Summer 8-17-2012

Effects of Timeframe on the Recall Reliability of ME/CFS Symptoms

Meredyth Evans
DePaul University, MEREDYTH.EVANS@GMAIL.COM

EFFECTS OF TIMEFRAME ON THE RECALL RELIABILITY OF ME/CFS SYMPTOMS

A Thesis Presented in

Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

DePaul University

BY

Meredyth Anne Evans

August, 2012

Department of Psychology

College of Science and Health

DePaul University

Chicago, Illinois

THESIS COMMITTEE

Leonard A. Jason, Ph.D.

Chairperson

Sheldon Cotler, Ph.D.

Reader

ACKNOWLEDGMENT

I would like to express my sincere appreciation to my thesis chair Dr. Leonard A. Jason and committee member Dr. Sheldon Cotler for the invaluable feedback and guidance they have provided me throughout this project. I would also like to thank my fellow team members for their assistance in recruitment and data collection, as well as Dr. Steve Miller for providing his expertise and direct assistance in statistical analysis. Furthermore, I would like to express my gratitude to the Graduate Research Funding Program at DePaul University for providing funding for participant compensation.

VITA

Meredyth was born in Sterling, Virginia, September 15, 1984. She graduated from Park View High School in 2003 and she received her Bachelor of Arts degree in Psychology from The University of Virginia in 2007. Afterward, she served as a research manager on a study at the University of Virginia Medical Center assessing issues related to fatigue, neurocognitive functioning, and quality of life in gastrointestinal cancer patients receiving chemo-radiation therapy. Additionally, Meredyth served as a project coordinator for an NICHD-funded study on adolescent peer relations. Meredyth received two funding awards in 2010, from the Graduate Research Funding Program and the Doctoral-Undergraduate Opportunities for Scholarship (DUOS) program at DePaul University.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Retrospective self report data is often used for a wide range of research purposes, and is especially prominent in the behavioral and medical fields. This method of self report has been particularly useful in research of the illness commonly referred to as myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). The precise definition of ME/CFS is often debated, but the most widely used case definition stipulates that to meet criteria for the illness, a person must have experienced six or more months of chronic fatigue of new or definite onset, that is not substantially alleviated by rest, not the result of ongoing exertion, and that results in substantial reductions in occupational, social, and personal activities (Fukuda, et al., 1994). Furthermore, a person must have four out of eight accompanying symptoms (e.g. sore throat, lymph node pain, muscle pain, joint pain, post-exertional malaise, headaches of a new or different type, memory and concentration difficulties, and unrefreshing sleep) that have also persisted for at least six months or longer (Fukuda et al., 1994). Retrospective self report methods are often used in research settings to determine whether an individual has experienced the required fatigue and accompanying symptoms for the length, frequency, and severity necessary to receive a diagnosis of ME/CFS (Hawk et al., 2007; Jason et al., 1999; King & Jason, 2005; Reeves et al., 2005).

More generally, self report data can help researchers and health care personnel gain information about the health status and quality of life of individuals suffering from various health problems, medical conditions and

chronic illnesses. Retrospective data have proved to be very useful because it is less costly and time intensive than prospective, longitudinal studies and allows researchers and health professionals to gather information about events that may have occurred weeks, months, or years prior to participation in a research study or formal health evaluation (Beckett, Da Vanzo, Sastry, Panis, & Peterson, 2001). Although there is clear utility in using retrospective self report data to assess health status, retrospective surveys (no matter how simple the question) often require complex mental processes. For instance, when a respondent is confronted with a question that asks about a past event, he/she implicitly goes into a series of steps in order to recall the information. The respondent must process and interpret the question, evaluate the question in terms of his/her individual knowledge and the general scope of the survey, understand the interviewer or administrator's expectations, and evaluate the response in terms of its social desirability (Bradburn, Rips, & Shevell, 1987).

Research on these psychological processes and their potential effects on the validity and reliability of survey data suggest that researchers need to be careful when designing and evaluating studies that involve retrospective methods for gathering information. Specifically these processes can contribute to the phenomenon known as recall bias, which occurs when the ability to accurately and reliably report an event is dependent on the strength of the memory for the event as well as environmental factors. For instance, if a person's memory for a specific event is distorted in any way, it will be more difficult to accurately report the event when asked to on a survey.

It is particularly important to assess the phenomenon of recall bias for health symptoms because the more a physician knows about a patient's symptoms, the more information they have when developing effective treatment plans. Furthermore, controlling for recall bias is especially important in the assessment of symptoms experienced by individuals with ME/CFS. Currently, ME/CFS is a poorly understood illness without any universally recognized biological markers. Therefore, self-report measures are heavily relied on for making diagnostic decisions. This reliance on patient report data allows for biases that may ultimately decrease the reliability and validity of diagnostic assessments. Furthermore, many measures used to assess health symptoms have varying reporting periods (recall timeframes) and although research has found that recall bias may increase with longer reporting periods, few studies have been conducted in this area (Broderick et al., 2008). It may be especially important to understand how varying reporting periods are more or less susceptible to recall bias when assessing symptoms of ME/CFS. Many researchers are hopeful that an improvement in the methods to diagnose the illness will lead to a more homogenous illness group which in turn would make it easier for scientists to find clear biological markers of the disease (Jason et al., 2010).

Recall Bias: A General Overview

Specific biases that can occur in retrospective data include either forgetting that an event ever occurred and thereby not reporting it on a survey (omission), or misremembering an event as having occurred more recently in time than it actually did (telescoping) (Sudman & Bradburn, 1973). Whether these biases occur depends on a variety of factors that have to do with the characteristics of the event, the way in which an individual perceives the event, and the context in which an event is remembered (Stull, Leidy, Parasuraman, & Chassany, 2009). Characteristics related to the event itself might include how recent in time the event occurred in relation to the evaluation or assessment (recency) as well as the complexity for the event in question. Furthermore, the personal significance of an event (saliency) (Gendreau, Hufford, & Stone, 2003), as well as the respondent's mood at the time an event is remembered (Stull, Leidy, Parasuraman & Chassany, 2009), may also impact an individual's memory for it. The degree to which certain factors will affect a person's memory for past events, as well as the circumstances in which these factors are most influential, are important for researchers to understand when designing and evaluating research studies.

Timeframe and Recall Bias

Much of the current research on cognitive theory and recall bias has been influenced by the early work of Ebbinghaus in 1885 and his influential "forgetting curve." Ebbinghaus famously demonstrated that the rate at which individuals forget information is more pronounced immediately after an event has occurred,

and that the degree of forgetting tends to plateau as more time goes by (Ebbinghaus, 1913). This early work by Ebbinghaus has influenced more recent studies looking at how the recency phenomenon influences our ability to accurately remember events. Studies have shown that the longer the recall period, the less accurate a person's memory is for distinctive and autobiographical events (Bradburn et al., 1987; Skowronski, Betz, Thompson, & Shannon, 1991). Furthermore, longer recall periods reduce response accuracy for hospitalizations, health events, symptom onset, and overall morbidity (Amjadi-Begvand et al., 2004; Celebrezze & Terry, 1965; Dedominicis & Grechi, 1965; Feikin et al., 2010). Other studies, however, have found that longer recall periods do not necessarily result in decreased accuracy or reliability. For instance, it has been found that people can consistently report pain severity (Brauer, Thomsen, Loft, & Mikkelson, 2003) as well as missed workdays (Rivicki, Irwin, Reblando, & Simon, 1994) at one month and three month reporting periods. Additionally, it has also been found that the Positive Affect-Negative Affect Scale (PANAS), has excellent test-retest reliability across multiple timeframes (e.g. current, today, past few days, past week, past few weeks, past year, general) (Watson, Clark, & Tellegen, 1988; Watson & Clark, 1994).

Overall, more support has been garnered for greater inaccuracies as recall period is increased, which might suggest that retrospective surveys should include very short reporting periods. On the other hand, shorter reporting periods are not without biases either. For instance, the phenomenon known as forward telescoping happens more often with shorter reporting periods such as asking

someone to recall information over the past week (Bradburn, 2000). Forward telescoping occurs when a person reports events that occurred prior to the reporting period designated on a survey item, and this in turn causes an overreporting of events (Bradburn, 2000; Sudman & Bradburn, 1973). Biases associated with longer reporting periods (e.g. the past six months; the past year) are more likely to involve the phenomenon of forgetting, which in turn causes omission of information and overall underreporting of events (Clarke, Fiebig, Gerdtham, 2008; Sudman & Bradburn, 1973).

Some studies have found that an underreporting of events is more common than overreporting. This phenomenon has been particularly recorded for reports of health care utilization (Evans & Crawford, 1999). It has been suggested that these over and underreporting biases may cancel each other out as long as the optimal reporting period is utilized on a survey. This optimal reporting period may be a timeframe that falls in between the extremely recent and very long (Sudman & Bradburn, 1973), but it is still unclear how one decides what reporting period is too long or too short. This issue may depend on other factors that have been shown to influence our memory for past events (e.g. complexity, saliency, and mood).

Recall Bias: Stability, Complexity, and Context

In addition to his research on the phenomenon of recency, Ebbinghaus also discovered that the rate of forgetting depends on the characteristics of the event or phenomena in question (Stull et al., 2009). For instance, events that are fairly stable over time or those that happen once in a lifetime are remembered

with greater ease than events that change and fluctuate over time (Stone & Shiffman, 2002). When choosing what timeframe to use in a survey, it is important to know as much about the phenomenon being measured as possible. For instance, a short recall period may enhance information accessibility for the specific time period, but if there is no stable pattern to the phenomena, a short recall period might not capture the true nature of the symptom's variability over time (Stone & Shiffman, 2002; Stull et al., 2009). Clarke et al. (2008) assert that there is a tradeoff between reporting accuracy and loss of information when deciding between a shorter or longer recall timeframe. Short timeframes may increase the accuracy of recall, but investigators risk losing valuable information about the true nature of the phenomena that would be better captured with a longer recall period. In other words, a short timeframe is not always compelling for certain phenomena and it is important to take variability and stability into account when thinking about an optimal timeframe (Clark et al. 2008).

In the case of health symptoms, Stone and colleagues (2002) assert that when someone reports about a highly variable symptom, they are making an overall assessment of their experience, but cannot indicate the variable nature of the symptom in such a short time period. However, when highly variable symptoms are reported over longer timeframes, an individual will attempt to summarize their experience, which can reduce reporting accuracy (Stone, Schwartz, Broderick, & Shiffman, 2005). The complexity of recalled information has been described by Converse and Presser (1986) and later by Reis and Gable (2000) as information accessibility (as cited in Stull et al., 2009). Certain

phenomena are accessed with greater ease than others. For instance the names of medications are less accurately recalled compared to a person's memory for hospitalization utilization (Evans & Crawford, 1999). Additionally, medical conditions are harder to recall if a survey includes scientific terms used by physicians rather than using more layman's terms for conditions (Madow, 1967); a finding that highlights the importance of thoughtful planning for constructing clear and culturally sensitive surveys.

Context can also have a profound effect on response bias. Context related to recalled information refers to the personal and social meaning the information has for the respondent, as well as how this information is perceived in the broader environment (Stull et al., 2009). One specific contextual factor that can affect recall accuracy is the saliency of the information being recalled. The saliency of a recalled event has to do with its significance or personal relevance to the respondent. In terms of recall bias, highly salient information is often recalled more accurately than less salient information. This finding has been reported for recall of major symptoms versus minor symptoms (Cannell, Marquis, & Laurent, 1977), for recall of pain intensity versus pain location (Dawson, Kanim, & Sra, 2002), and for recall of information when in in-patient hospitalization versus out-patient consultation (Stull et al., 2009).

Another factor affecting recall is the mood and health status of the respondent at the time of an assessment or evaluation (Broderick, Schwartz, Shiffman, Hufford, & Stone, 2003). Specifically, when people are in a negative mood, they are more likely to access and recall information that is also negatively

charged (Stull et al., 2009). This phenomenon of current mood bias is seen with recall of affect and attitude (Blaney, 1986) as well as in medical research. For instance, someone who is experiencing more pain during an assessment is more likely to recall past pain symptoms as more severe than they had initially reported at baseline (Eich et al., 1985).

Recall for Health Symptoms

Retrospective survey data is used in many different areas of study, but in order to assess the accuracy and reliability of this type of data in the context of an illness such as ME/CFS, it is important to understand recall bias that is associated with specific health symptoms. Two symptoms that are commonly cited in the recall literature and that are frequently experienced by individuals with ME/CFS are pain and fatigue.

