Evans & Jason, *Cogent Psychology* (2015), 2: 1079945
http://dx.doi.org/10.1080/23311908.2015.1079945

cogent psychology

HEALTH PSYCHOLOGY | RESEARCH ARTICLE

The impact of symptom stability on time frame and recall reliability in CFS

Meredyth Evans[1] and Leonard A. Jason[2*]

Received: 21 November 2014
Accepted: 28 July 2015
Published: 28 August 2015

*Corresponding author: Leonard A. Jason, Center for Community Research, DePaul University, Chicago, IL, USA
E-mail: ljason@depaul.edu

Reviewing editor:
Peter Walla, University of Newcastle, Australia

Additional information is available at the end of the article

Abstract: This study is an investigation of the potential impact of perceived symptom stability on the recall reliability of symptom severity and frequency as reported by individuals with chronic fatigue syndrome (CFS). Symptoms were recalled using three different recall time frames (the past week, the past month, and the past six months) and at two assessment points (with one week in between each assessment). Participants were 51 adults (45 women and 6 men), between the ages of 29 and 66 with a current diagnosis of CFS. Multilevel Model Analyses were used to determine the optimal recall time frame (in terms of test–retest reliability) for reporting symptoms perceived as variable and as stable over time. Headaches were recalled more reliably when they were reported as stable over time. Furthermore, the optimal time frame in terms of test–retest reliability for stable symptoms was highly uniform, such that all Fukuda CFS symptoms were more reliably recalled at the six-month time frame. Furthermore, the optimal time frame for CFS symptoms perceived as variable, differed across symptoms. Symptom stability and recall time frame are important to consider in order to improve the accuracy and reliability of the current methods for diagnosing this illness.

Subjects: Health & Society; Mental Health; Psychological Science

Keywords: chronic fatigue syndrome; recall reliability; symptom stability; assessment

ABOUT THE AUTHORS

Leonard A. Jason is the director of the Center for Community Research at DePaul University where he serves as the head of a research team involved in projects related to the assessment of chronic fatigue syndrome (CFS) and Myalgic Encephalomyelitis (ME). Meredyth A Evans received her doctorate in Clinical-Community Psychology at DePaul University and has served as a member of Jason's ME and CFS research team since 2009. The research team's activities are primarily focused on efforts to improve the diagnostic criteria of ME and CFS. The present study fits closely with the scope of the research team's key goals, as it serves as an investigation of factors that influence the reliability and validity of symptom reporting. Efforts to improve the validity of symptom assessment can lead to more valid diagnostic criteria and more effective treatments.

PUBLIC INTEREST STATEMENT

Chronic fatigue syndrome (CFS) is a debilitating illness and the cause is currently unknown. Developing an accurate and reliable way of diagnosing CFS remains a challenge for health professionals. Diagnosis typically requires patients' self-report of their symptoms and there is little research on factors that influence symptom reporting. This study is an investigation of whether symptom stability influences the reliability in which patients report their symptoms over time, and whether symptom stability can influence the optimal time frame for assessing symptoms on a questionnaire. Findings revealed that when headaches were perceived as more stable, they were recalled more reliably. Furthermore, CFS symptoms that were rated as stable over time were recalled with greater reliability when using longer time frames (the past six months vs. past week). More research on factors that influence the accuracy and reliability of symptom reporting may lead to improvements in the way CFS is assessed and diagnosed.

1. Introduction

The assessment of health symptoms associated with chronic fatigue syndrome (CFS), Myalgic Encephalomyelitis (ME), and Myalgic Encephalomyelitis/chronic fatigue syndrome (ME/CFS) often employ retrospective self-report measures, especially for the purpose of determining whether an individual is experiencing the required length, frequency, and severity of symptoms necessary to receive a diagnosis (Hawk, Jason, & Torres Harding, 2007; Jason et al., 1999; King & Jason, 2005; Reeves et al., 2005). Despite the utility in using retrospective methods, it is often not a simple task for individuals to recall specific health information over weeks, months, and years. Two factors that can impact memory for health symptoms are the length of the recall time frame (e.g. past week vs. past month) used on health surveys, and the perceived symptom stability (i.e. the pattern/consistency in which the health symptom is experienced) (Broderick et al., 2008; Stone & Shiffman, 2002).

Often there is no explanation as to why a particular time frame is used on a health survey (Broderick et al., 2008). Furthermore, the stability of symptoms is rarely taken into account on health surveys. Previous research has revealed that fairly stable events over time are remembered with greater ease than events that fluctuate (Stone & Shiffman, 2002). A greater awareness of how symptom stability can impact a person's recall of their health symptoms may ultimately improve the accuracy, reliability, and overall relevance of health assessments in the context of these chronic and complex illnesses.

It is important to know as much about the symptom experience of an illness as possible when choosing what time frame to use in a survey. For instance, a short recall period may enhance information accessibility for the specific time period, but if there is no stable pattern to the symptom, a short recall period might not capture the true nature of the symptom's variability over time (Stone & Shiffman, 2002; Stull, Leidy, Parasuraman, & Chassany, 2009). Clarke, Fiebig, and Gerdtham (2008) assert that symptom stability may influence which time frame is best for assessing a given phenomenon. While shorter time frames may reduce forgetting, they emphasize that a short time frame is not always compelling for certain phenomena (Stull et al., 2009). For example, when someone reports about a highly variable health symptom that individual is making an overall assessment of their experience, but cannot indicate the variable nature of the symptom in such a short time period (Stone & Shiffman, 2002). Whereas, when highly variable symptoms are reported over longer time frames, an individual will attempt to summarize their experience, which can reduce reporting accuracy (Stone, Schwartz, Broderick, & Shiffman, 2005).

