Published in final edited form as:
Pain. 2008 September 30; 139(1): 146–157. doi:10.1016/j.pain.2008.03.024.

The accuracy of pain and fatigue items across different reporting periods

Joan E. Broderick, Ph.D., **Joseph E. Schwartz, Ph.D.**, **Gregory Vikingstad, B.A.**, **Michelle Pribbernow, M.A.**, **Steven Grossman, M.S.**, and **Arthur A. Stone, Ph.D.**
Department of Psychiatry and Behavioral Science, Stony Brook University

Abstract

The length of the reporting period specified for items assessing pain and fatigue varies among instruments. How the length of recall impacts the accuracy of symptom reporting is largely unknown. This study investigated the accuracy of ratings for reporting periods ranging from 1 day to 28 days for several items from widely used pain and fatigue measures (SF36v2, Brief Pain Inventory, McGill Pain Questionnaire, Brief Fatigue Inventory). Patients from a community rheumatology practice (*N*=83) completed momentary pain and fatigue items on average 5.4 times per day for a month using an electronic diary. Averaged momentary ratings formed the basis for comparison with recall ratings interspersed throughout the month referencing 1-day, 3-day, 7-day, and 28-day periods. As found in previous research, recall ratings were consistently inflated relative to averaged momentary ratings. Across most items, 1-day recall corresponded well to the averaged momentary assessments for the day. Several, but not all, items demonstrated substantial correlations across the different reporting periods. An additional 7 day-by-day recall task suggested that patients have increasing difficulty actually remembering symptom levels beyond the past several days. These data were collected while patients were receiving usual care and may not generalize to conditions where new interventions are being introduced and outcomes evaluated. Reporting periods can influence the accuracy of retrospective symptom reports and should be a consideration in study design.

Keywords

pain; fatigue; recall; measurement; momentary assessment; outcome assessment; SF36; Brief Pain Inventory; McGill Pain Questionnaire; Brief Fatigue Inventory

Introduction

In recent years, a growing body of literature from our laboratory and others has suggested that patient reported outcomes (PROs) involving recall can suffer from bias [10,19,28]. The processes underlying these recall biases include various cognitive heuristics and the inherent difficulty of remembering all experiences over the reporting period [21]. Heuristics that give disproportionate weight to the peak and the most recent symptom experiences when constructing a recall rating are among the replicated biases observed [8,9,20,25,28].

Corresponding author: Joan E. Broderick, Ph.D., Department of Psychiatry and Behavioral Science, Putnam Hall, South Campus, Stony Brook University, Stony Brook, NY 11794-8790, Phone: 631-632-8083, Fax: 631-632-3165, Email: Joan.Broderick@StonyBrook.edu.

Reporting periods vary across patient reported outcome items and instruments. For example, the most commonly used version of the SF36 has a 4-week recall [31] and another version has a 1-week recall. Other instruments use multiple recall time frames within a single questionnaire; for example, some of the Brief Pain Inventory items [6] reference the last week, and other items omit a time frame. Other instruments, such as the McGill Pain Inventory [17], ask about symptoms "right now," or, as in the Brief Fatigue Inventory [18], during the last 24 hours. The rationale for selecting a particular time frame for the assessment of symptoms is often not described by instrument developers, although the choice is probably based on time frames of interest in clinical trials, "common sense," and the precedence established by existing assessment instruments.

While the reporting periods mentioned are not unreasonable, little attention has focused on issues pertaining to the validity of data obtained from different reporting periods. Among the important issues for researchers to consider are (1) whether length of recall influences ratings, (2) whether it would be advisable, for example, to use a 1-month recall for baseline assessment in a clinical trial to capture a more reliable estimate of the symptom and then use a 1-week recall at the end of the trial to capture patients' status at the point of optimal therapeutic benefit, and (3) can study results be compared with instruments that use different recall periods?

We conducted a study to investigate the accuracy of recalled pain and fatigue using items from widely used instruments. First, based upon previous research on recall bias, we hypothesized that there would be a negative monotonic relationship between length of recall and the correlation of recall ratings with averaged momentary ratings for the same time period. That is, as the length of recall extends from one day to a month, recall bias(es) would be increasingly activated and the correspondence between recall and momentary measures would decline. For the purposes of this investigation, aggregated momentary ratings were conceptualized as accurately representing symptom levels. Second, based on our prior studies [26,28], we hypothesized that mean ratings of pain and fatigue would increase as the reporting period increased, despite no actual change in levels observed in momentary ratings. A third analysis compared recalled ratings of each of the prior 7 days to prospectively collected momentary ratings; we hypothesized that accuracy would be lower for days that were earlier in the week (further in the past).

Method

Participants

Study participants (N=117) were recruited from two offices of a community rheumatology practice (see CONSORT statement, Figure 1). A research assistant informed patients in the waiting room that they might be eligible to participate in a study if they experienced pain or fatigue. The protocol was briefly described and contact information was obtained from interested patients in order to telephone them later, provide a more detailed description of the study procedures, and screen for eligibility. Patients were required to be available for 30 consecutive days and to satisfy the following eligibility criteria: ≥ 18 years of age; absence of significant sight, hearing, or writing impairment; fluent in English; normal sleep-wake schedule; diagnosed with a chronic disease; experienced pain or fatigue in the last week; able to come to the research office two times within a month; and, had not participated in another study using an electronic diary during the previous 5 years. The study protocol was approved by the Stony Brook University Institutional Review Board, and patients provided informed consent and were compensated $100 for participation in the study. An additional incentive of $150 was awarded to one participant who was randomly selected (by lottery) from those who completed all six recall questionnaires on the scheduled days. The study was conducted from September 2005 through June 2006.

Telephone screening of 279 patients determined that 86 (31%) were ineligible due to visual or hearing difficulties, inability to hold a pen, atypical sleep-wake schedule, no chronic illness diagnosis, or previous participation in a momentary assessment study. Of 193 eligible patients, 76 (39%) declined participation, and 117 (61%) participated. Physician-confirmed diagnoses indicated multiple co-morbidities. The most prevalent diagnoses were osteoarthritis (50%), rheumatoid arthritis (29%), lupus (18%) and fibromyalgia (11%). A majority were female (86%) and married (64%) with a mean age of 56 years (range 28–88). Most were high school graduates (95%), and 41% of the participants had completed college or more education. Eleven patients dropped out of the study resulting in an analysis sample of N=106.

Materials and Measures

Momentary ratings of pain and fatigue

For the purposes of this investigation, aggregated momentary measures of pain and fatigue are conceptualized as accurate measures of patients' experience of these symptoms across the reporting periods. While undoubtedly subject to sources of error, the fact that they do not require memory or mental aggregation of experience to make the rating means that they are not subject to the biases introduced by faulty memory or operation of cognitive heuristics. For this reason, we view them as accurate measures of patient symptom experience and use them as the standard with which to examine the accuracy of recall ratings as representations of patient experience.[1]

Momentary ratings of pain and fatigue were collected for 29–31 days on a hand-held computer (Palm Zire 31). The electronic diary (ED) utilized a software program provided by invivodata, inc. (Pittsburgh, PA) that featured auditory tones to signal the participant to complete a set of ratings. The questions were presented on the ED screen, and participants used a stylus to tap their responses. The ED was programmed to generate an average of 7 randomly-scheduled (within intervals) signals spread across the participant's waking hours (an average of one every 2 hours and 20 minutes, constrained to ensure a minimum of 30 minutes between prompts) determined by when the participant informed the ED that she was going to bed at night and set the wake up alarm the next morning. In addition to the random signals, an End-of-Day assessment was completed by the participant at the time the ED was put to sleep (results not reported here). Participants had three minutes to complete an assessment after being signaled; after that time assessments could not be completed and were classified as missed. When a signal occurred at an inconvenient time (driving a car, during a business meeting), the ED allowed the participant to delay an assessment for 5 to 20 minutes (5 minute increments). For a given assessment, the delay feature could be utilized only once, and when the time expired another signal was generated to prompt completion of the ratings. The software also included a suspend feature that helped subjects incorporate the diary into their daily routine by allowing them to suspend prompting for defined periods of sleep and uninterruptible activities such as church or a business meeting. Prompting could be suspended multiple times a day for 30 minutes to 2 hours (15-minute increments) at a time. If participants ended their activity earlier than expected, they could end the suspension and prompting would resume. If a prompt had been scheduled to occur during the time the ED was suspended, this assessment was coded as "skipped" and was not rescheduled for a later time.

