Contemporary Clinical Trials 31 (2010) 483–490

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Validity of average, minimum, and maximum end-of-day recall assessments of pain and fatigue

Arthur A. Stone *, Joan E. Broderick, Joseph E. Schwartz

Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY, USA

ARTICLE INFO

Article history:
Received 24 February 2010
Accepted 18 June 2010

Keywords:
PRO
Daily diary
Momentary measures
Pain
Fatigue

ABSTRACT

End-of-day (EOD) diary assessments of symptoms have the potential to reduce recall bias associated with longer recall periods, and therefore be useful for generating accurate patient reported outcomes (PROs). In this report we examine the relative validity of diary questions about the experience of daily pain and fatigue, including several questions about experience for the entire day and questions about minimum and maximum daily levels, with previously collected data [1]. Validity estimates are based on comparisons of EOD reports with momentary recordings of pain and fatigue from the same days. One hundred and six participants with rheumatologic diseases yielded 2852 days for analysis. Differences in levels as assessed by EOD and momentary reports were small (just a few points), although in many instances were significantly different. Correlational analyses indicated that "how much," "how intense," and "on average" EOD questions were more strongly associated with momentary reports ($rs = 0.85–0.90$ for pain and 0.81–0.83 for fatigue) than were minimum and maximum questions ($rs = 0.73–0.80$ for pain and 0.67–0.75 for fatigue). Overall, the pain measures had higher EOD-momentary correspondence than the fatigue measures. Analyses of difference scores between EOD and momentary reports confirmed the better correspondence of the average questions compared with minimum and maximum questions. There was little evidence of individual differences in level and correspondence analyses. The implication of these results is that over-the-day diary measures may yield superior PROs than those based on minimum or maximum daily levels.

Patient Reported Outcomes (PROs) are patients' self-reports of their symptoms, the impacts of their symptoms, and their behaviors. PROs have received considerable attention because they provide a unique perspective on patients' health and functioning [2]. One problem with self-report measures is the length of the recall period [3,4], the amount of time to be considered when completing an assessment. Long recall periods may stretch the ability of respondents to accurately recall and summarize information, leading to concerns about accuracy of reports [4].

By limiting the duration of recall period, daily diaries should reduce recall bias compared to assessments with longer recall periods (e.g., weeks), and they can be aggregated over days to cover reporting periods typically used by retrospective assessments [5]. Diary questions often ask about the entire day's symptoms, but can also include questions about the day's least or lowest level of a symptom or the day's worst or maximum level. Recently these alternatives were explicitly suggested in the FDA's PRO Guidance document [6]. There are two reasons why least/worst levels may be appealing candidates for assessment: 1) they avoid the potentially difficult cognitive process of summarizing experience and 2) least/worst may be the construct of interest as opposed to average experience over-

* Corresponding author. Department of Psychiatry and Behavioral Science, Putnam Hall, Stony Brook University, Stony Brook, NY 11794-8790 USA. Tel.: +1 631 632 8833; fax: +1 631 632 3165.
E-mail addresses: Arthur.stone@sunysb.edu (A.A. Stone), joan.broderick@stonybrook.edu (J.E. Broderick), Joseph.Schwartz@sunysb.edu (J.E. Schwartz).

1551-7144/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2010.06.004

the-day. For example, one can hypothesize that treatments could reduce maximum pain during a day, yet have only a modest impact on average levels. It is also notable that some weekly recall questionnaires for pain assessment ask about least and worst levels (e.g., BPI; [7]), indicating interest in these constructs.

Some validity data are available for EOD diaries with ratings of the over-the-day experience, and those results are encouraging. One study of post-surgical patients compared EOD recall of daily pain with the average, peak, and last-of-day variables based on 5 randomly selected momentary assessments [8]. EOD recall of pain intensity correlated about 0.70 with the average of momentary reports and only 4% recall bias from peak and end pain was found. Second, previous results from a subset of the current dataset showed good correspondence between EOD diaries and momentary reports for pain and fatigue measures [1]: correlations ranged from 0.75 to 0.85.

