

The Validity and Reliability of Pain Measures
for use in Clinical Trials in Adults:
Review Paper Written for the Initiative on Methods, Measurement, and Pain Assessment
in Clinical Trials (IMMPACT) meeting, April 12-13, 2003

Mark P. Jensen, Ph.D.^{1,2}

¹Department of Rehabilitation Medicine, Box 356490, University of Washington School of Medicine,
Seattle, Washington 98195-6490 (U.S.A.)

²Multidisciplinary Pain Center, Box 356044, University of Washington Medical Center, 1959 Northeast
Pacific, Seattle, Washington 98195-6044 (U.S.A.)

Correspondence to: M. P. Jensen, Tel.: 206-543-3185; fax: 206-685-3244; email:

mjensen@u.washington.edu; address: Department of Rehabilitation Medicine, Box 356490,
University of Washington School of Medicine, Seattle, Washington 98195-6490 (U.S.A.)

Abstract

In order to be most useful, clinical trials of pain treatments should use measures that are both reliable and valid. The purpose of this paper is to summarize the evidence concerning the validity and reliability of pain measures that can be used in clinical trials of pain treatments. The results of this review indicate that commonly used single-item ratings of pain intensity are all valid and adequately reliable as measures of pain intensity, although some scales appear to be easier for patients to understand and use than others. There is less research examining the psychometric properties of measures of pain dimensions other than pain intensity, such as pain relief, the temporal aspects of pain, and pain quality (including pain affect). This lack of evidence limits the conclusions that may be drawn concerning the reliability and validity of these other measures for use in clinical trials. The discussion includes specific recommendations for selecting from among the available pain measures, as well as recommendations for future research into the pain assessment for clinical trials.

1. Introduction

Clinical trials of pain treatment are essential for identifying and estimating the effectiveness of interventions that might provide pain relief. In order for the results of such trials to be deemed valid, the pain measures used should have proven reliability and validity. This paper is one of a series of papers written for the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) meeting scheduled for April 12-13, 2003, in preparation of a consensus statement concerning recommendations for measures that should be used in clinical trials of pain treatment.

This paper will focus on measures of pain, including those that assess the dimensions of pain intensity, pain relief, the temporal aspects of pain (including breakthrough pain), and pain quality. Other related dimensions, specifically pain behavior, pain interference, and composite pain measures that combine pain intensity ratings with other dimensions, will not be reviewed in this paper. Pain interference and composite measures that include pain intensity will not be reviewed here because they are covered in other papers written for IMMPACT. In addition, measures of pain behavior will not be covered in this review (except for the specific behavior of a request for a rescue dose, which is sometimes used as an outcome measure in clinical trials), because: (1) pain behaviors are behaviors that communicate pain to others and so represent “pain” from the view of an observer, not the patient; (2) measures of pain behavior may, or may not, reflect a person’s pain experience, depending on the patient and the setting; (3) measures of pain behavior tend to be used in specific pain populations (e.g., children, patients with chronic pain), and tend not to be used in other pain populations in clinical trials; and (4) the literature on measures of pain behavior is extensive, and therefore deserves a separate IMMPACT paper if this domain is to be considered as one of the domains to be assessed in pain clinical trials. Other measures that are not covered in this review include those developed for specific pain diagnoses or illness conditions (e.g., the Neuropathic Pain Scale, some scales of the European Organization for

Research & Treatment of Cancer (EORTC) quality of life measure) because these measures have yet to be validated for use across pain conditions.

In the first section of the paper, and for each pain dimension covered, the review includes a brief description of the measure, a review of the available evidence concerning the validity (in particular, its validity as a measure of change in pain as the result of treatment) and reliability of the measure, and a summary of the strengths and weaknesses of the measure. The second section discusses specific recommendations concerning the assessment of each outcome dimension, based on the results of the review presented in the first section. The third section discusses some specific issues concerning how pain measures can, and should, be used in clinical trials, including: (1) How frequently pain should be measured; (2) The validity of using retrospective (recall) measures of pain intensity; (3) Use of paper-and-pencil pain diaries versus other strategies for assessing pain over time in clinical trials; (4) The use of composite scores versus individual ratings in outcome research; (5) The use of rescue dose requests during a clinical trial as an outcome measure; (6) Whether (and how) to address the issue of multiple pain sites in clinical trials; and (7) Whether it is appropriate at this time to develop a consensus concerning the standardization of pain measures in clinical trials (e.g., to recommend a specific form of VAS, with specific instructions and endpoints, if investigators choose to use a VAS). The final section of this paper discusses a number of unanswered questions concerning pain assessment in clinical trials, and makes recommendations for future research to address these questions.

2. Review of Pain Measures

2.1. Ratings of pain intensity

Single-item ratings of pain intensity are the most commonly used measures in pain research and clinical settings. The three most commonly used pain intensity scales are the Visual Analogue Scale, Numerical Rating Scale, and Verbal Rating Scale. Table 1 summarizes the strengths and weaknesses of

these three types of pain intensity rating scales.

2.1.1. Visual Analogue Scale of pain intensity. A Visual Analogue Scale of pain intensity (VAS-I) consists of a line, usually 100mm long, with each end of the line labeled with descriptors representing the extremes of pain intensity (e.g., ‘no pain,’ ‘extreme pain’). Respondents place a mark on the line that represents his or her pain intensity level, and the distance measured from the ‘no pain’ end to the mark (possible range = 0 – 100mm) is that person’s VAS pain score.

The VAS-I has consistently demonstrated sensitivity to changes in pain associated with treatment or time (e.g., Joyce et al., 1975; Stambaugh and Sarajian, 1981; Seymour, 1982; Turner, 1982; Anderson et al., 1991; Sandouk et al., 1991; Moore et al., 1994; Ingham et al., 1996; Tannock et al., 1996; Talmi et al., 1997; Holland et al., 1998; Frost et al., 2000; Manfredi et al., 2000; Mercandante et al., 2000; Zeppetella, 2000; Lundeberg et al., 2001; Rice et al., 2001; Barton et al., 2002; Jensen et al., 2002; Steiner et al., 2003), and usually shows strong associations with other pain intensity ratings (e.g., Kremer et al., 1981; Seymour, 1982; Walsh and Leber, 1983; Ahles et al., 1984; Ekblom and Hansson, 1988; Littman et al., 1985; Jensen et al., 1986; Wilkie et al., 1990; Gaston-Johansson et al., 1992; Grossman et al., 1992; Soh and Ang, 1992; Paice and Cohen, 1997; Sze et al., 1998; Ramer et al., 1999; Chang, Hwang, and Feuerman, 2000; Klepstad et al., 2000; Freeman et al., 2001; Good, 2001; Singer et al., 2001). VAS measures of pain intensity have been shown to be distinct from VAS measures of pain unpleasantness, supporting the discriminative validity of both (Price et al., 1987).

The scores from VAS-Is appear to have the qualities of ratio data for groups of people (Price and Harkins, 1987; Price et al., 1983). This means that differences in pain intensity (for groups, not necessarily for individuals) as measured by VASs represent actual differences in magnitude. For example, a significant change in average pain intensity from 60mm to 30mm on a VAS computed from a group of individuals would indicate that perceived pain intensity was halved in this sample of patients.

Test-retest reliability of the VAS-I has been examined in a number of studies, with time periods ranging from five minutes ($r = .95$; Grossman et al., 1992) to one week ($r = .75$; Chang, Hwang, and Feuerman, 2000; $r = .85$, Fischer et al., 1999). These reliability coefficients are almost always very high (see also Padilla et al., 1983; Hollen et al., 1993; Roach et al., 1997; Bergh et al., 2000; Good et al., 2001; Lundeberg et al., 2001; Gallagher et al., 2002), and only very rarely drop below .70 (e.g., Love et al., 1989 reported VAS-I test-retest reliability coefficients as .77 for current pain but only .49 for “worst” pain and .57 for “best” pain over the course of several days).

One potential strength of VAS-I measures over other measures of pain intensity is the high number of response categories of VASs. Because they are usually measured in millimeters, a 100mm VAS can be considered as having 101 response levels. This high number of response categories makes the VAS potentially more sensitive to changes in pain intensity than measures with fewer numbers of response categories. Of course, there is an upper limit to the number of response categories necessary to fully characterize different levels of perceived pain intensity. For example, laboratory research indicates that people are able to identify, on average, about 21 noticeable differences between weak and intolerable experimental pain (Hardy, Wolff, & Goodell, 1952). Based on this alone, measures with more than 21 response levels are not likely to be any more sensitive than measures that have 21 response levels.

One way to empirically determine whether different measures are more, or less, sensitive to changes in pain is to administer the measures before and after a pain treatment, and determine whether there is a difference between measures in their ability to detect changes in pain. In such research, sensitivity to changes in pain can be operationalized as a statistic that reflects the effect size for detecting a change in pain (e.g., pretreatment to posttreatment) or a difference between treatment and control conditions. Relevant statistics include the t-statistic, the F-statistic, the p-value associated with these

statistics, or some measure of change divided by a measure of variance (e.g., lambda). Larger t- and F-statistics, and smaller p- or lambda values indicate greater sensitivity.

Using this strategy, Wallenstein (1991) performed a reanalysis of 11 RCTs of analgesics for cancer (2 RCTs) and post-operative (9 RCTs) pain. Ten of these studies included both a VAS-I and VRS-I (Verbal Rating Scale of pain intensity, see below) measure of pain. The VAS-I was more sensitive than the VRS-I in six of these studies, and the VRS-I was more sensitive than the VAS-I in the remaining four studies. Littman et al. (1985) similarly performed a reanalysis of 23 RCTs of analgesics for postoperative, cancer, acute trauma, or renal or urethral colic pain (total number of subjects 1,330). They found that three scales (VRS-I, VAS-I, and a VRS of pain relief) were similarly sensitive, although the relief ratings tended to show slightly greater sensitivity than VAS-I difference scores did, and VAS-I difference scores showed slightly greater sensitivity than VRS-4 difference scores did.

Other researchers have also found VASs are slightly more sensitive (Holland et al., 1998; Stockler et al., 1998; Bellamy et al., 1999; Frost et al., 2000; Graff-Radford et al., 2000; Bone et al., 2002) than other measures of pain intensity for detecting changes in pain. However, some studies have shown VAS-Is to be slightly less sensitive (Moore et al., 1994; Jenkinson et al., 1995; Magnusson et al., 1995; Bolton et al., 1998; Jensen et al., 1998) or essentially equivalent in sensitivity (Stambaugh and Sarajian, 1981; Kucuk et al., 2001; Jensen et al., 2002) to other pain measures. In one study, a VAS-I was equivalent in sensitivity to a NRS-I, but both were more sensitive than a 4-point VRS-I (Breivik et al., 2000). In short, studies comparing VAS-Is to other pain intensity measures (in particular VRS-Is and Numerical Rating Scales of pain intensity, or NRS-Is), suggest the possibility that they might be more sensitive to changes in pain than other measures under certain conditions, but they might also be less sensitive under other conditions. More importantly, for the most part, this research shows that when a significant treatment effect is found using one measure, it is almost always also found for other pain

intensity measures.

Although VAS-Is appear to be about as sensitive to changes in pain as other pain intensity measures are, there is evidence that VASs may be more difficult than other pain ratings for patients to understand and complete. For example, Bruera et al. (1991) found that 16% of 101 palliative care patients were unable to complete a VAS-I, even with nurse assistance, and that this number increased to 84% as disease progressed. Littman et al. (1985), who performed the reanalysis of 23 RCTs cited above, also reported on the frequency of missing data in these clinical trials. Of the 167 subjects in these studies who had missing data, 93 (56%) were missing data on all scales (VAS-I, VRS-I, VRS-Relief). However, most of the rest (63, or 44%) were missing data only for the VAS-I.

Kremer et al. (1981) examined the failure rates of and preferences for a VAS-I, a 0 – 100 NRS-I, and a 5-point VRS-I among 50 patients seen at a pain clinic. They found that the VAS had the highest failure rate (11%), and that the failure rates for the 0 – 100 NRS (2%) and VRS (0%) were very low. The mean age of the persons unable to complete the VAS (73.3 years) was significantly higher than those who were able to complete this measure (54.4 years). In this study, the VRS was the scale most preferred (by 59% of the patients), followed by the 0 – 100 NRS (25%); the VAS was least preferred by the patients (16%). Gagliese and Melzack (1997) also found that the failure rate of a VAS-I was much higher among the elderly (60 – 79 years; 30% failure rate) than among middle-aged participants (46 – 59 years, 19%), while young participants had no problem with the measure (27 – 45 years, 0%).

Mostly replicating the findings of Kremer et al. (1981), Paice and Cohen (1997) compared the preference and failure rates of a VAS-I, 0 – 10 NRS-I, and 5-point VRS-I in 50 patients with cancer-related pain. While 10 (20%) of their subjects were unable to complete the VAS, all were able to complete the VRS and NRS. Moreover, mean opioid intake was significantly higher for subjects unable to complete the VAS than for those who were able to complete this measure. They found that half

(50%) of the patients preferred the 0 – 10 NRS, but that many (28%) also preferred the VRS over the other scales. Only six (12%) of the subjects preferred the VAS over the other scales.

Shannon et al. (1995) administered the McGill Pain Questionnaire (MPQ; see below), three VAS scales (of pain intensity, pain relief, and mood), a VRS of pain intensity, and a Face Scale to 63 inpatients with cancer. Again, the VAS scales evidenced the highest failure rate, with 89% able to complete the VRS, 84% the MPQ, 81% the Face Scale, and 75% the VAS scales. Soh and Ang (1992) asked 79 patients with various cancer diagnoses to complete a VAS-I and a VRS-I. Although they did not report specific failure rates, they did comment that the VAS was more difficult to explain to patients than the VRS was. Sze et al. (1998) administered a VAS-I and a 0 –10 NRS-I to 95 patients with various cancer diagnoses. Again, the failure rate for the VAS-I (14%) was higher than for the NRS-I (3%). Stahmer et al (1998) also found the failure rate for a VAS-I (15%) to be much higher than that for a NRS-I (0%) in a sample of hospitalized patients with pain. On the other hand, Tannock et al. (1996) found a 6-point VRS-I and a VAS-I to have similar failure rates (8% and 11%, respectively) in a sample of 136 men with prostate cancer. Clearly, although patients do not always have difficulty using pain rating scales, when they do have with difficulty with pain measures, they tend to have more difficulty with the VAS than with other measures, including NRSs and VRSs.