Each ED assessment began with the participant responding to questions about his location, activity, whether alone or with others, and one positive and one negative affect rating ("happy"

[1]Recall ratings may include valuable patient perceptions and beliefs that go beyond the momentary experience of symptoms. For example, recall ratings of colonoscopy pain may correlate more highly with the decision to have a second colonoscopy than momentary ratings of pain during the colonoscopy.[19] This study was designed to investigate the accuracy of recall ratings as measures representing the average experience of that symptom across various reporting periods.

and “frustrated”) on a visual analogue scale (VAS). Our primary outcomes were questions drawn from well-established questionnaires used to assess pain and fatigue: SF36 version 2 [30], Brief Pain Inventory [5], Brief Fatigue Inventory [18], and McGill Pain Inventory[15, 16]. All ED item ratings were done on a 100-point VAS. With the exception of the SF36v2 Bodily Pain item (“none” to “very severe”), all of the other item VAS anchors were “not at all” to “extremely.” Because the protocol was that recall reports for six different periods would be compared to mean levels of momentary ratings for the same periods, the reporting time for the momentary assessments was modified to “right now.” Table 1 lists the items included in the momentary assessment.

Recall pain and fatigue items

The Interactive Voice Recording (IVR) recall assessments (described below) retained the wording and scaling of the original instrument with only the reporting period modified to correspond to 1, 3, 7, or 28 days (see Table 1).

SF36v2—The SF36 version 2 [30] is a global health measure comprised of eight important concepts that capture widely-valued human functioning and well-being and are directly affected by disease and treatment. It was designed as a health survey that could be self-administered, would accurately characterize the level of overall health functioning and well-being across the range from full health through chronic illness, could detect clinical change in health, and would be sufficiently brief to allow widespread application. The SF36v2 time frames available are 1-month or 7-day recall. General US population and disease-specific norms have been established for Version 2 of the SF36 allowing comparisons of disease burden across age groups and disease types. It also allows interpretation of an individual’s scores relative to various norms: healthy age- and sex-matched groups, or with same or other diagnostic groups.

Items from two of the eight scales are used in the primary analyses of this study: Bodily Pain (BP) and Vitality (VT). The Bodily Pain item is: “How much bodily pain have you had during the past 4 weeks” (6 point scale, “none” to “severe”). The four Vitality items were: “How much of the time during the past 4 weeks did you feel full of life?”, “…did you have a lot of energy?”, “…did you feel worn out?”, and “…did you feel tired?” (5 point scale, “none of the time” to “all of the time”).

Brief Pain Inventory and McGill Pain Questionnaire—The Brief Pain Inventory (BPI) is a 20-item inventory partially based on the McGill Pain Questionnaire and originally designed for use in cancer patients [4,5]. It has achieved widespread use in a broad variety of disorders with chronic pain. Pain intensity is measured in terms of both current pain at the time the rating is being made and “worst,” “least,” and “average” pain as well as 15 of the McGill pain adjectives. Functional limitations of pain are measured with 7 items relating to interference with mood, walking and other physical activity, work, social activity, enjoyment of life, and sleep. Cleeland and others have reported coefficient alphas greater than .70 for the intensity items and interference items [11,14]. Validity has been inferred from the ability of the BPI to detect expected differences in severity of pain between patients whose site of disease and need for analgesia is different, and on the correlation between intensity and interference scales [3, 11].

The McGill Pain Questionnaire [15] was developed in 1975 to measure pain intensity and three psychological dimensions of pain: sensory-discriminative, motivational-affective, and cognitive-evaluative and continues to be used in its original form. Melzack reported test-retest reliability of .70 for 10 patients across an interval of up to 7 days [15]. In a study of 40 post-surgical patients, correlations of the McGill scales with VAS ratings were .50 for affective pain

[29]. Other studies have reported strong associations between the McGill scales and other measures of pain [7].

This study reports on 1 of the BPI items: pain intensity item ("Please rate your pain by circling the one number that best describes your pain on the average (5 point scale, "none of the time" to "all of the time") and 3 adjectives from the McGill: aching (sensory), stabbing (sensory), and nagging (affective) based on the low, medium and high response rates to these items. The wording of the IVR ratings of these items were: "Some of the words below describe your pain. For each word, check the degree to which that word describes your pain" (4 point scale, "none" to "severe"). These adjectives are also used in the BPI.

Brief Fatigue Inventory—Like the BPI on which it was modeled, the Brief Fatigue Inventory (BFI) was designed to assess the severity and impact of fatigue in cancer patients, but its use has expanded to many diseases and community samples [1,18]. Fatigue is assessed "right now," and "usual" and "worst" fatigue for the past 24 hours. There are also six items measuring the impact of fatigue on activities of daily living. A factor analysis determined that a single dimension was being measured by the BFI, internal consistency was high, and correlations with other fatigue scales were high suggesting good construct validity [18,32]. The BFI item examined in this study was "Please rate your fatigue (weariness, tiredness) by circling the one number that best describes your USUAL level of fatigue (11-point scale, "no fatigue" to "as bad as you can imagine").

Procedure

First Laboratory visit

Following a telephone eligibility interview, participants came to the research office on two occasions, at the beginning and at the end of the study. During Visit 1, participants completed demographic and questionnaire measures and signed a release for us to obtain confirmation of a rheumatological diagnosis from their physicians. The questionnaires included the pain and fatigue items from the inventories being investigated in this study as well as several psychological measures not reported here. Participants were trained in the use of an electronic diary (ED) to collect momentary and end-of-day ratings of pain and fatigue.

Participants were provided with a packet of six numerically labeled envelopes holding recall questionnaires to use at home and mail back in stamped envelopes. To ensure that the interactive voice response system called when they were likely to be home, participants specified a one-hour window for IVR contacts close to bedtime. However, in order to avoid anticipatory monitoring of pain and fatigue, participants were not informed of the dates they would be contacted to complete the recall questionnaires.

A research assistant telephoned the participant 24 hours after the first laboratory visit to answer any questions and troubleshoot any problems with using the ED. Additionally, a follow-up call was made once a week for the next three weeks to ensure the ED was working properly and to answer any questions.

Interactive voice recording recall assessments

Each participant was randomly assigned to 1 of 10 different recall assessment schedules that specified when, within the 28–30 day study period, the two 1-day, two 3-day, and one 7-day recall assessments would take place; the one 28-day recall was fixed at the end of the study. The schedules were designed with no IVR assessments scheduled in the first 3–6 days and no overlap of days for the 1-day, 3-day, and 7-day recall assessments. One of the 1-day assessments took place on a weekend and the other on a weekday. The two 3-day assessments

took place on a Monday and a Tuesday or Friday. The 7-day recall took place on a weekday. Thus, each recall time frame included weekend and weekdays. Based on the schedule, the IVR computer system (Prosodie Interactive, Inc.) telephoned the participant between 6 PM -12 AM (the call was scheduled one hour before the participant's bedtime) to inform his to open a particular envelope (1–6) and to complete the designated recall assessment within the next hour. The IVR system dialed the participant at the beginning of the specified one hour time period, 15 minutes later if the system encountered a busy signal or reached an answering machine, and again at the end of the 1 hour window if no ratings had been entered yet. After the assessment was completed on a paper questionnaire, the participant was instructed to call the IVR system and enter his responses to each item, thus ensuring time- and date- stamped entries so that compliance with the protocol could be checked. The completed paper assessment was also mailed back to the research office. If no ratings were received on the first night, the IVR system telephoned the participant the next evening to request the assessment be completed. In the event that no ratings were received on the second night, a research assistant telephoned the participant on the third day to determine the problem and instigate an assessment that night.

End-of-study laboratory visit

Participants returned to the research office 29–31 days after the start of the study and completed two additional recall tasks during the visit. They first completed a VAS item asking for a rating of average pain over the last 7 days (anchors: no pain, worst possible pain). A follow-up question asked for a rating of how difficult it was to make the 7-day recall rating (4 choices: "not at all" to "quite a bit"). This was followed by a task in which the participant was asked to rate his pain for each of the last 7 days. Participants were given 7 index cards with the day of the week indicated on each card, and they were asked to make a VAS rating of their pain for that day (same anchors as 7-day recall). Each card included an option, "can't remember," if the participant could not remember his pain on that day. Participants were free to rate the 7 cards in any order.

