

Interference with activities due to pain and fatigue: accuracy of ratings across different reporting periods

Joan E. Broderick · Stefan Schneider ·
Joseph E. Schwartz · Arthur A. Stone

Accepted: 21 May 2010 / Published online: 11 June 2010
© Springer Science+Business Media B.V. 2010

Abstract

Purpose This study examined the impact of different reporting period lengths on the accuracy of items measuring interference due to pain and fatigue with work, walking, and relations with others.

Methods Six items from well-established instruments (Brief Pain Inventory, Brief Fatigue Inventory, SF-36) were investigated in a prospective study of 117 patients with chronic rheumatological illness. Daily ratings were compared with recall ratings of 1, 3, 7, and 28-day reporting periods.

Results The level of recall ratings (RRs) for reporting periods of 3 days or more were significantly higher than the level of aggregated end-of-day (EOD) ratings. Correspondence between aggregated EOD and RRs was good ($r \geq .80$) regardless of the length of the reporting period. Ratings of interference for a single day were highly correlated with aggregated EOD for up to 14 days prior to the single rating ($r \geq .76$).

Conclusions Recall ratings with reporting periods of up to a month yield good correspondence with aggregated daily ratings, although the absolute level of the rating will be inflated for recall periods of 3 days or longer.

Keywords Pain · Fatigue · Momentary assessment · Quality of life · Patient reported outcomes

Abbreviations

BFI	Brief fatigue inventory
BPI	Brief pain inventory
EOD	End of day assessment
IVR	Interactive voice recording
RR	Recall rating across a reporting period

Introduction

Interference with daily activities due to illness is a common focus of quality of life research. Questionnaires measuring interference include the domains of work, physical activity, activities of daily living, and social relations. Depending upon the assessment instrument, the reporting period can vary. For example, the Brief Fatigue Inventory (BFI) [1] asks about interference during the past 24 h, the Brief Pain Inventory (BPI) [2] and the EORTC-Quality of Life-C30 [3] ask about interference during the last week, the most common version of the SF36 Health Survey [4] asks about the past 4 weeks, and the Multidimensional Pain Inventory [5] does not specify a reporting period. Is the reporting period important for the accuracy of such instruments?

Previous research on average pain and fatigue intensity has demonstrated discrepancies between ratings based on real-time assessment and recall ratings of a week or more [6, 7]. Other research has suggested that respondents use cognitive heuristics in making recall ratings for reporting periods of a week or more [8, 9]. Many studies have also found that recall ratings indicate higher symptom levels for longer reporting periods [10–13]. In combination, these studies suggest that self-reports covering a week or more may include distortions that reduce the accuracy of a measure of average or usual experience. No research to

J. E. Broderick (✉) · S. Schneider · J. E. Schwartz ·
A. A. Stone
Department of Psychiatry and Behavioral Science, Putnam Hall,
South Campus, Stony Brook University, Stony Brook,
NY 11794-8790, USA
e-mail: Joan.Broderick@StonyBrook.edu

date has examined the accuracy of items measuring interference of functioning due to health conditions across different reporting periods. That is, does length of reporting period affect the accuracy of patients' reports of interference? Is there an optimal reporting period that should be used for these types of items?

Whereas some patient experiences, such as pain, are perceived on a moment-to-moment basis, this is probably not the case for interference with functioning, which naturally involves cognitive inference and judgment. Determining whether one's activities have been limited may be assessed across time as one tries to make plans and to engage in them. The experience of interference may also be closely bound to "memorable" incidents of activity limitation during the recall period that may inflate the perception of interference across all days in the reporting period (the Peak Heuristic) [9]. Alternatively, in chronic illness, patients may develop broad, stable judgments about the degree of interference that they have, which may not be influenced much by day-to-day or week-to-week fluctuations.