2.1.2. Numerical Rating Scale of pain intensity. A Numerical Rating Scale of pain intensity (NRS-I) consists of a range of numbers (usually 0 – 10, but sometimes 0 – 20, 0 – 100 or other ranges). Respondents are told that the lowest number represents ‘no pain’ and the highest number represents an extreme level of pain (e.g., ‘pain as intense as you can imagine’). They are asked to write down, circle, or state the single number that best represents their level of pain intensity, and the number they select is their pain intensity score.

Although traditionally NRS-I measures have been used less often than VAS-I or VRS-I (Verbal

Rating Scales of pain intensity, see below) measures, an increasing number of researchers have been using NRS-Is to test for treatment effects in pain clinical trials. The findings of the research that has been performed supports the validity and reliability of NRS scales, and indicates that their psychometric properties are very similar to those of VAS measures. For example, NRS-I scales tend to show strong associations with other pain rating scales (Kremer et al., 1981; Seymour, 1982; Jensen et al., 1986; Ekblom and Hansson, 1988; Jensen et al., 1989; Wilkie et al., 1990; Paice and Cohen, 1997; Sze et al., 1998; Singer et al., 2001). NRS-Is have also shown to be sensitive to changes (increases) in pain associated with treatments often associated with short-term increases in pain (e.g., radiotherapy for cancer, Trotti et al., 1998; physical therapy, Smith et al., 1998), and to decreases associated with pain treatment (e.g., Chesney and Shelton, 1976; Stenn et al., 1979; Keefe et al., 1981; Turner, 1982; Paice and Cohen, 1997; Backonja et al., 1998; Bolton and Wilkinson, 1998; Farrar et al., 1998; Rowbotham et al., 1998; Grond et al., 1999; Holzheimer et al., 1999; Jensen et al., 1999; Leksowski, 2001; Wilkie et al., 2000; Eisenbert et al., 2001; Lundeberg et al., 2001; Meuser et al., 2001; Rice et al., 2001; Palangio et al., 2002).

NRS-Is have demonstrated criterion-related validity through their significant and positive associations with analgesic medication use (Daut et al., 1983), perceived need to contact health care providers (Sandbloom et al., 2001), pain interference (Daut et al., 1983; Owen et al., 2000), dyspnea (Smith et al., 2001), and a number of additional specific symptoms sometimes associated with painful conditions, such as nausea, dry mouth, dyspnea, lack of appetite, fatigue, and constipation (Chang et al., 2000), and negative associations with treatment satisfaction (Lin, 2000) and measures of global quality of life (Wang et al., 1999; Chang et al., 2000; Owen et al., 2000; Poulos et al., 2001; Sandbloom et al., 2001).

Further support for the validity of 0 – 10 NRS-Is comes from Portenoy et al. (1999), who found

that the responses to this scale showed an appropriate dose-response to treatment with oral transmucosal fentanyl citrate. In another study, a 0 – 10 NRS-I completed on one occasion predicted subsequent decreases in functioning among 93 persons with various cancer diagnoses (Dodd et al., 2001).

However, average NRS-I scale scores may not have ratio qualities (Price et al., 1994), especially when compared to VASs, which do appear to have ratio qualities when averaged across a number of people.

De Wit, van Dam, Hanneman et al. (1999) showed that 86% of a sample of 156 patients with various cancer diagnoses were able to complete 2 month's worth of daily diaries that included a 0 – 10 NRS. They found that patient ratings of average pain provided during interviews every two weeks showed strong associations with actual diary averages (r s ranged from .80 to .91), which provides some support for the validity of retrospective ratings of average pain. However, patient retrospective ratings tended to be higher by about 0.5 on the 0 – 10 scale, on average, than their actual average pain intensity was (as calculated from the diaries), calling into question the accuracy of retrospective rating of past pain using 0 – 10 NRSs (see discussion of retrospective ratings in section 4.2., below).

Two studies found that a NRS-I had a very high degree of test-retest stability over a few minute period ($r = .82$, Bergh et al., 2000; $r = .91$, Lundeberg et al., 2001). Another study found that very good stability for NRS-I ratings of worst pain ($r = .93$) and average pain ($r = .78$), but not for current pain ($r = .59$) over about a 2-day period (Daut et al., 1983). The coefficients were much lower (.34, .24, and .22) when the time period was extended to about 91 days (Daut et al., 1983), although a high degree of stability in pain intensity ratings would not necessarily be expected over a 3-month period, since pain can change from one moment to the next.

Farrar et al. (2000) performed a study that provides important information concerning the meaning of change in pain as defined by a 0 – 10 NRS. They operationalized a meaningful change in pain as that level of change that is associated with a patient not requesting a rescue dose as part of a

titration phase of a clinical trial. They found that an absolute change of 2 points (out of 10) and a percent change of 33% in the 0 – 10 NRS showed the optimal sensitivity and specificity for detecting a meaningful change in pain in a sample of 130 patients with various cancer diagnoses. Although it will be important to replicate these findings in additional samples, these data do support the utility of 0 – 10 NRSs in particular, since such guidelines are not yet available for other measures of pain intensity (see below for their findings concerning a 4-point VRS of pain relief).

Consistent with the review of the relative sensitivity of VAS-I measures, research sometimes finds NRS-I measures to be a little more sensitive to changes in pain than other measures (Jensen et al., 1998; Du Pen et al., 1999; Eisenberg et al., 2001), sometimes a little less sensitive (Portenoy et al., 1999), and sometimes to show essentially the same level of sensitivity (Bolton et al., 1998; Ekblom and Hansson, 1988). One study, cited above, found a 11-point NRS-I to be about as sensitive as a VAS-I, and both of these to be more sensitive than a 4-point VRS-I (Breivik et al., 2000). In short, if a treatment has a significant effect on pain, the evidence indicates that the NRS-I (or VAS-I or VRS-I, for that matter) is about as likely as other measures of pain intensity to detect the effect of the treatment.

On the other hand, NRS-Is tend to be preferred over VASs by patients (Williams et al., 2000). Interestingly, among possible ranges of NRS, a 0 – 10 range is most preferred (by 54% of respondents), followed by a 0 – 100 range (16%), followed by 0 – 20 (1%; Williams et al., 2000), in a sample of patients with chronic pain. Twenty-nine percent of the respondents in this study did not have a preference for one pain measure over the others. In another study, there appeared to be a slight preference for a NRS-I over a VRS-I in a sample of English-speaking patients seeking care in an emergency room (59% preferred the NRS-I and 41% preferred the VRS-I; Puntillo and Neighbor, 1997). However, among the Spanish-speaking patients, there was a tendency for the VRS-I (55%) to be preferred over the NRS-I (45%). But these differences in preference rates were not statistically

significant for either sample (Puntillo and Neighbor, 1997).

Cognitive impairment may interfere with the comprehension and use of pain rating scales, although it may impact the use of some scales more than others. Radbruch et al. (2000) administered a Mini Mental Status Examine (MMSE) to 108 patients with advanced cancer in a palliative care unit, and also attempted to administer the Brief Pain Inventory (Cleeland and Ryan, 1994) intensity and interference items (all 0 – 10 NRSs) to these patients. If the patients were unable to complete the BPI items, they were asked to scale the intensity of their pain on a 4-point VRS (none, mild, moderate, severe). If they were unable to use the 4-point VRS, they were simply asked to confirm the presence or absence of pain (i.e., a 2-point VRS-I) along with other symptoms. Radbruch et al. found that only 75% of these patients with advanced cancer were able to complete the 0 – 10 intensity items, and 62% the 0 – 10 interference items. Moreover, the number of missing responses for the BPI intensity items ($r = -.64$) and interference items ($r = -.47$) were both associated significantly with the MMSE score, indicating that a patient's degree of cognitive impairment impacts his or her ability to respond appropriately to 0 – 10 NRS scales. However, many of the patients unable to complete the BPI 0 – 10 NRS items were able to complete a 4-point VRS of pain intensity, and all of the patients, even those who could not rate their pain using a 4-point VRS, were able to report on the presence or absence of pain.

2.1.3. Verbal Rating Scale of pain intensity. Verbal Rating Scales of pain intensity (VRS-I) consist of a list of descriptors or phrases (e.g., 'none,' 'some,' 'moderate,' 'severe') that represent varying degrees of pain intensity. Each word or phrase has a number associated with it (e.g., 'none' = 0, 'severe' = 3). Respondents are asked to select the single word or phrase that best represents his or her pain level, and the respondent's score is the number associated with the word chosen. In the pain literature, the number of descriptors in VRS-Is range can from 4 (e.g., Seymour, 1982) to as many as 15 (e.g., Gracely et al., 1978; and this latter scale would have 16 descriptors if "no pain" were added to it).

Like VAS-Is and NRS-Is, VRS-Is demonstrate sensitivity to changes in pain with treatment (Fox and Melzack, 1976; Rybstein-Blinchik, 1979; Stambaugh and Sarajian, 1981; Tannock et al., 1989; Bergman et al., 1992; Stelian et al., 1992; Bergman, et al., 1994; Murphy et al., 1994; Ellershaw et al., 1995; Ingham et al., 1996; Tannock et al., 1996; Hammerlid et al., 1997; Bolton and Wilkinson, 1998; Farrar et al., 1998; Rogers et al., 1998; Portenoy, Payne, Coluzzi et al., 1999; Molenaar et al., 2001; Doyle et al., 2002; Jensen et al., 2002), and show strong associations with other measures of pain intensity (Kremer et al., 1981; ; Walsh and Leber, 1983; Littman et al., 1985; Jensen et al., 1986; Fishman et al., 1987; Ekblom and Hansson, 1988; Paice and Cohen, 1997; Rogers et al., 1998; Klepstad et al., 2000). Concerning test-retest stability, one study found the VRS-I to be adequately stable over a matter of minutes ($Kappa = .71$, Ellershaw et al., 1995), and a second found the VRS-I to demonstrate relatively low stability ($r = .55$, Sneeuw, Aaronson, Osoba et al., 1997) over a 1-week period.

Although VRSs are usually scored by using a rank method (e.g., scoring “no pain” as 0, “mild pain as 1, “moderate pain” as 2, etc.), this scoring method has been criticized because it assumes equal intervals between the intensity descriptors, even though it is extremely unlikely that equal perceptual intervals exist. This characteristic of rank-scoring procedures can pose several problems with one is interpreting VRS data. For example, rank scoring does not allow for adequate interpretations of the magnitude of any differences found. For example, a change from 3 to 2 (on a 4-point VRS) might represent a 10% decrease in perceived pain or a 50% change, depending on the perceived interval represented by the words on the list. In addition, some investigators have raised the objection that ranked data should not be analyzed with the more common (and usually more powerful) parametric statistics. However, it has become increasingly recognized that most parametric techniques (such as analysis of variance and the t-test) are still valid when used with data that do not necessarily represent

equal-interval values, especially if the number of categories on the scale is five or more (Cicchetti et al., 1985; Philip, 1990; Rasmussen, 1989; see also Baker et al., 1966).

Cross-modality matching procedures have been used as a means of transforming VRS ratings to scale scores that are more likely to have ratio properties, that is, to scores with equivalent intervals (Gracely et al., 1978a, 1978b; Price et al., 2001). The matching procedure involves asking each patient to indicate the severity that each word represents in reference to one or more other modalities (such as the loudness of a tone, the length of a line, or handgrip force). The rating that the patient gives to a particular word (or the average of several, if the patient rates each word more than once) is then used as the score for that word for that patient. Because the modalities used by patients to match pain descriptors to can themselves be indexed using ratio scales, the numbers or scores derived from such a procedure are believed likely to have ratio properties and to reflect actual perceived differences in magnitudes.

There are two major limitations of cross-modality matching procedures. First, the procedure is time consuming and can be tedious, both of which can adversely affect patient compliance (Ahles et al., 1984). One way to address this problem is to assign standardized scores for each word based on data from groups of previously tested individuals (see Gracely et al., 1978a; Tursky et al., 1982; Urban et al., 1984 for standardized scores for specific words). Second, most of the standardized scores have been developed using non-patients in response to experimental pain. There is evidence that chronic pain patients may rate the intensity of pain words differently than do acute (i.e., postoperative) pain patients (Wallenstein et al., 1980). Even within diagnostic subgroups, the score given to a word by one patient has been shown to vary from that given by other patients, indicating that standardized scores for VRS adjectives may be less reliable than originally hoped (Urban et al., 1984).

Moreover, VRS-I scores obtained through cross-modality procedures may correlate so highly with those obtained by using the ranking method that they contain essentially the same degree of useful information (Hall, 1981; Levine and De Simone, 1991). Similarly, VRS-I scores created by either of the two methods show the same patterns of associations to other pain measures, again suggesting that the information contained in the scores derived from the two methods are comparable (Jensen et al., 1989). Therefore, until strong evidence emerges to support the validity and utility of cross-modality matching procedures or scores over ranked scores, it probably makes sense to use the simpler ranking method be used when relationships between pain intensity and other factors are examined. The more sophisticated cross-modality matching procedures should be used only when ratio-like scaling is needed (i.e., when one needs to know the specific magnitude of differences in pain ratings across time or between groups).

The strengths of VRS-Is include the ease with which they can be administered and scored, provided that scores are calculated using the ranking method or from data developed from previous cross-modality matching experiments. Because they are generally easy to comprehend, compliance rates for VRS-Is are as quite good and often better than those for other measures of pain intensity (Jensen et al., 1986; Jensen et al., 1989). For example, in the Radbruch et al. (2000) study, cited above, many of the patients with advanced cancer in a palliative care unit who were unable to complete 0 – 10 NRSs were able to complete a simple 4-point VRS. VRS-Is are also often preferred as much (Puntillo and Neighbor, 1997) or even more than NRS-Is (Kremer et al., 1981).