Data analytic strategy

Compliance with momentary ratings and IVR recall ratings

Momentary and recall data were screened for compliance with the protocol. An insufficient number of completed momentary ratings during a reporting period could fail to accurately capture the pain and fatigue experienced by the patient. Likewise, a recall rating that was not completed immediately at the end of the reporting period could be subject to memory distortions. Consequently, strict criteria were outlined for inclusion of data based upon compliance with the protocol.

Compliance criteria for momentary data were: (1) overall compliance across the 28 days of ≥ 75% (number of completed prompts/number completed + number missed), and (2) a minimum of 3 momentary ratings on each day covered by the 1-, 3-, and 7-day reporting periods. The compliance criterion for the IVR recall ratings was that the rating was made on the day that they were contacted to make a rating.

Adjustment for reliability of momentary means

Heuristically, the goal of this study was to compare each IVR report to the "true" mean of all possible momentary reports for the same period of time. We view the set of obtained momentary reports as a random sample of all possible reports that could have obtained while participants were awake, and therefore the observed or computed mean is considered a sample estimate of the average for that period. Given that the standard errors of the observed means will tend to be larger for shorter periods (e.g., 1 day or 3 days, compared to 7 days or 28 days) due to the smaller number of momentary reports used to calculate the mean, there would be a

methodological bias for the correlation of IVR recall reports with the average momentary rating to be attenuated (i.e. weaker) for shorter reporting periods. To circumvent this problem, we estimated a multi-level mixed model in which the set of person-specific "true" means for each reporting period was treated as a random factor (i.e., latent variable), and we simultaneously estimated the correlations among the 6 latent variables and the 6 corresponding IVR recall reports. The correlation estimates generated by this analytic approach are adjusted for the attenuation, due to unreliability, that would occur if we were simply to analyze computed averages of momentary ratings. The use of full-information maximum likelihood estimation permitted us to estimate correlations for all 83 subjects, even though some were missing data for one of the recall periods. While the analysis estimated the full 12×12 correlation matrix, our interest was exclusively in the 6 correlations of corresponding IVR and averaged momentary reports. The 6×6 matrix of variances (squared standard errors) and covariances for these 6 correlations was used to test pre-specified contrasts. (The details of how this model was specified and estimated may be obtained from the authors.) These analyses were performed is SAS, version 8.2, using Proc MIXED.

Results

Compliance

Comparison of the correlations between momentary and recall ratings across the different reporting periods is more interpretable if the same patients are contributing to each correlation; otherwise any differences in correlations might be due to different participants contributing to them. Consequently, we required that patients had to meet both the ED and IVR compliance criteria outlined above for 5 of the 6 recall assessments. This resulted in excluding 23 patients. Had we required very stringent compliance (meeting both sets of criteria met for all 6 recall assessments), 43 patients would have been excluded – yielding an unacceptably small sample. Thus, the analyses of correspondence of IVR recall reports with averages of momentary reports for the same period were performed on a sample of 83 patients. In this group, the average compliance with the ED protocol was very high (partially due to excluding those with low compliance), with participants collectively completing momentary ratings for 95% – 97% of the times they were prompted, depending on the recall assessment period. This yielded a mean of 5.4 (0.60) momentary ratings per day for the analyzed sample with 80% of days having 5 or more ratings.

The most common form of non-compliance was not completing the IVR recall assessment on the day they were contacted by the IVR computer (n=15); another 5 patients did not meet compliance criteria for the momentary assessments; and, 3 other patients had compliance problems with both IVR and momentary data, In addition, in two cases, the electronic diary malfunctioned resulting in lost data. Table 2 displays the characteristics of the analyzed sample. Comparisons were conducted to determine if the analyzed sample (*n*=83) differed significantly from those patients excluded (*n*=23) due to non-compliance. T-tests and Chi-square tests found neither substantial nor statistically significant differences between the analyzed and excluded participants on any of the demographic or health variables.

Initial assessments of Health, Pain and Fatigue

The means of the SF36v2 scores place the participants at approximately the 50th percentile for psychological well-being (Mental Component Scale) and at approximately the 25th percentile for physical well-being (Physical Component Scale) using the SF36v2 arthritis normative sample [31]. The mean ratings for pain and fatigue over the last month, reported at baseline, were both above 6 on a 0–10 scale.

Correlation between momentary and recall measures

The IVR recall ratings were correlated with the latent variable representing the true mean of the momentary assessments for the corresponding reporting period. Figure 2a displays these correlations for the two global pain items (BPI average pain, SB36v2 Bodily pain) and the global fatigue item (BFI usual pain) across the reporting periods. Assuming that the aggregated momentary ratings capture the actual levels of the symptom, then the correlations represent the extent to which the recall ratings reflect these levels. These correlations ranged from a high of .86 for 1-day recall of usual fatigue (BFI) to a low of .60 for recall of 7-day bodily pain (SF36v2). Tests for differences between the level of correlation for the 1-day recall versus the 7-day recall indicated a significant reduction for bodily pain (p=.01), usual fatigue (p=.02) and, marginally, average pain (p=.10). As will be seen in other items, a consistent trend was observed for 28-day correlations to exceed correlations for 7-day recall ratings. Figure 2b displays correlations for the three McGill pain adjectives. These ranged from a high of .81 for 1-day recall of nagging pain to a low of .59 for 7-day recall of aching pain. The 7-day correlations for both aching and nagging pain were significantly lower than the 1-day correlations (p=.05). Figure 2c displays the correlations for the four SF36v2 Vitality scale items. These ranged from . 80 for 1-day recall of tired to a low of .53 for 7-day recall of energy. Tests of the decrease in the correlation from 1-day and 7-day recall for two of the items trended toward statistical significance (energy, p=.07; tired, p=.09).

It must be noted that all of the momentary items were rated on 100-point VAS, while the range of response options for the recall items ranged from 11 to 4. Correlations between momentary ratings and items with fewer response options may be constrained relative to those with a higher number of response options. Both items with 11-point scales (BFI average pain, BPI usual fatigue) generally had the highest correlations with momentary ratings across the reporting periods, though several of the items with fewer options had correlations nearly as high (e.g., McGill “nagging” pain – 4 options, SF36v2 “worn out” – 5 options).

Comparison of levels of pain and fatigue rated by momentary and recall assessment

The mean ratings on the pain and fatigue recall items were compared for each reporting period to the estimated mean of the averaged momentary assessments and are shown in Table 3. The means of the IVR ratings and the latent variable representing person-specific means of the momentary ratings for each of the 6 reporting periods were generated within the context of the multi-level mixed model used to estimate the above correlations. Pre-specified contrasts were used to test the differences between IVR and latent variable means for each recall length, implicitly averaging the means for the two 1-day and two 3-day reporting periods. For every item, the IVR rating was significantly higher than the averaged momentary rating at each reporting period. Given the p values obtained, Bonferroni correction would not alter the conclusions.

We were also interested in the trend of mean levels of pain and fatigue as one moves from the 1-day to 3-day to 7-day to 28-day reporting periods. Once again, pre-specified contrasts were used to estimate the trend separately for the IVR means and momentary rating latent variable means; the contrast coefficients were −1.5, −1.5, −.5, −.5, 1, 3, respectively, for the two 1-day, two 3-day, 7-day, and 28-day reporting periods. There was not a significant trend in the average levels of the momentary ratings for 7 of 10 items (see significance levels for EMA slopes in Table 3), indicating that actual mean levels of these symptom remained relatively stable across the reporting periods. The momentary slopes increased across the reporting periods for BFI “usual fatigue” (p=.02), and the SF-36v2 Vitality items “Worn Out” (p=.03) and “Tired” (p=. 005). Specifically, as the length of the reporting period increased, the average momentary ratings for these items increased by up to 2.8 points across the reporting periods. In contrast, the slopes of 8 of 10 items were positive for recall IVR recall ratings, all indicating that as the

reporting period increased, the level of the symptom was rated higher. These increases from the 1-day recall to the 28-day recall ranged from 8.8 (SF36v2 Bodily Pain) to 3.6 (SF36v2 "Tired") points. Only the SF36v2 Vitality items "Full of Life" and "Energy" did not display a positive slope. Furthermore, 5 of 10 IVR recall slopes were significantly different from the corresponding momentary slopes and a sixth showed a trend (see Table 3) suggesting inflated recall ratings. Finally, visual inspection of the means in Table 3 suggests that the positive slopes were being created primarily by the smaller discrepancy between momentary and IVR recall ratings for the 1-day recall compared with the larger differences for the other reporting periods, i.e., the recall ratings did not systematically increase as the reporting period increased beyond 3 days. This was examined, post-hoc, by testing the IVR slopes against the momentary slopes while excluding the ratings for the 1-day time period. Only two items showed a positive difference in slopes when the 1-day recall was removed from the analysis. The BPI average pain item had a significant slope ($p \leq .001$) and the SF36v2 bodily pain item showed a trend ($p = .10$). Figures 3a and 3b display the momentary and IVR ratings for the BPI average pain intensity item and the SF36v2 bodily pain item for each reporting period. Thus, recall ratings on pain and fatigue items are consistently higher than averaged momentary ratings for all reporting periods, the difference is somewhat smaller for 1-day recall than for longer recall periods, and for pain intensity items (only) the magnitude of the difference increases as the reporting period increases across the 4 reporting periods.