It is likely that both symptom stability and recall time frame impact the reliability of health symptom reporting. Some studies in the field have examined the test–retest reliability of symptom reporting at different time frames. For instance, a study by Hawk et al. (2007) evaluated the test–retest reliability of a CFS diagnostic instrument (Jason et al., 1997), which assesses Fukuda et al. (1994) defined CFS symptoms using a six-month time frame. They also examined the test–retest reliability for recall of energy and fatigue using the time frames "past day" and "past week." Hawk et al. (2007) found that when using a six-month time frame, the average intraclass correlation scores for items assessing the eight CFS case-defining symptoms were very good (.77). Two symptoms (tender/sore lymph nodes and pain in multiple joints) had somewhat lower reliability scores (.58 and .49, respectively). Items asking participants to rate their perceived energy, expended energy, and fatigue experienced over the past day, had lower reliability scores (.59, .40, and .22, respectively) compared to the same items recalled over the past week (.77, .59, and .81, respectively) (Hawk et al., 2007). The authors have suggested that these symptoms likely fluctuate often and can be more consistently recalled over a longer time frame (Hawk et al., 2007). Additionally, a study by Evans and Jason (2013) used multi-level model analyses to examine whether there is an optimal time frame for reliably recalling key CFS symptoms, and found that when participants were asked to recall their CFS symptoms at two assessment points (with one week apart) they were most reliable when using a six-month recall time frame. Past literature suggests that as recall time frame becomes longer, the accuracy for reporting symptoms such as pain and fatigue is reduced (Broderick et al., 2008; Friedberg & Sohl, 2008;

Sohl & Friedberg, 2008); however, Broderick and colleagues (2008) assert that people with chronic illnesses may have a fairly good understanding of their symptom pattern over time, which may partially explain the findings by Evans and Jason (2013) and by Hawk and colleagues (2007).

It is possible that the stability of a health symptom can moderate the relationship between time frame length and the reliability of symptom reporting. Very little research has been done on the potential impact of perceived symptom stability on the reliability of symptom recall. While Hawk and colleagues (2007) investigated the test–retest reliability of CFS symptom reporting at different time frames, there are no published studies that have investigated the potential influence of symptom stability specifically on the recall reliability of CFS symptoms across varying time frames. The present study is a preliminary one that serves as an investigation of the impact of perceived symptom stability on the test–retest reliability of CFS symptoms across three different recall time frames (e.g. past week, past month, past six months). The authors hypothesize that CFS symptoms will be recalled with greater consistency (yield stronger reliability coefficients) when symptoms are perceived as stable over time rather than variable. Furthermore, the authors aim to better understand whether there is an optimal recall time frame in terms of test–retest reliability, for CFS symptoms that are perceived as variable over time, compared to symptoms perceived as stable. Research aiming to improve the methods for CFS symptom assessment could inform the methods used for diagnosing and treating CFS, and could also contribute to the field's understanding of individuals' experience with this illness.

2. Method

2.1. Participants

Participants were identified through the use of an Institutional Review Board (IRB) approved research advertisement published in a CFS Chicago newsletter. The current study group was also made up of individuals who participated in an earlier non-pharmacological intervention at DePaul University's Center for Community Research (Jason et al., 2007). Participants were required to be at least 18 years old and to meet criteria for the Fukuda et al. (1994) case definition for CFS as well as the Centers for Disease Control and Prevention's (CDC's) empiric case definition (Reeves et al., 2005). CFS participants received a five-dollar Amazon gift card upon completion of the study.

2.2. Procedure

Data collection occurred at two time points, with one week between the first and second assessment. Researchers received verbal consent from participants over the phone and scheduled two phone interviews. In order to ensure that all participants completed the questionnaires under the same conditions, both interviews took place over the phone at the same time, and on the same day of the week. During the first interview, participants were not told that they would be asked the same questions a week later, and instead were informed that they would be taking another short symptom survey during the second interview. This was to ensure that participant responses at the second interview were not primed by the first.

During the first phone interview, participants were read questions aloud from the Symptom Inventory Revised (SI-R), which was altered from the original Symptom Inventory developed by the CDC (Wagner et al., 2005) to provide more time frames. Participants were also read questions from a Symptom Stability Questionnaire and a short demographic survey that were developed by the authors of this manuscript. Phone interviewers repeated items for participants as necessary. During the second phone assessment, participants were read items from the SI-R and the Symptom Stability Survey a second time. Following completion of the second phone assessment, participants were debriefed on the purposes of the study.

2.3. Measures

The study measures were administered over the phone by IRB approved graduate students and staff members at the Center for Community Research at DePaul University. Interviewers read the same set of instructions to all participants and recorded responses as they were given.