Pain has been studied under a variety of different contexts, including rheumatoid arthritis (Stone, Broderick, Kaell, DelesPaul, & Porter 2000; Broderick et al., 2008), chronic pain (Stone, Broderick, Shiffman, Litcher-Kelly, & Calvanese, 2003; Stone et al., 2005), and the illness known as fibromyalgia (Williams et al., 2004). In order to compare recalled pain with averaged momentary pain assessments, researchers have used paper daily diaries, electronic diaries, ecological momentary assessment (EMA) techniques, and a combination of prospective and retrospective survey methods (McColl, 2004; Stull et al., 2009). Within this literature, researchers have consistently found that there is a tendency to recall higher pain levels than were previously reported at baseline. Stone, Broderick, Shiffman, and Schwartz (2004) suggest that the discrepancy

between recalled pain and momentary pain may be due to the tendency to only recall salient pain events and to ignore periods where pain is not experienced as intensely. Specifically, it has been shown that when attempting to make an overall pain assessment, patients are not merely averaging their pain; rather, they are using cognitive heuristics in which they rely on aspects such as the variability of the pain experience (Redelmeier, Katz, & Kahneman, 2003), peak periods of pain over time (Stone et al., 2000), the most recent pain experiences (recency) (Redelmeier et al, 2003; Stone et al., 2000), and/or a combination of peak and recency known as "peak-end" (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Redelmeier, & Kahneman, 1996). Despite the discrepancy between recalled pain and momentary pain, there is still a moderate correlation between the two reporting periods and this moderate correlation might suggest to researchers that either method is sufficient for assessing pain. However, when changes in recalled pain and momentary pain are compared in a "within subjects" design, the correlation is significantly lower (Stone et al., 2004). Furthermore, it has been found that people with chronic pain who perceived their pain as worsening over time, did not show a significant change in multiple reports of momentary pain assessments (Stone et al., 2004).

Research on pain across different reporting periods suggests that when patients are asked to recall pain, their recall accuracy weakens over the course of seven days. However it was also found that correlations between recalled pain ratings and momentary pain ratings were higher for a 28 day recall timeframe compared to a seven day recall timeframe (Broderick et al., 2008). Broderick and

colleagues (2008) theorize that individuals with chronic illnesses may have a good idea of their typical symptom pattern overtime, thus allowing them to make an overall assessment of the last 28 days based on their symptom beliefs. This explanation seems best to describe chronic symptoms that are fairly stable overtime, but it is still unclear if pain that is fluctuating over a long period of time would have a similar effect on recall. Beyond issues related to accuracy and reporting length, Williams, Davies, and Chadury (2000) found that chronic pain patients may not always be recalling pain severity, but may instead report on the perceived impact that the pain has on functioning. Additionally, they found that the way one respondent interprets a pain rating scale can be very different from the way another person will (Williams et al., 2000). These results underscore the subjective nature of pain as well as its multidimensionality.

A handful of studies have investigated issues of recall for chronic fatigue and fatigue related to the illness ME/CFS (Broderick et al., 2008; Friedberg & Sohl, 2008; Sohl & Friedberg, 2008). Similar to findings within the pain literature, participants report experiencing higher levels of fatigue when it is recalled retrospectively compared to multiple momentary fatigue assessments. Despite this discrepancy, there is still a moderate to high correlation between a person's retrospective rating of fatigue and the average of their momentary ratings of fatigue (Broderick et al., 2008; Friedberg & Sohl, 2008). It has also been found that the more variable the fatigue, the higher the discrepancy between recalled fatigue and averaged momentary fatigue (Sohl & Friedberg, 2008); further validating the proposed impact of symptom stability on recall accuracy.

Much of the literature on recall bias for health related symptoms focuses on general pain and fatigue. Only a couple of studies have assessed recall accuracy for these symptoms in the context of the illness ME/CFS (Friedberg & Sohl, 2008; Sohl & Friedberg, 2008). Furthermore, the authors of these studies limited their investigation to the extent that retrospective recall correlates with momentary recall (i.e. recall accuracy). The diagnostic measures used in research can only be useful if they are shown to be both accurate and reliable (Spitzer, Endicott, & Robins 1978); therefore, it is equally important to evaluate the effects of varying timeframes on the reliability of recall for health symptoms.

Reliability of ME/CFS Symptom Ratings

A few studies have evaluated the test-retest reliability of specific diagnostic instruments used in ME/CFS research (Hawk et al., 2007; Jason et al.,1997). For instance, Hawk et al. (2007) investigated the issue of diagnostic reliability and test-retest reliability of a CFS Questionnaire. The CFS Questionnaire is a revised version of the CFS Screening Questionnaire developed by Jason et al. (1997) and can be used as a diagnostic instrument based on the Fukuda et al. (1994) criteria. In order to determine the sensitivity, specificity, and reliability of the measure, Hawk and colleagues (2007) administered the measure at two time points and to individuals with ME/CFS, major depressive disorder (MDD), and to healthy controls. Items from the CFS Questionnaire assess aspects of functioning and symptom experience and incorporate varying timeframes that range from ‘the past day’ to the ‘past six months.’ Participants completed the CFS

Questionnaire twice with two weeks in between each assessment and independent raters evaluated each assessment to determine a diagnosis.

Based on evaluations from the independent raters, the researchers found that the average test-retest reliability of the measure was very good, with most intraclass correlation coefficients at .70 or higher; however, different items were found to have better reliability scores than others. For instance, some items on the CFS questionnaire were developed to assess the reported frequency and severity of the eight ME/CFS case defining symptoms (fatigue/sickness following mental or physical exertion, unrefreshing sleep, problems remembering or concentrating, muscle aches and pains, joint pain, sore throat, tender lymph nodes/swollen glands, and headaches) recalled over the past six months. Overall, the average intraclass correlation scores for these items were very good (.77); however two symptoms (tender/sore lymph nodes and pain in multiple joints) had somewhat lower reliability scores (.58 and .49 respectively) (Hawk et al., 2007). Some additional items also had lower test-retest reliability scores. For instance, items asking participants to rate (on a scale of 0 to 100) their perceived energy, amount of expended energy, and amount of fatigue experienced over the past 24 hours, had lower reliability scores (.59, .40, and .22 respectively). Interestingly, these same items were also recalled at a slightly longer timeframe (over the past week) and resulted in better reliability scores (.77, .59, and .81 respectively). The authors have suggested that these symptoms likely fluctuate often and can be more consistently recalled over a longer timeframe. Furthermore, these results suggest

that it is important to take into account the type of question being asked and the timeframe in which it is asked when developing reliable diagnostic instruments.

The most widely used case definition for ME/CFS (Fukuda et al., 1994) requires a person to have at least six months of disabling fatigue and also four out of eight core symptoms (impaired memory or concentration, headaches, sore throat, lymph node pain, muscle pain, joint pain, unrefreshing sleep, and post-exertional malaise) also lasting six months or longer (Fukuda et al., 1994). This definition has been criticized for being too vague and lacking objective criteria to reliably classify individuals with ME/CFS (Jason et al., 2010). In order to improve the objectivity and reliability of the diagnostic criteria, researchers have recommended the use of standardized measures for charting and assessing symptoms of ME/CFS (King & Jason, 2005).

The Centers for Disease Control and Prevention (CDC) has developed an empirical case definition that assesses fatigue, the eight accompanying symptoms of ME/CFS, and disability, using validated and standardized measures (Reeves et al., 2005). The authors of the empirical case definition use the Symptom Inventory (Wagner et al., 2005) to assess the occurrence, frequency, and severity of the eight accompanying symptoms of ME/CFS recalled over the past month. The Medical Outcomes Survey Short-Form-36 (SF-36) is used to assess disability, and utilizes either a four-week recall timeframe or a one-week recall timeframe (Keller et al., 1997). Lastly, the Multidimensional Fatigue Inventory (MFI) is used to assess fatigue and requires participants to rate symptoms over the previous days (Smets, Garssen, Bonke, & Haes, 1995).

Each standardized measure included in the empirical case definition employs different timeframes, and this is not uncommon in many research studies. Some instruments used in research of health symptoms use multiple recall timeframes on a single questionnaire and others do not even specify a timeframe (Broderick et al., 2008). Furthermore, there is rarely any justification given for why a particular timeframe is used (Broderick et al., 2008). Given, the strong need for objective and reproducible ME/CFS criteria, it would be beneficial to determine the degree to which varying timeframes impact recall for specific ME/CFS symptoms.

The empirical case definition has received considerable controversy, as some have found that the definition may erroneously include people with primary psychiatric conditions and may lack the appropriate sensitivity for selecting individuals with the illness (Jason, Najar, Porter, & Reh, 2009; Jason et al., 2010). Although the empirical case definition may not become the staple of diagnosis and assessment in ME/CFS research, many future efforts will likely be made to establish an empirically derived case definition to improve the accuracy and reliability of diagnoses; especially in the absence of unequivocal biological markers.

Overall, researchers in the field have recommended that measures used to chart and assess symptoms of ME/CFS be both comprehensive and sensitive to the variability of symptom experience across individuals with this illness (Jason et al., 1999). Also, it has been noted that differences in the criteria used to classify individuals with ME/CFS accounts for the largest proportion of diagnostic

unreliability (Jason, Helgerson, Torres-Harding, Carrico, & Taylor, 2003). In order to decrease criterion variance and enhance diagnostic reliability, researchers have suggested that ME/CFS criteria incorporate specific standardized instruments to use as well as explicit guidelines regarding the number, frequency, and severity of symptoms required for a diagnosis. In light of these issues as well as the extensive reliance on self-report measures in ME/CFS research, the impact of different recall timeframes should also be investigated when developing and evaluating the diagnostic criteria for ME/CFS. Furthermore, it has been argued that the validity and reliability of symptom recall is important for developing appropriate treatments (Fienberg, Loftus, & Tanur, 1985). More research needs to be done to determine the optimal recall length for assessing specific symptoms of ME/CFS.

In sum, retrospective self-report measures are often used in research to assess symptoms that are commonly reported by individuals with ME/CFS. This reliance on patient report data allows for biases that may negatively impact the reliability and validity of diagnostic and treatment decisions. Furthermore, many measures used to assess ME/CFS symptoms have varying recall timeframes. It is unclear what the optimal reporting period is for tracking health symptoms, especially for a complex chronic illness such as ME/CFS where certain symptoms may fluctuate overtime. Only a small number of studies have investigated the issue of recall bias for symptoms of ME/CFS, and these studies limited their investigation to the extent of agreement between patients' reports of momentary fatigue versus fatigue that was recalled over a week-long timeframe (Friedberg &

Sohl, 2008; Sohl & Friedberg, 2008). There are a few documented studies that have assessed the test-retest reliability of an ME/CFS diagnostic instrument that includes multiple timeframes for different items; however, to this authors' knowledge there are no reported studies in the ME/CFS literature that have assessed the test-retest reliability of each case-defining symptom on a standardized instrument across varying timeframes. For instance, Hawk, et al. (2007) assessed test-retest reliability for the eight case-defining symptoms of ME/CFS recalled over a six month timeframe; however, it is unclear whether test-rest reliability would be stronger for these symptoms at shorter timeframes.

The ME/CFS literature is lacking information on the potential impact of timeframe on symptom recall in individuals with ME/CFS. In response to this lack of important psychometric information, this study served as an evaluation of the test-retest reliability of a revised Symptom Inventory that includes four different timeframes (right now, past week, past month, and past six months). Research has shown that both short (e.g. past week) and long recall periods (e.g. past six months) can negatively impact recall accuracy for health symptoms in different ways. Furthermore, contextual factors such as symptom stability and momentary symptom severity have also been found to impact reporting accuracy. Very little research has been done on the ways varying timeframes and contextual factors can influence the test-retest reliability of health symptom reports; particularly in the context of ME/CFS.

Statement of Hypotheses and Research Questions

Hypothesis I. ME/CFS symptoms will be recalled with greater consistency (yield stronger reliability coefficients) when symptoms are perceived as stable over time rather than variable.

Hypothesis II. An increase in momentary (right now) symptom severity ratings from baseline to assessment two (occurring one week later) will significantly predict an increase in past week, past month, and past six month symptom ratings from baseline to assessment two. A decrease in momentary (right now) symptom severity ratings from baseline to assessment two (occurring one week later) will significantly predict a decrease in past week, past month, and past six month symptom scores from baseline to assessment two.

Research Question Ia. Is there an optimal recall timeframe in terms of test-retest reliability, for ME/CFS symptoms that are perceived as variable over time?

Supplemental Research Question 1b. Is there an optimal recall timeframe in terms of test-retest reliability, for ME/CFS symptoms that are perceived as stable over time?

Research Question II. Does the optimal recall timeframe in terms of test-retest reliability, differ by the ME/CFS symptom being measured?

Research Question III. What is the optimal recall timeframe in terms of test-retest reliability, for ME/CFS symptoms, in the absence of contextual factors (e.g. symptom stability and momentary symptom severity)?

CHAPTER II

METHOD

This section presents information on participant recruitment, study procedures, and measurement tools. Data were collected over the phone at two time points, from individuals with a current diagnosis of ME/CFS.

Participants

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 ME/CFS. The majority of participants identified as White (94%), one participant identified as Asian/Pacific Islander, and two identified as "other." Two participants identified as Latin/Hispanic origin. Approximately half of all participants were married (N=27), 13 were never married, and 11 were divorced. The majority of participants received a standard college degree or higher (70.6 %) and all 51 participants reported at least a high school degree. Over half of the participants were on disability (58.8 %), with the large majority citing chronic fatigue syndrome 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 (78%) were confirmed with letters of documentation by independent physicians. All 51 participants met criteria for the Fukuda et al. (1994) case definition. Participants were identified through the use of an IRB approved research advertisement published in an ME/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 (Brown & Jason, 2007). Participants received a five dollar Amazon gift card upon completion of the study.

