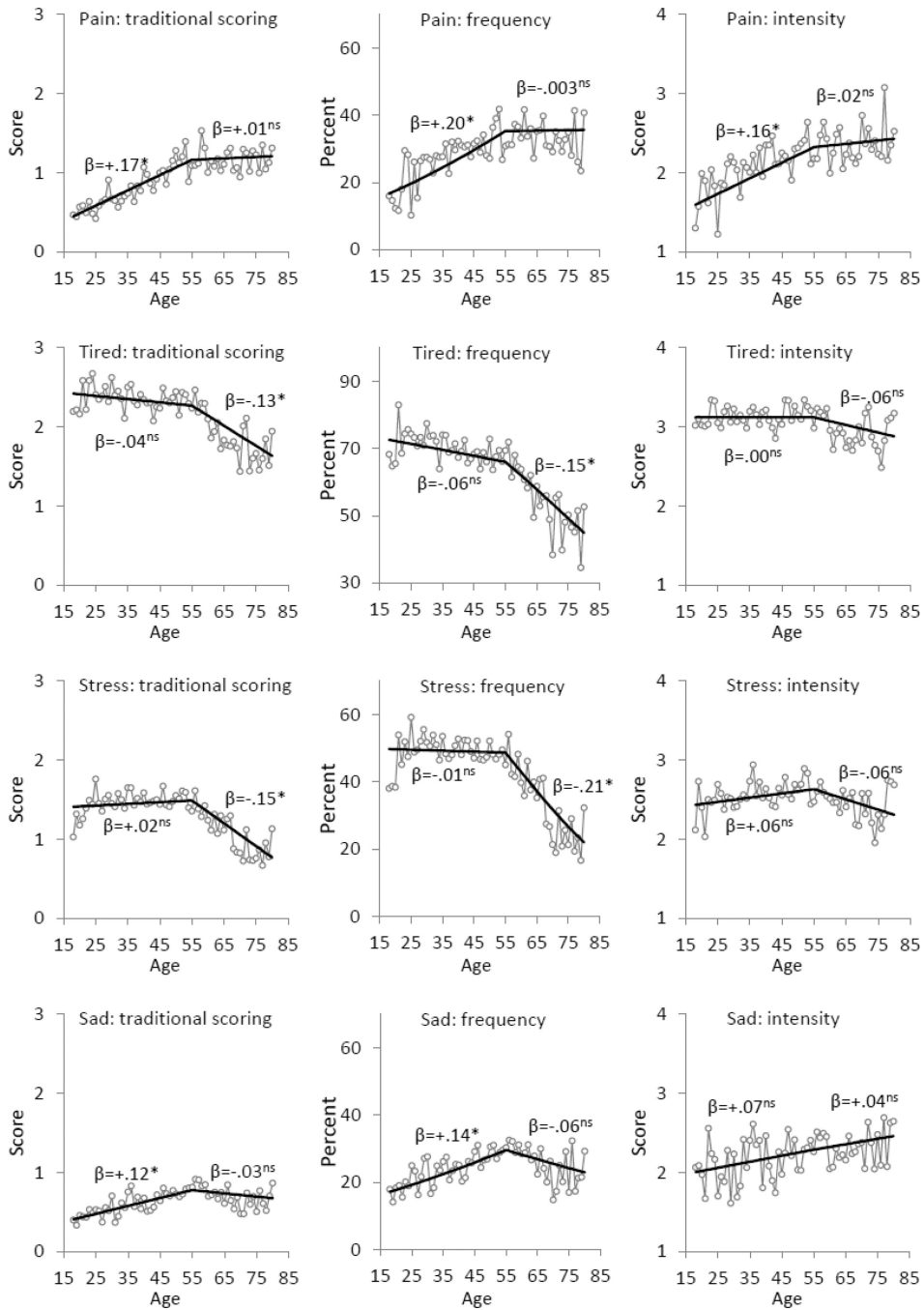


Figure 3. Age differences in pain, tiredness, stress, and sadness based on traditional average scores and based on two-part model distinction of frequency and intensity. Solid lines represent estimated slopes based on piecewise linear regression models (inflection point of 54 years of age), and open circles represent scores plotted for each year of age (average $n = 191$ per year of age, range of $n = 73$ to 283). * $p < .001$.