2.3.1. CFS symptom assessment

The SI-R assesses the presence, frequency, and severity of the Fukuda et al. (1994) case-defining symptoms of CFS (post-exertional malaise, unrefreshing sleep, problems with memory and/or concentration, muscle aches and pains, joint pain, sore throat, tender lymph nodes/swollen glands, and headaches). The CDC Symptom Inventory (Wagner et al., 2005) assesses the frequency and severity of symptoms over the past month and has been shown to have good internal consistency, with a Chronbach's alpha coefficient of .88 for the total inventory score, and .87 for the total score from a short form version that includes only six symptoms (fatigue after exertion, unrefreshing sleep, muscle aches, sleeping problems, problems with memory, and problems with concentration). For the purposes of this study, revisions to the SI-R included the addition of four time frames: right now, past week, past month, and past six months. Participants' frequency and severity ratings on the SI-R were multiplied to create a composite score for each symptom at the past week, past month, and past six-month intervals, with scores ranging from 0 to 25 (Wagner et al., 2005). The authors are not aware of any previous studies on the test–retest reliability of the CDC Symptom Inventory (Wagner et al., 2005).

2.3.2. Symptom stability

The Symptom Stability Survey was administered to participants at both phone assessments and is a measure of a participant's perceived stability of the case-defining symptoms. For each symptom listed on the Symptom Stability Survey, respondents indicated whether they perceived each symptom to have been relatively stable, fluctuating/variable, or not present over the course of the past six months. The Symptom Stability Survey was developed by this author at DePaul University for the purposes of the current study.

2.3.3. Demographic information

Participants were administered a short demographic survey during the first assessment and following completion of the SI-R. The demographic survey included eight questions, which assessed age, gender, weight, height, race, marital status, occupational status, number of children, and highest grade level.

2.3.4. Statistical analyses

The present study utilized a multilevel modeling (MLM) approach within a repeated measures design in order to assess the reliability of symptom reports across two interview assessments. In order to assess the reliability of symptom reports using an MLM approach, the slope coefficients were observed, and those coefficients observed to be closest to 1.0, represented more reliable symptom reporting. Additionally, MLM allows for the assessment of nested data; thus providing a way to quantify the extent to which slope coefficients vary as a function of time frame.

3. Results

3.1. Participant traits

The study population consisted of 51 adults (45 women and 6 men), between the ages of 29 and 66 (M = 50.39) with a current diagnosis of CFS. The majority of participants self-identified as White (94%), one participant self- identified as Asian/Pacific Islander, and two were identified as "other." Two participants self-identified as Latin/Hispanic origin. Approximately half of all participants reported that they were married (N = 27), 13 identified as never married, and 11 identified as divorced. The majority of participants reported holding a standard college degree or higher (70.6%) and all 51 participants reported at least a high school degree. Over half of the participants identified that they were on disability (58.8%), with the large majority reporting CFS as the cited reason for their disability claim. Only one participant reported working full time and six reported working part time. A large proportion of participant diagnoses (84%) were confirmed with letters of documentation by independent physicians.

3.2. Primary results

A multilevel statistical model was used to test the hypothesis that CFS symptoms perceived as stable would yield stronger reliability coefficients when reported across two assessment points (with a one-week period between assessments). Level 1 of the model tested the extent that symptom composite scores (symptom frequency times symptom severity) at interview one predicted symptom composite scores at interview two. Level 2 of the model tested whether the perceived stability of each symptom (dummy coded as 1 = stable and 0 = variable): (1) predicted the symptom composite scores at interview two, and (2) moderated the reliability between interview one and interview two (see Table 1 for descriptive information of the nine ME/CFS symptoms rated as variable; see Table 2 for descriptive information of the nine ME/CFS symptoms rated as stable). The symptom scores at the three recall time frames were not analyzed separately in the analysis but were grouped to represent a single variable referred to as Interview One Scores Collapsed Across time frame. Group mean centering was conducted for the Level 1 variables, so as to control for the influence of between-person variance on the slope coefficients. In the equation below, γ_{00} represents the average scores on the symptom composite for those average in stability (stable vs. non-stable symptoms), γ_{01} represents the average increase on the symptom composite based on stable vs. non-stable symptoms, and γ_{11} represents the average increase of impact of interview 1 scores on the symptom composite based on stable vs. non-stable symptoms.

$$\text{Level 1: } y_{ij} = b_0 + b_1\text{Interview One Scores Collapsed Across Time Frame}_{ij} + r_{ij}$$

$$Y = \text{Symptom composite scores at all three time frames, at interview two}$$

$$\text{Level 2: } b_{0i} = \gamma_{00} + \gamma_{01}\text{Stability}_i + v_i$$

$$b_{1i} = \gamma_{10} + \gamma_{11}\text{Stability}_i$$

When using an alpha cut-off of .05, the hypothesis that stable symptoms would be recalled with greater consistency was supported for three of the eight Fukuda et al. (1994) case-defining CFS symptoms: post-exertional malaise (PEM), headaches, and memory problems. For PEM, there was a significant main effect of symptom stability $F(1, 49) = 5.93$, $p = .019$; however, there was not a main effect of PEM composite scores at interview one, $F(1, 100) = .087$, $p = .768$, in predicting PEM composite scores at interview two. There was a significant interaction effect $F(1, 100) = 4.16$, $p = .044$, such that the relationship between PEM composite scores at the first interview and PEM composite scores at the second interview was significantly stronger for those who rated their symptoms as stable than for those who did not, $b = .48$, SE = .23, $t(100) = 2.04$, $p = .044$. The within variance of the distribution residuals was 6.98 and the between variance of distribution residuals was 14.50. The ICC score was calculated as .68, suggesting that 68% of the variance in predicting PEM scores at interview two is explained by the nesting of both individual factors and symptom stability.