Procedure

Data collection occurred on two separate occasions 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, the interviews took place over the phone and were scheduled with one week in between the first and second interview, 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 a revised Symptom Inventory (SI-R: See Appendix A) which was altered by this author from the original Symptom Inventory developed by Wagner et al. (2005). Participants were also read a Symptom Stability Survey (See Appendix B), a short demographic survey, and a significant events questionnaire, all developed by this author and others at DePaul University. Phone interviewers repeated items for participants as necessary. During the second phone assessment, participants were read items from the SI-R, the Symptom Stability Survey, and the significant events questionnaire a second time. Following completion of the

second phone assessment, participants were debriefed on the purposes of the study.

Measures

All 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.

ME/CFS Symptom Assessment

The Symptom Inventory-Revised (SI-R) (see Appendix A) assesses the presence, frequency and severity of the case-defining symptoms of ME/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) according to Fukuda et al. (1994). The SI-R is a revision of an earlier Symptom Inventory that was developed by the Center for Disease Control and Prevention (CDC). The CDC's symptom inventory 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 0.88 for the total inventory score and 0.87 for the total score from a short-form version, including only six symptoms (fatigue after exertion, unrefreshing sleep, muscle aches, sleeping problems, problems with memory, and problems with concentration). The CDC Symptom Inventory has also been found to have excellent convergent validity with standardized measures of fatigue and functioning (Wagner et al., 2005). For the purposes of this study, revisions to the

Symptom Inventory included the addition of four timeframes: right now, past week, past month, past six months. Additionally, 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 momentary (right now) recall timeframe does not measure frequency; therefore, a composite score could not be created and instead, the right now timeframe served as a moderating variable in the analysis of hypothesis II. There is currently no information on the test-retest reliability of the CDC Symptom Inventory (Wagner et al., 2005) or on the unpublished revised version (SI-R) altered by this author for the purposes of the present study. The CFS Questionnaire, developed by Hawk et al. (2007) also measures the Fukuda et al. (1994) case-defining symptoms using a six month timeframe, and these items were found to have very good test-retest reliability, with an average intraclass correlation score of .77 across all eight case-defining symptoms (Hawke et al., 2007).

Symptom Stability

The Symptom Stability Survey (see Appendix B) was administered to participants at both phone assessments and is a measure of the perceived stability of each case defining symptom. 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 and there is currently

no information available on the internal consistency or test-retest reliability of this measure. Future analyses will be conducted to obtain data on the psychometric properties of the Symptom Stability Survey as well as the revised Symptom Inventory (SI-R).

Demographic Information

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

Significant Events

At the end of each phone interview, participants were administered a significant events questionnaire which asked questions related to the typicality of participant mental and physical health over the week of the interview and whether any recent significant events occurred that might have impacted mental and physical health at the time of the interview. Responses to the significant events questionnaire were not taken into account for the analyses presented in this paper, but will be evaluated in future analyses of the presented study sample.

CHAPTER III

RESULTS

The current study is an investigation of the effects of recall timeframe, symptom stability, and momentary symptom severity on the reliability of ME/CFS symptom reports. 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 symptom stability, changes in momentary (i.e. right now) symptom severity, and timeframe. Presented below are the re-stated hypotheses and research questions as well as subsequent MLM analyses and results.

For hypothesis I, it was expected that ME/CFS symptom composite scores would be recalled with greater consistency (yield stronger reliability coefficients) across interview assessments, when symptoms were perceived as stable over time rather than variable. A multilevel statistical model was used to test hypothesis I. Level 1 of the model tested the extent that interview one symptom composite scores predicted interview two symptom composite scores (see Table 1 for descriptive information of all nine symptom composite scores across two waves

and the three timeframes). 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 2 for descriptive information of the nine ME/CFS symptoms rated as variable; See Table 3 for descriptive information of the nine ME/CFS symptoms rated as stable). The symptom scores at the three recall timeframes were not analyzed separately in the analysis but were grouped to represent a single variable referred to as Interview One Scores Collapsed Across Timeframe. 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.

Level 1: $y_{ij} = b_0 + b_1$ Interview One Scores Collapsed Across Timeframe $_{ij} + r_{ij}$

Y= Symptom composite scores at all three timeframes, at interview two

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

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

Table 1

Means and Standard Deviations of Symptom Composites on the SI-R at Interviews 1 and 2, N=51

		Interview 1	Interview 2
Symptom	Timeframe	M (SD)	M (SD)
Sore Throat	Week	4.25 (5.03)	4.88 (6.14)
	Month	4.24 (4.94)	4.10 (5.05)
	Six Months	5.35 (6.07)	5.21 (6.06)
Lymph Nodes	Week	5.71 (6.04)	5.14 (5.92)
	Month	5.53 (6.18)	4.76 (5.18)
	Six Months	6.43 (6.86)	5.45 (5.75)
Post Exertional Malaise (PEM)	Week	16.82 (4.93)	16.86 (4.81)
	Month	17.00 (4.96)	16.53 (4.95)
	Six Months	17.90 (5.69)	17.14 (5.10)
Muscle Pain	Week	12.00 (6.26)	11.00 (5.87)
	Month	11.73 (6.89)	11.33 (6.07)
	Six Months	12.18 (6.64)	11.35 (5.99)
Joint Pain	Week	8.94 (6.78)	9.35 (6.87)
	Month	9.45 (7.16)	8.90 (6.26)
	Six Months	9.63 (7.59)	9.24 (6.66)
Unrefreshing Sleep	Week	16.92 (6.00)	17.41 (6.34)
	Month	16.25 (6.29)	15.82 (5.97)
	Six Months	16.20 (7.27)	15.94 (6.65)
Headaches	Week	7.39 (6.53)	7.25 (6.26)
	Month	7.37 (5.40)	6.76 (5.56)
	Six Months	8.41 (6.57)	7.41(5.37)
Memory Problems	Week	10.47 (6.52)	10.12 (7.14)
	Month	10.47 (6.44)	10.25 (6.66)
	Six Months	10.90 (6.72)	10.76 (6.89)
Difficulty Concentrating	Week	11.75 (5.88)	11.96 (6.75)
	Month	11.86 (6.20)	11.84 (6.25)
	Six Months	12.71 (6.30)	12.20 (6.39)

Table 2

Means and Standard Deviations of Symptom Composites on the SI-R at Interviews 1 and 2 for Symptoms Rated as Variable

		Interview 1	Interview 2
Variable Symptoms	Timeframe	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 3

Means and Standard Deviations of Symptom Composites on the SI-R at Interviews 1 and 2 for Symptoms Rated as Stable

		Interview 1	Interview 2
Stable Symptoms	Timeframe	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)

Hypothesis I was supported for three of the eight case-defining ME/CFS symptoms: PEM, headaches, and memory problems. For post-exertional malaise, 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 = 0.48, SE = 0.23, 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 percent 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 memory composite scores at interview one F(1, 92) = 0.01, 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 = 0.66, SE = 0.18, 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 percent 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 $F(1, 98) = 0.91$, $p = .343$, 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, $b = 0.53$, $SE = 0.17$, $t(98) = 3.07$, $p = .003$. 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 percent of the variance in predicting headache scores at interview two is explained by the nesting of individual factors and symptom stability.

For Hypothesis II, it was expected that an increase in momentary (right now) symptom severity ratings from interview one to interview two would significantly predict an increase in past week, past month, and past six month symptom scores from interview one to interview two. A decrease in momentary (right now) symptom severity ratings from interview one to interview two will significantly predict a decrease in past week, past month, and past six month symptom scores from interview one to interview two (for a display of the means and standard deviations of the symptom severity scores at all four timeframes and across interviews one and two, please refer to Table 7 under Appendix C).

A multilevel statistical model was used to test Hypothesis II. Level 1 of the model tested the extent that past week, past month, and past six month symptom composite scores (variable notation in the model is Timeframe) predicted symptom composite scores at interview two. Level 2 of the model tested whether the change (increase or decrease) in Right Now severity ratings over a one week interval predicted symptom composite scores (i.e. collapsed across timeframe) at interview two. In order to determine the change in momentary severity at each symptom from interview one to interview two, a Right now Difference Score (Right Now Diff.) was calculated (Right Now severity score at interview two minus Right Now severity score at Interview one). The model presented below was re-estimated changing the reference group for timeframe, in order to test the significance of the slope for each timeframe. Timeframe is a categorical variable that is dummy coded in the estimation of the model. Two dummy vectors are used for each time that Timeframe appears in the model.

Level 1: $y_{ij} = b_0 + b_1$ Timeframe $_{ij} + r_{ij}$

y= Symptom composite scores at interview two

Level 2: $b_{0i} = \gamma_{00} + \gamma_{01}$ Right Now Diff. $_i$

$b_{1i} = \gamma_{10} + \gamma_{11}$ Right now Diff. $_i$

Hypothesis II was supported for two of the case-defining ME/CFS symptoms, sore throats and difficulty concentrating, but only when the reference group was six months.

Sore Throat

The right now difference score significantly predicted sore throat scores at interview two, when the reference group was the six month timeframe, *b* = 4.21, *SE* = 1.86, *t*(22.09) = 2.26, *p* =.03, such that, an increase in the right now severity score from interview one to interview two, significantly predicted an increase in the sore throat composite score for interview two. The right now difference score does not significantly predict sore throat composite scores at interview two, when the reference group is the past month timeframe, the *b* = 1.38, *SE* = 1.86, *t*(22.09) = 0.74, *p* = .466. The right now difference score did not significantly predict sore throat composite scores at interview two, when the reference group is the past week timeframe, the *b* = 2.15, *SE* = 1.86, *t*(22.09) = 1.15, *p* = .262. There was no difference in the relationship between the right now difference score and the outcome based on the timeframes; the relationship between past week did not differ from past six months (*b* = -2.07, *SE* = 1.63, *t*(26) = -1.27, *p* = .22) and the relationship between past month did not differ from past six months (*b* = -2.83, *SE* = 1.63, *t*(26) = -1.74, p = .09). The within variance of

the distribution residuals was 14.50 and the between variance of distribution residuals was 23.52. The ICC score was calculated as .58, suggesting that 58 percent of the variance in predicting sore throat scores at interview two was explained by the nesting of individual factors and the right now difference score.

Concentration

The right now difference score significantly predicted concentration composite scores at interview two, when the reference group was the six month timeframe, $b = 2.11$, $SE = 0.93$, $t(56.40) = 2.27$, $p = .03$. The right now difference score only marginally predicted concentration composite scores at time two, when the reference group was the past month, $b = 1.68$, $SE = 0.93$, $t(56.40) = 1.80$, $p = .078$. The right now difference score did not significantly predict concentration scores at time two, when the reference group was the past month timeframe, $b = 1.35$, $SE = 0.93$, $t(56.40) = 1.45$, $p = .153$. There was no difference in the relationship between the concentration difference score and the outcome scores based on the timeframes; past week did not differ from past six months ($b = 1.68$, $SE = 0.59$, $t(86) = -0.746$, $p = .46$) and past month does not differ from past six months, $b = 1.35$, $SE = 0.59$, $t(86) = -1.30$, $p = .196$. The within variance of the distribution residuals was 7.49 and the between variance of distribution residuals was 30.40. The ICC score was calculated as .80, suggesting that 80 percent of the variance in predicting sore throat scores at interview two is explained by the nesting of individual factors and the right now difference score.

For research question Ia, it was asked, what is the optimal recall timeframe in terms of test-retest reliability, for ME/CFS symptoms that are perceived as variable/unstable over time? Research question Ib was supplementary to Ia and reads, what is the optimal recall timeframe in terms of test-retest reliability, for ME/CFS symptoms that are perceived as stable over time?

A multilevel model was used in the analysis of research question Ia and Ib. 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, and (2) how timeframe moderated the way symptom composites at interview one predicted scores at interview two. Level 1 of the model tested the main effect of timeframe. Analyses for research question Ia and Ib were both 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.

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

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

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

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

Variable Sore Throats

When sore throats were rated as variable, sore throat composite scores at interview one significantly predicted sore throat composite scores at interview two, for the six month reference, $b = 0.80$, $SE = 0.17$, $t(62.39) = 4.81$, $p < .001$. The relationship between interview one and interview two is significantly stronger for the six month timeframe compared to the past week ($b = 0.20$, $SE = 0.28$, $t(65.03) = -2.1$, $p = .04$) and the past month ($b = -0.04$, $SE = -0.84$, $t(60.62) = -2.28$, $p=.026$). The slope coefficient for the six month reference was closest to 1.0 at .80, suggesting that six months is the optimal timeframe for variable sore throats. The within variance of the distribution residuals was 3.83 and the between variance of distribution residuals was 21.26. The ICC score was calculated as .85, suggesting that 85 percent of the variance in predicting variable sore throat scores at interview two is explained by the nesting of both individual factors and the sore throat scores at interview one.