In this study, patients provided recall ratings (RRs) of interference due to pain and fatigue for different reporting periods: 1-, 3-, 7- and 28 days. They also reported interference at the end of the day (EOD) for 28 consecutive days. The EOD ratings were conceptualized as being less subject to memory bias than the longer recall periods and were used as an accuracy reference for the RRs. Interference items from the Brief Pain Inventory [2], Brief Fatigue Inventory [1], and SF36 [14] were used.

The primary analysis addressed the question, does the accuracy of interference ratings decline as the reporting period extends? We hypothesized that RRs would be significantly higher than EOD ratings and that this difference would increase as the reporting period increases. In addition, we hypothesized that the correlations between RRs and EOD ratings would decrease as the reporting period increased from 1 to 28 days. A secondary analysis based upon the available data asked the question, when a patient rates interference for a single day, how well does that rating correspond to interference for the 2 weeks prior to that day? We hypothesized that a rating for 1 day would have the highest correspondence with aggregated ratings of the few days immediately preceding it, and the correlation would decline for aggregates of longer time periods.

Methods

Participants

Patients ($N = 117$) from two offices of a community rheumatology practice were recruited during 2005–2006.

Eligibility criteria included being available for 30 consecutive days, ≥ 18 years of age, absence of significant sight, hearing, or writing impairment, fluent in English, normal sleep-wake schedule, diagnosis of chronic rheumatological disease, pain or fatigue in the last week; able to come to the research office two times within a month, and, no participation in another study using an electronic diary within the previous 5 years. The study protocol was approved by the Stony Brook University Institutional Review Board (Approval #20045609), and patients provided informed consent and were compensated \$100 for participation.

Interested patients ($N = 279$) were screened by telephone, and 86 (31%) did not meet one or more of the eligibility criteria. Of 193 eligible patients, 76 (39%) declined participation, and 117 (61%) participated. Eleven patients dropped out of the study, and two patients' hand-held computer malfunctioned resulting in loss of data; thus, the study sample was $N = 104$. The most prevalent diagnoses were osteoarthritis (49%), rheumatoid arthritis (29%), lupus (17%) and fibromyalgia (11%). Participants tended to be female (87%), White (91%), married (64%), and had a mean age of 55 years (range 28–88). Most were high school graduates (96%), with 72% having completed some college. At baseline, the majority of participants reported at least moderate levels of pain (84%), and symptoms of fatigue most or all of the time (64%) during the past month (see Table 1).

Materials

Three items *each* for interference due to pain and due to fatigue were assessed in a month-long protocol involving daily end-of-day (EOD) ratings and recall ratings (RR). The items studied involved interference with work, walking, and social interactions (see Table 2) and were taken from the Brief Pain Inventory (BPI) [2], the Brief Fatigue Inventory (BFI) [1], and the SF36 [14]. These are widely used instruments, and a number of other instruments have similar items [3, 5, 15]. The recall ratings (RRs) used the original instrument item wording and 0–10 response scales with only the reporting period altered for 1, 3, 7, and 28-day reporting periods. The only exception regarding the response scale was the SF36 item that presents 5 response options. There was no overlap of days for the recall reporting periods except for the 28-day reporting period.

EOD ratings were collected on hand-held computers (Palm Zire 31) that were triggered by the patient shutting down the computer before sleep. The hand-held computer utilized software provided by invivodata, inc. (Pittsburgh, PA). The item wording was identical to the recall ratings, except the item began with the stem, "during the day." EOD ratings were made on a 100-point VAS using the anchors "not at all" and "extremely." The protocol from

Table 1 Demographic and symptom characteristics of the sample ($N = 104$)

Age (mean, SD)	55.4 (11.0)
Female	87%
Race (white)	91%
Married	64%
Education	
Did not finish high school	4%
High school graduate	24%
Some college	29%
College graduate or more	43%
Employed	52%
Disability benefits	25%
Family income	
Less than \$20,000	11%
\$20,000–\$49,999	25%
\$50,000–\$74,999	28%
Greater than \$75,000	36%
Baseline pain ratings (mean, SD) ^a	3.3 (0.8)
None	0%
Very mild	1%
Mild	15%
Moderate	42%
Severe	38%
Very severe	4%
Baseline fatigue (mean, SD) ^b	2.9 (1.0)
None of the time	1%
A little of the time	7%
Some of the time	28%
Most of the time	32%
All of the time	33%