For headaches, there was a significant main effect of symptom stability $F(1, 45) = 9.62$, $p = .003$, but not a main effect of headache composite scores at interview one $F(1, 92) = .01$, $p = .931$, in predicting headache composite scores at interview two. There was a significant interaction effect $F(1, 92) = 13.74$, $p < .001$, such that the relationship between headache composite scores at interview one and headache composite scores at interview two was significantly stronger for those who rated their symptoms as stable than for those who did not, $b = .66$, SE = .18, $t(92) = 3.71$, $p < .001$. The within variance of the distribution residuals was 9.90 and the between variance of distribution residuals was 16.24. The ICC score was calculated as .45, suggesting that 45% of the variance in predicting headache scores at interview two is explained by the nesting of both individual factors and symptom stability.

For memory problems, there was a significant main effect of symptom stability $F(1, 48) = 7.94$, $p = .008$, but not a main effect of memory composite scores at interview one $F(1, 98) = .91$, $p = .343$, in predicting memory composite scores at interview two. There was a significant interaction effect

F(1, 98) = 9.45 *p* = .003, such that the relationship between memory composite scores at interview one and memory composite scores at interview two was significantly stronger for those who rated their symptoms as stable than for those who did not, *b* = .53, SE = .17, *t*(98) = 3.07, *p* = .003. The within variance of the distribution residuals was 5.32 and the between variance of distribution residuals was 35.84. The ICC score was calculated as .87, suggesting that 87% of the variance in predicting headache scores at interview two is explained by the nesting of individual factors and symptom stability.

Table 1. Means and standard deviations of symptom composites on the SI-R at interviews 1 and 2 for symptoms rated as variable

Variable symptoms	Time frame	Interview 1	Interview 2
		M (SD)	M (SD)
Sore throats		*n* = 30	*n* = 30
	Week	4.30 (4.20)	4.80 (5.01)
	Month	4.10 (3.43)	4.33 (4.44)
	Six months	5.10 (4.20)	5.40 (5.75)
Lymph nodes		*n* = 23	*n* = 23
	Week	5.83 (4.56)	4.70 (4.26)
	Month	5.43 (4.83)	5.04 (4.83)
	Six months	6.17 (4.52)	5.22 (3.43)
PEM		*n* = 5	*n* = 5
	Week	12.00 (2.12)	13.60 (5.55)
	Month	14.00 (6.36)	10.60 (3.71)
	Six months	15.20 (7.79)	13.60 (3.05)
Muscle pain		*n* = 20	*n* = 20
	Week	8.75 (4.66)	7.45(3.85)
	Month	7.60 (4.92)	7.85 (3.91)
	Six months	8.50 (4.76)	7.75 (4.04)
Joint pain		*n* = 23	*n* = 23
	Week	7.30 (5.45)	7.26 (5.15)
	Month	7.26 (5.15)	7.61 (5.79)
	Six months	7.78 (5.66)	7.22 (4.60)
Unrefreshing sleep		*n* = 10	*n* = 10
	Week	10.10 (5.92)	11.30 (6.53)
	Month	9.70 (4.72)	10.00 (5.29)
	Six months	7.50 (4.79)	7.70 (3.86)
Headache		*n* = 31	*n* = 31
	Week	6.19 (5.21)	6.68 (5.75)
	Month	6.74 (5.11)	5.65 (4.05)
	Six months	7.19 (5.94)	6.45 (3.80)
Memory		*n* = 19	*n* = 19
	Week	7.16 (5.27)	7.63 (7.27)
	Month	6.5 (5.2)	7.32 (5.31)
	Six months	7.95 (6.51)	7.37 (4.78)
Concentration		*n* = 16	*n* = 16
	Week	10.69 (5.92)	10.06 (5.73)
	Month	10.13 (5.71)	9.81 (4.07)
	Six months	9.81 (4.07)	10.75 (5.36)

Table 2. Means and standard deviations of symptom composites on the SI-R at interviews 1 and 2 for symptoms rated as stable