Single day recall ratings conducted at the last laboratory visit

At the beginning of the last laboratory visit, participants completed a 7-day recall rating of average pain. The correlation with averaged momentary ratings for the week was .78 – similar to what was observed for the 7-day IVR recall rating (r=.74) collected during the month. When asked how difficult it was to make the 7-day recall rating, 60% responded "not at all difficult", 23% responded "a little bit," and 17% responded "moderately" or "quite a bit."

To investigate the nature of retrospective daily recall ratings, the accuracy of the 7 day-by-day ratings collected during the last visit to the laboratory was examined. For each of the 7 days, the averaged momentary ratings for that day were correlated with the recall rating for that day. The observed order of card completion by 96% of participants was starting with yesterday and working back to 7 days ago. Among the 106 participants, there was a systematic increase in the percent saying they could not remember their pain as the length of recall increased. All 106 patients gave a rating for yesterday's pain, and starting from 2 days ago through 7 days ago the percent of participants checking "can't remember" were 2%, 6%, 14%, 19%, 25%, and 34%. Similar to the previous analyses, in order to compare correlations across days in a constant sample of individuals, only the subset of participants with complete recall data for all 7 days, who also had a minimum of 3 momentary reports for each day, was included in the analyses (n=52). These between-subject correlations ranged from .70 to .83 (see Figure 4) indicating reasonably good correspondence with momentary ratings and, contrary to expectation, no tendency for the correlation to decrease for pain ratings of days that were farther in the past.

However, the data were further explored by correlating each day's recall rating with the average of the momentary ratings for each of the non-corresponding days, that is, with every other day except the day being recalled, and averaging these six correlations. These correlation averages are also shown in Figure 4 (open circles) and indicate that except for the most recent two days, the recall ratings corresponded as well to pain levels on other days as they do to pain on the targeted recall day.

Although the between-subject correlations indicate fairly high correspondence between the 7 single-day recalls and the average of corresponding momentary ratings, there is an alternative view of accuracy that is based on how well patients are able to recall and differentiate their daily pain levels across the week. That is, can patients remember which days of the last week

were relatively high in pain (or fatigue) versus which were lower? This is a within-subject question that is not addressed by the between-subject correlations presented above. To estimate these associations, the pooled within-subject correlation was computed. The pooled correlation was .29 ($F(1, 312)=29.5$, $p<.0001$; 95% CI: −.40 to 18). We realized that the magnitude of this correlation might be reduced by inclusion of patients who exhibited very little day-to-day variability in recall or momentary ratings. To ensure that this was not accounting for the low correlation, we took the 50% of participants ($n=26$) above the median for the variability ($SD=6.95$) of the means of the momentary ratings across the 7 days. The pooled within-person correlation for this subset of patients was .31 ($F(1, 156)= 16.2$, $p<.0001$; 95% CI: −.46 to .16), which was only marginally greater than for the entire sample. Finally, one could argue that by conducting the analysis on participants who gave all 7 ratings, we biased the correlations downward, since they may have given ratings when they really could not remember 6 or 7 days ago. When we recalculate the within-subject correlation for the most recent 5 days only, we found an pooled within-person correlation of .33 ($F(1, 208)=26.0$, $p<.0001$; 95% CI: −.46 to .20), similar to the correlation for all 7 days.

Moving to a level comparison of the day-by-day ratings, mean levels of day-by-day recall ratings and averaged momentary ratings replicated the pattern found in the primary analyses. That is, recall ratings were significantly higher than averaged momentary ratings. Across the 7 day-by-day ratings, the discrepancy was 11.8 points (momentary: 46.3 vs. recall: 58.1, $p=.0001$) for yesterday through 10.0 points (momentary: 45.9 vs. recall: 55.9, $p=.0001$) for 7 days ago.

To examine whether averaging 7 day-by-day recall ratings yields a more accurate score than a 7-day recall rating, the 7 day-by-day ratings were averaged and the mean was correlated with the average of the momentary ratings for the week. This correlation was .85, compared with the correlation of .74 for the 7-day recall rating reported above. The difference between these correlations is significant ($p=.015$).

Discussion

This is the first study to systematically investigate the accuracy of pain and fatigue questionnaire items using different reporting periods. Items from some of the most widely used instruments of pain and fatigue were evaluated. The standard reporting periods for these instruments vary, and it is not uncommon for researchers to alter the reporting period to better match the needs of a particular research question [13]. However, very little is known about whether the length of the reporting period impacts the accuracy of the rating obtained. Three types of accuracy were evaluated: 1) the correlation between recall and momentary reports, which assesses the consistency of relative rank orderings of pain and fatigue ratings across patients, 2) the difference in mean levels of symptom ratings between recall and momentary reports, and (3) within-subject accuracy of symptom levels across 7 days.

This study showed a gradual decline in the correlation between the momentary ratings and the IVR recall ratings as the reporting period was increased from 1 to 7 days for several of the items. For all items, the highest correspondence was for 1-day recall (.70–.85). Some of the 3-day recalls also were high – 6 of 10 correlations were > .70, while the other 4 were > .60. Across all items and reporting periods, the amount of shared variance between the momentary and the recall measures ranged from 28% to 73%, indicating there can be substantial differences between recall and momentary assessment. One factor that may have attenuated the correlations is a limited number of response options for some of the recall items. While the two items with 11 response items had some of the highest correlations, there was not a linear relationship as the number of options dropped down to 4 or 5. Nevertheless, this observation bears further scrutiny.

Our largest correlations are similar to those of the two-week recall for pain intensity and momentary ratings obtained by Salovey and colleagues; they observed a .74 correlation [24]. Salovey viewed accounting for half of the variance between recall and momentary pain intensity as reasonable. As there is no definitive threshold for judging the validity of a measure, investigators will need to consider what degree of validity is adequate for the purposes of a particular study.

The observation that some correlations for 28-day recall were higher than those for 7-day recall was unexpected, since we hypothesized that 28-day recall would have the lowest correlation. We speculate that patients with a chronic illness have beliefs that are based on their extensive experience with the symptoms; that is, they probably have a good idea of their typical pain and fatigue levels. Given the difficulty of recalling many previous days, as in the case of the 28-day reporting period, patients may refer to these beliefs, which provide a reasonably good estimate of average symptom levels [22]. However, when asked to construct a rating to represent the past 7 days, patients may attempt to retrieve some memories from the week, which are subject to the distortion of the cognitive heuristics mentioned earlier (e.g., peak effect), thereby decreasing accuracy. These results are consistent with our findings from a cognitive interview study of how patients go about generating a 7-day pain recall rating, suggesting that patients are unable to remember pain levels across the week and that they utilize related sources of information to decide on a rating [2]. Furthermore, when patients engaged in a task of recalling each of the last 7 days, the within-subject correlations are very poor indicating that patients cannot remember fluctuations in daily pain levels across several days. And, Figure 4 demonstrated that while the recall of pain “yesterday” and the “day before” exhibited some specificity, recall for days prior to that corresponded as well to the other days of the week as it did to the target day, suggesting that these ratings represent general evidence-based beliefs about their pain as opposed to actual memories.

Our results are based on patients whose symptoms were stable over the study period, and may not generalize to recall when patients are engaged in a clinical trial where symptoms levels change. If beliefs about one’s pain levels are driving recall reports, then recall ratings made in the context of changing pain levels may not be accurate, because changes in beliefs may lag behind actual change in pain. Alternatively, if beliefs about change in pain are altered by expectations generated by adopting a new treatment regimen, then it is possible that recall ratings of pain could change without any actual change in pain [23]. These possibilities deserve further investigation and caution should be exercised in generalizing the findings of this study to clinical trials.