^a Assessed with SF36 v-2 item: “How much bodily pain have you had during the past 4 weeks?”; possible scores 0–5

^b Assessed with SF36 version-2 item: “How much of the time during the past 4 weeks did you feel tired?”; possible scores 0–4

which these data are drawn included momentary ratings of pain and fatigue/energy intensity items as well as recall ratings of these items; these data are reported elsewhere [7].

Design and procedure

Patients began the study at the research laboratory where they completed consent forms, were trained to use the hand-held computer to collect EOD ratings, and were instructed in procedures to complete the RRs across the month. They were given paper copies of the recall questionnaire that contained the interference items and were told that they would be telephoned by an interactive voice recording (IVR) computer system (Prosodie, North

America) in the evening about an hour before bedtime on several occasions during the month. If the IVR system did not reach the patient, another call was made 15 min later, and a final reminder call was made at the end of the hour. The system informed the patient that a recall questionnaire was to be completed that night and, once completed, the patient was to call back the system to record the item responses. This emulated the typical procedure for paper questionnaire assessments while providing time- and date-stamped data capture that precluded concerns about unknown time of completion [16]. If a patient did not enter RR item responses on the night of the IVR call, the patient was called again on the next evening. Patients did not know which days the IVR system would call them. They were randomized to 1 of 10 schedules for the dates of the RRs to counter-balance the sequence of the reporting periods and weekday versus weekend. Two 1-day, two 3-day, one 7-day, and one 28-day RR were collected. None took place in the first 3–6 days of the protocol. The day after training, the research staff telephoned the patient to answer questions and troubleshoot any problems with the hand-held computer, and a follow-up call was made once a week for the next 3 weeks for the same purpose. Patients returned to the research office at the end of the month-long protocol to deliver the hand-held computer and to complete a pain and fatigue assessment that is not reported here.

Analytic strategy

To compare mean levels of EOD averages and RR ratings, repeated measures analysis of variance was used with assessment type (averaged EOD versus RR) and reporting period as within-subject factors. A priori contrasts were used to examine mean level differences between EOD and RR ratings for each of the reporting periods (average mean differences were tested for the two 1-day and the two 3-day periods). Pre-specified contrasts were also used to examine whether the differences between EOD and RR means became more (or less) pronounced when moving from the 3-day to the 7-day and to the 28-day reporting period by testing for a linear trend in the difference between EOD and RR means across reporting periods (i.e., the interaction between type of report and length of reporting period).

To examine the level of correspondence between RR and EOD across reporting periods, EOD ratings were averaged for each reporting period and correlated with the corresponding RR.¹ Pooled average correlations were

¹ In contrast to analyses of the correspondence between aggregated *momentary* reports and RRs in which we have adjusted average momentary reports for unreliability [7], we did not adjust averages of EOD reports for unreliability. We consider a set of obtained momentary reports as a random and incomplete sample of all possible moments of a reporting period. By contrast, the combination

Table 2 Mapping of pain and fatigue interference questionnaire items with corresponding End-of-Day (EOD) items