Stable Sore Throats

Sore throat scores at interview one did not significantly predict sore throat scores at interview two, when the reference was the past six months, $b = 0.18$, SE = 0.76, $t(16.67) = 0.23$, $p = .819$, the past month, $b = 1.84$, $SE = 1.34$, $t(15.13) = 1.38$, $p = .188$, or the past week, $b = -0.97$, $SE = 0.85$, $t(16.25) = -1.15$, $p = .268$. There is no difference in the relationship between the sore throat scores at interview one and interview two based on the timeframes; the relationship between past six months does not differ from the past month ($b = 1.84$, $SE = 1.63$, $t(16.73) = 1.02$, $p = .322$) or from the past week ($b = -0.97$, $SE = 1.29$, $t(17.85) = -$

0.90, p = .380). The slope coefficient for the six month reference was closest to 1.0 at 0.18, suggesting that six months is the optimal timeframe for variable sore throats. Although six months was determined as optimal, all three timeframes produced slope coefficients with poor predictive validity relative to the other symptoms, suggesting that within the study population, there is poor reliability in reporting sore throats that are experienced as stable. The within variance of the distribution residuals was 19.76 and the between variance of distribution residuals was 17.01. The ICC score was calculated as .46, suggesting that 46 percent of the variance in predicting stable sore throat scores at interview two is explained by the nesting of individual factors and sore throat scores at interview one.

Variable Lymph Node Pain.

When lymph node pain was rated as variable over time, lymph node scores at interview one significantly predicted lymph node scores at interview two for the six month reference, b = 0.77, SE = 0.28, t(47.43) = 2.71, p = .009. Interview one did not significantly predict lymph node scores at interview two for the past month reference, b = 0.07, SE = 0.33, t(46.08) = 0.22, p = .824, or the past week interval, b = 0.03, SE = 0.29, t(47.25) = 0.12, p = .91. There is no difference in the relationship between the lymph node scores at interview one and interview two based on the timeframes; the relationship between past six months does not differ from the past month (b = 0.07, SE = 0.47, t(47.90) = -1.48, p = .145) or from the past week (b = 0.03, SE = 0.45, t(49.67) = -1.619, p = .112). The six month slope coefficient was closest to 1.0 at .77, suggesting that six months is the optimal timeframe for lymph node pain experienced as variable. The within

variance of the distribution residuals was 3.44 and the between variance of distribution residuals was 11.52. The ICC score was calculated as .77, suggesting that 77 percent of the variance in predicting variable sore lymph node scores at interview two is explained by the nesting of individual factors and sore throat scores at interview one.

Stable Lymph Node Pain

When lymph node pain was rated as stable over time, lymph node scores at interview one significantly predicted lymph node scores at interview two, for the six month reference, $b = 1.30$, $SE = 0.49$, $t(28.53) = 2.63$, $p = .014$. Lymph node scores at interview one did not significantly predict scores at interview two for the past month, $b = 0.30$, $SE = 0.86$, $t(26.62) = 0.35$, $p = .731$, or the past week, $b = 0.12$, $SE = 0.60$, $t(29.83) = 1.29$, $p = .847$. There is no difference in the relationship between the lymph node scores at interview one and interview two based on the timeframes; the relationship between past six months does not differ from the past month ($b = 0.30$, $SE = 1.12$, $t(28.70) = -0.89$, $p = .380$) or from the past week, $b = 0.12$, $SE = 0.92$, $t(29.83) = -1.29$, $p = .207$. The six month slope coefficient was closest to 1.0 at 1.30, suggesting that six months is the optimal timeframe for lymph node pain experienced as stable. The within variance of the distribution residuals was 7.86 and the between variance of distribution residuals was 31.44. The ICC score was calculated as .80, suggesting that 80 percent of the variance in predicting stable lymph node scores at interview two is explained the nesting of individual factors and stable lymph node scores at interview one.

Variable Post-Exertional Malaise

When PEM was rated as variable, PEM scores at interview one did not significantly predict PEM scores at interview two, for the past six month reference, $b = 0.31$, $SE = 0.45$, $t(6.93) = 0.68$, $p = .518$, the past month, $b = 0.44$, $SE = 0.63$, $t(6.35) = 0.71$, $p = .505$, or the past week, $b = -0.29$, $SE = 0.47$, $t(6.84) = -.62$, $p = .557$. There was no difference in the relationship between PEM scores at interview one and at interview two based on the timeframes; the relationship between six months did not differ from past month ($b = 0.44$, $SE = 0.82$, $t(6.94) = 0.16$, $p = .875$) or from the past week ($b = -0.29$, $SE = 0.73$, $t(7.72) = -0.82$ $p = .438$). The slope coefficient for the past month reference is closest to 1.0 at .44, suggesting that the past month timeframe is optimal for PEM that is experienced as variable. All three timeframes produced slope coefficients with poor predictive validity relative to the other variable symptoms, suggesting that within this study population, there is poor reliability in reporting PEM that is experienced as variable. The within variance of the distribution residuals was 7.23 and the between variance of distribution residuals was 9.66. The ICC score was calculated as .57, suggesting that 57 percent of the variance in predicting variable PEM scores at interview two is explained by the nesting of individual factors and variable PEM scores at interview one.

Stable Post-Exertional Malaise

When PEM scores were rated as stable over time, PEM scores at interview one significantly predicted PEM scores at interview two for the six month reference, $b = 0.79$, $SE = 0.19$, $t(106.85) = 4.09$, $p < .001$ and at the past week reference, $b = 0.44$, $SE = 0.21$, $t(104.93) = 2.13$, $p = .04$. There was no difference in the relationship between PEM scores at interview one and at interview two based on the timeframes; the relationship between six months does not differ from past month, $b = 0.30$, $SE = 0.33$, $t(109.37) = -1.5$, $p = .142$, or from the past week, $b = 0.44$, $SE = 0.32$, $t(112.91) = -1.09$, $p = .275$. The slope coefficient for the six month reference is closest to 1.0 at .79, suggesting that the six month timeframe is optimal for reporting PEM experienced as stable. The within variance of the distribution residuals was 7.01 and the between variance of distribution residuals was 14.82. The ICC score was calculated as .68, suggesting that 68 percent of the variance in predicting stable PEM scores at interview two is explained the nesting of individual factors and stable PEM scores at interview one.

Variable Muscle Pain

When muscle pain was rated as variable over time, muscle pain scores at interview one significantly predicted muscle pain scores at interview two for the six month reference, $b = 0.56$, $SE = 0.25$, $t(40.21) = 2.25$, $p = .03$. Interview one did not significantly predict muscle pain scores at interview two for the past month reference, $b = -0.22$, $SE = 0.36$, $t(38.12) = -0.60$, $p = .553$, or the past week $b = -0.23$, $SE = 0.24$, $t(40.54) = -0.98$, $p = .335$. The relationship between

interview one and interview two for the six month timeframe is marginally stronger compared to the past week, b = -0.23, SE = 0.40, t(42.70) = -1.99, p = .054, and the past month b = -0.22, SE = .45, t(38.79) = -1.73, p = .09. The slope coefficient at the past six month interval is closest to 1.0 at .56, suggesting that the past six month timeframe is optimal for reporting muscle pain experienced as variable. The within variance of the distribution residuals was 3.28 and the between variance of distribution residuals was 12.65. The ICC score was calculated as .79, suggesting that 79 percent of the variance in predicting variable muscle pain scores at interview two is explained the nesting of individual factors and variable muscle pain scores at interview one.

Stable Muscle Pain

When muscle pain scores were rated as stable over time, muscle pain scores at interview one marginally predicted scores at interview two for the six month reference, b = 0.82, SE = 0.44, t(72.40) = 1.86, p = .067. Interview one scores did not predict interview two scores for the past month, b = 0.75, SE = 0.55, t(69.15) = 1.35, p = .182, or the past week, b = -0.23, SE = 0.24, t(40.54) = -.98, p = .335. The relationship between interview one and interview two was significantly stronger at the six month reference compared to the past week (b = -0.56, SE = 0.66, t(79.89) = -2.08, p = .041). The relationship between interview one and interview two for the past six months was no different than the past month (b = 0.75, SE = 0.72, t(69.93) = -1.02, p = .919). The slope coefficient at the six month interval is closest to 1.0 at .82, suggesting that the past six month timeframe is optimal for reporting muscle pain experienced as stable. The within

variance of the distribution residuals was 13.75 and the between variance of distribution residuals was 16.70. The ICC score was calculated as .55, suggesting that 55 percent of the variance in predicting stable muscle pain scores at interview two is explained by the nesting of individual factors and stable muscle pain scores at interview one.

Variable Joint Pain

When joint pain was rated as variable, scores at interview one marginally predicted scores at interview two, for the past week reference, $b = 0.58$, $SE = 0.29$, $t(47.81) = 2.00$, $p = .051$. Interview one does not significantly predict joint pain scores at interview two, for the past month, $b = 0.15$, $SE = 0.33$, $t(47.35) = 0.45$, $p = .655$, or the past six months, $b = 0.22$, $SE = 0.21$, $t(49.54) = 1.04$, $p = .304$. There is no difference in the relationship between joint pain scores at interview one and at interview two based on the timeframes; the relationship between past week does not differ from past month ($b = 0.15$, $SE = 0.38$, $t(42.77) = 1.14$, $p = .260$) or from the past six months, ($b = 0.22$, $SE = 0.43$, $t(51.05) = 0.85$, $p=.399$). The slope coefficient for the past week is closest to 1.0 at .58, suggesting that the past week timeframe is optimal for reporting joint pain experienced as variable. The within variance of the distribution residuals was 3.29 and the between variance of distribution residuals was 19.35. The ICC score was calculated as .85, suggesting that 85 percent of the variance in predicting variable joint pain scores at interview two is explained by variable joint pain scores at interview one.

Stable Joint Pain

When joint pain was rated as stable over time, joint pain scores at interview one significantly predicted joint pain scores at interview two for the six month reference, $b = 0.95$, $SE = 0.35$, $t(46.75) = 2.72$, $p = .009$, and the past month reference, $b = 1.82$, $SE = .53$, $t(43.62) = 3.43$, $p=.001$. Interview one scores did not significantly predict interview two scores for the past week, $b = -0.64$, $SE = 0.37$, $t(46.25) = -1.74$, $p = .089$. The relationship between joint pain scores at interview one and interview two was significantly different by timeframe; the relationship at the past six months was significantly stronger compared to the past week ($b = -0.64$, $SE = 0.58$, $t(49.57) = -2.72$, $p = .009$) but not significantly stronger than the past month ($b = 1.82$, $SE = 0.68$, $t(45.98) = 1.28$, $p = .206$). The slope coefficient for the past six month interval is closest to 1.0 at .95, suggesting that the past six month timeframe is optimal for reporting joint pain experienced as stable. The within variance of the distribution residuals was 8.48 and the between variance of distribution residuals was 25.99. The ICC score was calculated as .75, suggesting that 75 percent of the variance in predicting stable joint pain scores at interview two is explained by the nesting of individual factors and stable joint pain scores at interview one.

Variable Unrefreshing Sleep

When unrefreshing sleep was rated as variable, unrefreshing sleep scores at interview one did not significantly predict scores at interview two for the past six month timeframe, $b = 0.30$, $SE = 0.56$, $t(17.71) = 0.54$, $p = .598$, the past month, $b = 0.82$, $SE = 0.67$, $t(17.22) = 1.22$, $p = .238$, or the past week

timeframe, $b = 0.64$, $SE = 0.48$, $t(18.35) = 1.33$, $p = .199$. There was no difference in the relationship between unrefreshing sleep scores at interview one and at interview two based on the timeframes; the relationship between interview one and two at the past month timeframe does not differ from past six months ($b = 0.30$, $SE = 0.86$, $t(16.82) = 0.60$, $p = .557$) or from the past week ($b = 0.64$, $SE = 0.91$, $t(18.64) = -.20$ $p = .844$). The slope coefficient at the past month reference is closest to 1.0 at .82, suggesting that the past month timeframe is optimal for reporting unrefreshing sleep experienced as variable. The within variance of the distribution residuals was 6.90 and the between variance of distribution residuals was 20.81. The ICC score was calculated as .75, suggesting that 75 percent of the variance in predicting variable unrefreshing sleep scores at interview two is explained by the nesting of individual factors and unrefreshing sleep scores at interview one.

Stable Unrefreshing Sleep

When unrefreshing sleep was rated as stable, unrefreshing sleep scores at interview one significantly predicted unrefreshing sleep scores at interview two when the interval was six months, $b = 0.41$, $SE = 0.17$, $t(95.17) = 2.34$, $p = .021$. Unrefreshing sleep at interview one did not significantly predict unrefreshing sleep at interview two, for the past month reference, $b = 0.29$, $SE = 0.23$, $t(91.14) = 1.24$, $p = .218$ or the past week, $b = 0.18$, $SE = 0.26$, $t(89.80) = 0.67$, $p = .504$. There was no difference in the relationship between unrefreshing sleep scores at interview one and at interview two based on the timeframes; the relationship between interview one and two at the six month timeframe does not differ from

past month (*b* = 0.29, *SE* = 0.34, *t*(98.72) = -0.37, *p* = .713) or from the past week (*b* = 0.18, *SE* = 0.36, *t*(97.29) = -0.64, *p* = .525). The slope coefficient at the six month reference was closest to 1.0 at .41, suggesting that the six month timeframe is optimal for reporting unrefreshing sleep experienced as stable. The within variance of the distribution residuals was 5.90 and the between variance of distribution residuals was 22.79. The ICC score was calculated as .79, suggesting that 79 percent of the variance in predicting stable unrefreshing sleep scores at interview two is explained by the nesting of individual factors and stable unrefreshing sleep scores at interview one.