Stable symptoms	Time frame	Interview 1	Interview 2
		M (SD)	M (SD)
Sore throats		*n* = 8	*n* = 8
	Week	10.88 (5.00)	12.38 (7.73)
	Month	11.63 (5.45)	9.75 (5.83)
	Six months	14.38 (7.01)	12.50 (4.41)
Lymph nodes		*n* = 14	*n* = 14
	Week	11.14 (6.24)	10.57 (6.72)
	Month	11.14 (6.41)	8.86 (6.01)
	Six months	13.14 (7.43)	11.21 (6.42)
PEM		*n* = 46	*n* = 46
	Week	17.35 (4.87)	17.22 (4.65)
	Month	17.33 (4.76)	17.17 (4.65)
	Six months	18.20 (5.45)	17.52 (5.15)
Muscle pain		*n* = 30	*n* = 30
	Week	14.57 (5.86)	13.73(5.39)
	Month	14.87 (6.27)	14.03 (5.76)
	Six months	15.03 (6.20)	14.10 (5.54)
Joint pain		*n* = 22	*n* = 22
	Week	13.09 (5.89)	13.72 (6.23)
	Month	13.95 (6.03)	12.91 (5.65)
	Six months	14.18 (7.20)	13.73 (5.96)
Sleep		*n* = 41	*n* = 41
	Week	18.59 (4.76)	18.90 (5.39)
	Month	17.85 (5.58)	17.24 (5.26)
	Six months	18.32 (6.12)	17.95 (5.55)
Headache		*n* = 16	*n* = 16
	Week	11.56 (7.20)	9.94 (6.69)
	Month	10.44 (4.41)	10.50 (6.48)
	Six months	12.88 (5.49)	11.00 (6.21)
Memory		*n* = 31	*n* = 31
	Week	12.77 (6.28)	11.94 (6.58)
	Month	13.16 (5.75)	12.35 (6.45)
	Six months	12.94 (6.21)	13.13 (7.05)
Concentration		*n* = 35	*n* = 35
	Week	12.23 (5.88)	12.83 (7.08)
	Month	12.66 (6.32)	12.77 (6.88)
	Six months	13.60 (6.57)	13.17 (6.97)

The current study involved the assessment of nine symptoms across three different time frames, resulting in 27 statistical tests of significance. Due to the large number of tests and relatively small sample size, a Bonferroni correction to the alpha level was made by dividing .05 by 27, resulting in a more conservative alpha level of .002. When using the Bonferroni corrected alpha level of .002, the hypothesis that stable symptoms would be recalled with greater consistency was supported for headaches only. The main effect of symptom stability for headaches $F(1, 45) = 9.62$, $p = .003$, is

Table 3. Slope coefficients of variable CFS symptoms at three time frames

Variable symptoms	Time frame	*b*	SE	*df*	*t*	*p*
Sore throat	Week	.20	.17	61.99	1.19	.239
	Month	−.04	.31	58.68	−.14	.888
	Six months*	.80	.17	62.39	4.81	<.001
Lymph node pain	Week	.03	.29	47.25	.12	.91
	Month	.07	.33	46.08	.22	.824
	Six months*	.77	.28	47.43	2.71	.009
PEM	Week	−.29	.47	6.84	−.62	.557
	Month*	.44	.63	6.35	.706	.505
	Six months	.31	.45	6.93	.68	.518
Muscle pain	Week	−.23	.24	4.54	−.98	.335
	Month	−.22	.36	38.12	−.60	.553
	Six months*	.56	.25	40.21	2.25	.03
Joint pain	Week*	.58	.29	47.81	2.00	.051
	Month	.15	.33	47.35	.45	.655
	Six months	.22	.21	49.54	1.04	.304
Unrefreshing sleep	Week	.64	.48	18.35	1.33	.199
	Month*	.82	.67	17.22	1.22	.238
	Six months	.30	.56	17.71	.54	.598
Headaches	Week	−.08	.22	76.77	−.37	.711
	Month	−.51	.36	69.80	−1.42	.159
	Six months*	.32	.24	74.54	1.30	.197
Memory	Week	−.14	.34	38.47	−.41	.688
	Month*	−.07	.45	37.23	−.15	.881
	Six months	−.14	.29	39.71	−.50	.62
Concentration	Week	.04	.27	39.80	.27	.786
	Month	−.17	.37	36.39	−.45	.654
	Six months*	.20	.43	35.81	.46	.647

*Optimal time frame (coefficients closest to 1.0).

approaching significance. With the more conservative alpha level, there continues to be a significant interaction effect $F(1, 92) = 13.74$, $p < .001$, such that the relationship between headache composite scores at interview one and headache composite scores at interview two was significantly stronger for those who rated their symptoms as stable than for those who did not, $b = .66$, SE = .18, $t(92) = 3.71$, $p < .001$.

In order to assess the impact of perceived symptom stability on test–retest reliability at each of the three time frames (past week, past month, and past six months), the slope coefficients were assessed at each time frame for symptoms rated as variable and also for symptoms rated as stable. In the model presented below, the outcome variable represents the symptom composite scores reported at interview two. For ease of description, level 2 of the model tested (1) the extent that symptom composite scores at interview one predicted composite scores at interview two (see Table 1 for descriptive information of the nine CFS symptoms rated as variable; see Table 2 for descriptive information of the nine CFS symptoms rated as stable) and (2) how time frame moderated the way symptom composites at interview one predicted scores at interview two. Level 1 of the model tested the main effect of time frame. Analyses were conducted using the formula listed below, selecting out for variable and stable symptoms. Grand mean centering was conducted for the Level 2 variables, so as to ease interpretation. In the equation below, γ_{00} represents the score on y for six months for those with average scores on the symptom at interview 1, γ_{10} represents the difference in average scores between the past week and six months for those average on the symptom at interview 1, γ_{20} represents the difference in average scores between the past month and six months for those average on the symptom at interview 1, γ_{01} indicates the degree of increase in average scores that are expected based on a 1 point increase in the Symptom Score at Interview 1, γ_{11} represents the degree of increase in the difference between past week and six months expected based on a 1 point increase in Symptom Score at Interview 1, γ_{21} represents the degree of increase in the difference between past month and six months expected based on a 1 point increase in Symptom Score at Interview 1, and r_i represents the variability around average scores as a function of person.