One weakness of VRSs is that patients need to read over, or be familiar with, the entire list of pain adjectives before they can select the one that most closely describes their pain. For longer lists (e.g., 15 or more items), this requirement can make the task time consuming, and the clinician or researcher cannot be assured that the patient or subject adequately reviewed the entire list of adjectives. Also, because VRS-Is require patients to select from a finite number of descriptors, patients may be

unable to find one that accurately describes their perceived pain intensity (Joyce et al., 1975). Among illiterate patients, VRSs are less reliable than other pain intensity measures (Ferraz et al., 1990). When differences are found in the relative sensitivity of scales for detecting treatment effects, VRS-Is tend not to fare as well as VAS-Is or NRS-Is (e.g., Breivik et al., 2000). Finally, a clinician or researcher using a Verbal Rating Scale must select a scoring procedure; and, as already discussed, each scoring method has its drawbacks. Possibly because of the relative weaknesses of VRSs, and the availability of other measures of pain intensity, VRSs are being used less often as primary outcome measures than they have previously in pain treatment outcome research.

2.1.4. Other single-item measures of pain intensity. Single-item measures other than VAS-Is, NRS-Is, and VRS-Is are used much less often to assess pain intensity in pain research. Measures that have been used include Mechanical Visual Analogue Scales, Graphic Rating Scales, Faces Scales, and various combination scales.

A Mechanical Visual Analogue Scale of pain intensity (M-VAS-I) is very similar to the VAS-I, except that instead of making a pencil or pen mark on a line on a paper, the respondent moves a slider between the two extremes of pain on a plastic or cardboard scale. The scale administrator then looks on the back of the scale and directly reads the distance that the slider was moved from a ruler. M-VAS-Is are very strongly associated with VAS-Is ($r = .99$, Grossman et al., 1992; $r = .77$, Ramer et al., 1999) and other pain intensity ratings (Geddes et al., 1990; Ramer et al., 1999). They are also highly reliable over a 5-minute period ($r = .95$, Grossman et al., 1992). In short, they appear to share many of the properties of VAS-Is.

Graphic Rating Scales of pain intensity (GRS-I) are also similar to VAS-Is. The primary difference is that GRS-Is add specific markers along the VAS continuum with labels associated with each marker. For example, the GRS-I used by Greenwald et al. (1987) consisted of a 100mm line with

the numbers 1 through 5 evenly spaced along the line, and descriptors ('no pain,' 'slight pain,' 'moderate pain,' 'very bad pain,' 'pain as bad as can be') below each number. Depending on the specific instructions, respondents to GRSs might circle the number or descriptor or make a mark on the line (using the numbers or descriptors as guidelines) that best represents their pain intensity. McMillan et al. (1988) showed that a 0 – 10 GRS was sensitive to decreases in pain that occurred when a pain monitoring system in an inpatient cancer treatment center was established. Ekblom and Hansson (1988) found that a GRS-I showed a similar sensitivity to change in pain with treatment as did a VAS-I and NRS-I.

Face Scales of pain intensity present the respondent with drawings of facial expressions representing increasing levels of pain intensity and suffering. Respondents select the single drawing that best represents their pain level, and their score is the number (rank order) of the expression chosen. Although Face Scales were originally developed, and have been primarily used, for assessing pediatric pain, Face Scales have also been used to assess pain intensity in adults. Evidence supports the ability of Face Scales to detect changes in pain with treatment (Bellamy et al., 1999). Two studies found the Face Scale to show strong associations with a VAS-I in two separate samples of patients with pain (e.g., $r = .82$, Ramer et al., 1999; $r = .92$, Freeman et al., 2001), and Shannon et al. (1995) found that about 81% of their sample with various cancer diagnoses were able to complete the Face Scale (compared with 75% who were able to complete a VAS-I and 89% a VRS-I). These preliminary studies suggest that Face Scales could potentially be valid as measures of pain intensity. However, Ramer et al. (1999) did comment that some of the male patients in their study were uncomfortable with rating their pain at the highest level using the Face Scale because the expression representing the most severe level of pain had tears on the face of the drawing. This raises the possibility that the Face Scale (or at least one that includes tears) may under-estimate pain intensity in some patients with severe pain.

Finally, different components of pain intensity measures can be combined into single scales (e.g., combine numbers with descriptors making a NRS/VRS-I, see Grossman et al., 1992; Campbell et al., 2000; Maunsell et al., 2000; or a diagram with descriptors, see Sneeuw et al., 1999; Sneeuw, Aaronson, Sprangers et al., 1997). The evidence from studies looking at NRS/VRS-Is suggests that they, too, are valid as measures of pain intensity, as shown by their strong associations with other measures of pain intensity (Grossman et al., 1992), association with analgesic use, pain interference, and measures of global quality of life (Maunsell et al., 2000), and association with treatment history and concern about cancer (Campbell et al., 2000).

2.2. Measures of pain relief

Whereas pain intensity ratings ask patients to rate the intensity of felt pain, pain relief ratings ask patients to rate how much ‘relief’ from pain they have experienced, usually in reference to a specific treatment or intervention. Table 2 lists and summarizes the primary findings of this review concerning the assessment of pain relief in clinical trials.

Relief ratings have been shown to be sensitive to the effects of treatment (VAS relief ratings, Wallenstein., 1991; Shannon et al., 1995; Manfredi et al., 2000; VRS relief ratings: Stambaugh and Saragian, 1981; Littman et al., 1985; Wallenstein, 1991; Farrar et al., 1998; Barton et al., 2002; Steiner et al., 2003; 0 – 100% percent relief rating: Hwang et al., 2003). Also, in one study, relief ratings were strongly and negatively associated with pain intensity ratings (VAS relief rating, Fishman et al., 1987). However, in two other studies, the associations between pain relief and pain intensity measures were weak (VAS rating, Ramer et al., 1999; NRS rating, Daut and Cleeland, 1982).

As when pain intensity measures are compared in terms of their sensitivity, pain relief ratings appear to be somewhat more sensitive to changes in pain associated with treatment when compared to pain intensity change scores (e.g., pretreatment - posttreatment) (Ohnhaus and Adler, 1975; Littman et

al., 1985; Wallenstein, 1999; Fischer et al., 1999). However, the differences in sensitivity are rarely large. Also, some studies have found relief ratings to be less sensitive (Hwang et al., 2003), and some about as sensitive (Ekblom and Hansson, 1988; Jensen et al., 2002) as pain intensity change scores. Thus, the use of relief ratings over pain intensity change scores will not usually result in substantial increases in ability to detect treatment effects.

Supporting the validity of relief ratings as indicants of change in pain intensity, some studies have shown positive associations between pain intensity change scores and relief ratings (VAS, Angst et al., 1999; NRS, De Conno et al., 1994). Interestingly, however, the association between pain relief and change in pain intensity is not always strong, so ratings of these two constructs (change in pain, pain relief) appear to measure related but also distinct constructs (Haas et al., 2002). For example, Angst et al. (1999) found that when pain intensity and pain relief were assessed 10-, 20-, and 30-minutes following an infusion (pain intensity was also assessed pre-infusion), pain relief ratings tended to increase as pain intensity decreased. However, for many patients, pain relief ratings remained above 0 (indicating at least some relief) even when pain intensity returned to pre-infusion levels (see also Dalton et al., 1988). Similarly, Jensen et al. (2002) found that the strength of the association between pain relief and change in pain ratings decreased as time since pretreatment increased (see also Feine et al., 1998). This correlation coefficients between pain relief and pain intensity change scores were 0.75 and 0.80 in two samples of post-operative patients 15 minutes following treatment, but dropped to between 0.56 and 0.65 24 hours after treatment. Moreover, a close examination of their data indicated that there was always a subset of patients who reported that they experienced some pain relief even when their pain was higher posttreatment relative to pretreatment levels (Jensen et al., 2002).

Further support for the distinction between pain relief and pain intensity change scores was found by Fischer et al. (1999), who reported that perceived pain relief was more strongly associated with

satisfaction with treatment than was actual change in pain intensity (change in pain intensity ratings). Consistent with this findings, De Wit, van Dam, Abu-Saad et al. (1999) demonstrated the distinction between a VRS rating of pain relief from a measure of pain intensity by performing a factor analysis of pain intensity ratings, a VRS rating of pain relief, and other measures. They found that the pain relief rating loaded with measures of treatment satisfaction and perceived adequacy of analgesia, but not with the pain intensity ratings. In short, the data strongly support the conclusion that perceived pain relief and change in pain intensity are related but also distinct dimensions of pain.

Farrar et al.'s (2000) findings concerning the meaningfulness of change in pain as measured by a change in a 0 – 10 NRS were described above. These investigators also identified the specific rating of relief (using a 5-point VRS-R scale: none, slight, moderate, lots, complete) best associated with a meaningful change in pain. They found that a rating of 'moderate' relief best represented meaningful change to the participants with cancer pain in their study, supporting this rating as a reasonable treatment outcome goal if relief ratings are included as an outcome measure in a clinical trial.

2.3. Measures of the temporal aspects of pain

The temporal aspects of pain include its frequency, variability, and duration, as well as its pattern across time (over minutes, hours, days, or months). Table 3 lists and summarizes the primary findings of this review concerning the assessment of the temporal aspects of pain in clinical trials.

Temporal aspects of pain can be assessed by asking patients to rate their pain on multiple occasions over time in the form of daily diaries. Data from such diaries can be coded to score many of the temporal aspects of pain. Variability can be operationalized as the standard deviation of pain intensity ratings, frequency as the number of times pain intensity is above specific thresholds (e.g., number of times pain intensity is greater than 0, or even greater than some level indicating 'moderate' or 'severe' pain, Serlin et al., 1995), and average duration as the average amount of time patients

experience pain levels above specific cutoffs. Specific time patterns of pain within or across days can also be coded from these data (e.g., no change over time, increases or decreasing over time, Jamison & Brown, 1991). However, not every clinician or researcher has the resources to be able to administer and code diary data, and at least one study calls into question the veracity of diary data when each entry is not observed by the clinician or investigator (or electronically by an electronic diary) (Stone et al., 2002; see more discussion of this issue in section 4.3., below).

One temporal pain dimension is the frequency of pain. Kaasa et al. (1995) used a 5-point VRS to measure the frequency of cancer pain that ranged from “All day” to “Not at all.” They found that responses to this measure were strongly associated with a composite measure of pain intensity and pain interference. Rathmell et al. (1991) asked patients with head or neck cancer to rate the frequency of their pain on 4-point VRS with 1 = “Never” and 4 = “Daily.” Pain frequency, but not pain intensity (also measured by a 4-point VRS-I) was associated with type of treatment received, with patients who received both radiation and surgery reporting greater pain frequency than those who received radiation alone. Samarel et al. (1996) showed that a combination 5-point NRS/VRS of pain frequency (“1. Never” to “5. Always”) loaded with measures of pain intensity and pain upsetness into a single scale. This scale was subsequently found to be significantly associated with other symptoms, such as fatigue, and with treatments received (chemotherapy vs. no treatments). These preliminary findings indicate that pain frequency is both related to, but also might be distinct from, measures of pain intensity.

Temporal dimensions of pain are also included in the McGill Pain Questionnaire (MPQ; Melzack, 1975; see description of the MPQ pain quality descriptors and scales in the next section). On the MPQ, patients are allowed to select one from three temporal categories (i.e., select up to three words total): “brief,” “momentary,” and “transient,” representing the occurrence of brief periods of pain; “rhythmic,” “periodic,” and “intermittent,” representing the occurrence of changing pain; and

“continuous,” “steady,” and “constant,” representing the occurrence of constant pain. However, the validity and reliability of these temporal descriptors have not yet been systematically evaluated.

Another type of pain that is related to its temporal aspect is breakthrough pain. Portenoy and Hagen (1990) defined breakthrough pain as an episode of severe or excruciating pain that occurs in the context of an ongoing background moderate (or less) pain. A series of questions first described by Portenoy and Hagen (1990), and used in subsequent studies (Portenoy et al., 1999; Hwang et al., 2003) identifies the presence/absence of breakthrough pain and, if present, asks about its severity, location, frequency, onset, duration, relationship to fixed analgesic dose, precipitating events, predictability, pathophysiology, and etiology. One study found that the presence of breakthrough pain was associated with other important pain-related variables such as average intensity of background pain, pain interference, and measures of both depression and anxiety (Portenoy, et al., 1999). More recently, Hwang et al. (2003) demonstrated that the frequency of the presence of breakthrough pain was sensitive (i.e., it decreased significantly) to the effects of use of the Agency for Health Care Policy and Research cancer pain management guidelines in a sample of 74 consecutive patients with cancer-related pain.

While measures of pain frequency, variability, and duration assess qualities of the pain that may be associated with the specific diagnosis or cause of pain. Two additional temporal aspects of pain assessed in some outcome studies are time to analgesia onset (since administration of the treatment, usually a medication) and time to meaningful pain relief. While these temporal domains may also be related to the pain diagnosis or underlying cause of pain, they reflect primarily qualities of the specific treatment under consideration, and may be used to compare the speed at which two more or treatments produce analgesia. One method for assessing these qualities is the “stop watch” technique, in which two stop watches are started at the time of analgesic administration, one labeled “time to analgesia onset” and the other labeled “time to meaningful pain relief.” Study participants are instructed to stop the first

watch when (and if) they first notice any analgesic effect, and the second when they notice pain relief that is “meaningful” to them (cf. Barton et al., 2002).

2.4. Measures of the qualitative and affective components of pain

Pain has many sensory and affective qualities in addition to its intensity component. The most common measure of these aspects of pain is the McGill Pain Questionnaire, but the short-form McGill Pain Questionnaire and single-item ratings have also been used. Table 4 lists and summarizes the primary findings of this review concerning the assessment of pain quality (including pain affect) in clinical trials.