In clinical trials, if one wished to optimize the accuracy of symptom reports without having to engage in momentary assessment, these data would suggest use of multiple end-of-day ratings. One-day recalls, especially for the global pain and fatigue items (BPI “average pain” and BFI “usual fatigue”), correlated greater than .80 with momentary assessment. One-day recalls would permit observation of day-to-day variability in symptoms and could capture change over time, thus being useful in clinical trials. We also found that averaging ratings of pain for 7 days recalled at the end of a week yielded a higher correlation with aggregated momentary assessment than did a single 7-day recall task.

Turning to possible differences in symptom levels from recall versus momentary assessment, we replicated the observation of higher ratings of symptoms on recall [26,28]. In every instance, recall ratings were significantly higher than momentary assessments, and often the difference between the two ratings was substantial, ranging from 5 to up to 20 points on a 100-point scale. For example, momentary ratings for a single day of “average” pain yielded a mean of 41, while the mean of the IVR recall ratings was 47, a 6-point difference - not a large discrepancy. Yet, there was a 17-point difference for the SF36v2 bodily pain item for the 28-day reporting period.

As expected in a chronic illness in steady state, the averaged momentary ratings of pain and fatigue remained fairly constant across the different reporting periods. However, particularly as the reporting period increased from 1-day to longer periods, recall ratings became higher and more discrepant with the stable momentary ratings. One possible criticism of this finding is that the momentary ratings may have "missed" peaks in pain resulting in lower mean values. However, given the random sampling of moments in this protocol and the high compliance rates, there is no reason to suspect a systematic bias of missing peaks in the sampling, making this a very unlikely explanation. A study comparing responses by asthma patients on the SF36 using 1-week versus 4-week reporting periods also found that the longer reporting period resulted in scores indicating poorer health.[12] Likewise, a recent survey study showed higher ratings of pain and fatigue for one-month recall versus reports of current symptoms [27]. These data also suggest that using different reporting periods at different assessment points in a study could generate apparent differences in symptom levels that are due to recall length but not actual differences in symptom level. Specifically, if one administered a 4-week recall at baseline, and a 7-day recall at the conclusion of treatment, one might observe apparent reductions in pain in both control and treatment groups that were at least partially attributable to the change in reporting period.

In summary, the primary goal of this study concerned the accuracy of recall pain measures as a function of the length of reporting period. Relative to the average of momentary reports for the same period, we expected less accuracy in recall with longer reporting periods. Overall, the data followed this pattern to the 7-day recall period, but an increase in accuracy was observed for the 28-day recall. Our general conclusion is that a number of recall ratings for pain and fatigue have substantial correspondence with momentary reports even for 28-day recalls. Second, we found that not all items tolerate increasing reporting periods with the same degree of accuracy. Third, we replicated previous research that finds that recall ratings are consistently "inflated" compared to the average of momentary ratings for the same period. Fourth, we found that reporting periods of 3, 7, and 28 days generated similar ratings of pain and fatigue levels suggesting that these reporting periods may be interchangeable, perhaps due to taping into beliefs about pain and fatigue rather than veridical memory of symptom levels. Fifth, we found that a 1-day recall corresponded well with the momentary ratings for that day, suggesting that for some applications, measurement of pain and fatigue may not require momentary assessment and can be done with multiple single end-of-day ratings. Finally, it is important to note that these data were collected while patients were receiving usual care, and the results may not generalize to clinical trial conditions where change in symptoms is expected.

Acknowledgements

This research was supported by grants from the National Institutes of Health (1 U01-AR052170-01; Arthur A. Stone, principal investigator) and by GCRC Grant no. M01-RR10710 from the National Center for Research Resources. We would like to thank Pamela Calvanese, Doerte Junghaenel, and Leighann Litcher-Kelly for their assistance in collecting data.

Software and data management services for the electronic diary assessments were provided by invivodata, inc (Pittsburgh, PA). Broderick and Stone have a financial interest in invivodata, inc.

References

1. Anderson KO, Getto CJ, Mendoza TR, et al. Fatigue and sleep disturbance in patients with cancer, patients with clinical depression, and community-dwelling adults. J Pain Symptom Manage 2003;25:307–318. [PubMed: 12691682]
2. Broderick JE, Stone AA, Calvanese P, Schwartz JE, Turk DC. Recalled pain ratings: a complex and poorly defined task. Journal of Pain 2006;7:142–149. [PubMed: 16459280]
3. Cleeland, C. Pain assessment in cancer. In: Osoba, E., editor. Effect of cancer on quality of life. Boca Raton, Florida: CRC Press; 1991.

4. Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994;23:129–138. [PubMed: 8080219]
5. Daut RL, Cleeland CS. The prevalence and severity of pain in cancer. Cancer 1982;50:1913–1918. [PubMed: 7116316]
6. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. Pain 1983;17:197–210. [PubMed: 6646795]
7. De Conno F, Caraceni A, Gamba A, et al. Pain measurement in cancer patients: a comparison of six methods. Pain 1994;57:161–166. [PubMed: 8090512]
8. Eich E, Reeves JL, Jaeger B, Graff-Radford SB. Memory for pain: relation between past and present pain intensity. Pain 1985;23:375–380. [PubMed: 4088698]
9. Fredrickson B. Extracting meaning form past affective experiences: The importance of peaks, ends, and specific emotions. Cogn Emot 2000;14:577–606.
10. Gorin, AA.; Stone, AA. Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In: Baum, A.; Revenson, T.; Singer, J., editors. Handbook of Health Psychology. Mahwah, NJ: Erlbaum; 2001. p. 405-413.
11. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the Brief Pain Inventory for use in documenting the outcomes of patients with noncancer pain. Clin J Pain 2004;20:309–318. [PubMed: 15322437]
12. Keller SD, Bayliss MS, Ware JE Jr, Hsu MA, Damiano AM, Goss TF. Comparison of responses to SF-36 Health Survey questions with one-week and four-week recall periods. Health Serv Res 1997;32:367–384. [PubMed: 9240286]
13. Litcher-Kelly L, Martino S, Broderick J, Stone A. A systematic review of measures used to assess chronic pain in randomized clinical trails and controlled trials. Journal of Pain 2007;8:906–913. [PubMed: 17690014]
14. McDowell, I.; Newell, C. Measuring health: A guide to rating scales and questionnaires. 2nd Ed.. New York: Oxford University Press; 1996.
15. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975;1:277–299. [PubMed: 1235985]
16. Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30:191–197. [PubMed: 3670870]
17. Melzack, R.; Katz, J. The McGill Pain Questionnaire: Appraisal and current status. In: Turk, D.; Melzack, R., editors. Handbook of Pain Measurement. 2nd ed.. New York: Guilford Press; 2001. p. 35-52.
18. Mendoza TR, Wang XS, Cleeland CS, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer 1999;85:1186–1196. [PubMed: 10091805]
19. Redelmeier D, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. Pain 2003;104:187–194. [PubMed: 12855328]
20. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996;66:3–8. [PubMed: 8857625]
21. Robinson M, Clore G. Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychol Bull 2002;128:934–960. [PubMed: 12405138]
22. Robinson M, Clore G. Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. J Pers Soc Psychol 2002;83:198–215. [PubMed: 12088126]
23. Ross M. The relation of implicit theories to the construction of personal histories. Psychol Rev 1989;96:341–357.
24. Salovey P, Smith A, Turk DC, Jobe JB, Willis GB. The accuracy of memories for pain: not so bad most of the time. American Pain Society Journal 1993;2:184–191.
25. Stone A, Broderick J, Kaell A, DelesPaul P, Porter L. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics? Journal of Pain 2000;1:212–217. [PubMed: 14622620]
26. Stone A, Broderick J, Shiffman S, Schwartz J. Understanding recall of weekly pain from a momentary assessment perspective: Absolute accuracy, between- and within-person consistency, and judged change in weekly pain. Pain 2004;107:61–69. [PubMed: 14715390]

27. Stone AA, Broderick JB, Schwartz JE, Schwarz N. Context effects in survey ratings of health, symptoms, and satisfaction. Med Care. in press

28. Stone AA, Schwartz JE, Broderick JE, Shiffman S. Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. Personality and Social Psychology Bulletin 2005;31:1340–1346. [PubMed: 16143666]

29. Taenzer, P. Postopertive pain: Relationships among measures of pain, mood, and narcotic requirements. In: Melzack, R., editor. Pain measurement and assessment. New York: Raven Press; 1983.

30. Ware, JE., Jr; Kosinski, M.; Turner-Bowker, D.; Gandek, B. How to score version 2 of the SF-12 Health Survey. Lincoln, RI: Quality Metric Incorporated; 2002.