$$\text{Level 1: } y_{ij} = b_{0i} + b_{1i}\ \text{Past Week vs. Six Months}_{ij} + b_{2i}\ \text{Past Month vs. Six Months}_{ij} + r_{ij}$$

$$\text{Level 2: } b_{0i} = \gamma_{00} + \gamma_{01}\ \text{Symptom Score at Interview One}_i + r_i$$

$$b_{1i} = \gamma_{10} + \gamma_{11}\ \text{Symptom Score at Interview One}_i$$

$$b_{2i} = \gamma_{20} + \gamma_{21}\ \text{Symptom Score at Interview One}_i$$

Results of the above analyses revealed that the optimal time frame for CFS symptoms perceived as variable over time, differed across symptoms (see Table 3 for slope coefficients of CFS symptoms

Table 4. Slope coefficients of stable CFS symptoms at three time frames

Stable symptoms	Time frame	*b*	SE	*df*	*t*	*p*
Sore throat	Week	−.97	.85	16.25	−1.15	.268
	Month	1.84	1.34	15.13	1.38	.188
	Six months*	.18	.76	16.67	.23	.819
Lymph node pain	Week	.12	.60	29.83	1.29	.847
	Month	.30	.86	26.62	.347	.731
	Six months*	1.30	.49	28.53	2.63	.014
PEM	Week	.44	.21	104.93	2.13	.04
	Month	.30	.23	102.45	1.31	.195
	Six months*	.79	.19	106.85	4.09	<.001
Muscle pain	Week	−.56	.40	74.44	−1.41	.164
	Month	.75	.55	69.15	1.35	.182
	Six months*	.82	.44	72.40	1.86	.067
Joint pain	Week	−.64	.37	46.25	−1.74	.089
	Month	1.82	.53	43.62	3.43	.001
	Six months*	.95	.35	46.75	2.72	.009
Unrefreshing sleep	Week	.18	.26	89.80	.67	.504
	Month	.29	.23	91.14	1.24	.218
	Six months*	.41	.17	95.17	2.34	.021
Headaches	Week	.00	.24	30.17	.012	.991
	Month	.30	.48	27.75	.62	.543
	Six months*	1.40	.23	30.36	6.04	<.001
Memory	Week	−.17	.21	60.32	−8.26	.412
	Month	.04	.29	59.13	.14	.893
	Six months*	1.03	.19	60.72	5.32	<.001
Concentration	Week	−.43	.31	71.52	−1.41	.164
	Month	.14	.30	71.73	.48	.634
	Six months*	.47	.23	74.15	2.07	.042

*Optimal time frame (coefficients closest to 1.0).

rated as variable at each time frame). Specifically, the past six months was observed as the optimal time frame for five symptoms (e.g. sore throat, lymph node pain, muscle pain, headaches, and difficulty with concentration) whereas the past month was observed to be optimal for reporting PEM, unrefreshing sleep, and memory problems. Lastly, the past week time frame was found to be optimal for variable joint pain. For symptoms perceived as stable, the optimal time frame in terms of test–retest reliability was highly uniform, such that all CFS symptoms were more reliably recalled at the six-month time frame compared to the past week and past month time frames (see Table 4 for slope coefficients of CFS symptoms rated as stable at each time frame).

4. Discussion

The present study serves as a preliminary investigation of the impact of symptom stability on the test–retest reliability of CFS symptom composite scores across two assessment points, and at three different recall time frames (past week, past month, past six months). Consistent with prior research showing that greater stability can improve recall (Stone & Shiffman, 2002; Stull et al., 2009), results of the current study revealed that headaches experienced as stable over time were recalled with greater consistency than headaches perceived as variable. These findings suggest that the degree to which symptom stability impacts the reliability of symptom reporting may depend on symptom type. While it is unclear why this finding was not replicated in more CFS symptoms, it is possible that the relatively small sample size and need for a more conservative alpha cut-off (.002) were factors. Future studies should investigate the impact of CFS symptom stability on test–retest reliability using more robust sample sizes.