2.4.1. McGill Pain Questionnaire (MPQ). The MPQ consists of 78 pain descriptors classified into 20 categories of pain that can be scored to assess four major dimensions of pain: sensory, affective, evaluative, and miscellaneous pain, as well as a total pain severity score (“Pain Rating Index” or “PRI” scores MPQ represent the sum of the ranked values of descriptors selected within each pain dimension and “Number of Words Chosen” or “NWC” MPQ scores represent the total number of words selected within each pain dimension; Melzack, 1975). Data support the conclusion that the MPQ qualitative scale scores assess something other than pain intensity. For example, Chung et al. (2001) found very weak associations between a pain intensity rating and both the total MPQ-NWC ($r = -.09$) and total MPQ-PRI ($r = .00$). Other investigators have found stronger associations between MPQ scale scores and pain intensity ratings (Graham et al., 1980, r s up to 0.40, lowest r -value not specified; Ahles et al., 1984, r s ranged from 0.49 to 0.57; Wilkie et al., 1992; r s up to 0.58, lowest r -value not specified; Zalon, 1999; r s between 0.33 and 0.76). While these associations are usually positive, indicating that the MPQ scales and pain intensity usually assess related dimensions, they are not strong enough to support the conclusion that MPQ scales and pain intensity rating scales assess the same thing.

Further evidence for a distinction between the MPQ scale scores and pain intensity ratings was

found by De Conno et al. (1994). They performed two factor analyses using a VAS-I, a NRS-I, a VRS-I, the MPQ-PRI score, and a composite measure of the frequency of five different qualities of pain obtained at two different points in time in 53 patients with various cancer diagnoses. A single factor emerged from each factor analysis, with the three pain intensity measures loading most strongly on this factor (factor loadings ranged from 0.79 to 0.92), and the MPQ-PRI showing a weak loading in one analysis (0.39; but it showed a stronger loading in the second analysis, 0.72). Similarly, a factor analysis of change scores in these measures from one time point to the next, plus a 5-point rating of pain relief, resulted in a single factor with the pain intensity change scores showing stronger loadings (range = 0.80 to 0.83) and the MPQ-PRI score showing a weaker loading (0.47) on this factor (De Conno et al., 1994).

The MPQ scales have been found to be positively associated with analgesic medication use (Ahles et al., 1983), illness conviction (Dalton et al., 1988), and reported quality of life (Schipper et al., 1984). Also, the MPQ scales have shown expected sensitivity to the effects of pain treatments (Briggs, 1996; Burchiel et al., 1996; Eija et al., 1996; Nikolajsen et al., 1996; Pozehl et al., 1995; Tannock et al., 1989; Tesfaye et al., 1996; Plesh et al., 2000; Naeser et al., 2002), supporting the validity of the MPQ scales as measures of pain. Support for the validity of the MPQ-Affective scale to assess the affective component of pain, specifically, was reported by Ahles et al. (1983), who found that this scale was more strongly associated with measures of psychological distress than with measures of pain intensity. Also, Kremer et al. (1982) reported that cancer patients with low pain intensity report a greater affective component of their pain on the MPQ-Affective scale than patients with low back pain do, consistent with the hypothesis that cancer pain may have a greater affective associations (e.g., be more worrisome and cause more fear) than low back pain.

Responses to the MPQ have also been found to discriminate between different pain diagnoses.

For example, in an early study, Dubuisson and Melzack (1976) found that patients with each of eight types of pain (e.g., menstrual pain, toothache, cancer pain) used different words from the MPQ to describe their pain experience. Subsequent investigators have found that the MPQ scales and/or items can discriminate between patients whose pain can be ascribed to physical causes from patients whose pain had no detectable physical cause (Leavitt and Garron, 1980; Perry et al., 1988, 1991), patients who carried a diagnosis of trigeminal neuralgia from patients who carried a diagnosis of atypical facial pain (Melzack et al., 1986), patients with leg pain caused by diabetic neuropathy from patients with leg pain from other origins (Masson et al., 1989), patients with cluster headache from patients with other (migraine and mixed) headache (Jerome et al., 1988), patients with temporal mandibular joint-related pain from patients with myogenous facial pain (Mongini and Italiano, 2001), and patients with nociceptive from patients with neuropathic pain (Wilke et al., 2001). However, in one study, the overlap in pain description between diagnostic groups was so great that the MPQ descriptors were described as having only limited value as a diagnostic tool among patients with dental pain (Seymour et al., 1983).

Several studies have examined the reliability of the MPQ in patients with different types of pain problems. In studies with patients with cancer pain, studies have found that responses to the MPQ are generally consistent over the time span of several days (Graham et al., 1980; Love et al., 1989; Walsh and Leber, 1983). In a study with patients with low back pain, Love et al. (1989) found adequate test-retest stability for the MPQ scale scores (Total: $r = .83$; Sensory: $r = .76$; Affective: $.78$) over the course of several days. Concerning utility, one study found the MPQ to be difficult for most persons with terminal cancer receiving palliative care to use (Talmi et al., 1997). However, a second study found that 84% of a sample of patients with cancer were able to complete the MPQ (Shannon et al., 1995).

Despite some evidence that the MPQ scales are sensitive to the effects of pain treatments, when differences are found in the sensitivity of MPQ scales compared to the more simple pain intensity ratings, the MPQ scales tend to be less able to detect changes in pain than the intensity ratings are (Jenkinson et al., 1995; Bellamy et al., 1999; Graff-Radford, 2000). Interestingly, however, the failure rate of the MPQ (among the elderly) is higher for the traditional VAS measure (30.4%) than for the MPQ (13%) (although the failure rate for a simple 6-point VRS was even less among the elderly in this sample: 9%)(Gagliese and Melzack, 1997).

2.4.2. Short-Form McGill Pain Questionnaire (SF-MPQ). The SF-MPQ consists of a subset of 15 descriptors from the MPQ drawn from the sensory and affective categories (Melzack, 1987). Responses to the 15 SF-MPQ items can be scored to form a total SF-MPQ score as well as both Sensory and Affective SF-MPQ subscale scores.

While not a great deal of research has been performed with the SF-MPQ, the research that has been performed is promising. The SF-MPQ Sensory, Affective, and Total scores are strongly associated with the original MPQ scales (Dudgeon et al., 1993; Melzack, 1987). Also, preliminary data suggest that the SF-MPQ items, like the items from the original MPQ, may be useful in discriminating patients with different types of pain problems from one another (Melzack, 1987), although they do not appear to be useful for distinguishing between different types or etiologies of spinal cord injury-related pain (Putzke et al., 2002).

In one study, the internal consistency (Cronbach's alpha) of the SF-MPQ items was shown to be excellent in a sample of persons with cancer pain (0.91; Hollen et al., 1994). However, in a sample of post-operative patients, the internal consistency of the total scale score (0.72 for describing current pain and 0.85 for describing pain in the past 24 hours), and especially the scale scales (Sensory: 0.64 for current pain and 0.81 for pain in the past 24 hours; Affective: 0.41 for current pain and 0.63 for pain in

the past 24 hours) was diminished. Also, the two SF-MPQ subscales were strongly associated with one another, suggesting the possibility that the two SF-MPQ scale scores may tap into a similar underlying construct (Hollen et al., 1994).

The SF-MPQ Total and scale scores have also been shown to be sensitive to the effects of pain treatments (Serrao et al., 1992; King et al., 1993; Fowlow et al., 1995; Thomas et al., 1995; Backonja et al., 1998; Rowbotham et al., 1998; Rice et al., 2001), but like the MPQ, the SF-MPQ scales do not appear to be as sensitive to the effects of pain treatments as more traditional single-item pain intensity rating scales (Stelian et al., 1992; Frost et al., 2000; but see Harden et al., 1991 for a study in which the SF-MPQ was about as sensitive as a VAS-I for detecting a treatment effect).

2.4.3. Assessing pain affect with single-item rating scales. The MPQ and SF-MPQ assess a variety of pain qualities, including the affective component of pain. In addition to these multiple-item measures of pain affect, several investigators have advocated the use of single-item rating scales to assess the affective dimension of pain. Assessment of pain affect, or pain unpleasantness, is supported by the evidence that the affective component of pain is conceptually and empirically distinct from pain intensity (Gracely et al., 1978a, 1978b; Jensen et al., 1989; Jensen, Karoly, & Harris, 1991; Melzack and Wall, 1983; Turskey, 1976), although it is important to remember that pain affect is not completely independent from pain intensity (Fernandez and Turk, 1992; Gracely, 1992). Whereas pain intensity may be defined as the magnitude of the pain (how much a person hurts), pain affect may be defined as the emotional arousal, or distress, caused by pain. The most common single-item measures of pain affect have been VASs (VAS-A), VRSs (VRS-A), and NRS-As.

VASs of pain affect are very similar to VASs of pain intensity. Only the end-point descriptors are different. Examples of the extremes used in VAS-affect scales are "not bad at all" and "the most unpleasant feeling possible for me" (Price, Harkins, & Baker, 1987). There is evidence that supports the

validity of VAS affect measures. Studies have shown that they are more sensitive than VAS intensity measures to treatments that should influence pain affect more than pain intensity (Price, Barrell, & Gracely, 1980; Price, 1984; Price et al., 1987). Also, as with VASs of pain intensity, VAS-As appear to have the qualities of ratio scales (Price & Harkins, 1987; Price et al., 1983). VAS-As are also sensitive to treatment effects (Price & Barber, 1987; Price, Harkins, Rafii, & Price, 1986; Price, Von der Gruen, Miller, Rafii, & Price, 1985).

Price et al. (1987) examined the ability of a VAS of pain intensity and pain affect to distinguish between different diagnostic groups. They found that a sample of patients with cancer (and patients with low back pain and causalgia) showed a significantly larger difference between the intensity and unpleasantness ratings than patients with upper back pain, myofascial pain, labor pain, or orofacial pain did. This further supports the distinction between the affective and intensity components of pain, and the ability of the VAS to assess each pain component separately. However, patients may not always be able to distinguish between the sensory and affective components of pain (Turk et al., 1985; Williams et al., 2000), and the association between measures of each pain dimension may be so strong that in many situations they may appear to be measuring the same thing (Turk et al., 1985; Good et al., 2001).

Other weaknesses of VAS-affect measures are likely to be similar to those of VAS-intensity measures. Most of the research using these measures has been conducted with young or middle-aged subjects. The utility of such measures in geriatric populations has not yet been examined; it may be that older people have difficulty with VAS-A scales as they do with VAS-I scales. Because VAS-A measures are single-item scales, they may be less reliable and less valid for examining the full spectrum of affective responses relative to multiple-item measures, such as the Affective subscale of the MPQ or SF-MPQ. Also, there is limited research comparing VAS-A measures to other measures of pain affect. A single experiment suggests that a VAS-A may be less able than a VRS-A (see below) to discriminate

between pain intensity and pain affect (Duncan et al., 1989), perhaps because words are so often used to describe emotional reaction, whereas VASs (and NRSs for that matter) may pull for more of the intensity (magnitude) component of the pain experience.

NRS measures of pain affect (NRS-As) are uncommon in the pain research literature; only two studies that report data concerning the psychometric properties of NRS-As were identified. Spiegel et al. (1983) administered a 0 – 10 NRS of pain intensity and a 0 – 10 NRS of pain suffering to 86 women with breast cancer. They found that the two NRS scales were very strongly associated with one another ($r = .81$). They also found that the NRS of pain affect was significantly associated with measures of maladaptive coping, emotional distress, and use of analgesics. Smith et al. (1998) also administered 0 – 10 scales of pain intensity and pain affect (0 = 'Not unpleasant at all,' 10 = 'As unpleasant as you can imagine'), to 32 patients with various cancer diagnoses, and found that physical therapy increased the intensity rating but not the unpleasantness rating of pain. Such a finding supports the distinction between pain intensity and pain unpleasantness, even though measures of these two dimensions of pain may be strongly associated with one another (Gracely, 1992).

Similar to VRS-Is, affect VRSs (VRS-As) consist of adjectives describing increasing amounts of discomfort and suffering. Respondents select a single word from the list that best describes the degree of unpleasantness of their pain. Like VRS-I measures, VRS-A scales may be scored in three ways: (a) the ranking method, (b) the cross-modality matching method, or (c) the standardized score method (using scores developed from cross-modality matching procedures with a standardization group). The advantages and disadvantages of these methods have already been discussed with respect to VRSs of pain intensity, and the same cautions are offered here. That is, the simpler ranking method is recommended if the investigator wishes to examine the relation between pain intensity and other

constructs, and the use of standardized scores developed from cross-modality matching procedures if the investigator requires a measure more likely to have ratio properties.

Evidence for the validity of VRS-As is mixed. On the positive side, VRS-As appear to be more sensitive than measures of pain intensity to treatments designed to impact the emotional component of pain (Fernandez & Turk, 1994; Gracely, Dubner, & McGrath, 1979; Gracely et al., 1978a; 1978b; Heft, Gracely, & Dubner, 1984). In another study, a VRS-A was only moderately associated with a VAS-I, which itself was strongly associated with a NRS-I (Ahles et al., 1984). This finding provides additional support for a distinction between pain intensity and pain affect, and supports the validity of a VRS-A for assessing pain affect.

On the other hand, several other factor analytic and correlational investigations among patients with chronic pain, patients with postoperative pain, and laboratory volunteers indicate that, like VAS-As, VRS-As are not always distinct from measures of pain intensity (Gaston-Johansson et al., 1992; Jensen et al., 1989; Jensen and Karoly, 1987; Levine and De Simone, 1991). This pattern of overlap between similar measures of pain intensity and affect may have something to do with the relatively low level of reliability of single-item measures. Alternatively, a lack of independence among measures of these two dimensions may reflect the simple fact that they are not completely independent; presumably, some degree of pain intensity is necessary someone to experience pain affect, and pain affect should increase as pain intensity increases (more intense pain is usually more bothersome). Pain intensity and pain affect may be conceptually distinct, but often closely related to one another in the same way that height and weight are distinct by closely associated with each other (Gracely, 1992). Another drawback to VRS-As affect is that they force respondents to choose only one descriptor, even when none of the available descriptors (or more than one of the available descriptors) captures their affective response to pain.

3. Recommendations for Assessing Pain in Clinical Trials

The results of this review summarize evidence concerning the validity and reliability of the most commonly used pain measures. The findings support the multidimensional nature of pain, and provide varying degrees of support for the validity and reliability of measures of pain intensity, pain relief, temporal pain patterns, and pain quality (including affective qualities of pain). The findings also provide guidance for researchers and clinicians concerning which measures may have the most utility, and suggest avenues of future research that will help to clarify the psychometric properties of cancer pain measures. Table 5 presents a list of some specific recommendations concerning the assessment of pain in clinical trials.