31. Ware, JE.; Kosinski, M.; Dewey, JE. How to Score Version 2 of the SF-36 Health Survey. Lincoln, RI: Quality Metric Incorporated; 2000.

32. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage 1997;13:63–74. [PubMed: 9095563]

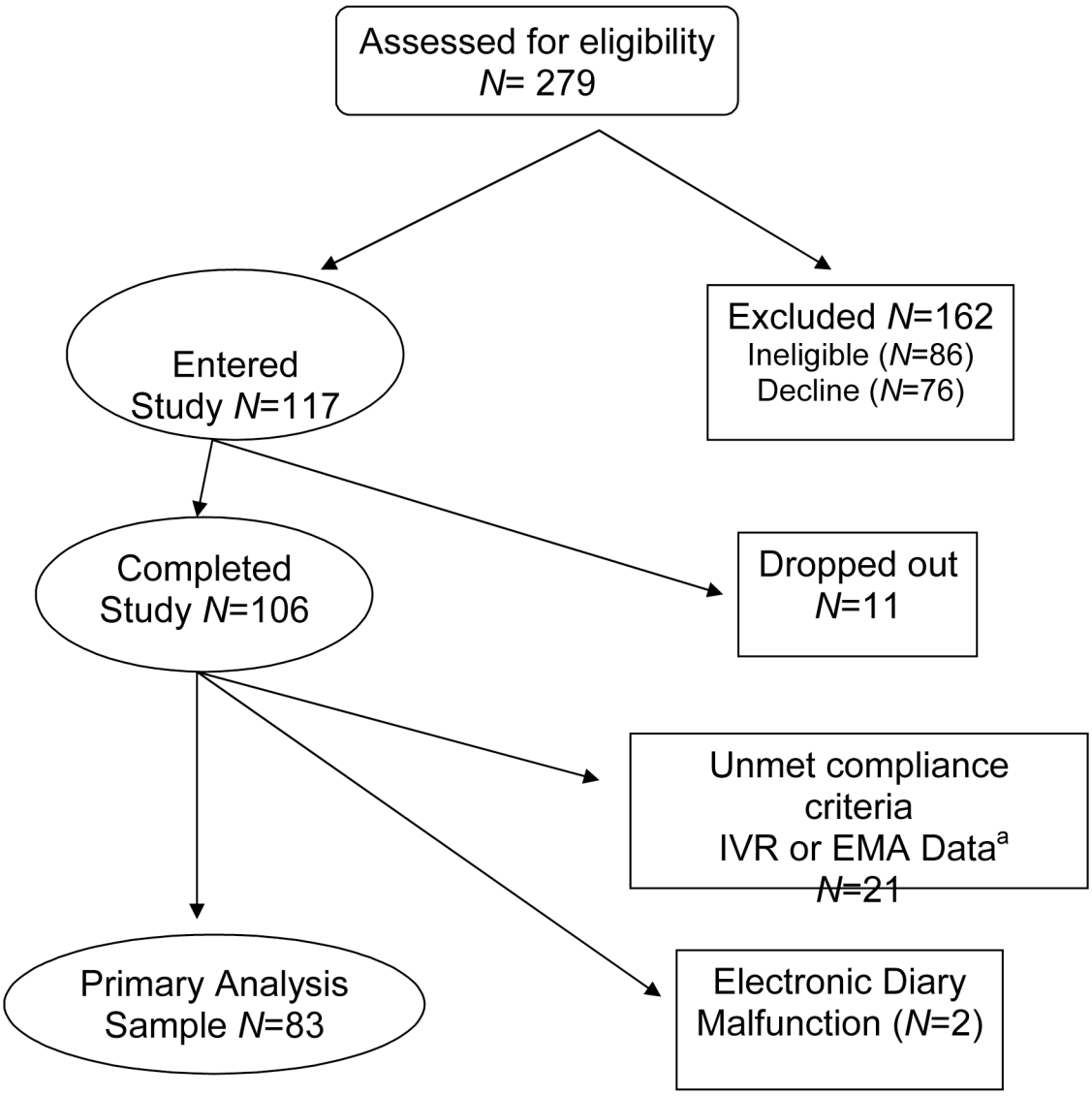
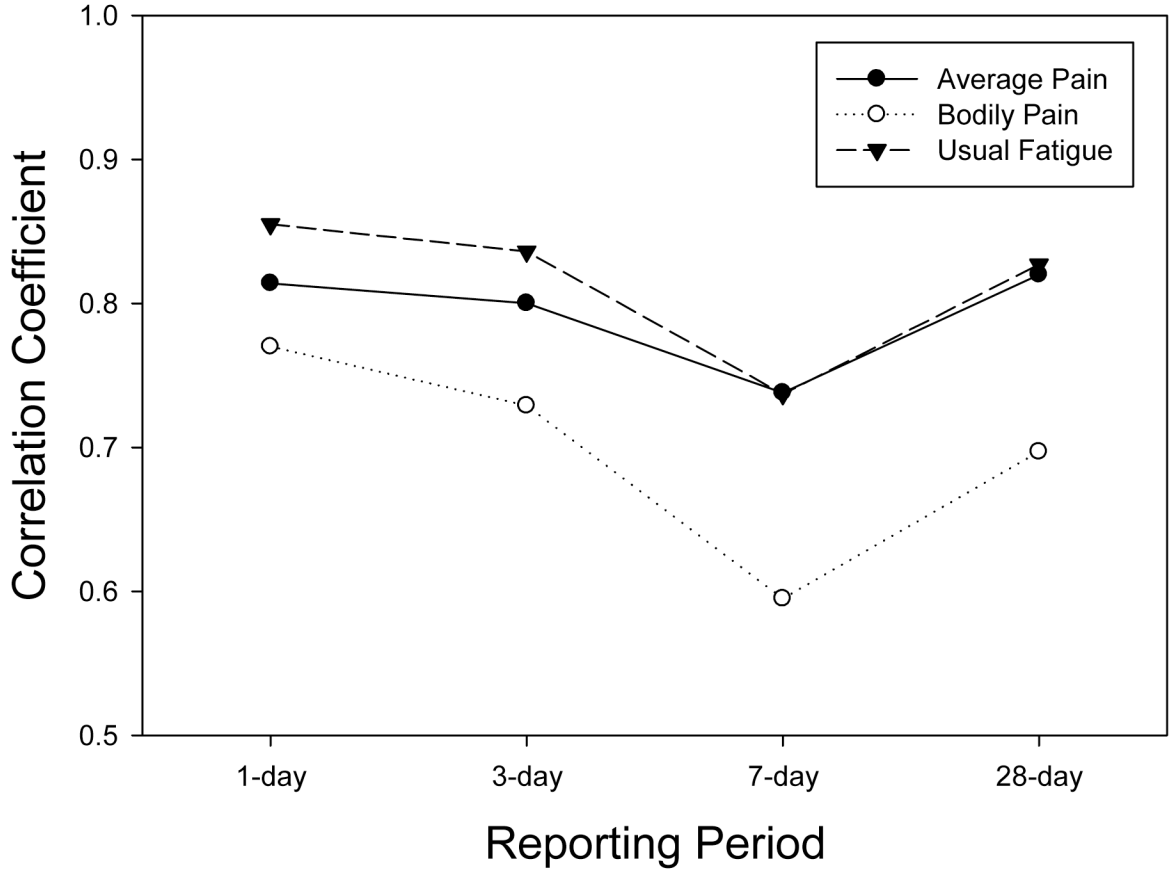
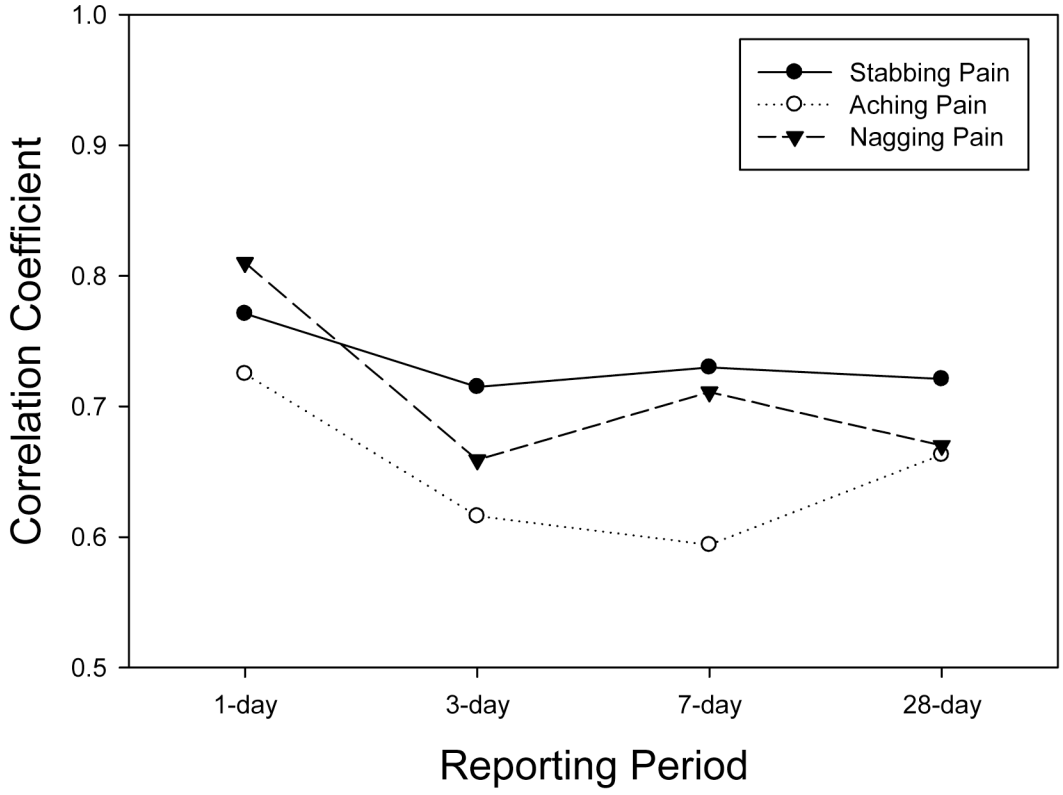