The current study also served as an investigation of whether there is an optimal time frame, in terms of test–retest reliability for CFS symptoms perceived as variable vs. stable over time. Of the recall time frames assessed in the current study (the past week, the past month, and the past six months), the optimal time frame for CFS symptoms that were perceived as variable over time differed across symptoms; whereas, the optimal time frame for stable symptoms was highly uniform (optimal at the past six months for all nine CFS symptoms measured). For variable symptoms, the past six months was observed as the optimal time frame for five of the nine symptoms measured (e.g. sore throat, lymph node, muscle pain, headaches, and concentration) whereas the past month was observed to be optimal for recall of PEM, unrefreshing sleep, and memory problems. Lastly, the past week time frame was found to be optimal for variable joint pain. While an optimal time frame could be identified, it is important to note that four symptoms (e.g. PEM, headache, memory, and concentration) had relatively weak slope coefficients, suggesting that when these symptoms are perceived as variable over time, they are not reliably recalled from one week to another. PEM and cognitive difficulties including memory and concentration problems are often cited as cardinal symptoms of CFS and ME (Carruthers et al., 2003, 2011; Jason, Evans, Porter, et al., 2010). Only five of the total 51 participants in this study reported that their PEM was variable over time (see Table 2) and less than half of all participants reported that memory and concentration were variable (19 and 16, respectively; see Table 2). It is possible that when key symptoms of this illness are experienced as variable and fluctuating, they are more difficult to recall consistently. It is also possible that individuals who report these symptoms as variable may represent a unique subset. The majority of participants reported headaches as variable over time (31 out of 47; see Tables 2 and 3); however the optimal slope coefficient for this symptom was still weak at .32, suggesting that headaches have poor recall reliability when perceived as variable over time. These findings may be explained by the tendency for people to use cognitive heuristics (i.e. mental shortcuts) when assessing variable symptoms over a longer time frame; which in turn affects reliability and accuracy of reporting (Bradburn, 2000; Bradburn, Rips, & Shevell, 1987). Due to the fact that the majority of the study population reported their headaches as variable, it is recommended that researchers and physicians are knowledgeable of the fluctuating nature of this symptom as well as the weak reliability in reporting the frequency and severity of this symptom over long time periods.

Stable sore throats had the weakest slope coefficients at all three time frames compared to the other eight symptoms, suggesting that sore throats are not as reliably recalled when perceived as stable over time. Interestingly, the optimal slope coefficient for variable sore throats was higher than the optimal slope coefficient for stable sore throats. It is unclear why sore throats are recalled more consistently when variable and at the past six-month time frame. Sore throats are not widely considered a cardinal symptom of CFS, which is supported by this study data, showing that only 38 of the total 51 participants reported experiencing sore throats over the course of their illness and 58.8% of these respondents reported their sore throats as variable rather than stable over time. Stone and Shiffman (2002) assert that when a respondent reports about a highly variable symptom, they are making an overall assessment of their experience, and cannot indicate the variable nature of the symptom in a short time period. However, when highly variable symptoms are reported over longer time frames, an individual will attempt to summarize their experience. Summarizing variable events over a long time frame has been found to reduce reporting accuracy (Stone et al., 2005); however, in the case of this study, when reporting on particular symptoms, such as sore throats over longer time frames, variability may actually improve recall reliability.

A limitation of this study is reflected in the sample used, as it was not selected through random assignment and thus participants may share certain characteristics that are different from the larger population of individuals affected by CFS. For instance, a large majority of the participants were White women and middle aged. Based on research by Jason and colleagues (1999), we know that CFS occurs at higher rates in African-American and Latino populations; therefore, the current sample may not be generalizable to the entire CFS population.

Another limitation of this study was the uneven frequency of stable vs. variable ratings for certain symptoms. These symptoms were either unevenly rated as stable (e.g. PEM) or variable (e.g. sore throat). Kahn (2011) asserts that establishing a rule of thumb for sample size in achieving statistical power can be difficult because it is important to take sample size into consideration at two levels of data. Kahn reports that a large number of cases in each group improve reliability of Level 1 estimates (Kahn, 2011). Monte Carlo research conducted by Maas and Hox (2005) reveal that samples with at least 30 Level 2 units provide sufficiently unbiased estimates; however, they also report that samples with only 10 Level 2 units maybe also be sufficient. The majority of symptom cases in this study met the 10 unit limit at the Level 2 grouping. However, even when there are 30 units in the Level 2 grouping, the variance components will be biased. Therefore, Kahn (2011) argues that the more cases at Level 2, the better. This concern of sample size at Level 2 and the subsequent impact on power is most prominent for PEM and sore throats, which have very uneven stable vs. variable ratings and also have groups with cases below 10.

Overall, findings from the presented study reveal that perceived symptom stability can influence the optimal time frame for reliably reporting CFS symptoms. The degree of impact that perceived stability has on test–retest reliability may depend on the specific CFS symptom being measured as well as individual characteristics of the respondent. It is suggested that future studies investigate additional factors that may influence the reliability of symptom reporting for this illness. These additional factors may include recent stressful life events, social support, or an individual's stage/progression of illness. It is recommended that future diagnostic and assessment research in this field take contextual factors into account, especially for the purpose of improving upon current methods used to diagnose and understand the symptom pattern and course of this highly complex and often misunderstood illness.

Funding
This work was supported by The National Institute of Allergy and Infectious Diseases [grant number R01AI105781]; The Eunice Kennedy Shriver National Institute of Child Health and Human Development [grant number R01HD072208].

Competing interests
The authors declare no competing interest.

Author details
Meredyth Evans[1]
E-mail: mevans24@depaul.edu
Leonard A. Jason[2]
E-mail: ljason@depaul.edu
[1] Clinical-Community Psychology, DePaul University, Chicago, IL, USA.
[2] Center for Community Research, DePaul University, Chicago, IL, USA.

Citation information
Cite this article as: The impact of symptom stability on time frame and recall reliability in CFS, Meredyth Evans & Leonard A. Jason, *Cogent Psychology* (2015), 2: 1079945.