3.1. Measuring pain intensity

There are several conclusions that may be drawn from the findings of the research on the psychometric properties of pain intensity measures. First, and most importantly, each of the commonly used ratings of pain intensity, including the VAS-I, the NRS-I, the VRS-I, all appear adequately valid and reliable as measures of pain intensity among the many different samples of persons with pain. Other pain intensity rating scales (e.g., Mechanical Visual Analogue Scales, Graphic Rating Scales) are used less often, but the research that has been performed using these measures generally supports their validity as well. Moreover, no one scale consistently shows greater sensitivity than any other in their ability to detect changes in pain.

While reliability is an important issue for pain intensity measures, as it is for any measure, reliability can be difficult to determine for single-item measures of pain. Internal consistency, one of the most common measures of reliability, cannot be computed from single-item rating scales. Also, test-retest stability coefficients for measures of pain may not always reflect reliability, since pain can, and often does, change from one moment to the next. Such changes in pain can reduce the test-retest

reliability coefficient even for pain measures that are highly reliable. However, as it turns out, when examined, the single-item measures of pain intensity appear to have adequate test-retest stability (often, but not always, greater than 0.80) over short periods of time.

While assessing the reliability of pain measures poses challenges, assessing the validity of pain measures for detecting change associated with treatment is relatively straightforward. Measures that show expected (and statistically significant) decreases following pain treatments known to be effective can be judged to be valid for detecting changes in pain in pain clinical trials. Thankfully, the findings from the studies reviewed support the validity of all commonly used ratings of pain intensity for this purpose.

However, there do appear to be consistent and important differences between VRS-Is, NRS-Is, and VAS-Is in terms of their failure rates and in patient preference. VASs usually show higher failure rates than NRS-Is and VRS-Is, and NRS-Is tend (when differences are found) to show slightly greater failure rates than VRS-Is. Similarly, VRS-Is and NRS-Is tend to be preferred over VAS-Is by patients. Higher failure rates with VAS-Is have been shown to be associated with greater age and amount of opioid intake, and mental impairment has been shown to be associated with inability to complete 0 – 10 NRS ratings of pain intensity. Many patients unable to complete 0 – 10 NRS-Is appear to be able to complete 4-point VRS-Is, however.

As a group, these findings suggest that VAS-I ratings is not be the best choice for assessing pain intensity in clinical trials, especially among patients who are elderly or who may be using opioid medications. NRS-Is, on the other hand, appear to be well tolerated by most patients, and appear to be at least as sensitive and valid as the more traditional VAS-I rating scales. Eleven-point (i.e., 0 – 10) NRS-Is also have the advantage of the existence of data that help clarify the meaning of specific ratings and NRS-I change scores (Farrar et al., 2000; 2001; Serlin et al., 1995; Jensen et al., 2001). However, if

the population is expected to include patients with significant cognitive impairment, a simple 4-point VRS-I (e.g., no, mild, moderate, or severe pain) may be the best choice as the primary outcome measure in a pain clinical trial. Also, even if a 0 – 10 NRS-I is selected as the measure of choice in a particular clinical trial (or even as the primary outcome measure in a trial), investigators should consider including a VRS-I as a secondary measure to help describe the effects of the pain treatment in terms of changes in these descriptors (e.g., the percentages of study participants in each condition who described their pain as decreasing from one level, such as severe, to a lower level, such as moderate), to help ensure fewer failure rates in pain assessment if there are any study participants who have difficulty with the NRS-I measure, and to be able to compare findings across studies that also include a VRS-I as one of the outcome measures.

3.2. Measuring pain relief

On the surface, many clinicians or researchers might assume that a rating of pain relief following a treatment represents, or should represent, the same thing as a pretreatment to posttreatment decrease in pain intensity. If this were true, then asking patients to rate pain relief following a pain treatment could be seen as an alternative to assessing change in pain intensity pretreatment to posttreatment. However, even though pain relief ratings are sensitive to the effects of treatment (and sometimes more sensitive than pain intensity change scores are), pain relief ratings are not always strongly associated with pre-to-posttreatment changes in pain intensity ratings. Moreover, some patients rate themselves as having experienced ‘relief’ even when posttreatment pain returns to, or even becomes higher than, pretreatment levels. These findings suggest that pain relief ratings should not be interpreted to represent the same thing as pretreatment to posttreatment changes in pain (Jensen et al., 2002). Supporting the use of pain relief as a secondary outcome measure, however, there is evidence that perceived pain relief may be more meaningful to patients than actual change in pain (Fischer et al., 1999). Often, but not always,

pain relief measures are more sensitive to the effects of pain treatment than pain intensity change scores are.

Regarding the selection of pain relief rating scales, the available evidence does not clearly support the use of any one type of relief rating (e.g., VRS, NRS, or VAS) over any other. However, practical considerations might suggest that a VRS of pain relief (e.g., ‘no relief,’ ‘slight relief,’ ‘moderate relief,’ ‘lots of relief,’ ‘complete relief’) may help limit the chances that patients will confuse the relief rating with pain intensity ratings, since NRS and VAS pain intensity measures can look very similar to NRS and VAS measures of pain relief.

3.3. Measuring the temporal aspects of pain

Measures of the temporal aspects of pain, including its variability, frequency, and duration, have not received adequate attention in pain research. The available evidence indicates that measures of pain frequency have shown criterion-related validity through their association with pain intensity and interference composite scores, type of treatment received, and pain affect (the level of ‘upsetness’ caused by pain). In at least one of the studies reviewed, pain frequency was associated with the type of treatment received, whereas the pain intensity rating used in the study was not, suggesting that pain frequency and pain intensity can be considered distinct dimensions of cancer pain. Presence and frequency of ‘breakthrough’ pain (periods of excruciating pain in the context of ongoing background pain), another important temporal aspect of pain, was similarly shown to be associated with pain interference, as well as psychological functioning. Frequency of breakthrough pain has also been used in at least one study as an outcome measure.

It is possible, even likely, that temporal aspects of pain such as the frequency and unpredictability of breakthrough pain (or even, alternately, the frequency of pain-free periods), may have an impact on patient functioning over and above any effects of global average pain intensity. It is

also possible that pain treatments that impact such variables may have a greater impact on patient quality of life than treatments that focus exclusively on background or baseline pain might. To test these important hypotheses, valid and reliable measures of the temporal aspects of cancer pain are needed. Unfortunately, although the studies that have been performed indicate that pain frequency and variability can be assessed, there is a paucity of research that evaluates the psychometric properties of measures of the temporal aspects of pain, or that develops additional measures of this important pain dimension that can then be evaluated.

Certainly, the “stop watch” technique, or any other strategy for assessing the time to analgesia onset and meaningful pain relief in response to an analgesic is an appropriate and relatively simple domain to assess in clinical trials of analgesic treatments for acute (e.g., post-operative) or breakthrough pain. Such measures may be particularly important when comparing a new analgesic against an established standard analgesic or intervention for providing quick pain relief.

When appropriate, investigators should strongly consider using (or developing if needed) measures of the temporal aspects of pain as secondary measures in clinical trials. The temporal dimensions that should be particularly considered include the presence/absence, frequency, and intensity of breakthrough pain, and time to analgesia and meaningful pain relief.

3.4. Measuring the qualitative aspects of pain

Pain is known to have qualities in addition to its intensity. It can be experienced as hot, cold, tingly, deep, dull, worrisome, or any one (or more) of many other qualities. Measures of the qualitative and affective components of pain may be used to more fully describe a patient’s pain experience. Such measures could also potentially contribute to improved evaluation and treatment of pain. Given the likelihood that some pain treatments will be found to impact some pain qualities more than others, inclusion of pain quality measures in clinical trials might help determine the specific qualities of pain

that would most benefit from each pain treatment that is evaluated (Galer and Jensen, 1997). Moreover, to the extent that a treatment might impact a relatively few subset of pain qualities, ratings of specific pain qualities may turn out to be more sensitive to the effects of some treatments than ratings of global pain intensity. If so, then systematic use of pain quality measures in clinical analgesia trials may help identify effective treatments that might otherwise have been determined to have little effect on pain.

The MPQ, described above, is the measure most often used to assess the qualitative aspects of pain in pain research. Discriminative validity of the MPQ is evidenced by the moderately strong associations between the MPQ scale scores and measures of pain intensity. These associations are strong enough to indicate that the MPQ scores assess pain, but also not so strong to suggest that MPQ scores assess only pain intensity. The findings also show that the MPQ scales are associated with measures of quality of life, and are sensitive to the effects of pain treatment. Evidence supports the validity of the MPQ-Affective subscale, in particular, for assessing pain-related distress, given the stronger associations of this scale with measures of psychological distress than with measures of pain intensity, and the relatively high scores on the MPQ-Affective scale among persons with cancer pain compared with persons with low back pain.

However, the MPQ is a relatively lengthy measure (listing 78 descriptors), and many of the descriptors may not be appropriate or needed in many groups of patients with pain. In addition, despite the possibility that certain pain qualities may be more strongly affected by a particular analgesic or pain treatment than others, more often than not, the MPQ scale scores are less sensitive to the effects of pain treatments than the simple single-item measures of pain intensity. This may be due to the fact that the MPQ scale scores include a large variety of pain qualities – some of which might be affected by an intervention and others that might not be affected. The inclusion of so many pain qualities into single scales may therefore weaken the ability to detect treatment effects on specific pain qualities.

The Short-Form MPQ has some strengths that may make it more practical than the MPQ to use in pain clinical trials. First, it includes only 15 descriptors instead of 78, markedly reducing the assessment burden on subjects. In addition, it retains descriptors from two of the MPQ primary categories (sensory and affective), making it possible to assess these dimensions of pain quality if needed. Finally, unlike the MPQ, which requires patients to select no more than a single word from each of 20 categories of pain, respondents to the SF-MPQ are allowed to rate the severity of each pain descriptor on a 0 – 3 scale. This allows for scoring and analysis of each specific quality of pain.

However, like the MPQ scales, the SF-MPQ scales have not demonstrated any greater sensitivity to the effects of pain treatments than simple pain intensity rating scales. Moreover, to date, the SF-MPQ has not been used to determine whether pain treatments affect some pain qualities (i.e., the specific MPQ descriptors) and not others. Also, although the SF-MPQ total scale score has been found to have adequate to excellent internal consistency, the SF-MPQ Sensory and SF-MPQ-Affective scales evidence less reliability, and these two scales have also been found to be so strongly associated with one another that they may be assessing a similar underlying construct. Additional research is needed to determine the utility of the SF-MPQ as an outcome measure in clinical trials (see discussion of future research on pain measures in section 5.4. below).

Until more evidence concerning the reliability and validity of pain quality measures is obtained, investigators should at least consider assessing pain quality as a secondary variable in clinical trials to describe any impact of treatment on pain qualities, and explore whether the treatment impacts some pain qualities more than others. At this point, the SF-MPQ appears to be the most appropriate measure for this purpose.

4. Issues concerning the use of pain measures

Evidence for (or against) the validity and reliability of measures of pain intensity, pain relief, the

temporal aspects of pain, and pain quality as outcome measures in pain clinical trials provide the investigator with information that may be used to select from among specific pain domains and measures. However, once the pain domains and measures have been selected, the investigator must then determine how these measures should be used in any one particular study. Questions that need to be addressed include the following: (1) How often and for how long should pain be measured?; (2) When measuring pain intensity, should study participants only be asked to rate their current pain, perhaps on multiple occasions, or can recalled pain (e.g., worst, least, and average over a specified period of time) can be trusted?; (3) To what extent should unsupervised (e.g., at-home) pain diaries be used, and what measures, if any, do investigators need to take to ensure that these are completed as instructed?; (4) Can single-item measures be used exclusively, or are there any situations when composite measures of pain would be appropriate?; (5) To what extent can, or should, rescue dose requests be used as outcome measures in pain clinical trials?; (6) How should the fact that many patients with pain problems experience multiple pain complaints be taken into account when assessing pain in clinical trials?; and (7) Should (or is it practical for) there be standardization in the format of and endpoints for pain intensity and affect measures in clinical trials? In this section, each of these questions will be addressed in order, taking into account available empirical evidence that speaks to the issues whenever possible. Table 6 lists the specific question and a bottom-line recommendation concerning the question.

4.1. How often and for how long should pain be measured?

Even a brief review of published pain clinical trials shows that the frequency and timing of pain assessment varies tremendously from one study to another. One study examining the effects of a fast acting analgesic for cancer related breakthrough pain had study participants rate pain just before they took the study medication, and then 15, 30, 45, and 60 minutes after taking the study medication (Farrar et al., 1998). A study examining the effects of IV Parecoxib Sodium for relieving postoperative pain

asked participants to rate pain intensity pretreatment and then again 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 10, 12, 16, and 24 hours after administration of the study medication (Barton et al., 2002). Yet another study examining the effects of a cognitive-behavioral intervention for back pain interviewed participants pretreatment and then 3, 6, and 12 months following treatment (Moore et al., 2000).

Despite the variability in the frequency and length of time that pain is assessed in pain clinical trials, most of the published studies (probably in part because they must pass peer review) can be judged to have assessed pain for an appropriate period of time at an appropriate frequency, given the goals of the study and the expected effects of the pain treatment. Almost always, and appropriately, pain is assessed pretreatment and over the time period that the effects of the treatment or intervention are expected to last. In the case of a fast acting analgesic, the window of assessment may be only 60 minutes or less. In the case of chronic pain, if evidence is to be obtained that the intervention is beneficial, the assessment window must clearly be much larger; a matter of months or even years.

Although having only two (pretreatment and posttreatment) or three (pretreatment, posttreatment, and follow-up) assessments in a controlled trial will provide the basic data needed to judge a treatment as being effective (usually relative to a placebo or appropriate control condition), more assessments obtained during the time period that the intervention is thought to be effective will provide important information concerning the pattern(s) of changes in pain following treatment. Such information might be considered essential if the clinical trial is comparing two active interventions, since the treatments might have similar effects overall (in average pain across a specific time window) but have very different effects on the pattern of pain; one might be faster acting but the other might last longer. Hence, and up to some reasonable limit (e.g., it would not make sense to ask study participants to rate their pain every minute for 24 hours), more than three assessments provide for a better picture of the effects of an analgesic or other intervention.