Figure 1.
Study CONSORT statement
[a]Insufficient compliance with ratings for primary analyses: IVR (*N*=15), EMA (*N*=3), both (*N*=3).





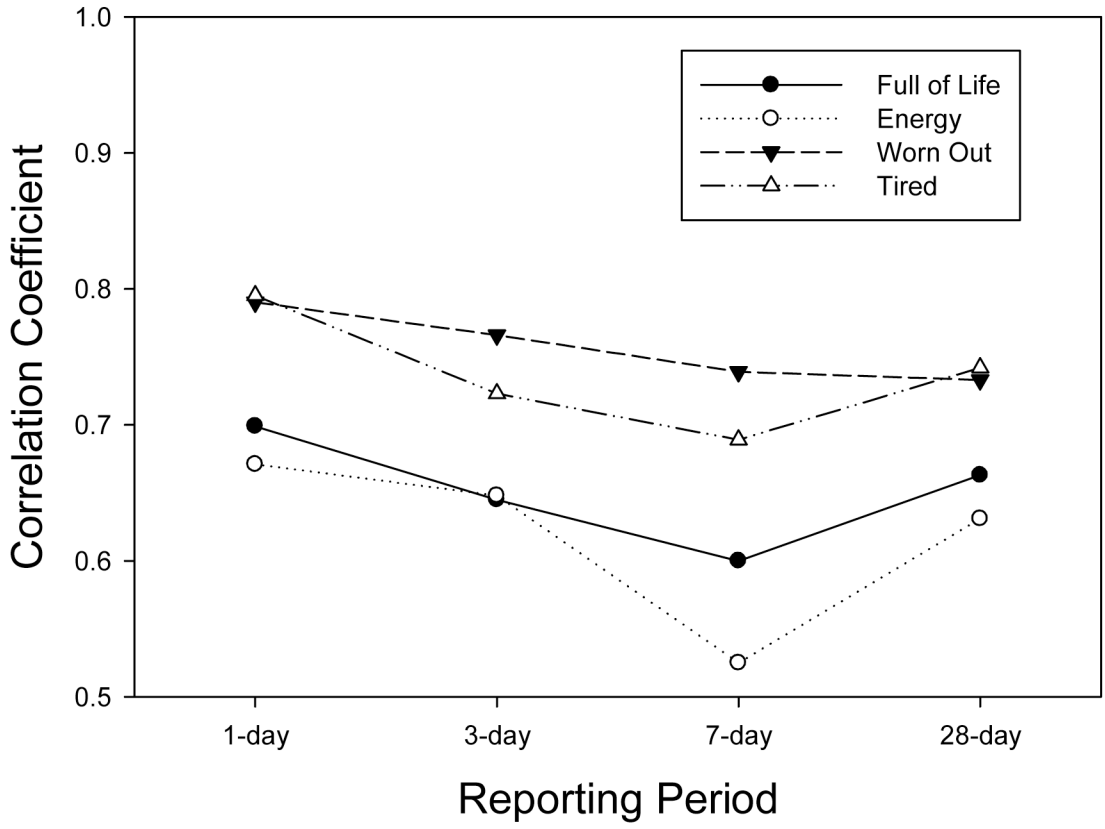


Figure 2.

Figure 2a. Reliability-adjusted correlations of IVR recall with the average of momentary assessments for pain intensity (BPI) bodily pain (SF36v2) and usual fatigue (BFI).

Figure 2b. Reliability-adjusted correlations of IVR recall with the average of momentary assessments for McGill pain adjectives.

Figure 2c. Reliability-adjusted correlations of IVR recall with the average of momentary assessments for SF36v2 Vitality scale items.

Figure 3(a)

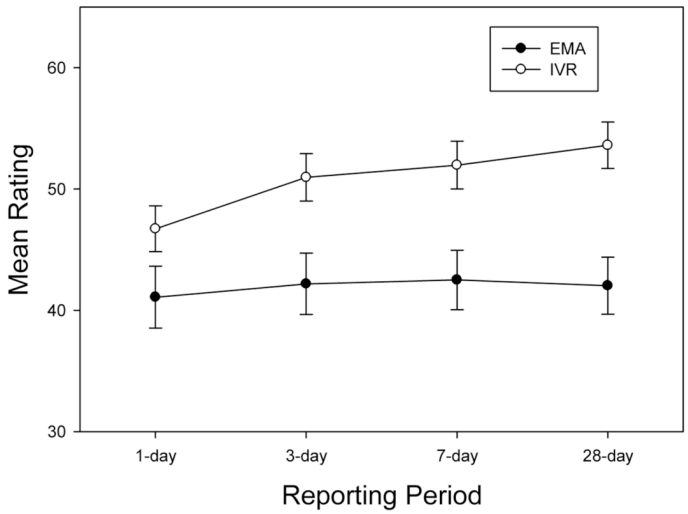


Figure 3(b)

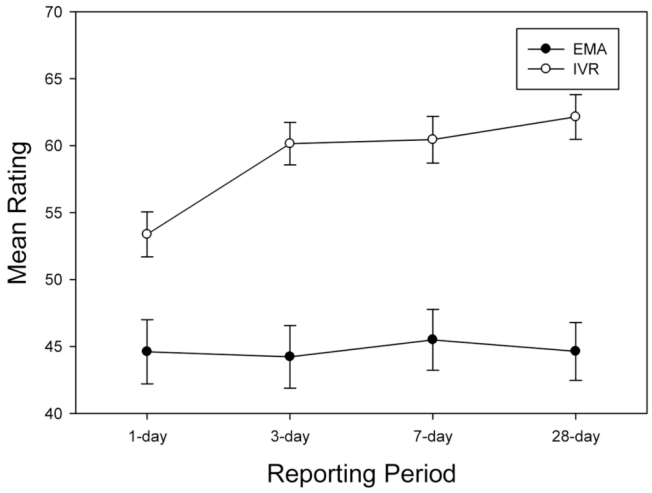


Figure 3.
Mean ratings for momentary (EMA) and IVR recall ratings of BPI average pain (2a) and SF36v2 bodily pain (2b) items.