Source of item	Original questionnaire item and study recall item ^a	Original Questionnaire Scaling ^b	Electronic EOD item	Electronic item 100-point VAS anchors
<i>Fatigue items</i>				
BFI	Circle the one number that describes how, during the past 24 h , fatigue has interfered with your normal work (includes both work outside the home and daily chores)	Does not interfere, completely interferes (11)	During the day, how much did fatigue interfere with your work (job or home)?	Not at all, extremely
BFI	Circle the one number that describes how, during the past 24 h , fatigue has interfered with your walking ability	Does not interfere, completely interferes (11)	During the day, how much did fatigue interfere with your walking ability?	Not at all, extremely
BFI	Circle the one number that describes how, during the past 24 h , fatigue has interfered with your relations with other people	Does not interfere, completely interferes (11)	During the day, how much did fatigue interfere with your relations with other people?	Not at all, extremely
<i>Pain items</i>				
SF36	During the past 4 weeks , how much did <i>pain</i> interfere with your normal work (including both work outside the home and housework)?	Not at all, extremely (5) ^c	During the day, how much did pain interfere with your work (job or home)?	Not at all, extremely
BPI	Circle the one number that describes how, during the past week , pain has interfered with your normal work (includes both work outside the home and daily chores)	Does not interfere, completely interferes (11)	During the day, how much did pain interfere with your walking ability?	Not at all, extremely
BPI	Circle the one number that describes how, during the past week , pain has interfered with your walking ability	Does not interfere, completely interferes (11)	During the day, how much did pain interfere with your relations with other people?	Not at all, extremely
BPI	Circle the one number that describes how, during the past week , pain has interfered with your relations with other people	Does not interfere, completely interferes (11)		

Brief Fatigue Inventory (BFI), SF36 version 2 (SF36v2) Brief Pain Inventory (BPI)

^a Recall item reporting period was varied to indicate 1, 3, 7, or 28 days

^b Numerical scales with endpoint anchors and number of points in parentheses

^c Original scaling changed to 0–10 point scale to be consistent with other items and anchors

estimated for the two 1-day and 3-day reporting periods. To test differences between correlations for significance, rather than comparing all correlation coefficients individually, a structural equation modeling approach (using the CALIS procedure in SAS, version 9.1) was used to examine the null-hypothesis that correlations between RRs and EOD ratings did not differ across the four reporting periods. Specifically, for each item, we compared the goodness of fit chi-square of a model in which all correlations were freely estimated versus a model in which they were constrained to be equal across all reporting periods. Using the same approach, we also examined for each reporting period whether the correlations between RRs and EOD ratings differed between the interference items.

To examine the level of correspondence between a single RR and EOD ratings covering a longer time period, we averaged increasing numbers of EOD ratings—starting with the two sequential EOD ratings taken at and prior to

the day of the second 1-day RR, then adding a third day, a fourth day into the past, up to a 14-day period—and correlated these various EOD averages with the 1-day RR rating. Within the month-long protocol, RRs were scheduled in a counter balanced fashion across participants. Using the second 1-day RR provided at least 14 days of data prior to that RR for this analysis.

Compliance criteria

Testing of the hypotheses requires (1) that EOD and RR data are collected at the specified time and (2) that there are sufficient EOD data for any given reporting period to adequately represent the experience of interference during that time. Compliance criteria were specified for each hypothesis.

Primary analysis: does accuracy of RRs decline as the reporting period becomes longer?

Not all patients completed RRs on the first evening that they were contacted. If they completed a RR on the second

Footnote 1 continued

of EOD reports can be regarded as an exhaustive representation of all possible moments within a reporting period.

night for any given reporting period, it is possible that, by virtue of receiving the call on the first night, they were more observant of their interference experience the second day when they then completed their RR. Thus, we chose to use only RRs that were collected on the first (unexpected) IVR contact night. The compliance criteria for the EOD ratings were that none was missing on any day covered by the 1-, 3-, and 7-day reporting periods, and that a minimum of 26 EOD ratings were obtained during the 28-day reporting period. Since 1-day and 3-day RRs were collected twice during the protocol, meeting the compliance criterion was required for at least one out of the two 1- and 3-day recall periods. Of the 104 patients, 17 were excluded from analysis of this hypothesis because they did not meet one or both of these compliance criteria, resulting in an analysis sample of $n = 87$.

Secondary analysis: how well does a 1-day RR of interference generalize to longer time periods?

The second 1-day RR was used in this analysis, given that it was always obtained later in the protocol, following a minimum of 14 preceding EOD ratings. As in Hypothesis 1, only RRs made on the targeted date were used. For the EOD ratings, the compliance criterion was that no EOD report was missing on any day during the period of 2 weeks preceding the 1-day RR (14 days, including the day of the RR). Out of the 104 patients, 16 did not meet one or both of these criteria and were excluded, resulting in an analysis sample of $n = 88$.