Variable Headaches

Headache scores at interview one did not significantly predict headache scores at interview two, for the six month reference, *b* = 0.32, *SE* = 0.24, *t*(74.54) = 1.30, *p* = .197, the past month, *b* = -0.51, *SE* = 0.36, *t*(69,80) = -1.42, p = .159, or the past week reference, *b* = -0.08, *SE* = 0.22, *t*(76.77) = -.37, *p* = .711. The relationship between headaches at interview one and two at the past six months was marginally different from the past month (*b* = -0.51, *SE* = 0.43, *t*(68.40) = -1.93, *p* = .058) but was not significantly different from the past week (*b* = -0.08, *SE* = 0.37, *t*(82.31) = -1.07, *p* = .289). The slope coefficient at the past six month reference is closest to 1.0 at .32, suggesting that the six month timeframe is optimal for reporting headaches experienced as variable. The within variance of the distribution residuals was 8.21 and the between variance of distribution residuals was 12.50. The ICC score was calculated as .60, suggesting that 60 percent of the variance in predicting variable headache scores at interview two is

explained by the nesting of individual factors with variable headache scores at interview one.

Stable Headaches

When headaches were rated as stable over time, headache scores at interview one significantly predicted headache scores at interview two, when the reference was six months, $b = 1.40$, $SE = 0.23$, $t(30.36) = 6.04$, $p < .001$. Headache scores at interview one did not significantly predict headache scores at interview two at the past month, $b = 0.30$, $SE = 0.48$, $t(27.75) = 0.62$, $p = .543$, or the past week, $b = 0.00$, $SE = 0.24$, $t(30.17) = 0.01$, $p = .991$. The relationship between Headache scores at interview one and headache scores at interview two was significantly stronger at the past six month reference compared to the past week ($b = 0.00$, $SE = 0.39$, $t(32.27) = -3.55$, $p = .001$) and marginally stronger than the past month ($b = 0.30$, $SE = 0.56$, $t(28.85) = -1.98$, $p = 0.058$). The slope coefficient at the past six month reference is closest to 1.0 at 1.40, suggesting that the six month timeframe is optimal for reporting headaches that are stable. The within variance of the distribution residuals was 8.63 and the between variance of distribution residuals was 36.75. The ICC score was calculated as .81, suggesting that 81 percent of the variance in predicting variable headache scores at interview two is explained by the nesting of individual factors and headache scores at interview one.

Variable Memory Problems

Memory scores at interview one did not significantly predict Memory scores at interview two for the six month reference, $b = -0.14$, SE = 0.29, $t(39.71)$

= -0.50, p = .618, the past month, b = -0.07, SE = 0.45, t(37.23) = -0.15, p = .881, or the past week interval, b = -0.14, SE = 0.34, t(38.47) = -.41, p = .688. There was no difference in the relationship between memory scores at interview one and at interview two based on the timeframes; the past six months was not significantly different from the past month, b = -0.07, SE = 0.59, t(40.01) = 0.13, p = .898 or from the past week, b = -0.14, SE = 0.52, t(41.43) = 0.01, p = .991. The slope coefficient for the past month reference is closest to 1.0 at -.07, suggesting that the six month timeframe is optimal for reporting variable memory problems. Although the past month reference was determined as optimal, all three timeframes produced slope coefficients with poor predictive validity, suggesting that within this study population, there is poor reliability in reporting memory problems that are experienced as variable. The within variance of the distribution residuals was 7.22 and the between variance of distribution residuals was 28.07. The ICC score was calculated as .80, suggesting that 80 percent of the variance in predicting variable memory scores at interview two is explained by the nesting of individual factors and memory scores at interview one.

Stable Memory Problems

When memory problems were rated as stable over time, memory scores at interview one significantly predicted memory scores at interview two, when the reference was six months, b = 1.03, SE = 0.19, t(60.72) = 5.32, p < .001. Memory scores at interview one did not significantly predict memory scores at interview two at the past month, b = 0.04, SE = 0.29, t(59.13) = .14, p = .893, or the past week, b = -0.17, SE = 0.21, t(60.32) = -8.26, p = .412. The relationship between

memory scores at interview one and memory scores at interview two was significantly stronger at the past six months compared to the past week (b = -0.17, SE = 0.33, t(61.88) = -3.61, p = 0.001) and the past month (b = 0.04, SE = 0.38, t(60.42) = -2.63, p = .011). The slope coefficient at the past six month interval is closest to 1.0 at 1.03, suggesting that the six month timeframe is optimal for reporting memory problems experienced as stable. The within variance of the distribution residuals was 3.62 and the between variance of distribution residuals was 41.44. The ICC score was calculated as .92, suggesting that 92 percent of the variance in predicting stable memory scores at interview two is explained by stable memory scores at interview one.

Variable Concentration Problems

When concentration scores were experienced as variable, concentration scores at interview one did not significantly predict concentration scores at interview two for the six month reference, b = 0.20, SE = 0.43, t(35.81) = 0.46, p = .647, the past month, b = -0.17, SE = 0.37, t(36.39) = -0.45, p = .654, or the past week reference, b = 0.04, SE = 0.27, t(39.80) = 0.27, p = .786. There was no difference in the relationship between concentration scores at interview one and at interview two based on the timeframes; the past six months was not significantly different from the past month (b = -0.17, SE = 0.51, t(29.98) = -0.73, p = .473) or the past week (b = 0.04, SE = 0.58, t(39.80) = -0.27, p = .786). The slope coefficient at the six month reference is closest to 1.0 at .20, suggesting that the six month timeframe is optimal for reporting variable concentration problems. Although the six month interval was determined as optimal, all three timeframes

produced slope coefficients with poor predictive validity, suggesting that within the current population, there is poor reliability in reporting concentration problems that are experienced as variable. The within variance of the distribution residuals was 8.00 and the between variance of distribution residuals was 15.87. The ICC score was calculated as .66, suggesting that 66 percent of the variance in predicting variable concentration scores at interview two is explained by the nesting of individual factors and concentration scores at interview one.

Stable Concentration Problems

When concentration problems were rated as stable over time, concentration scores at interview one significantly predicted concentration scores at interview two, when the reference was six months, $b = 0.47$, $SE = 0.23$, $t(74.15) = 2.07$, $p = .042$. Concentration scores at interview one did not significantly predict concentration scores at interview two for the past month reference, $b = 0.14$, $SE = 0.30$, $t(71.73) = 0.48$, $p = .634$, or the past week, $b = -0.43$, $SE = 0.31$, $t(71.52) = -1.41$, $p = .164$. The relationship between concentration scores at interview one and concentration scores at interview two was significantly stronger at the past six month interval compared to the past week ($b = -0.43$, $SE = 0.44$, $t(75.46) = -2.05$, $p = .044$) and the past month ($b = 0.14$, $SE = 0.43$, $t(75.68) = -0.76$, $p = .451$). The slope coefficient for the six month reference is closest to 1.0 at .47, suggesting that the six month timeframe is optimal for reporting stable concentration problems. The within variance of the distribution residuals was 19.76 and the between variance of distribution residuals was 17.00. The ICC score was calculated as .46, suggesting that 46 percent of the variance in

predicting stable concentration scores at interview two is explained by the nesting of individual factors and stable concentration scores at interview one (see Table 4 for slope coefficients of variable ME/CFS symptoms at all three timeframes; see Table 5 for slope coefficients of stable ME/CFS symptoms at all three timeframes).

Table 4

Slope Coefficients of Variable ME/CFS symptoms at Three Timeframes

Variable Symptoms	Timeframe	*b*	*SE*	*df*	*t*	*p*
Sore Throat	Week	0.20	0.17	61.99	1.19	.239
	Month	-0.04	0.31	58.68	-.14	.888
	Six Months*	0.80	0.17	62.39	4.81	<.001
Lymph Node Pain	Week	0.03	0.29	47.25	0.12	.91
	Month	0.07	0.33	46.08	0.22	.824
	Six Months*	0.77	0.28	47.43	2.71	.009
PEM	Week	-0.29	0.47	6.84	-0.62	.557
	Month*	0.44	0.63	6.35	0.706	.505
	Six Months	0.31	0.45	6.93	0.68	.518
Muscle Pain	Week	-0.23	0.24	40.54	-0.98	.335
	Month	-0.22	0.36	38.12	-0.60	.553
	Six Months*	0.56	0.25	40.21	2.25	.03
Joint Pain	Week*	0.58	0.29	47.81	2.00	.051
	Month	0.15	0.33	47.35	0.45	.655
	Six Months	0.22	0.21	49.54	1.04	.304
Unrefreshing Sleep	Week	0.64	0.48	18.35	1.33	.199
	Month*	0.82	0.67	17.22	1.22	.238
	Six Months	0.30	0.56	17.71	0.54	.598
Headaches	Week	-0.08	0.22	76.77	-0.37	.711
	Month	-0.51	0.36	69.80	-1.42	.159
	Six Months*	0.32	0.24	74.54	1.30	.197
Memory	Week	-0.14	0.34	38.47	-0.41	.688
	Month*	-0.07	0.45	37.23	-0.15	.881
	Six Months	-0.14	0.29	39.71	-0.50	.62
Concentration	Week	0.04	0.27	39.80	0.27	.786
	Month	-0.17	0.37	36.39	-0.45	.654
	Six Months*	0.20	0.43	35.81	0.46	.647

Note. The symbol * refers to the optimal timeframe (coefficients closest to 1.0)

Table 5

Slope Coefficients of Stable ME/CFS symptoms at Three Timeframes

Stable Symptoms	Timeframe	*b*	*SE*	*df*	*t*	*p*
Sore Throat	Week	-0.97	0.85	16.25	-1.15	.268
	Month	1.84	1.34	15.13	1.38	.188
	Six Months*	0.18	0.76	16.67	0.23	.819
Lymph Node Pain	Week	0.12	0.60	29.83	1.29	.847
	Month	0.30	0.86	26.62	.347	.731
	Six Months*	1.30	0.49	28.53	2.63	.014
PEM	Week	0.44	0.21	104.93	2.13	.04
	Month	0.30	0.23	102.45	1.31	.195
	Six Months*	0.79	0.19	106.85	4.09	<.001
Muscle Pain	Week	-0.56	0.40	74.44	-1.41	.164
	Month	0.75	0.55	69.15	1.35	.182
	Six Months*	0.82	0.44	72.40	1.86	.067
Joint Pain	Week	-0.64	0.37	46.25	-1.74	.089
	Month	1.82	0.53	43.62	3.43	.001
	Six Months*	0.95	0.35	46.75	2.72	.009
Unrefreshing Sleep	Week	0.18	0.26	89.80	0.67	.504
	Month	0.29	0.23	91.14	1.24	.218
	Six Months*	0.41	0.17	95.17	2.34	.021
Headaches	Week	0.00	0.24	30.17	0.012	.991
	Month	0.30	0.48	27.75	0.62	.543
	Six Months*	1.40	0.23	30.36	6.04	<.001
Memory	Week	-0.17	0.21	60.32	-8.26	.412
	Month	0.04	0.29	59.13	0.14	.893
	Six Months*	1.03	0.19	60.72	5.32	<.001
Concentration	Week	-0.43	0.31	71.52	-1.41	.164
	Month	0.14	0.30	71.73	.48	.634
	Six Months*	0.47	0.23	74.15	2.07	.042

Note. The symbol * refers to the optimal timeframe (coefficients closest to 1.0)

For research question II it was asked, does the optimal recall timeframe in terms of test-retest reliability, differ significantly by the ME/CFS symptom measured? A multilevel statistical model was used to test Research question II. Level 1 of the model tested (1) the main effect of timeframe, (2) the extent that symptom composite scores at interview one predicted composite scores at interview two, (3) the interaction of timeframe and symptom composite scores at interview one in predicting scores at interview two, and (4) how symptom type moderated the way symptom composite scores at interview one predicted composite scores at interview two. Level 2 of the model tested the main effect of symptom type. The variable Symptom Type included in the model below, represents all nine ME/CFS symptoms, each with a designated code (e.g. Sore throat = 1, Lymph Node = 2…etc). Group mean centering was conducted for the Level 1 continuous variables, so as to control for the influence of between-person variance on the slope coefficients.

Level 1: $y_{ij} = b_{0i} + b_1$ Timeframe + b_2 Symptom Composite Score at Interview One

+ b_3 Timeframe (Symptom Composite Score at Interview One) + r_{ij}

y= Symptom Score at Interview Two

Level 2: $b_{0i} = \gamma_{00} + \gamma_{01}$ Symptom Type + υ_{ij}

$b_{1i} = \gamma_{10} + \gamma_{11}$ Symptom Type

$b_{2i} = \gamma_{20} + \gamma_{21}$ Symptom Type

$b_{3i} = \gamma_{30} + \gamma_{31}$ Symptom Type

The omnibus F test revealed that symptom scores at interview one significantly predicted symptom scores at interview two, $F(1, 1273.30) = 1435.34$, $p < .001$. There was a significant main effect of symptom type ($F(8, 1275.66) = 11.17$, $p < .001$) but not a significant main effect of timeframe, ($F(2, 1273.88) = 1.97$, $p = .139$). There were no significant two-way interactions, such that timeframe by interview one scores ($F(2, 1275.92) = 1.79$, $p = .168$), symptom type by interview one scores ($F(8, 1282.22) = 1.71$, $p = .091$), and timeframe by symptom type, ($F(16, 1273.32) = 0.36$, $p = .990$) were all insignificant. Additionally, There was not a significant three way interaction between timeframe, symptom type, and interview one symptom scores, $F(16, 1273.67) = 0.343$, $p = .993$; hence, the optimal recall timeframe does not differ by the ME/CFS symptom being measured.