In short, a recommended minimum pain assessment schedule to determine if an intervention is at all effective relative to a placebo would include a pretreatment and posttreatment assessment, as well as a follow-up assessment at some time period after treatment that still lies within the window of time that the treatment is expected to be effective (perhaps at the time that the intervention is thought to have its peak efficacy). However, at least one additional assessment between the posttreatment assessment and the follow-up assessment, as well as one beyond the time period of effects of the intervention would provide a preliminary estimate of the pattern of effects of the intervention. More assessments than this (e.g., hourly for a treatment that is expected to be effective over several hours; monthly for treatments that are expected to be effective over a period of months or years) would provide a clearer picture of the effects of the intervention over time, and also provide for a better (more reliable and valid) measure of the overall effects of the intervention if a single score of pain intensity differences, pretreatment to posttreatment and each follow-up point (e.g., sum of pain intensity differences, or SPID) is to be used as an outcome variable.

4.2. When measuring pain intensity, should study participants only be asked to rate their current pain, perhaps on multiple occasions, or can recalled pain (e.g., worst, least, and average over a specified period of time) be trusted?

While the evidence supports the validity and reliability of single-item pain intensity rating scales for measuring current pain intensity, this does not necessarily mean that ratings of past pain (e.g., worst, least, or usual pain over a specified time period) are valid and reliable. If a measure of average pain over a specified time period is needed or desired, then the most valid measure of actual average pain would be the sum (or average) of multiple measures of current pain administered during that time period. For such ratings, for example, an average of three ratings of current pain per day for four days (12 ratings in all) appears to provide a measure that has adequate reliability and excellent validity as

measure of average pain during a 7-day window, at least among persons with chronic pain (Jensen and McFarland, 1993). Similarly, if multiple measures of current pain were obtained over time, then these data could be used to estimate the participant's worst pain and least pain.

However, not all investigators have the resources to administer, code, and score diary data. Moreover, unless some method is used to ensure that the study participants recorded pain intensity on a diary as instructed, the veracity of diary data can always be called into question (Stone et al., 2002). An easy solution to this dilemma would be to simply ask study participants to provide their own estimates of worst, least, and average pain over the specified time period, say over the last week (Dworkin and Siegfried, 1994). Such measures, if accurate (or accurate enough to detect treatment effects), would make it extremely easy and less costly to assess average pain.

The available evidence suggests that pain recall is not specifically accurate. For the most part, people tend to overestimate previous pain (recall of chronic pain four to nine weeks later: Linton and Götestam, 1983; recall of post-delivery pain one to two days later: Rofé and Algom, 1985; recall of chronic pain during the previous week: Jamison et al., 1989; recall of low back pain 18 months later: Linton, 1991; recall of postoperative pain up to 12 months later: Tasmuth et al., 1996; recall of pain associated with myocardial infarction six months later; Everts et al., 1999; recall of preoperative pain three months later: Lingard et al., 2001). But some studies report that patients can sometimes underestimate pain as well (recall of labor pain two days later: Norvell et al., 1987; recall of postsurgical pain about six days later: Valdix et al., 1995), and at least one study found that patients can provide accurate measures of previous pain (recall of average back pain over the previous week: Bolton, 1999). Concerning painful medical procedures, one study found that recall of pain after the procedure was strongly associated with the peak pain intensity during the procedure as well as the pain reported during the last three minutes of the procedure (Redelmeier and Kahneman, 1996).

Despite the fact that patients are not accurate in their reports of previous pain (in general, overestimating previous pain), this does not speak to the question of whether recall measures are valid for detecting changes in pain over time. Even if there is a systematic bias in pain recall (e.g., if all patients report previous pain as being one point higher on a 0 – 10 scale than it actually was), but patient recall is still strongly associated with actual pain (e.g., the correlation between recalled pain and previous pain is .80 or higher), then recall measures could still potentially be used for detecting changes in pain. For example, if actual average pain (computed from daily diaries over the course of a week) dropped from 8/10 to 5/10, but recalled usual pain (assessed at the end of the week pretreatment and posttreatment) dropped from 9/10 to 6/10, on average, the 9 to 6 change would not be accurate (the real change in average intensity was 8 to 5 in this example), but would still reflect the decrease in usual pain. In short, the fact that pain recall is not specifically accurate (in terms of it represent the exact pain intensity experienced) is not a great concern when considering recall ratings as indicants of usual (or worst or least pain) in clinical trials. Of much greater importance are (1) the validity of recall measures for reflecting previous pain (as indicated by the correlation between the recall measure and actual previous pain levels) and (2) the validity of recall measures for detecting changes in pain.

Concerning the question of association with previous pain, most (but not all, see Valdix et al., 1995) studies show strong associations between usual and least pain recall measures (but not necessarily worst pain recall measures) and actual ratings of previous pain. For example, Salovey et al. (1993) asked a group of persons with chronic pain to provide hourly pain ratings for two weeks. At the end of this two-week period, they phoned the study participants and asked them to rate their current and recalled (usual, worst, least) pain levels over the last two weeks. The correlation coefficients between the average of all of the pain ratings from the diaries (representing actual average pain) and recalled usual pain was 0.83. The correlation between recalled worst pain and actual worst pain from the diaries

was 0.68, and that between recalled least pain and actual least pain was 0.87. Using regression analyses, Salovey et al. found that current pain made a significant contribution to the prediction of recalled usual pain over and above the effects of the actual average pain, indicating that pain at time of recall had at least some effect on the patients' memories of usual pain (see also Eich et al., 1985, 1990; Smith and Safer, 1993).

In a similar study, Jensen et al. (1996) found associations that were very similar to those found by Salovey et al. between recalled usual, least, and most pain and diary ratings of these variables ($r_s = 0.78, 0.81, \text{ and } 0.64$, respectively), and also found that current pain had a biasing impact on pain recollection. It is interesting that in both Salovey et al. (1993) and Jensen et al. (1996), using similar procedures but different samples, recall for worst pain was relatively poor. However, Jensen et al. (1996) found that the ability to predict actual average pain could be improved by combining recalled usual and least pain into a single score (correlation with actual average = 0.87), and that each made an independent contribution to the prediction of actual average pain when controlling for the other. Other studies have also found strong associations between recalled and actual pain (correlation coefficients greater than 0.80 between recall and actual measures of pain: Jamison et al, 1989; Babul et al., 1993; Singer et al., 2001).

Further support for the validity of recalled pain for detecting treatment effects comes from studies that have shown such measures to be sensitive to the effects of pain treatment. For example, Jensen et al. (1999) compared the relative sensitivity of single measures of current pain, average, worst, and least pain over the past two weeks, and various combinations of these measures (see discussion below concerning the use of composite measures in clinical trials) for detecting the effects of multidisciplinary pain treatment. Interestingly, each one of the single ratings (and composites) were able to detect the effects of treatment, with no statistically significant difference in sensitivity between

the measures. Similarly, at least one other study has shown that pain recall ratings are sensitive to treatment outcome (e.g., recalled usual pain in the past three months: Moore et al., 2002).

In summary, although the evidence indicates that patient recall of pain is usually not accurate in the sense of being representative of the exact level of pain experienced (usually as assessed using pain diaries), recall measures (especially of usual and least pain) are often strongly associated with actual pain reported previously, and therefore reflect (carry the variance of) actual previous pain. Moreover, such measures, even those that ask patients to recall pain during a three month window (e.g., Moore et al., 2002), have been shown to be sensitive to the effects of pain treatment. Therefore, the evidence supports the validity (and use, if selected) of usual and least pain ratings (or composites of these, see Jensen et al., 1996) as summary measures of previous pain in clinical trials. This means that investigators need not necessarily use pain diaries or more expensive methods for obtaining pain ratings over time in order to have valid estimates of usual pain for detecting treatment effects. A single measure of usual pain should due in most situations.

The only situation where more time-consuming or expensive options are needed would be when there is a need for highly accurate (accurate in terms of the specific pain ratings obtained, rather than valid in terms of the measure reflecting actual pain levels) estimates of usual, least, and worst pain within a specified time period. However, such highly accurate measures would rarely, if ever, be needed in a clinical trial whose major goal is to determine if an intervention has an impact on usual pain.

4.3. To what extent should unsupervised (e.g., at-home) pain diaries be used, and what measures, if any, do investigators need to take to ensure that these are completed as instructed?

As indicated in the previous section, the evidence suggests that there may not be a need to obtain multiple pain ratings over time with the use of pain diaries or even more complicated and expensive procedures in clinical trials where the primary outcome measure is usual or average pain over a

specified period of time. However, there may be situations in clinical trials when multiple assessments of current pain are indicated. The most common situation would include trials of analgesics designed to treat acute pain conditions; for example, in comparing a new analgesic for postoperative or breakthrough pain to a placebo or another analgesic. In such a situation, highly accurate estimates of the effects of the analgesic over time (and over the effective period of the analgesic) is needed. There is no evidence, yet, to support the accuracy of patient recall for the time patterns of the effects of analgesics; in fact, the evidence indicates that patient recall is rarely accurate (see above). Moreover, data indicate that people, if not supervised in some way, may not provide ratings on paper-and-pencil diaries as instructed (Stone et al., 2002). Thus, and until data are provided to support the veracity of unsupervised diary data, or methods are used that ensure the veracity of such data in a particular study, any study that uses unsupervised diary data should consider the findings from analyses that use such data as preliminary.

On the positive side, there are several ways to supervise study participants in the completion of diary data. Patient participants can be contacted by phone and asked to provide ratings of current pain via phone interview; participants can be asked to provide ratings via email (which is automatically time- and date-stamped); participants can be asked to complete a diary once/day (not necessarily at a specific time) and asked to mail the diary in on a daily basis (using the postmark as a means of establishing that the diary was likely completed as instructed); patient participants can be interviewed by study personnel at the specified assessment times (e.g., if the participant is hospitalized and the study design requires hourly or even less than hourly ratings); or participants can be given palmtop computers programmed to alert the participant when an assessment is needed, ask the participant the assessment questions, and record patient responses for later down-loading. Many of these procedures have been or are being used successfully in research studies (e.g., phone interviews: Cardenas et al., 2002; diary ratings emailed to

the investigators are one option given to participants in an ongoing clinical trial of hypnotic analgesia for SCI-related pain, NIH R01 HD042838-01, Mark P. Jensen, PI; mailed diaries: Keefe et al., 1997; patient interview or direct supervision/observation of responses to pain ratings: Doyle et al., 2002; palmtop computers: Honkoop et al., 1999) and each method appears reasonable. The investigator need only consider which option is optimal (in terms of cost and required staffing) to select from among these given the hypotheses and planned analyses associated with the clinical trial.

4.4. Can single-item measures be used exclusively, or are there any situations when composite measures of pain would be appropriate?

In addition to the more frequently used single ratings of pain intensity, pain intensity may also be assessed using multiple-item scales. For example, the four pain intensity rating scales from the Brief Pain Inventory (of worst, least, average, and current pain; Cleeland and Ryan, 1994) can be combined to form a single composite score representing pain severity (e.g., Shacham et al., 1984). Similarly, the pain intensity items of the Chronic Pain Grade (of current, worst, and average pain), a measure used in survey research, are usually combined to form a single pain severity score that is used to classify patients into various levels or grades pain severity (Von Korff et al., 1992).

When tested, measures that combine three or more of the domains of worst, least, average, and current pain show excellent internal consistency; almost always they have Cronbach's alpha greater than 0.85 (Cleary et al., 1995; Serlin et al., 1995; Mystakidou et al., 2001; Klepstad et al., 2002). Also, as would be expected based on the review above, individual items assessing current, worst, least, and average pain are all sensitive to the effects of pain treatment (e.g., Jensen et al., 1999; Hwang et al., 2003). Therefore, composites made up of these measures should also be sensitive (see Shiffmann et al., 2001; Thie et al., 2001).

In addition, there are theoretical reasons to expect that composite measures of pain might be

more sensitive than individual ratings for detecting pain treatment effects. According to psychometric theory, every measure is an imperfect estimate that contains both valid and invalid (or error) components (Nunnally 1978). Composites that are created from multiple measures, each containing valid variance, should have greater validity than the individual component measures because the valid components of each measure contribute to the composites while the error components, being random, tend to average to zero (Cronbach 1970). On average, as the number of measures used to create the composite score increases, the reliability and validity of the composite score should increase.

However, although, as expected, composite measures tend to be slightly more reliable and valid (i.e., sensitive to change in pain) than individual ratings on average (Bolton, 1999), they are not always superior, and when they are shown to have greater reliability or validity, the differences in sensitivity between composite measures and individual ratings are rarely large. For example, in the Jensen et al. (1996) study cited above, the correlations between actual average pain (as estimated by the average of hourly daily diary ratings) and individual ratings of usual, least, most, and current pain were 0.78, 0.81, 0.64, and 0.64, respectively. As a group, the composite measures (all 11 possible combinations of the four ratings) were more strongly associated with actual average pain than the individual ratings were (correlation coefficients ranged from 0.70 to 0.87). But the strongest association (0.87) was not with a composite made up of all four ratings (as might be expected based on psychometric theory), but a composite of usual and least pain (Jensen et al., 1996).

Similarly, composite measures of (1) average and least pain, (2) average, current, and worst pain, and (3) average current, worst, and least pain tended to be more sensitive to the effects of multidisciplinary pain treatment (F values pretreatment to 2-week follow-up range = 31.05 to 40.28, pretreatment to 1-month follow-up range = 30.34 to 41.51, and pretreatment to 2-month follow-up range = 26.36 to 42.29) than the individual ratings were (F values ranged from 21.01 to 30.36, 14.26 to 38.16,

and 16.56 to 28.42, respectively; Jensen et al., 1999). However, some of the individual ratings, for some of the analyses, were more sensitive than some of the composites (Jensen et al., 1999). More importantly, all of the individual and composite measures showed a statistically significant change in pain, and the F values associated with the change analyses were not significantly different across the various individual and composite measures.