Results

Mean level differences between RRs and aggregated EOD

The mean RRs and aggregated EOD ratings ($N = 87$) for corresponding reporting periods are displayed in Fig. 1 for the pain and fatigue interference items. As expected, the 1-day RR and EOD ratings were similar. No significant mean differences on 1-day recall were found for pain and fatigue interference with walking and for pain interference with social relations ($P > .20$). However, the 1-day RR ratings exceeded EOD ratings by approximately 3 points on a 100-point scale for pain and fatigue interference with work and for fatigue interference with social relations ($P < .01$), suggesting the possible influence of a measurement method effect.

Consistent with our hypothesis, for all items the 3-, 7- and 28-day RRs were significantly higher than the aggregated EOD rating for those reporting periods ($P < .01$ in each case). Although this pattern of results was observed across the items, the degree of differences varied by area of

functioning. RRs were approximately 10 points higher for interference with walking (pain and fatigue), 8 points higher for interference with work (pain and fatigue), and 3 points higher for interference with social relations (pain and fatigue). However, contrary to the hypothesis that the difference would increase across longer reporting periods, these mean differences remained relatively constant across the longer reporting periods (for all items and all slopes, $P > .15$).

Correspondence between RRs and aggregated EOD ratings across different reporting periods

The correlations of 1-, 3-, 7-, and 28-day RRs with the average of EOD reports ($N = 87$) for the corresponding periods are shown in Fig. 2. The correlations are generally high, in all cases exceeding .80. Contrary to the hypothesis, tests for differences among the correlations found no significant differences within items across reporting periods for the pain items, and for 2 of the 3 fatigue items. Only the fatigue interference with walking item showed evidence of a reporting period effect with the 28-day recall being significantly higher than the 3-day and 7-day reporting periods (P 's $< .001$)—a finding in the opposite direction of the hypothesis. To determine if correspondence was better for some items, tests for differences across items and within reporting period found 7 of 8 comparisons significant (P 's $< .05$); only the fatigue interference item was not significant. These differences were largely being driven by the interference with work items (both pain and fatigue) showing a lower correspondence than the interference with walking items (4 of 8 comparisons, $P < .01$). Thus, the hypothesis that increasing reporting periods would yield less accurate RRs was not supported.

How well does a single 1-day recall rating represent average interference during the preceding 2 weeks?

If interference is assessed for a single day, it may be the case that the rating is generalizable to more than just the last 24 h. To address this question, we aggregated increasing numbers of EOD ratings across the prior 2 weeks—starting with the day of the RR, then adding the prior day, then adding the two prior days and so on up to the 13 days prior to the 1-day RR. These EOD aggregates were correlated with the 1-day RR ($N = 88$). The correlations of the pain interference items range from .82 to .95, indicating that a 1-day RR provides a very good to extremely good representation of interference due to pain across the past 2 weeks. Indeed, the correlation of 14 aggregated EOD ratings with the 1-day RR was .89 for interference with walking, .85 for social activities, and .84 for working. Likewise, for the items measuring interference due to

Fig. 1 Mean end-of-day ratings and recall ratings of pain interference (left panel) and fatigue interference (right panel) with **a** walking, **b** work, and **c** social activities. Error bars represent standard errors