References

Bradburn, N. M. (2000). Temporal representation and event dating. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self report* (pp. 49–61). Mahwah, NJ: Erlbaum.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*, 157–161. http://dx.doi.org/10.1126/science.3563494

Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *Pain, 139*, 146–157. http://dx.doi.org/10.1016/j.pain.2008.03.024

Carruthers, B. M., Jain, A., De Meirleir, K. L., Peterson, D., Klimas, N., Lerner, A. M., ... van de Sande, M. (2003). Myalgic encephalomyelitis/chronic fatigue syndrome: Clinical working case definition, diagnostic and treatment protocols. *Journal Of Chronic Fatigue Syndrome, 11*, 7–115. http://dx.doi.org/10.1300/J092v11n01_02

Carruthers, B. M., van de Sande, M. I., De Meirleir, K. L., Klimas, N. G., Broderick, G., Mitchell, T., ... Stevens, S. (2011). Myalgic encephalomyelitis: International consensus criteria. *Journal of Internal Medicine, 270*, 327–338. http://dx.doi.org/10.1111/jim.2011.270.issue-4

Clarke, P. M., Fiebig, D. G., & Gerdtham, U. (2008). Optimal recall length in survey design. *Journal of Health Economics, 27*, 1275–1284. http://dx.doi.org/10.1016/j.jhealeco.2008.05.012

Evans, M., & Jason, L. A. (2013). Effects of timeframe on the recall reliability of CFS symptoms. *Evaluation & the Health Professions*. Retrieved from http://ehp.sagepub.com.ezproxy.depaul.edu/content/early/2013/09/09/0163278713497014.full.pdf+html doi:10.1177/0163278713497014

Friedberg, F., & Sohl, S. J. (2008). Memory for fatigue in chronic fatigue syndrome: The relation between weekly recall and momentary ratings. *International Journal of Behavioral Medicine, 15*, 29–33. http://dx.doi.org/10.1007/BF03003071

Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., & Komaroff, A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Annals of Internal Medicine, 121*, 953–959. http://dx.doi.org/10.7326/0003-4819-121-12-199412150-00009

Hawk, C., Jason, L. A., & Torres Harding, S. (2007. Reliability of a chronic fatigue syndrome questionnaire. *Journal Of Chronic Fatigue Syndrome, 13*, 41–66. http://dx.doi.org/10.1300/J092v13n04_05

Jason, L., Ropacki, M., Santoro, N., Richman, J., Heatherly, W., Taylor, R., ... Plioplys, S. (1997). A screening instrument for chronic fatigue syndrome: Reliability and validity. *Journal Of Chronic Fatigue Syndrome, 3*, 39–59. http://dx.doi.org/10.1300/J092v03n01_04

Jason, L. A., King, C. P., Frankenberry, E. L., Jordan, K. M., Tryon, W. W., Rademaker, F., & Huang, C.-F. (1999). Chronic fatigue syndrome: Assessing symptoms and activity level. *Journal of Clinical Psychology, 55*, 411–424. http://dx.doi.org/10.1002/(ISSN)1097-4679

Jason, L. A., Torres-Harding, S., Friedberg, F., Corradi, K., Njoku, M. G., Donalek, J., ... Papernik, M. (2007). Non-pharmacologic interventions for CFS: A randomized trial. *Journal of Clinical Psychology in Medical Settings, 14*, 275–296. http://dx.doi.org/10.1007/s10880-007-9090-7

Jason, L. A., Evans, M., Porter, N., Brown, M., Brown, A. A., Hunnell, J., & Friedberg, F. (2010). The development of a revised Canadian myalgic encephalomyelitis chronic fatigue syndrome case definition. *American Journal of Biochemistry and Biotechnology, 6*, 120–135. http://dx.doi.org/10.3844/ajbbsp.2010.120.135

Kahn, J. H. (2011). Multilevel modeling: Overview and applications to research in counseling psychology. *Journal of Counseling Psychology, 58*, 257–271. http://dx.doi.org/10.1037/a0022680

King, C. P., & Jason, L. A. (2005). Improving the diagnostic criteria and procedures for chronic fatigue syndrome. *Biological Psychology, 68*, 87–106. http://dx.doi.org/10.1016/j.biopsycho.2004.03.015

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86–92. http://dx.doi.org/10.1027/1614-2241.1.3.85

Reeves, W. C., Wagner, D., Nisenbaum, R., Jones, J. F., Gurbaxani, B., Solomon, L., ... Heim, C. (2005). Chronic fatigue syndrome: A clinically empirical approach to its definition and study. *BMC Medicine, 3*(1), 1–9.

Sohl, S. J., & Friedberg, F. (2008). Memory for fatigue in chronic fatigue syndrome: Relationships to fatigue variability, catastrophizing, and negative affect. *Behavioral Medicine, 34*, 29–38. http://dx.doi.org/10.3200/BMED.34.1.29-38

Stone, A. A., Schwartz, J., Broderick, J., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin, 31*, 1340–1346. http://dx.doi.org/10.1177/0146167205275615

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 226–243.

Stull, D., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion, 25*, 929–942. http://dx.doi.org/10.1185/03007990902774765

Wagner, D., Nisenbaum, R., Heim, C., Jones, J. F., Unger, E. R., & Reeves, W. C. (2005). Psychometric properties of a symptom-based questionnaire for the assessment of chronic fatigue syndrome. *Population Health Metrics, 3*, 1–8.