Finally, Jensen et al. (2002) compared the relative sensitivity of a VAS-I difference score, VRS-I difference score, VRS pain relief rating, and a composite made up of all three ratings for detecting the effects of morphine, ketorolac, and placebo on post-surgical pain in two different samples of patients (knee surgery and laparotomy). They found that the composite score was sometimes, but not always, more sensitive than the individual scores for detecting change in pain over time or as a result of the medication condition. In fact, no single rating or composite score emerged as consistently more sensitive than the others, and all measures detected expected changes.

Practically, what these findings mean is that, although it is reasonable combine measures of pain intensity (or changes in pain intensity with pain relief) into composite scores that represent pain severity (or change in pain intensity), and such measures will likely, on average (but not always), result in pain severity estimates that are slightly more reliable and more valid than individual ratings, only a limited improvement in the reliability and validity of pain measurement is attained by using such composite scores. For most purposes, then, a single rating of usual pain (in studies requiring the assessment of pain over a matter of days, weeks, or months, i.e., most trials involving patients with chronic pain) or current pain (in studies requiring the assessment of current pain over a period of minutes or hours, i.e., most trials involving patients with acute pain) is adequate for assessing pain intensity in clinical trials.

Composite measures might be considered if reliability and sensitivity is of great concern (e.g., in studies comparing two analgesics thought to have similar effects or in studies with very few subjects),

given the evidence that reliability and sensitivity tend to be improved with the use of composite measures. But the available evidence suggests that investigators should not count on a great improvement in the psychometric properties of pain assessment with the use of composite scores.

4.5. To what extent can, or should, rescue dose requests be used as outcome measures in pain clinical trials?

Some recent pain clinical trials have been using patient request for a rescue dose as a secondary outcome measure (e.g., Farrar et al., 1998; Eisenberg et al., 2001; Doyle et al., 2002; Chrubasik et al., 2003). As Farrar et al. (1998) point out, allowing patients in clinical trials access to a rescue dose of an analgesic provides an ethical way to incorporate placebo controls into a treatment efficacy study, since anyone not obtaining adequate pain relief would be provided with an analgesic known to be effective if they requested it. Allowing for rescue doses is also ethical when the adequacy of the experimental analgesic is not well established.

Of course, once a rescue dose is taken, pain ratings obtained from the time after the rescue dose was taken can no longer be used to compare the drug/placebo that patient is taking against the other drug conditions; that pain would presumably decrease as a result of the rescue dose and not because of the experimental drug. The usual way to address problem this is to “carry forward” the last pain rating (prior to taking the rescue dose) as subsequent pain ratings for that subject in the analyses (e.g., Farrar et al., 1998; Barton, 2002).

In addition to providing one solution to the ethical problem of using placebos in pain research, allowing for, and keeping track of, the frequency of rescue dose requests in study participants allows for another important measure of the efficacy of the pain treatments being examined. If the experimental analgesic is very effective, for example, one would expect that patients in a placebo condition or receiving an analgesic that is less effective will request rescue doses more often. Moreover, the average

time to rescue dose provides yet another secondary measure of outcome that can, and should, be reported in clinical trials (see Doyle et al., 2002). In short, monitoring requests for rescue doses (including when they occur) can be a very useful secondary outcome measure in pain clinical trials.

4.6. How should the fact that many patients with pain problems experience multiple pain complaints be taken into account when assessing pain in clinical trials?

While pain from multiple sites may be less of an issue in clinical trials involving acute pain episodes (e.g., postsurgical pain), virtually all populations of patients with chronic pain conditions include a significant subgroup who experience pain at more than one site. For example, in an interview survey of 93 adults with cerebral palsy, Schwartz et al. (1999), found that 67% of the respondents reported chronic pain, and that these participants reported an average of three ($SD = 1.73$) different pain sites. Turner and Cardenas (1999), in a survey of adults with spinal cord injuries, found that 132 (81%) of the respondents indicated that they had at least one pain problem, with 83% of these reporting more than one separate pain problem (41% reported more than three pain problems). Ehde et al. (2000) performed a similar survey of persons with acquired amputation, and found that over a third (36%) of their sample reported pain phantom limb pain, residual limb pain, and back pain. Another large subgroup (34%) experienced pain in two of these three locations, and 22% experienced pain in only one location. Only eight percent of their sample was pain free. Williams et al. (2000) interviewed 78 patients who were embarking on a chronic pain management course, and found that all but one had two or more distinct pain problems.

When asked to rate “current” (or “least,” “worst,” or “average” pain), a reasonable question that any participant in a clinical trial who experiences pain in multiple locations (and this may be a large subset of some samples of study participants) is, “Which pain should I rate?” Williams et al. (2000) found that patients with multiple pain site dealt with this dilemma in three main ways: 69% indicated

that they sometimes rated their “main” pain and ignored the others, 69% indicated that they sometimes rate whichever pain was worst at the time of assessment, and 64% indicated that they sometimes combine the pains into a single rating. Obviously, based on these responses, many patients use more than one strategy at different times when responding to a simple pain intensity rating scale. Similarly, if treatment reduces one type of pain (for example, neuropathic pain associated with a spinal cord injury) from 8/10 to 4/10, but another type of pain (for example, musculoskeletal shoulder pain associated with repetitive movements) pain stays at 8/10 from pretreatment to posttreatment, and the study participant is only asked to rate his or her usual, worst, least, and current pain, before and after treatment, he or she may not know whether to report a reduction from 8 to 4, 8 to 6 (8 and 6 being averages of the two pains before and after treatment), or no reduction, since there remained a pain intensity at the 8/10 level after treatment.

Ideally, investigators would screen study participants for having only one type of pain, or at least assess and report on the presence of multiple pain sites in their sample. However, rarely, if ever, do investigators assess the presence or absence of multiple pain sites in samples of clinical trial participants, much less report on the impact of pain treatments on separate pain problems. Future clinical trials would do well to consider assessing the number of pain sites in the sample (at least for descriptive purposes), and, if possible, the level(s) of pain intensity at each site both before and after treatment, as secondary measures. A straightforward way to do this would be to provide the study participant with a list of possible pain sites (e.g., head, neck, shoulder, upper back, lower back, arms, buttocks/hips, chest, stomach/pelvis, legs, other), and ask the participant to rate the level of pain intensity at each pain site both before and after treatment. Investigators should also provide study participants with clear instructions regarding how to rate the participants’ pain when they experience pain in more than one site (e.g., “Always use these measures to assess your back pain, regardless of any

other pain you experience” or “Use these scales to assess you pain overall, taking into account all of the pains that you experience”) to help clarify the meaning of responses to pain scales.

4.7. Should (or is it practical for) there be standardization in the format of and endpoints for pain intensity?

Given that the format and specific endpoints selected for pain intensity rating scales impacts the response to those scales (e.g., Sriwatanakul et al., 1983; Seymour et al., 1985; Breivik and Skoglund, 1998), and tens if not hundreds of different formats and endpoint descriptors are used in the VAS-Is, NRS-Is, and VRS-Is in pain clinical trials, it is usually not possible to directly compare the findings from one study to another. In order to make the results of different trials more comparable (which would be useful, for example, for determining the “average” impact of a particular analgesic or pain treatment across settings and populations, or for determining if a particular analgesic or treatment tends to be more effective for one type of pain problem than others), it is reasonable to consider whether it makes sense to recommend standard formats and endpoint descriptors for assessing pain intensity in clinical trials.

To develop such a recommendation probably requires a detailed review, and then discussion, of the many format and descriptor options that exist for available measures. There is not the time (or space) to include such a review in this paper. However, the IMMPACT group may wish to consider whether it is reasonable to recommend specific formats and endpoints for assessing pain in future clinical trials. As a starting point of this (possible) discussion, I would propose the following standard forms of VAS-I, NRS-I, and VRS-I pain measures.

Standard VAS-I: A horizontal 100mm line with small (between 2 – 4 mm) demarcation lines at each end, with the descriptor “No pain” centered below the demarcation line at the left end and “Pain as bad as/you can imagine” (taking up two lines) centered below the demarcation line at the right. One

hundred mm is the most common length of VASs in the literature, and the proposed descriptor of extreme pain is that used in the Brief Pain Inventory Pain Intensity items (Cleeland and Ryan, 1994) which are commonly used in pain research. The proposed standard instructions could read, “Please rate your pain by placing a mark on the line below that best describes your pain (at its worst, at its least, on average) in the last (24 hours, week, month, three months)” or “... your pain right now” (instructions adapted from the Brief Pain Inventory, Cleeland and Ryan, 1994).

Standard NRS-I: All eleven numbers from 0 through 10 presented horizontally (between 5 and 7 inches from the “0” to the “10” with the descriptor “No pain” centered below the “0” at the left end and “Pain as bad as/you can imagine” (taking up two lines) centered below the “10” at the right. The proposed standard instructions could read, “Please rate your pain by circling the number below that best describes your pain (at its worst, at its least, on average) in the last (24 hours, week, month, three months)” or “... your pain right now” (instructions adapted from the Brief Pain Inventory, Cleeland and Ryan, 1994).

Standard VRS-I: Four words (“none,” “mild,” “moderate,” and “severe”) presented in ascending order vertically, with the instructions, “Please select the one word that best describes your pain (at its worst, at its least, on average) in the last (24 hours, week, month, three months)” (or “... your pain right now”) from the list of words below.”

5. Recommendations for Future Research on Measures of Pain in Clinical Trials

This section discusses possible directions for future research on the reliability, validity, and use of pain measures in clinical trials. These recommendations are summarized in Table 7.

5.1. Research on pain intensity measures

There does not appear to be a strong need for future studies to determine the psychometric properties of single-item ratings of pain intensity. The extensive evidence that is available provides a

fairly clear picture concerning their validity and reliability. However, although there is preliminary evidence that composite measures of pain intensity (e.g., the average of rated “usual,” “least,” and “worst” pain over a specified time period as measures of “usual” over that time period) show only modest improvements in reliability and validity when compared with individual ratings, composite measures could potentially prove to be superior to individual ratings in settings and situations where reliability and validity (as determined by sensitivity to treatment effects) must be maximized, such as in case series (examining changes in pain over time in a series of $N = 1$ studies), preliminary clinical trials with very few subjects, or when comparing two active treatments that might show only subtle differences in efficacy. Thus, future research to help determine further whether composite measures of pain intensity are, or are not, superior to individual ratings, would be helpful.

In addition, the available evidence suggests that recall ratings of usual pain (even during very large time windows of up to three months) are valid for use in pain treatment outcome clinical trials. However, there has not yet been research to determine if such recall ratings are as sensitive to the effects of pain treatments as averages of pain intensity created from real-time assessments (using supervised diaries, see section 4.3. above). It is possible that recall of usual pain is less sensitive, perhaps significantly so, than pain scores created by averaging multiple assessments of current pain over time. Evidence concerning this issue would provide important information that would lend support for or against the use of ratings of recalled usual pain in clinical trials. In the meantime, investigators who chose to use recall ratings cannot be certain that their outcome measure will be as sensitive as the (more expensive, but perhaps also more valid) use of multiple pain measures over time.

5.2. Research on measures of pain relief

Although the evidence indicates that measures of pain relief are related to, but also statistically distinct from, changes in pain intensity from pretreatment to posttreatment, measures of pain relief are

sometimes (but not always) more sensitive to the effects of pain treatment than pretreatment to posttreatment pain intensity change scores. In one study, perceived pain relief was also found to be more strongly associated with satisfaction with treatment than pretreatment to posttreatment pain intensity change scores were (Fischer et al., 1999).

Although we can say that a pain relief rating is not the same as change in intensity, it is not clear what factors contribute to a pain relief score. Given the likelihood that measures of perceived pain relief will continue to be used in pain clinical trials, at least as secondary measures, it would be useful to have a better sense of the meaning of pain relief ratings. Are relief ratings related to changes in pain affect or changes in other qualities of pain in addition to changes in global pain intensity? Are relief ratings related to hope engendered by a modest (albeit perhaps brief) decrease in pain, so that small changes in pain intensity translate to larger ratings of pain relief in some patients (see Carlsson, 1983)? Are relief ratings more closely associated with the total amount of time spent experiencing relatively less pain during a specified window of time (e.g., the sum of pain intensity differences, or SPID, computed from multiple measures of current pain after a pain treatment), even if the pain intensity at the time of rating had returned to (or even become higher than) pretreatment pain levels? Research helping to clarify the meaning of pain relief ratings would help interpret the findings when an analgesic or other pain treatment is shown to have significant effects on such ratings in clinical trials.

5.3. Research on measures of the temporal aspects of pain

Measures of the temporal aspects of pain (e.g., frequency, duration, time to analgesic effect and/or to meaningful decreases in pain, presence, intensity, frequency, and duration of breakthrough pain) have been under-utilized and inadequately studied in pain research. However, these components of pain would clearly be important to patients experiencing pain (a treatment that decreases the frequency of pain or breakthrough pain, even if it does not alter the usual intensity of that pain would be

welcome to most patients). There is a great need for researchers and clinicians to develop additional measures of these temporal aspects of pain, and determine their reliability and validity in the context of pain clinical trials.

In particular, because of the problems and expense associated with collecting pain ratings over time, it would be useful to determine whether patient memory for the temporal aspects of pain are adequately valid and reliable; if they are, then it would be much easier to assess these aspects pain. For example, studies are needed that assess the actual frequency of pain (in patients with intermittent pain, such as headache patients) or of breakthrough pain (in patients with breakthrough pain) using adequately supervised pain diaries and also using patient recollection of pain or breakthrough pain frequency. Both types of measures (measures computed from supervised diaries versus recall measures) could be compared with respect to their ability to detect the effects of effective pain treatment. Research showing the association (if any) between the temporal aspects of pain and other pain-related measures (e.g., psychological functioning, pain interference) would also help to clarify the importance and meaning of measures of the temporal components of pain.

5.4. Research on measures of the qualitative aspects of pain, including pain affect

While there have been strong advocates for the use of measures of pain affect in clinical trials over the years, such measures are only rarely included in the published literature. Perhaps this is due to the fact that when pain intensity is eliminated (and this is a goal, even if rarely achieved, of most pain treatments), issues of pain quality and pain affect becomes moot. Perhaps the infrequent use of pain affect measures in clinical trials is also due in part to the fact that research shows that the pain intensity measures appear to be at least as, and sometimes more, sensitive to the effects of pain treatments than measures of pain quality.