Research question III is supplemental, and speculates, what the optimal recall timeframe is in terms of test-retest reliability, in the absence of contextual factors (e.g. stability and momentary symptom severity scores)? 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, and (2) how timeframe moderated the way symptom composite scores at interview one predicted scores at interview two. Level 1 of the model tested the main effect of timeframe. Analyses for research question III were conducted using all ME/CFS symptom scores regardless of stability ratings. Grand mean centering was conducted for the Level 2 variables, so as to ease interpretation.

Level 1: $y_{ij} = b_0 + b_1$Past Week Vs Six Months $_{ij} + b_2$ Past Month Vs Six Months $_{ij} + r_{ij}$

y= Symptom Score at Interview Two

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

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

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

Results of the above analyses revealed that the slope coefficients for all but one symptom (e.g. all except joint pain) were optimal at the six month timeframe in reliably reporting ME/CFS symptoms, in the absence of contextual level two factors (stability and momentary severity). The slope coefficient for joint pain scores reveal that the past month is optimal for reliably reporting joint pain (Please see Table 6 for slope coefficients of ME/CFS symptoms rated at all three timeframes without contextual factors).

Table 6

Slope Coefficients of ME/CFS symptoms across timeframe Sans Stability

All Symptoms	Timeframe	*b*	*SE*	*df*	*t*	*p*
Sore Throat	Week	-0.06	0.19	11.40	-0.30	.763
	Month	0.33	0.33	105.04	1.01	.316
	Six Month*	0.75	0.18	112.63	4.25	<.001
Lymph Node Pain	Week	-.10	0.24	105.42	-0.42	.678
	Month	.12	0.30	103.33	0.39	.698
	Six Month*	1.15	0.22	106.95	5.28	<.001
PEM	Week	0.30	0.19	116.47	1.57	1.20
	Month	0.28	0.22	112.99	1.29	.201
	Six Month*	0.72	0.18	118.12	3.98	<.001
Muscle Pain	Week	-0.48	0.25	117.12	-1.87	.064
	Month	0.43	0.36	109.84	1.19	.236
	Six Month*	0.74	0.28	114.81	2.68	.009
Joint Pain	Week	-0.09	0.25	108.39	-.358	.721
	Month*	0.81	0.32	105.77	2.53	.013
	Six Month	0.54	0.21	111.03	2.60	.010
Unrefreshing Sleep	Week	0.29	0.23	107.74	1.27	.207
	Month	0.35	0.21	108.34	1.63	1.63
	Six Month*	0.47	0.16	112.28	2.84	.005
Headaches	Week	-0.01	0.16	118.48	-0.09	.932
	Month	-0.33	0.28	108.24	-1.18	.240
	Six Month*	0.92	0.17	117.52	5.43	<.001
Memory	Week	-0.09	0.20	105.52	-0.45	.656
	Month	0.02	0.26	103.15	0.09	.933
	Six Month*	0.38	0.17	107.16	2.20	.03
Concentration	Week	-0.18	0.19	107.62	-0.96	.338
	Month	0.06	0.22	105.14	0.29	.770
	Six Month*	0.42	0.19	107.33	2.24	.027

Note. The symbol * refers to the optimal timeframe (coefficients closest to 1.0)

CHAPTER IV

DISCUSSION

The following chapter provides a review of the major findings from this study as well as implications for future work and research in the ME/CFS field. Limitations of the study are also identified and recommendations for future research in this area are presented.

Major Findings and Implications

The present study served as an investigation of the impact of contextual factors (e.g. timeframe, symptom stability, and momentary symptom severity) on the test-retest reliability of ME/CFS symptom composite scores (frequency multiplied by severity) across two assessment points. Results of hypothesis I, which tested the impact of symptom stability on reliability, revealed that symptom stability significantly and positively impacted test-retest reliability for post-exertional malaise (PEM), headaches, and memory problems, such that the more stable the symptom was perceived to be over time, the better participants' symptoms scores at interview one were in predicting scores at interview two.

Prior research supports the finding that greater stability can improve recall (Stone & Shiffman, 2002; Stull et al., 2009); however, it is unclear why this impact of stability was found for some and not all ME/CFS symptoms. These differential findings suggest that symptom stability can have a significant impact on the reliability of symptom reporting and that the size of the impact may depend on symptom type. Based on these findings, it is important for researchers and

health care professionals to take into account the potential impact of stability on symptom experience.

Results of hypothesis II revealed that increases in the momentary severity scores for concentration problems and sore throats across the two interviews, significantly predicted increases in concentration and sore throat composite scores at interview two when the reference group was the six month timeframe. These results were not observed at the other two timeframes (e.g. past week and past month) or for any of the additional ME/CFS symptoms measured.

Prior research has shown that mood, attitude, and health status at the time of an assessment can impact recall (Blaney, 1986; Broderick, Eich et al., 1985; Schwartz, Shiffman, Hufford, & Stone, 2003; Stull et al., 2009). Specifically, Eich et al. found that respondents with increased pain at the time of an assessment were more likely to recall their past pain symptoms as more severe than they had originally reported (1985). Based on this research, it was expected that momentary symptom severity would have a wider impact on the reliability of symptom reports assessed at the longer timeframes (past week, past month, past six months). Results of the present study suggest that for the majority of ME/CFS symptoms, the reliability of the composite scores are not largely affected by momentary symptom severity. Sore throats and concentration problems however, do appear to be impacted by a person's momentary status. It is unsurprising that an increase over the course of one week in momentary concentration severity could influence a persons' ability to reliably recall their concentration scores at the longer timeframes. It is possible that this impact on reliability is due to the fact

that people are having difficulty concentrating on what the question is asking them; thus having difficulty reliably recalling their concentration problems over the longer timeframe. Alternatively people may be using cognitive heuristics by adjusting their concentration problems as worse at the longer timeframes because it is experienced as more severe in the moment. Results showed that the shift in scores at the longer timeframes were in the same positive direction as the momentary changes in score, suggesting that the latter explanation is plausible.

One possible explanation for why changes in momentary sore throat severity impact recall at the six month period, is that the majority of participants rated their sore throat scores as variable over the six month period (see Table 2 and Table 3) and this instability in symptom experience may make sore throats more susceptible to cognitive biases. For instance, Bradburn, Rips, and Shevell (1987), and Bradburn (2000) assert that when respondents are asked to report on highly fluctuating symptoms at a longer timeframe, they are more likely to use cognitive heuristics for this highly complex task. Furthermore, these adjustments and short-cuts may be more susceptible to contextual factors such as momentary severity. As the recall timeframe gets longer, the task becomes more complex for the respondent, and it becomes more likely that a respondent will rely on cognitive short-cuts to answer the question; thus providing a possible explanation for why changes in momentary severity only significantly impacted symptom reports at the six month timeframe rather than the past month or past week. Given these findings, it may be important for researchers and physicians to take into account current health status when acquiring retrospective reports of certain

symptoms, especially when symptoms are highly fluctuating or reported over a longer timeframe.

Results of research question Ia revealed that the optimal timeframe for ME/CFS symptoms perceived as variable over time, differed across symptoms. The past six months was observed as the optimal timeframe 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 reporting PEM, unrefreshing sleep, and memory problems. Lastly, the past week timeframe was found to be optimal for variable joint pain. While an optimal timeframe could be identified for each ME/CFS symptom, 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 the illness ME/CFS (Carruthers et al., 2003; Jason 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 was 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 my represent a unique subset. The majority of participants reported headaches as variable over time (31 out of 47; see Table 2 and Table 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 when assessing variable symptoms over a longer timeframe; which in turn affects reliability and accuracy of reporting (Bradburn, Rips, & Shevell, 1987; Bradburn, 2000). Due to the fact that the majority of the study population reported their headaches as variable, it is recommended that researchers and physicians be 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.

Results of research question Ib (supplementary) revealed that the optimal timeframe in terms of test-retest reliability for ME/CFS symptoms perceived as stable over time was highly uniform, such that all nine ME/CFS symptoms were more reliably recalled at the six month timeframe compared to the past week and past month timeframes. Stable sore throats had the weakest slope coefficients at all three timeframes 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 timeframe. Sore throats are not widely considered a cardinal symptom of ME/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 percent of these respondents reported their sore throats as variable rather than stable over time. Stone and colleagues (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 timeframes, an individual will attempt to summarize their experience. Summarizing variable events over a long timeframe has been found to reduce reporting accuracy (Stone, Schwartz, Broderick, & Shiffman, 2005); however, in the case of this study, when reporting on particular symptoms, such as sore throats over longer timeframes, variability may actually improve recall reliability.

Results of research question II revealed that in the absence of contextual factors (e.g. stability, momentary severity), recall reliability across timeframe does not differ by symptom type. This is supported by the results of the supplementary research question III, which showed that in the absence of contextual factors (symptom stability and momentary severity), the optimal timeframe for reliably reporting ME/CFS symptoms appeared to be six months for all but one symptom (e.g. joint pain), which had an optimal timeframe of one month. While past literature shows a reduction in reporting accuracy when using longer recall timeframes, the results of this study show that longer timeframes may actually improve reliability. As mentioned previously, individuals with chronic illnesses may have a good grasp of their symptom pattern over time (Broderick et al., 2008), which may at least partially explain why individuals in this study were able to reliably make a global assessment of their symptoms at the

six month timeframe. People afflicted with a chronic illness such as ME/CFS may be more reliable in making a broad and global estimate of their symptoms over a longer timeframe because the shorter timeframe may be more susceptible to small changes that deviate from the normal symptom pattern. Clarke et al. (2008) assert that there is a tradeoff between reporting accuracy and loss of information when deciding between a shorter or longer recall timeframe. Short timeframes may increase the accuracy of recall, but investigators risk losing valuable information about the true nature of the phenomena that would be better captured with a longer recall period. More work is needed in this area in order to determine if the six month timeframe is optimal in understanding the experience of ME/CFS symptoms.

Limitations of Research

There are notable limitations of the current study. The study sample used was not selected through random assignment and thus participants may share certain characteristics that are different from the larger population of individuals affected by ME/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 samples; therefore, the current sample may not be generalizable to the entire ME/CFS population.

Another limitation of this study was the uneven frequency of stable versus 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. 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 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 versus variable ratings and also have groups with cases below 10.

Lastly, another possible limitation of this study is the potential for the "adjustment and anchoring" heuristic (Tversky & Kahneman, 1974) in influencing recall reliability across the timeframes. The adjustment and anchoring effect explains how people take information that they know and use that information as an anchor to help estimate information that they do not know. Steen et al. (1994) showed that individuals rating their asthma symptoms over a three month timeframe, first rated their asthma over the past month and used this rating as an anchor in order to estimate their asthma over the three month timeframe. It is possible that this anchoring effect was present in the current study; however, in an attempt to control this effect, the timeframes were spaced

out so that symptom ratings were not organized by symptom groupings but rather by timeframe groupings. For example, participants did not rate their sore throats at each timeframe all at once, but rather participants rated all nine symptoms at the first timeframe (right now) and then all nine symptoms at the second timeframe (past week) and so on. (see Appendix A for a visual representation of the questionnaire). By not positioning the different timeframes directly after the other for each symptom, it seems likely that the tendency for "adjustment and anchoring" heuristics is greatly reduced.

Conclusion and Future Directions

Overall, findings from the presented study reveal that contextual factors do influence the reliability of reporting ME/CFS symptoms; however, not as dramatically as might be expected. Furthermore, the degree of impact that these contextual factors have on test-retest reliability may depend on the ME/CFS symptom being measured as well as individual characteristics of the respondent. Furthermore, results showed that in general, individuals with this illness are capable of reliably recalling the frequency and severity of their symptoms over longer timeframes (e.g. six months), which is contrary to what might be expected based on literature documenting reduced accuracy of reports using longer timeframes.

It is recommended that future research in this area explore the potential tradeoff between reduced reporting accuracy and gaining more information about a phenomenon using longer timeframes. For instance, one way of assessing the validity of the longer six-month timeframe is by comparing the degree of

convergent validity that symptom scores measured at longer timeframes have with other diagnostic measures. One criterion that is necessary for receiving a diagnosis of ME/CFS is the experience of substantial reductions in occupational, social, and personal activities (Fukuda, et al., 1994). Future research might assess the degree to which symptom ratings at each timeframe correlate with or predict measures of substantial reduction.

In terms of the influence of contextual factors, it may also be conducive to understand additional factors that influence the reliability of symptom reporting. These additional factors may include recent stressful life events, social support, or the participants' stage/progression of illness. Participants of the current study answered questions regarding recent life events, stress, and additional health factors on the significant events questionnaire. Although these issues were not explored for the purposes of the present paper, these potentially influential factors will be explored in future research.