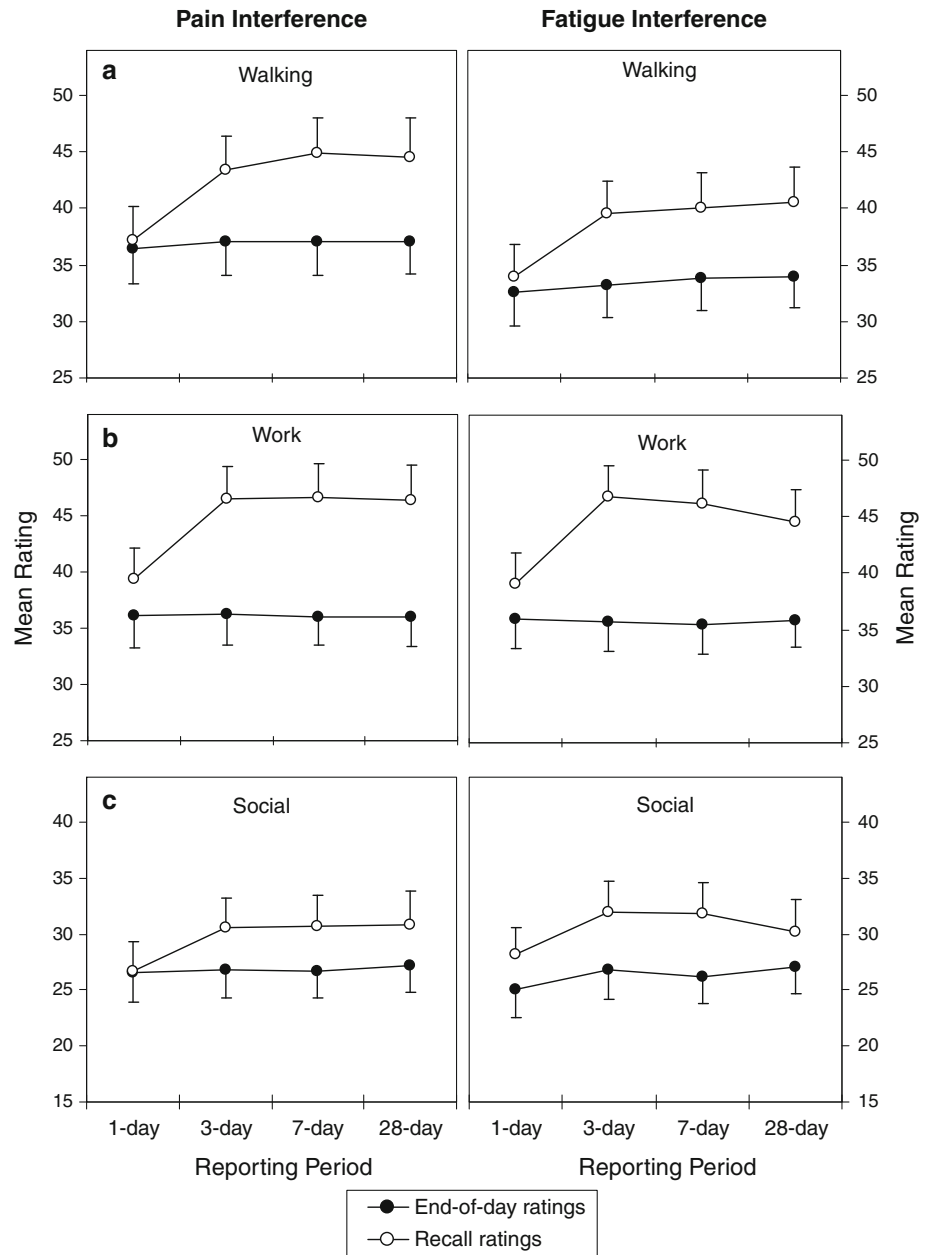
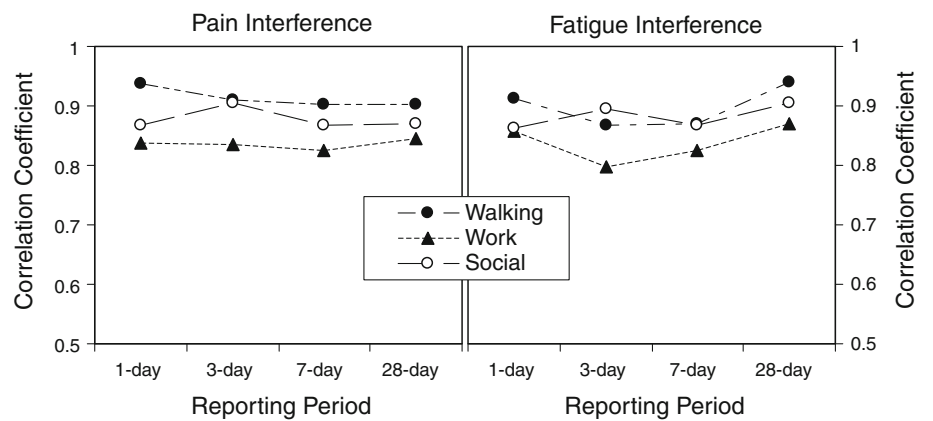