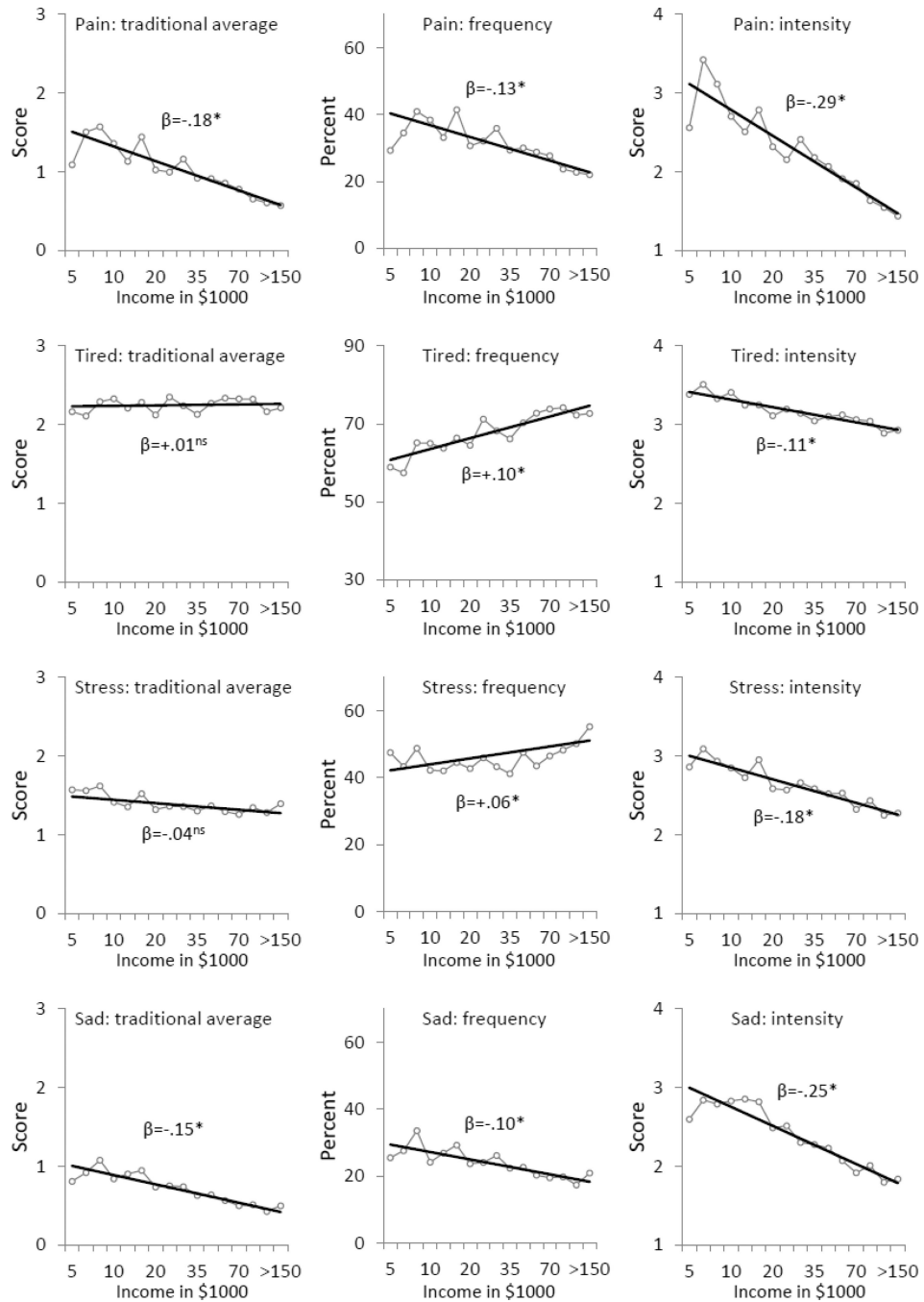


Figure 4.

Income in relation to pain, tiredness, stress, and sadness based on traditional average scores and based on two-part model distinction of frequency and intensity. Solid lines represent estimated linear effects, and open circles represent scores plotted for each income category (average $n = 768$ per income category, range of $n = 254$ to 1434). * $p < .001$.

Table 1

Log likelihood and Bayesian Information Criterion values for one-factor models (traditional average) and two-part models (distinguishing frequency and intensity factors)

	One-factor models (traditional average) ^a	Two-part models ^b
Pain		
Log likelihood	-61921.77	-35207.20
Bayesian Information Criterion	123871.91	70471.15
Tired		
Log likelihood	-75426.31	-65525.91
Bayesian Information Criterion	150880.99	131089.51
Stress		
Log likelihood	-71570.27	-53955.24
Bayesian Information Criterion	143168.92	107967.24
Sad		
Log likelihood	-60718.53	-31645.71
Bayesian Information Criterion	121465.44	63348.17

Note:

^aNumber of parameters = 3;

^bNumber of parameters = 6.

Table 2

Means, standard deviations, reliabilities, and intercorrelations of symptoms based on traditional average scores and two-part models

	Traditional model	Two-part model		
	Composite score	Frequency	Intensity	Frequency with intensity ^a
Pain				
Mean	0.90	29.7%	2.03	
SD	0.93	6.3% -68.0% ^b	1.51	
Reliability	.89	.85	.91	
Correlation				.60
Tired				
Mean	2.25	68.5%	3.10	
SD	1.43	30.2% -93.1% ^b	1.09	
Reliability	.77	.73	.76	
Correlation				.41
Stress				
Mean	1.35	46.8%	2.50	
SD	1.30	14.0% -82.1% ^b	1.14	
Reliability	.77	.74	.77	
Correlation				.48
Sad				
Mean	0.62	22.8%	2.15	
SD	1.05	4.0% -60.1% ^b	1.31	
Reliability	.80	.78	.84	
Correlation				.39

Note: $N = 12,829$.

^aCorrelations are adjusted for unreliability.

^bValues are -1SD and +1SD around the mean frequency.

Table 3

Gender differences in pain, tiredness, stress, and sadness based on traditional average scores and based on two-part models

	Traditional model			Two-part model		
	Composite score		Frequency	Intensity		Difference
	Mean	ES (SE)	Mean %	Mean	ES (SE)	ES (SE)
Pain						
Men	0.83	--	29.8	--	1.88	--
Women	0.96	+09 (.02)*	29.6	-01 (.03)	2.21	+22 (.04)*
Tired						
Men	2.07	--	67.0	--	2.90	--
Women	2.42	+25 (.03)*	70.4	+10 (.03)	3.28	+36 (.03)*
Stress						
Men	1.27	--	46.8	--	2.32	--
Women	1.43	+13 (.03)*	46.9	+00 (.03)	2.67	+30 (.04)*
Sad						
Men	0.59	--	23.0	--	1.98	--
Women	0.66	+07 (.03)	22.6	-02 (.03)	2.30	+25 (.04)*

Note:

* $p < .001$.

Effect sizes (ES) represent the standardized group difference calculated as the difference in means relative to the latent factor variance of each measure; SE = standard error. Sample sizes are $n = 5,634$ women and $n = 7,195$ men.