To our knowledge, this is the first report to compare the validity of EOD recalled over-the-day, least, and worst pain and fatigue diary questions with multiple momentary assessments from the same day. For EOD questions of average and "how much" pain/fatigue, we use the average of moments for the same day as the validity criterion; for the EOD measures of least pain/fatigue, we use the minimum value of the day's momentary reports; and, for EOD measures of worst pain/fatigue, we use the maximum value of the day's momentary reports. Evidence for the validity of EOD measures would be 1) that their levels are similar to the corresponding momentary measure and 2) that the correspondence over days between EOD and the momentary measure was high.

1. Methods

1.1. Participants

Patients were recruited from two offices of a community rheumatology practice. Participants were required to be available for 30 consecutive days and to meet the following eligibility criteria: ≥18 years of age; physician-confirmed diagnosis of a chronic rheumatological illness; experienced symptoms of pain or fatigue during the last week; no significant sight, hearing, or writing impairment; fluency in English; normal sleep–wake schedule; ability to come to the research office twice within a month; had not participated in another electronic diary study in the last 5 years. A total of 279 patients were telephone screened, and 86 (31%) were excluded due to one or more of the above eligibility criteria. Of the 193 eligible patients, 76 (39%) declined participation, and 117 (61%) participated. We examined the demographic characteristics of those who were eligible and participated versus those who were eligible and declined participation. Age, sex, educational achievement, marital status, race, and reported pain and fatigue at screening were examined by participation status. A near-significant difference was found for age where those who participated (56.3 years) were older than those who declined participation (52.8 years; $t(191) = 1.94$, $p = 0.053$); none of the other comparisons were significant. Over the course of the study eleven participants dropped out, and 106 completed the study. The final sample was middle-aged (X = 55.5 years), predominantly female (91%), white (92%), married (65%), and well-educated (63% had at least some college).

1.2. Procedure

The study protocol was approved by the Stony Brook University Institutional Review Board. Participants provided informed consent and were compensated $100. Data were collected from September 2005 through June 2006. Eligible patients came to the research office to complete demographic and questionnaire measures and to be trained in the use of an electronic diary (ED). Momentary and daily recall ratings of pain and fatigue intensity were collected for 29–31 days on a hand-held computer (Palm Zire 31). The ED utilized a software program provided by invivodata, inc. (Pittsburgh, PA) that featured auditory tones to signal the participant to complete a set of momentary ratings. It was programmed to generate an average of 7 randomly-scheduled (within intervals) prompts spread across the participant's waking hours (an average of one every 2 h and 20 min, constrained to ensure a minimum of 30 min 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, the ED prompted the participant to complete a daily recall assessment at the time the ED was put to sleep at night, the "End-of-Day" assessment. A research assistant telephoned the patient 24 h after the initial research office visit to answer any questions and troubleshoot potential problems with using the ED. A follow-up call was made once per week for the following 3 weeks to ensure the ED was working properly and to answer any questions. At the end of the month, patients returned the ED to the research office.

1.3. Measures

Items for this study were drawn from the Brief Pain Inventory (BPI) [9] and the Brief Fatigue Inventory (BFI) [10], with wordings modified to correspond to the different reporting periods. Zero to 100-point Visual Analog Scales were used, but scale endpoints varied according to question content. For the "how much" bodily pain question the anchors were "none" (0) and "very severe" (100), whereas for all other questions the anchors were "not at all" (0) and "extremely" (100). The EOD questionnaire contained several questions that were used to address the aims of this paper. Three asked about over-the-day levels of pain: How much bodily pain did you have?, How intense was your bodily pain?, and What was your average level of pain today? Another two questions asked about the lowest (What was the lowest level of your pain today?) and highest (What was the worst level of your pain today?) levels of pain for the day. A parallel set of questions was available for the construct of fatigue/tiredness: How fatigued (weary, tired) did you feel? and How tired did you feel? There were also questions about the lowest (What was the lowest level of your fatigue today?) and highest (What was the worst level of your fatigue today?) levels of fatigue for the day. Each of these EOD questions began with the stem "DURING THE DAY." These questions were also asked on a momentary basis. Each of these included the stem "BEFORE PROMPT." From each of these four momentary questions, the average, the minimum, and the maximum were derived.