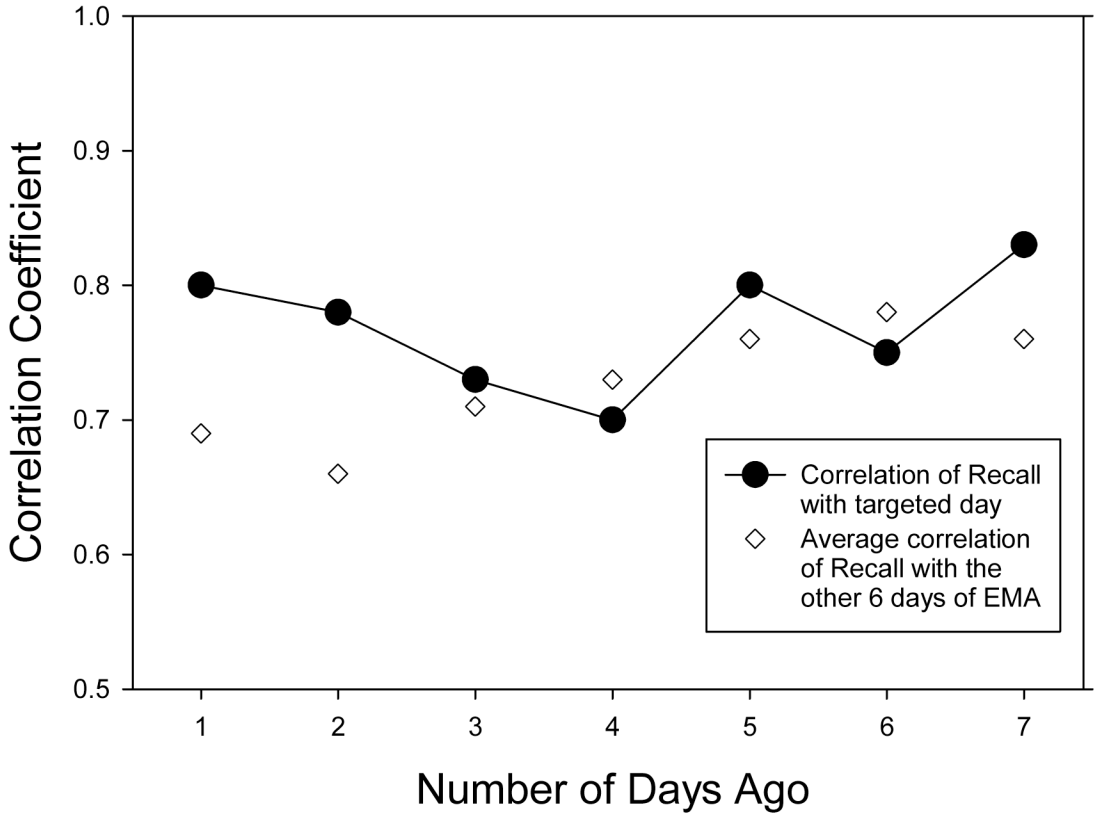


Figure 4.
Correlations of 7 day-by-day recall pain ratings with averaged momentary ratings (EMA) for that day and for the mean of non-corresponding days (N=52).

Table 1

Mapping of pain and fatigue items on the momentary ratings

Original Questionnaire/ IVR Recall Items[1]			Random Momentary Ratings	
Source of item[2]	**Item**	**Item Scaling[3]**	**Electronic momentary item**	**Electronic item VAS[4] anchors**
	Pain			
SF36v2	How much bodily pain have you had during the **past 4 weeks**?	none, very severe (6)	How much bodily pain did you have?	none, very severe
BPI	Please rate your pain by circling the one number that best describes your pain **on the** *average*.	no pain, as bad as you can imagine (11)	How intense was your bodily pain?	not at all, extremely
McGill BPI	Aching (sensory factor)	None, severe (4)	How aching was your pain?	not at all, extremely
McGill BPI	Stabbing (sensory factor)	None, severe (4)	How stabbing was your pain?	not at all, extremely
McGill BPI	Nagging (affective factor)	None, severe (4)	How nagging was your pain?	not at all, extremely
	Fatigue			
SF36v2	How much of the time during the **past 4 weeks** did you feel full of life?	all of the time, none of the time (5)	How full of life did you feel?	not at all, extremely
SF36v2	How much of the time during the **past 4 weeks** did you have a lot of energy?	all of the time, none of the time (5)	How energetic did you feel?	not at all, extremely
SF36v2	How much of the time during the **past 4 weeks** did you feel worn out?	all of the time, none of the time (5)	How worn out did you feel?	not at all, extremely
SF36v2	How much of the time during the **past 4 weeks** did you feel tired?	all of the time, none of the time (5)	How fatigued (weary, tired) did you feel?	not at all, extremely
BFI	Please rate your fatigue (weariness, tiredness) by circling the one number that best describes your USUAL level of fatigue during the **past 24 hours**.	no fatigue, as bad as you can imagine (11)		

[1] IVR recall items are identical to the original questionnaire wording except for the reporting period; the table lists the original questionnaire reporting period; for this study, the IVR items replaced the original reporting period with 1-day, 3-days, 7-days and 28-days to correspond to the protocol recall periods.

[2] SF36 version 2 (SF36v2), Brief Fatigue Inventory (BFI), Brief Pain Inventory (BPI), McGill Pain Questionnaire (McGill).

[3] Numerical scales with endpoint anchors listed and number of points in parentheses.

[4] Electronic items used 100-point VAS ratings with reference to pain or fatigue at the moment of the signal.

Table 2

Demographic and illness characteristics of the analyzed sample (N=83)

Age (Mean, SD)	56.2 (11.1)
Female	87%
Married	65 %
Race (White)	92%
Education	
Did not finish high school	5%
High school graduate	24%
Some college	29%
College graduate or more	41%
Employed	51%
Disability benefits	27%
SF36v2	
Physical Component	32.7 (10.5)
Mental Component	46.6 (12.8)
Initial telephone screening ratings (0–10) for past month	
Pain	6.3 (1.9)
Fatigue	6.3 (2.3)

Table 3

Mean (standard error) for aggregated momentary ratings (EMA) and IVR recall ratings for each reporting period.

Item	Reporting period								IVR Slope	EMA Slope
	1-day		3-day		7-day		28-day			
	EMA	IVR	EMA	IVR	EMA	IVR	EMA	IVR	*p* value	*p* value
BPI "average" pain §***	41.1 (2.5)	46.7 (1.9)	42.2 (2.5)	51.0 (2.0)	42.5 (2.4)	52.0 (2.0)	42.0 (2.4)	53.6 (1.9)	< .0001	ns
SF36v2 bodily pain §***	44.6 (2.4)	53.4 (1.7)	44.2 (2.3)	60.2 (1.6)	45.5 (2.3)	60.5 (1.7)	44.6 (2.2)	62.2 (1.7)	< .0001	ns
BPI & McGill: stabbing §***	23.0 (2.5)	33.6 (2.2)	24.4 (2.6)	39.6 (2.6)	23.7 (2.6)	37.9 (2.5)	23.8 (2.4)	40.2 (2.5)	< .0001	ns
BPI & McGill: aching §***	44.0 (2.6)	57.7 (1.7)	44.6 (2.6)	63.1 (1.7)	45.6 (2.5)	64.3 (1.9)	44.6 (2.4)	64.2 (1.8)	< .0001	ns
BPI & McGill: nagging	39.0 (2.9)	54.1 (2.1)	40.7 (2.9)	57.6 (2.1)	41.1 (2.8)	58.2 (2.4)	40.4 (2.7)	58.4 (2.1)	.005	.096
BFI "usual" fatigue	44.6 (2.3)	49.6 (2.0)	45.4 (2.3)	53.0 (2.2)	47.0 (2.2)	52.9 (2.0)	46.3 (2.1)	53.4 (2.1)	.0005	.023
SF36v2 full of life	47.1 (2.0)	55.5 (1.7)	48.4[a] (2.1)	53.3 (1.9)	47.1 (2.0)	53.5 (2.1)	47.7 (1.9)	53.1 (1.8)	.11	ns
SF36v2 energy	41.0 (1.8)	46.3 (1.7)	42.1[b] (1.9)	45.3 (1.7)	41.0[a] (1.8)	46.0 (1.8)	41.2[a] (1.7)	45.4 (1.9)	ns	ns
SF36v2 worn out §**	43.5 (2.2)	48.4 (1.8)	44.0 (2.2)	53.2 (1.8)	46.1 (2.2)	54.2 (2.0)	45.0 (2.1)	55.0 (1.9)	< .0001	.033
SF36v2 tired	43.4 (2.1)	54.8 (1.8)	44.1 (2.1)	58.1 (1.8)	46.2 (2.2)	59.4 (1.8)	45.4 (2.0)	58.4 (1.8)	.002	.005

Note. The data for the two 1-day and two 3-day reporting periods have been averaged.

All comparisons of EMA vs. IVR ratings within reporting periods (1-day, 3-day, etc.) are significant at ≤.001, except for [a] ≤.01, [b] ≤.05.

§ indicates that the IVR and EMA slopes are significantly different (***≤.001, **≤.01).