In sum, timeframe, symptom stability, and momentary severity do appear to influence the reliability in reporting ME/CFS symptoms. Furthermore, in the absence of stability and momentary severity, individuals were most reliable in reporting the majority of the nine ME/CFS symptoms over a six month timeframe. It will be important for researchers who are interested in the assessment of ME/CFS to take these contextual factors into account, especially if the intended goal of the research is in standardizing and improving the methods used to reliably and accurately diagnose this complex illness. Accurate and

reliable assessment is a crucial first step in understanding and treating this debilitating and often misunderstood illness.

CHAPTER V

SUMMARY

Retrospective self report measures are often used in research and diagnostic assessment of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) (Hawk, Jason, & Torres-Harding, 2007; Jason, King, Frankenberry, & Jordan, 1999; King & Jason, 2005; Reeves et al., 2005). These retrospective self-report measures are susceptible to recall bias, which has the potential to impact the reliability and validity of diagnostic decisions. One factor that can influence the magnitude of recall bias in symptom reporting, is the length of the recall timeframe used. Previous research has found that recall bias may increase when longer reporting periods are used, but very little research has been done on this area (Broderick et al., 2008), making it unclear what the optimal reporting period is for tracking health symptoms, especially for a complex chronic illness such as ME/CFS.

In order to contribute to the literature on the effects of timeframe length on symptom recall in individuals with ME/CFS, this study served as an investigation of the reliability of symptom data assessed at three recall timeframes (the past week, the past month, and the past six months) and at two assessment points (with one week in between each assessment). Symptoms that are experienced as more stable in nature have been found to be recalled with greater accuracy than symptoms that are highly fluctuating and variable; therefore, it was predicted that the test-retest reliability of ME/CFS symptoms measured at the different recall timeframes would be strongest for those symptoms that are stable overtime. This

hypothesis was supported for only three of the nine ME/CFS symptoms measured (e.g. post-exertional malaise, headaches, and memory).

Another aim of the study was to investigate the influence that an individual's current symptom severity has on symptom recall at longer timeframes. It was predicted that an increase (worsening) in momentary symptom severity ratings from baseline to assessment two, would predict an increase in the recall of symptom frequency and severity scores at longer timeframes. Similarly, it was expected that a decrease in momentary symptom severity ratings from week one to week two would result in a decrease in recall for symptom frequency and severity scores at longer timeframes. This hypothesis was only supported for two of the nine symptoms (e.g. sore throats and concentration problems) when the reference group was six months.

In order to further understand the influence of symptom stability on recall reliability, the present study investigated the optimal recall timeframe for symptoms rated as variable versus symptoms rated as stable. Results suggested that the optimal timeframe for variable ME/CFS symptoms differed across symptoms, such that, the past six months was observed as the optimal timeframe 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 reporting PEM, unrefreshing sleep, and memory problems. Lastly, the past week timeframe was found to be optimal for variable joint pain. Results revealed that the optimal timeframe for reliably reporting stable ME/CFS symptoms is highly uniform, such that all nine ME/CFS symptoms measured

were more reliably recalled at the six month timeframe compared to the past week and past month timeframes. In the absence of contextual factors (e.g. stability, momentary severity), recall reliability across timeframes did not differ by symptom type. Supplemental analyses revealed that in the absence of the contextual factors mentioned above, the optimal timeframe for reliably reporting ME/CFS symptoms appear to be six months for all but one symptom (e.g. joint pain), which had an optimal timeframe of one month.

REFERENCES

Amjadi-Begvand, S., Khanna, D., Park, G.S., Bulpitt, K.J., Weng, K.W. & Paulus, H.E. (2004). Dating the 'window of therapeutic opportunity' in early rheumatoid arthritis: accuracy of patient recall of arthritis symptom onset. *Journal of Rheumatology, 31*(9), 1686-1692.Retrieved rom http://www.jrheum.org/content/31/9/1686.abstract

Beckett, M., Da Vanzo, J. Sastry, N., Panis, C., & Peterson, C. (2001). The quality of retrospective data: An examination of long-term recall in a developing country. *The Journal of Human Resources, 36*(3), 593-625. Retrieved from http://www.jstor.org/stable/3069631

Blaney, PH. (1986). Affect and memory: A review. *Psychological Bulletin, 99* (2), 229-45.

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* (4798), 157-161. doi: 0.1126/science.3563494

Brauer, C., Thomsen, J.F., Loft, I.P., Mikkelsen, S. (2003). Can we rely on retrospective pain assessments? *American Journal of Epidemiology, 157* (6), 552-557. doi: 10.1093/aje/kwg017

Broderick, J.E., Schwartz, J.E., Shiffman, S., Hufford, M.R., & Stone, A.A.

(2003). Signaling does not adequately improve diary compliance. *Annals of Behavioral Medicine, 26,* 139-148. doi: 10.1207/S15324796ABM2602_06

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*(1), 146-157. doi:10.1016/j.pain.2008.03.024

Brown, Molly M (2007). Functioning in individuals with chronic fatigue syndrome: Increased impairment with co-occurring multiple chemical sensitivity and fibromyalgia. *Dynamic medicine, 6*(1), 6-15. doi: 10.1186/1476-5918-6-6

Cannell, C.F., Marquis, K.H., & Laurent, A. (1977). A summary of studies of interviewing methodology. (DHEW Publication No. (HRA) 77-1343). *Vital Health Statistics,* Series 2, Number 69. Rockville, Md.: National Center for Health Statistics.

Carruthers, B. M., Jain, A., de Meirleir, K. L., Peterson, D. L., Klimas, N. G., Lerner, A., & ... van de Sande, M. I. (2003). Myalgic encephalomyelitis/chronic fatigue syndrome: Clinical working case definition, diagnostic and treatment protocols. *Journal Of Chronic Fatigue Syndrome, 11*(1), 7-115. doi:10.1300/J092v11n01_02 Retrieved from http://www.cdc.gov/nchs/data/series/sr_02/sr02_069.pdf

Celebrezze, A.J., & Terry, L.J. (1965). Reporting of hospitalization in the health interview survey. (DHEW Publication No. HSM 73-1261).

Vital Health Statistics, Series 2, Number 6. Washington, DC: National Center for Health Statistics. Retrieved from http://www.cdc.gov/nchs/data/series/sr_02/sr02_006.pdf

Clarke, P.M., Fiebig, D.G., & Gerdtham, U. (2008). Optimal recall length in survey design. *Journal of Health and Economics, 27*(5), 1275-1284. doi: 10.1016/j.jhealeco.2008.05.012

Dawson, E.G., Kanim, L.E., Sra, P. et al. (2002). Low back pain recollection versus concurrent accounts: Outcomes analysis. *Spine, 27*(9), 984-993.

Dedominicis R., Grechi, G. (1965). Comparison of hospitalization reporting in three survey procedures. A study of alternative survey methods for collection of hospitalization data from household respondents. *Vital Health Statistics, 115*, 1-48.

Ebbinghaus, H. (1913). A Contribution to Experimental Psychology. New York: Teachers College, Columbia University. Retrieved from 26&dq=Ebbinghaus,+H.+(1913).+A+Contribution+to+Experimental+Psychology&source=bl&ots=RiAWNG70ay&sig=P4Ysl2jqqwQmM4nbocN2aiJ1Stw&hl=en&ei=_7ZTJHbOMH9nAfItnvAQ&sa=X&oi=book_result&ct=result&resnum=9&ved=0CEYQ6AEwCA#v=onepage&q&f=false

Eich, E., Reeves, J.L., Jaeger, B., & Graff-Radford, S.B. (1985). Memory for pain: relation between past and present pain intensity. *Pain, 23*(4), 375-380. doi:10.1016/0304-3959(85)90007-7

Evans, C., & Crawford, B. (1999). Patient self-reports in pharmacoeconomic studies: Their use and impact on study validity. *Pharmacoeconomics, 15*(3), 241-256. doi: 1170-7690/99/0003-0241

Feikin, D.R., Audi, A., Olack, B., Bigogo, G.M., Polyak, C., Burke, H., Williamson, J., & Brieman, R.F. (2010). Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. *International Journal of Epidemiology, 39*(2), 1-9. doi:10.1093/ije/dyp374

Fienberg, S.E., Loftus, E.F., & Tanur, J.M. (1985). Recalling pain and other symptoms. *Health and Society, 63*(3), 582-597. doi: 10.2307/3349849

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*(1), 29-33. doi: 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*(12), 953-959. Retrieved from http://web.ebscohost.com.ezproxy1.lib.depaul.edu/ehost/pdfviewer/pdfviewer?vid=2&hid=10&sid=414b5d9f-a698-4eda-b92e-ef4f6ba2c167%40sessionmgr12

Gendreau, M., Hufford, M., & Stone, A. (2003). Measuring clinical pain in chronic widespread pain: Selected methodological issues. *Best Practice &*

Research Clinical Rheumatology, 17(4), 575-592. doi:10.1016/S1521-6942(03)00031-7

Hawk, C., Jason, L.A., Torres Harding, S. (2007). Reliability of a chronic fatigue syndrome questionnaire. *Journal of Chronic Fatigue Syndrome 13*(4), 41-66. doi: 10.1300/J092v13n04_05

Jason, L.A., Evans, M., Brown, A., Brown, M., Porter, N., Hunnell, J., Anderson, V., Lerch, A. (2010). Sensitivity and specificity of the CDC empirical chronic fatigue syndrome case definition. *Psychology, 1*, 9-16. doi:10.4236/psych.2010.11002

Jason, L.A., Evans, M., Porter, N., Brown, M., Brown, 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(2). 120-135. Retrieved from http://www.scipub.org/fulltext/ajbb/ajbb62120-135.pdf

Jason, L. A., Helgerson, J., Torres-Harding, S.R., Carrico, A.W., & Taylor,mR.R. (2003). Variability in diagnostic criteria for chronic fatigue syndrome may result In substantial differences in patterns of symptoms and disability. *Evaluation & the health professions*, *26*(1), 3-22. doi: 10.1177/0163278702250071

Jason, L.A., King, C.P., Frankenberry, E.L., & Jordan, K.M. (1999). Chronic fatigue syndrome: Assessing symptoms and activity level. *Journal of Clinical Psychology*, *55*(4), 411-424. doi: 10.1002/(SICI)1097-4679(199904)

Jason, L.A., Najar, N., Porter, N., & Reh, C. (2009). Evaluating the Centers for Disease Control's Empirical Chronic Fatigue Syndrome Case Definition. *Journal of Disability Policy Studies, 20*(2), 93-100. doi: 10.1177/1044207308325995

Jason, L.A., Richman, J.A., Rademaker, A.W., Jordan, K.M., Plioplys, A.V., Taylor, R.R. , …Plioplys, S. (1999). A community-based study of chronic fatigue syndrome. *Archives of Internal Medicine, 159*(18), 2129- 2137. Downloaded from www.archinternmed.com on January 8, 2012.

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(1), 39-59. doi:10.1300/J092v03n01_04.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a Better End. *Psychological Science, 4*(6), 401-405. doi:10.1111/j.1467-9280.1993.tb00589.x

Keller, S.D., Bayliss, M.S., Ware, J.E., Hsu, M., Damiano, A.M., Goss, T.F. (1997). Comparison of responses to SF-36 health survey questions with one-week and four-week recall periods. *Health Services Research, 32*(3). 367-384. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1070196/

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

Maas, C.J.M., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*,85-92. Doi: 10.1027/1614-2241.1.3.85

Madow, W. G. (1967). Interview data on chronic conditions compared with information derived from medical records. Vital and health statistics. Series 2 (23). Washington D.C.: U.S. Government Printing Office

McColl, E. (2004). Best practice in symptom assessment: A review. Gut, 53(4), 49-45. doi: 10.1136/gut.2003.034355

Meng, X., Rosenthal, R., & Rubin, D. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172-175. doi:10.1037/0033-2909.111.1.172.

Redelmeier, D., & Kahneman, D. (1996). Patients' memories of painful medical treatments:real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3-8. doi:10.1016/0304-3959(96)02994-6

Redelmeier, D., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, *104*(1-2), 187-194. doi:10.1016/S0304-3959(03)00003-4

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. doi: 10.1186/1741-7015-3-19.

Rivicki, D.A., Irwin, D. Reblando, J., & Simon, G.E. (1994). The accuracy of self-reported disability days. *Medical Care, 32*(4), 401-404. Retrieved from http://www.jstor.org/stable/3766027

Skowronski, J.J., Betz, A.L., Thompson, C.P., & Shannon, L. (1991). Social memory in everyday life: Recall of self-events and other-events. *Journal of Personality and Social Psychology, 60*(6), 831-843. doi: 10.1037/0022-3514.60.6.831

Smets, E.M., Garssen, B.J., Bonke, B.,& DeHaes, J.C.J.M. (1995). The multidimensional fatigue inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research, 39*(5), 315-325. doi:10.1016/0022-3999(94)00125-O

Sohl, S.J., Friedberg, F. (2008). Memory for fatigue in chronic fatigue syndrome: Relationships to fatigue variability, catastrophizing, and negative affect. *Behavioral Medicines, 34(*1), 29-38. doi: 10.3200/BMED.34.1.29-38

Spitzer, R., Endicott, J., Robins, E. (1978). Research diagnostic criteria. *Archives of General Psychiatry 35*(6), 773-782.