Fig. 2 Correlations of recall ratings with the average of end-of-day ratings for pain interference and fatigue interference items



fatigue, a 1-day RR is a good representation of the prior 2 weeks with correlations ranging from .76 to .92. Correlations with the 14-day aggregates were .86 for walking, .78 for social activities and .76 for working.

Discussion

This study was designed to determine if recall ratings of health-related items measuring interference with daily activities are biased for longer reporting periods when the average of daily experience with interference is the construct of interest. That is, are measures based on shorter recall periods more accurate than those with recall periods of 7-days or a month? It should be noted that there is much discussion in the field about the possibility of unique contributions to recall ratings only accessible to respondents over longer reporting periods. This proposition will require empirical evidence. For the immediate purpose of determining if recall ratings of interference reflect the average or usual experience of daily experience of patients across different reporting periods, the daily ratings are used as the standard.

In what is becoming an often replicated finding, items with longer recall periods yielded higher symptom ratings than items with shorter recall periods [10, 12, 13, 17]. When the study participants rated interference for the past 3, 7, or 28 days, those ratings were significantly higher than the aggregated EOD ratings for those same days. A small component of this effect may be the method of measurement (IVR or hand-held computer) as evidenced by the 1-day RRs that were higher by 3 points for IVR compared with the EOD computer ratings for 3 of the 6 interference items. We are not sure why levels and correspondence statistics for the two 1-day reports were not higher. We suspect it may be due to response option differences (Numeric Rating Scale, in most cases, versus Visual Analog Scale), the slight difference in timing of item completion (IVR completed slightly earlier in the day), or to the mode of administration (IVR versus electronic diary). These possibilities should probably be considered in the design of future studies of this type.

The second part of the primary hypothesis—that as the reporting period lengthened, RRs would correlate more poorly with aggregated EOD ratings—was not supported. Recall ratings were moderately to highly correlated with interference measured daily, irrespective of the length of recall. Thus, people with chronic illness can provide reports about the impact of their disease with a good degree of accuracy across periods of time when actual, detailed memory is probably not available [7]. This suggests that what is being reported is probably a general view of their

disability that is, in fact, based upon their experiences over time.

There are several implications of these findings for clinical research. When all assessments in a given trial use the same recall period, the level differences associated with different recall periods would not be of concern. The only potential implication of level differences is that symptom ratings in trials using different recall periods are not comparable, especially if one of the studies uses 1-day recall (which should have relatively lower symptom levels than studies with longer recall periods). Regarding the correspondence findings, our conclusion is that there is no differential impact of recall period (at least through a 28-day period) on measurement validity. Overall, these results may be viewed as “good news” for those wishing to employ assessments with longer recall periods, at least for interference-type measures.

The secondary analysis focused on the ability of a rating of interference on a single day to represent the interference experienced across the two-week period prior to the rating. The high level of correspondence observed suggests that for this population of patients with a chronic rheumatological illness, a single end-of-day rating is a good proxy for the experience over a longer proximal time period. This is consistent with our supposition that these types of items are tapping a general, stable state.

One important limitation of this study is that it is observational and conducted with a patient sample whose symptoms were unlikely to change substantially across the 1 month reporting period. This can be contrasted with clinical trials where symptom levels of pain and fatigue are changing due to treatment. It is possible that the accuracy of RRs of interference due to pain and fatigue would not be as predictable and, thus, not as high in such situations; and a single EOD assessment, in particular, would be less likely to correlate as highly as in this study with interference experienced in the recent days and weeks. A second issue that is inherent in this type of study design is that patients were completing both daily and recall ratings across the same reporting periods. This “double measurement” would rarely be done in typical clinical or research settings. It is possible that by virtue of making EOD ratings, the subsequent recall ratings were influenced in ways that would not be evident in the absence of EOD ratings. The likely effect would be to create greater correspondence between the measures.

In sum, accuracy of recall ratings was examined in two ways in this study: level differences and correlations. Reporting period affected the *level* of the symptom reported, but did not influence the *correspondence* with the standard used for accuracy (aggregated EOD ratings). Taken together, these results suggest that researchers can have confidence in the accuracy of RRs of interference due

to pain or fatigue without much regard for the reporting period when conducting between-subject analyses. In that case, the level issue would be a constant across subjects and measurement points, and therefore not be an issue. Furthermore, a rating for a single day was highly representative of a proximal period of at least 2 weeks. The question of the impact on interference items of reporting period under conditions of systematic change due to treatment or other factors awaits further research.

Acknowledgments 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 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). JEB and AAS have a financial interest in invivodata, inc. and AAS is a senior scientist for the Gallup Organization.

References

- Mendoza, T. R., Wang, X. S., Cleeland, C. S., Morrissey, M., Johnson, B. A., Wendt, J. K., et al. (1999). The rapid assessment of fatigue severity in cancer patients: Use of the Brief Fatigue Inventory. *Cancer*, *85*(5), 1186–1196.
- Daut, R. L., & Cleeland, C. S. (1982). The prevalence and severity of pain in cancer. *Cancer*, *50*(9), 1913–1918.
- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, *85*(5), 365–376.
- Ware, J. E. (1993). *SF-36 health survey: Manual & interpretive guide*. Boston: The Health Institute, New England Medical Center.
- Kerns, R., Turk, D., & Rudy, T. (1985). The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain*, *23*, 345–356.
- Salovey, P., Smith, A., Turk, D. C., Jobe, J. B., & Willis, G. B. (1993). The accuracy of memories for pain: Not so bad most of the time. *American Pain Society Journal*, *2*, 184–191.
- Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *Pain*, *139*, 146–157.
- 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 arthritis? *Journal of Pain*, *1*(3), 212–217.
- Fredrickson, B. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, *14*(4), 577–606.
- Keller, S. D., Bayliss, M. S., Ware, J. E., Jr., Hsu, M. A., 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.
- Stone, A. A., Broderick, J. E., Schwartz, J. E., & Schwarz, N. (2008). Context effects in survey ratings of health, symptoms, and satisfaction. *Medical Care*, *46*(7), 662–667.
- Stone, A., Broderick, J., Shiffman, S., & Schwartz, J. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute accuracy, between- and within-person consistency, and judged change in weekly pain. *Pain*, *107*, 61–69.
- Williams, D. A., Gendreau, M., Hufford, M. R., Groner, K., Gracely, R. H., & Clauw, D. J. (2004). Pain assessment in patients with fibromyalgia syndrome: A consideration of methods for clinical trials. *Clinical Journal of Pain*, *20*(5), 348–356.
- Ware, J. E., Kosinski, M., & Dewey, J. E. (2000). *How to Score Version 2 of the SF-36 Health Survey*. Lincoln, RI: QualityMetric Incorporated.
- Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol*, *46*(10), 1121–1123.
- Stone, A., Shiffman, S., Schwartz, J., Broderick, J., & Hufford, M. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, *324*(7347), 1193–1194.
- van den Brink, M., Bandell-Hoekstra, E. N., & Abu-Saad, H. H. (2001). The occurrence of recall bias in pediatric headache: A comparison of questionnaire and diary data. *Headache*, *41*(1), 11–20.