Steen, N., Hutchinson, A., McColl, E., Eccles, M. P., Hewison, J., Meadows, K. A., & ... Fowler, P. (1994). Development of a symptom based outcome measure for asthma. *BMJ: British Medical Journal (International Edition), 309*(6961), 1065.

Stone, A., Broderick, J., Kaell, A., DelesPaul, P., & Porter, L. (2000). Does the peak-end phenomenon observed in laboratory pain studies apply to real-

world pain in rheumatoid arthritics? *The Journal of Pain, 1*(3), 212-217. doi:10.1054/jpai.2000.7568

Stone, A., Broderick, J., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain, 104*(1-2), 343-351. doi:10.1016/S0304-3959(03)00040-X

Stone, A.A. , Broderick, J.E., Shiffman, S.S., & Schwartz, J.E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between-and within-person consistency, and judged change in weekly pain. *Pain, 107*(1-2), 61-69. doi:10.1016/j.pain.2003.09.020

Stone, A. A., Schwartz, J., Broderick, J., & Shiffman, 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*(10), 1340-1346. doi:10.1177/0146167205275615

Stone, A.A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*(3), 226-243. doi: 10.1207/S15324796ABM2403_09

Stull, D., Leidy, N.K., Prasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion, 25*(4), 929-942. doi: 10.1185/03007990902774765

Sudman, S., & Bradburn, N.M. (1973). Effects of time and memory factors on response in surveys. *Journal of American Statistical Association, 68*(344), 805-815. Retrieved from http://www.jstor.org/stable/2284504?origin=JSTOR-pdf&

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1130.

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*(8), 1-8. doi:10.1186/1478-7954-3-8

Ware, J.E., Sherbourne, C.D. (1992).The MOS 36-item short form health survey (SF-36): Conceptual framework and item selection. *Medical Care , 30*(6), 473-483. Retrieved from http://www.jstor.org/stable/3765916

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. doi: 10.1037/0022-3514.54.6.1063

Watson, D., & Clark, L.A. (1994). The PANAS-X Manual for the positive and negative affect schedule-expanded form. Iowa, The University of Iowa. Retrieved from http://www.psychology.uiowa.edu/faculty/clark/PANAS-X.pdf

Williams, A., Davies, H., & Chadury, Y. (2000). Simple pain rating scales hide complex idiosyncratic meanings. Pain, 85(3), 457-463. doi:10.1016/S0304-3959(99)00299-7

Williams, D. A., Gendreau, M., Hufford, M. R., Groner, K., Gracely, R. H., & Clauw, D. J. (2004). Pain Assessment in Patients With Fibromyalgia Syndrome. *The Clinical Journal of Pain, 20*(5), 348-356. doi:10.1097/00002508-200409000-00010

Appendix A

Symptom Inventory-Revised

Interviewer Script: The following questions are about physical symptoms that you may be experiencing currently or have experienced over the past week, the past month, and/or the past 6 months. Please answer each question to the best of your ability

1. *Interviewer Script:* The symptoms that I am about to read to you refer to how you are feeling **RIGHT NOW**. When I say a symptom out loud, please tell me whether you are experiencing the symptom **RIGHT NOW** and **how bad** it is on a scale from 1-5, where 1=Very Mild, 2=Mild, 3=Moderate, 4=Severe, and 5=Very Severe

Symptom	Is the Symptom Present **right now**?		How bad is this symptom **right now**? Please circle on a scale from 1-5 **1 Very Mild** **2 Mild** **3 Mod.** **4 Severe** **5 Very Severe**				
Sore Throat	Yes	No	1	2	3	4	5
Tender Lymph Nodes or swollen glands in your neck or armpits	Yes	No	1	2	3	4	5
Fatigue after physical or mental exertion	Yes	No	1	2	3	4	5
Muscle aches or pains	Yes	No	1	2	3	4	5
Pain in several joints	Yes	No	1	2	3	4	5
Feeling un-refreshed from last night's sleep	Yes	No	1	2	3	4	5
Headache	Yes	No	1	2	3	4	5
Memory Problems that have caused you to substantially cut back on activities today?	Yes	No	1	2	3	4	5
Difficulty with thinking or concentrating that have caused you to substantially cut back on activities today?	Yes	No	1	2	3	4	5

2. *Interviewer Script:* Now I will read you the same list of symptoms but this time I am referring to your s ymptoms over the **PAST WEEK**. For each symptom I say out loud, please tell me whether you have experienced each symptom over the course of the **PAST WEEK**

Also, please tell me **how often** you experienced each symptom over the **PAST WEEK** on a scale of 1-5, where 1= A little of the time, 2= Some of the time, 3=A good bit of the time, 4=Most of the time, and 5= All of the time. Also please tell me **how bad** it is on a scale from 1-5, where 1=Very Mild, 2=Mild, 3=Moderate, 4=Severe, and 5=Very Severe

Symptom	Symptom Present over the **past week**?	**How often** did you experience this symptom over the **past week**? Please circle on a scale from 1-5					**How bad** was this symptom over the **past week**? Please circle on a scale from 1-5				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
		A little	**Some**	**A good bit "of the time"**	**Most**	**All**	**Very Mild**	**Mild**	**Mod.**	**Severe**	**Very Severe**
Sore Throat	Yes No	1	2	3	4	5	1	2	3	4	5
Tender Lymph Nodes or swollen glands in your neck or armpits	Yes No	1	2	3	4	5	1	2	3	4	5
Fatigue after physical or mental exertion	Yes No	1	2	3	4	5	1	2	3	4	5
Muscle aches or pains	Yes No	1	2	3	4	5	1	2	3	4	5
Pain in several joints	Yes No	1	2	3	4	5	1	2	3	4	5
Feeling un-refreshed from sleep	Yes No	1	2	3	4	5	1	2	3	4	5
Headache	Yes No	1	2	3	4	5	1	2	3	4	5
Memory Problems that caused you to substantially cut	Yes No	1	2	3	4	5	1	2	3	4	5

3. *Interviewer Script:* Now I will read you the same list of symptoms but this time I am referring to your symptoms over the **PAST MONTH**. For each symptom I say out loud, please tell me whether you have experienced each symptom over the course of the **PAST MONTH**

Also, please tell me **how often** you experienced each symptom over the **PAST MONTH** on a scale of 1-5, where 1= A little of the time, 2= Some of the time, 3=A good bit of the time, 4=Most of the time, and 5= All of the time. Also please tell me **how bad** it is on a scale from 1-5, where 1=Very Mild, 2=Mild, 3=Moderate, 4=Severe, and 5=Very Severe

Symptom	Symptom Present over the **past month**?	**How often** did you experience this symptom over the **past month**? Please circle on a scale from 1-5 **1 A little 2 Some 3 A good bit 4 Most 5 All "of the time"**	**How bad** was this symptom over the **past month**? Please circle on a scale from 1-5 **1 Very Mild 2 Mild 3 Mod. 4 Severe 5 Very Severe**
Sore Throat	Yes No	1 2 3 4 5	1 2 3 4 5
Tender Lymph Nodes or swollen glands in your neck or armpits	Yes No	1 2 3 4 5	1 2 3 4 5
Fatigue after physical or mental exertion	Yes No	1 2 3 4 5	1 2 3 4 5
Muscle aches or pains	Yes No	1 2 3 4 5	1 2 3 4 5
Pain in several joints	Yes No	1 2 3 4 5	1 2 3 4 5
Feeling un-refreshed from sleep	Yes No	1 2 3 4 5	1 2 3 4 5
Headache	Yes No	1 2 3 4 5	1 2 3 4 5
Memory Problems that caused you to	Yes No	1 2 3 4 5	1 2 3 4 5

4. *Interviewer Script:* Now I will read you the same list of symptoms but this time I am referring to your symptoms over the **PAST 6 MONTHS**. For each symptom I say out loud, please tell me whether you have experienced each symptom over the course of the **PAST 6 MONTHS**

   Also, please tell me **how often** you experienced each symptom over the **PAST 6 MONTHS** on a scale of 1-5, where 1= A little of the time, 2= Some of the time, 3=A good bit of the time, 4=Most of the time, and 5= All of the time. Also please tell me **how bad** it is on a scale from 1-5, where 1=Very Mild, 2=Mild, 3=Moderate, 4=Severe, and 5=Very Severe

symptom	Symptom Present over the **past 6 months**?		**How often** did you experience this symptom over the **past 6 months**? Please circle on a scale from 1-5 **1 A little 2 Some 3 A good bit "of the time" 4 Most 5 All**					**How bad** was this symptom over the **past 6 months**? Please circle on a scale from 1-5 **1 Very Mild 2 Mild 3 Mod. 4 Severe 5 Very Severe**				
Sore Throat	Yes	No	1	2	3	4	5	1	2	3	4	5
Tender Lymph Nodes or swollen glands in your neck or armpits	Yes	No	1	2	3	4	5	1	2	3	4	5
Fatigue after physical or mental exertion	Yes	No	1	2	3	4	5	1	2	3	4	5
Muscle aches or pains	Yes	No	1	2	3	4	5	1	2	3	4	5
Pain in several joints	Yes	No	1	2	3	4	5	1	2	3	4	5
Feeling un-refreshed from sleep	Yes	No	1	2	3	4	5	1	2	3	4	5
Headache	Yes	No	1	2	3	4	5	1	2	3	4	5

Appendix B

Symptom Stability Survey

Interviewer Script: Now I am going to ask you about the stability of your health symptoms. For each symptom I say out loud, please say "yes" if you have experienced the symptom over the past 6 months.

AND

If you have experienced the symptom, please tell me whether the symptom has been constant
or if it has been fluctuating and inconsistent over the past 6 months

*A symptom that is **constant** is one that occurs regularly and does not change much in how bad or severe it is over time.*

*A symptom that is **fluctuating** and inconsistent is one that does not occur regularly and there is no pattern to how bad or severe it is.*

More Examples:

*A **constant** symptom is one that is experienced every week or every day and with the same intensity or severity*

*A **fluctuating** symptom is one that is experienced some weeks but not others and there is no pattern to how often it is experienced or how bad it is experienced*

Symptom over the past 6 months?	Symptom Stability
Sore Throat	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Tender Lymph Nodes or swollen glands in your neck or armpits	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Fatigue after physical or mental exertion	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Muscle aches or pains	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Pain in several joints	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Feeling un-refreshed from sleep	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Headache	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Memory Problems that cause you to substantially cut back on activities	☐ Constant ☐ Fluctuating ☐ Symptom not Present
Difficulty with thinking or concentrating that causes you to substantially cut back on activities	☐ Constant ☐ Fluctuating ☐ Symptom not Present

Appendix C

Table 7

Table 7

Means and Standard Deviations of Symptom Severity Scores on the SI-R at Interviews 1 and 2, N=51

		Interview 1	Interview 2
Symptom	Timeframe	M (SD)	M (SD)
Sore Throat	Now	0.75 (1.11)	0.92 (1.15)
	Week	1.51 (1.39)	1.53 (1.46)
	Month	1.60 (1.29)	1.51 (1.27)
	Six Months	2.04 (1.35)	1.92 (1.43)
Lymph Nodes	Now	1.16 (1.24)	1.24 (1.45)
	Week	1.82 (1.41)	1.63 (1.46)
	Month	1.84 (1.43)	1.76 (1.35)
	Six Months	6.43 (6.86)	5.45 (5.75)
Post Exertional	Now	3.45 (1.25)	3.57 (1.25)
Malaise (PEM)	Week	3.86 (0.69)	3.94 (0.73)
	Month	3.90 (0.64)	3.86 (0.63)
	Six Months	4.10 (0.67)	3.90 (0.61)
Muscle Pain	Now	2.84 (1.21)	2.78 (1.22)
	Week	3.18 (0.91)	3.06 (0.93)
	Month	3.14 (1.08)	3.10 (0.94)
	Six Months	3.29 (1.06)	3.10 (0.85)
Joint Pain	Now	2.20 (1.54)	2.20 (1.48)
	Week	2.45 (1.38)	2.66 (1.33)
	Month	2.57 (1.35)	2.57 (1.25)
	Six Months	2.63 (1.48)	2.69 (1.29)
Unrefreshing Sleep	Now	3.78 (1.22)	3.61 (1.25)
	Week	3.78 (0.90)	3.84 (0.92)
	Month	3.71 (0.90)	3.61 (0.90)
	Six Months	3.75 (1.07)	3.69 (0.99)
Headaches	Now	1.41 (1.49)	1.25 (1.47)
	Week	2.45 (1.42)	2.47 (1.43)
	Month	2.67 (1.28)	2.45 (1.22)
	Six Months	2.84 (1.39)	2.78(1.19)
Memory Problems	Now	2.31 (1.57)	2.24 (1.49)
	Week	2.96 (0.10)	2.84 (1.24)
	Month	2.94 (1.01)	2.84 (1.10)
	Six Months	3.02 (1.03)	3.00 (1.15)
Difficulty	Now	2.69 (1.17)	2.84 (1.27)
Concentrating	Week	3.29 (0.88)	3.20 (1.06)
	Month	3.22 (0.83)	3.18 (0.91)
	Six Months	3.37 (0.96)	3.31 (0.99)