However, it is likely, and some research supports the conclusion that, pain treatments impact

some pain qualities more than others (e.g., Galer et al., 2002). Also, as reviewed above, some pain treatments are known to have a greater impact on pain affect than they have on pain intensity. Thus, it seems reasonable to include measures of pain quality and affect as secondary measures in clinical trials to determine the pattern of effects of the intervention on pain quality.

At this point, because of its brevity and simplicity, and application across a number of different pain conditions, the SF-MPQ appears to be the pain quality measure is the most practical to use in assessing pain quality in clinical trials. However, more research is needed to determine the utility and validity of the SF-MPQ for assessing pain quality. One of the potential weaknesses of the SF-MPQ is the relatively few number of response choices to each of the 15 SF-MPQ descriptors (“none,” “mild,” “moderate,” and “severe”), despite the fact that it is likely that patients are able to distinguish more than four different levels of each pain quality. Research could determine whether the relative sensitivity of the SF-MPQ items could be improved by using a 0 – 10 or 0 – 100 numerical format (e.g., 0 = “no throbbing pain;” 10 = “the most intense throbbing pain I can imagine”; see Galer and Jensen, 1997).

In addition, because of its brevity (which is a strength of the MPQ), there are many common pain qualities that the SF-MPQ does not assess (e.g., “dull” and “sensitive” pain, among others), so that the SF-MPQ might not assess the pain quality(ies) most important to a particular patients or population of patients. Therefore, research is needed to determine the pain qualities most often reported by patients with pain across conditions, to ensure that the SF-MPQ, or alternative pain quality measure, has adequate content validity.

Although single-item measures of pain affect (“pain unpleasantness” or “pain distress”) exist, pain affect is clearly more complex than pain intensity. In view of the multidimensional nature of pain affect, it is possible that single global measure of the distress associated with pain may not be adequate to assess pain affect. Research is needed to determine if there a need for separate indices that tap

distinct affective dimensions of pain. Similarly, single-item measures of pain affect may be less reliable than multiple-item measures, as would be suggested by the complexity of pain affect. Research is also therefore needed to compare the relative reliability of single- versus multiple-item measures of pain affect (e.g., MPQ or SF-MPQ Affective scales versus VAS, NRS, or VRS pain affect measures).

Finally, it was suggested above that VRS measures of pain affect (i.e., word lists of affective responses such as “not at all bothersome,” “unpleasant,” “annoying,” “distressing,” and “intolerable”) may be more able than NRS or VAS measures to discriminate between pain intensity and pain affect, because words are usually used to describe emotional reactions, whereas VAS-As and NRS-As, due to their similarity in appearance to measures of pain intensity, may elicit ratings that reflect pain intensity more than pain affect. Research is needed to test this hypothesis, and whether VRS-A measures may be more valid than VAS-A or NRS-A measures for assessing the affective component of pain.

In short, there is a need for research that directly compares the psychometric properties of existing single-item (i.e., VAS, NRS, VRS measures with varying endpoint descriptors) and multiple-item measures of pain affect (i.e., SF-MPQ and MPQ affect scales). These measures should be compared regarding their ability to be distinct from measures of pain intensity (e.g., show sensitivity to treatments that are known to effect pain affect and not pain intensity; show relative insensitivity to treatments known to effect pain intensity and have a minimal effect on pain affect; show stronger associations with other measures of affect, such as general anxiety and depression, than measures of pain intensity) and also in terms of their reliability (over the course of minutes and also over the course of days). Such additional evidence concerning the psychometric properties of existing measures of pain affect would provide the information needed for investigators to make informed choices concerning the need to include a measure or measures of pain affect in a study, as well as for selecting the specific measure(s) to use, if it is determined that a measure of this domain would assist in the evaluation of a pain

treatment.

6. Summary and Conclusions

A great deal of research has been performed that provides data concerning the psychometric properties of pain measures. The findings from this research support and confirm the multidimensional nature of pain. The results also support validity of a number of measures, especially the most commonly used measures of pain intensity. Measures of other dimensions of pain, such as pain relief and the temporal and qualitative aspects of pain, are less often used and studied. Yet measures of these and other pain dimensions may prove to be invaluable for assessing pain and the efficacy of pain treatment. Future research that develops, refines, and evaluates such measures will provide important information that investigators and clinicians may then use to select specific scales for their research and clinical work. By increasing knowledge about and options for pain assessment, investigators will ultimately contribute to a better understanding and alleviation of pain.

Acknowledgements

This review was supported by funding from IMMPACT.

References

- Ahles TA, Blanchard JB, Ruckdeschel JC. The multidimensional nature of cancer-related pain: Pain 1983; 17:277-288.
- Ahles TA, Ruckdeschel JC, Blanchard JB. Cancer-related pain – Assessment with visual analogue scales. J Psychosom Res 1984; 28:121-124.
- Anderson PE, Cohen JI, Everts JC, Bedder MD, Burchiel KJ. Intrathecal narcotics for relief of pain from head and neck cancer. Arch Otolaring Head Neck Surg 1991; 1117:1277-1280.
- Angst MS, Brose WG, Dyck JB. The relationship between the Visual Analogical Pain Intensity and Pain Relief Scale changes during analgesic studies in chronic pain patients. Anesthesiol 1999; 91:34-41.
- Babul N, Darke AC, Johnson DH, Charron-Vincent K. Using memory for pain in analgetic research. Ann Pharmacother 1993; 27:9-12.
- Backonja M, Beydoun A, Edwards KR, Schwartz SL, Fonseca V, Hes Marykay, LaMoreaux L, Garofalo E. Gabapentin for the treatment of painful neuropathy in patients with diabetes mellitus. JAMA 1998; 280:1831-1836.
- Baker BO, Hardyck C D, Petrinovich LF. Weak measurement vs strong statistics: An empirical critique of S. S. Stevens' prescriptions on statistics. Educational and Psychological Measurement 1966; 26:291-309.
- Barton SF, Langeland FF, Snabes MC, LeComte D, Kuss ME, Dhadda SS, Hubbard RC. Efficacy and safety of intravenous parecoxib sodium in relieving acute postoperative pain following gynecologic laparotomy surgery. Anesthesiol 2002; 97:306-314.
- Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scale in osteoarthritis patients. Current Med Res Opinion 1999; 15:113-119.
- Bergh B, Sjöström B, Odén A, Steen B. An application of pain rating scales in geriatric patients. Aging Clin Exp Res 2000; 12:380-387.
- Bergman B, Aaronson N, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: A modular supplement to the EORTC Core Quality Of Life Questionnaire (QLQ-30) for use in lung cancer clinical trials. Eur J Cancer 1994; 30:635-642.
- Bergman B, Sullivan M, Sörenson S. Quality of life during chemotherapy for small cell lung cancer. Acta Oncologica 1992; 31:19-28.
- Bolton J. Accuracy of recall of usual pain intensity in back pain patients. Pain 1999; 83:533-539.
- Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiop Ther 1998; 21:1-7.
- Bone M, Critchley P, Buggy DJ. Gabapentin in postamputation phantom limb pain: a randomized, double-blind, placebo-controlled, cross-over study. Pain 2002; 27:481-486.
- Breivik EK, Björnsson GA, Skovlund E. A comparison of pain ratings scales by sampling from clinical trial data. Clin J Pain 2000; 16:22-28.

- Breivik EK, Skoglund LA. Comparison of present pain intensity assessment on horizontally and vertically oriented visual analogue scales. *Meth Find Exp Clin Pharmacol* 1998; 20:719-724.
- Briggs M. Surgical wound pain: A trial of two treatments. *J Wound Care* 1996; 5:456-460.
- Burchiel KJ, Anderson VC, Brown FD, Fessler RG, Friedman WA, Pelofsky S, Weiner RL, Oakley J, Shatin D. Prospective, multicenter study of spinal cord stimulation for relief of chronic back and extremity pain. *Spine* 1996; 21:2786-2794.
- Bruera E, Kuehn N, Miller M, Selmsler P, MacMillan K. The Edmonton Symptom Assessment System (ESAS): A simple method for the assessment of palliative care patients. *J Pall Care* 1991; 1:6-9.
- Campbell BH, Marbella A, Layde PM. Quality of life and recurrence concern in survivors of head and neck cancer. *Laryngoscope* 2000; 110:895-906.
- Caraceni A, Mendoza T, Mencaglia E, Baratella C, Edwards K, Forjaz M, Martini C, Serlin R, de Conno F, Cleeland C. A validation study of an Italian version of the Brief Pain Inventory (Breve Questionario Per La Valutazione Del Dolore). *Pain* 1996; 65:87-92.
- Caraceni A, Portenoy R, a working group of the IASP Task Force on Cancer Pain. An international survey of cancer pain characteristics and syndromes. *Pain* 1999; 82:263-274.
- Cardenas DD, Warms CA, Turner JA, Marshall H, Brooke MM, Loeser JD. Efficacy of amitriptyline for relief of pain in spinal cord injury: results of a randomized controlled trial. *Pain* 2002; 96:365-373.
- Carlsson AM Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. *Pain* 1983; 16:87-101.
- Chang, VT, Hwang SS, Feuerman M. Validation of the Edmonton Symptom Assessment Scale. *Cancer* 2000; 88:2164-2171.
- Chang, VT, Hwang SS, Feuerman M, Kasimis BS. Symptom and quality of life survey of medical oncology patients at a veterans affairs medical center. *Cancer* 1999; 88:1175-1183.
- Chesney MA, Shelton JL. A comparison of muscle relaxation and electromyogram biofeedback treatments for muscle contraction headache. *J Behavior Ther Exper Psych* 1976; 7:221-225.
- Chrubasik S, Model A, Black A, Pollak S. A randomized double-blind pilot study comparing Doloteffin[®] and Vioxx[®] in the treatment of low back pain. *Rheumatol* 2003; 42:141-148.
- Cicchetti DV, Showalter D, Tyrer PJ. The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *ApplPsych Measure* 1985; 9:31-36.
- Cleary PD, Morrissey G, Oster G. Health-related quality of life in patients with advanced prostate cancer: A multinational perspective. *Qual Life Res* 1995; 4:207-220.
- Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. *Ann Acad Med* 1994; 23:129-138.
- Cronbach LJ. *Essentials of Psychological Testing*. New York: Harper and Row, 1970.
- Dalton JA, Toomey T, Workman MR. Pain relief for cancer patients. *Can Nurs* 1988; 11:322-328.
- Daut RL, Cleeland CS, Flannery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. *Pain* 1983; 17:197-210.

- De Conno F, Caraceni A, Gamba A, Mariani L, Abbatista A, Brunelli C, La Mura A, Ventafridda V. Pain measurement in cancer patients: A comparison of six methods. *Pain* 1994; 57:161-166.
- de Wit R, Dam, van F, Abu-Saad HH, Loonstra S, Zandbelt L, Buuren, van A, Heijden, van der K, Leenhouts G. Empirical comparison of commonly used measures to evaluate pain treatment in cancer patients with chronic pain. *J Clinical Oncology* 1999; 17:1280-1287.
- de Wit R, Dam, van F, Hanneman M, Zandbelt L, Buuren, van A, Heijden, van der K, Leenhouts G, Loonstra S, Abu-Saad HH. Evaluation of the use of a pain diary in chronic cancer pain patients at home. *Pain* 1999; 79:89-99.
- Dodd MJ, Miaskowski C, Paul SM. Symptom clusters and their effect on the functional status of patients with cancer. *Oncol Nurs Forum* 2001; 8:465-470.
- Doyle G, Jayawardena S, Ashraf E, Cooper SA. Efficacy and tolerability of nonprescription ibuprofen versus celecoxib for dental pain. *J Clin Pharmacol* 2002; 42:912-919.
- Dubuisson D, Melzack R. Classification of clinical pain descriptors by multiple group discriminant function analysis. *Exper Neurol* 1976; 51:480-487.
- Dudgeon D, Ranbertas RF, Rosenthal S. The short-form McGill Pain Questionnaire in chronic cancer pain. *J Pain Sympt Manage* 1993; 8:191-195.
- Duncan GH, Bushnell MC, Lavigne GJ. Comparison of verbal and visual analogue scales for measuring the intensity and unpleasantness of experimental pain. *Pain* 1989; 37:295-303.
- Du Pen SL, Du Pen AR, Polissar N, Hansberry J, Kraybill BM, Stillman M, Panke J, Everly R, Syrjala K. Implementing guidelines for cancer pain management: results of a randomized controlled clinical trial. *J Clin Oncol* 1999; 17:361-370.
- Dworkin, RH, Siegfried RN. Are all those pain ratings necessary? *Pain* 1994; 58:279.
- Eich E, Rachman S, Lopatka C. Affect, pain, and autobiographical memory. *J Abnorm Psychol* 1990; 99:174-178.
- Eich E, Reeves JL, Jaeger G, and Graff-Radford, SB. Memory for pain: relation between past and present pain intensity. *Pain* 1985; 3:375-380.
- Eija K, Tasmuth T, Pertti NJ. Amitriptyline effectively relieves neuropathic pain following treatment of breast cancer. *Pain* 1996; 4:293-302.
- Eisenberg E, Lurie Y, Daoud D, Ishay A. Lamotrigine reduces painful diabetic neuropathy: a randomized controlled study. *Neurology*, 2001; 57:505-509.
- Eklblom A, Hansson P. Pain intensity measurements in patients with acute pain receiving afferent stimulation. *J Neuro Neurosurg Psych* 1988;5 1:481-486.
- Ellershow J, Peat SJ, Boys LC. Assessing the effectiveness of a hospital palliative care team. *Palliative Med* 1995; 9:145-152.
- Everts B, Karlson B, Währborg, Abdon N-J, Herlitz J, Hedner T. Pain recollection after chest pain of cardiac origin. *Cardiology* 1999; 2:115-120.
- Farrar JT, Cleary J, Rauck R, Busch M, Nordbrock E. Oral transmucosal fentanyl citrate: randomized, double-blinded, placebo-controlled trial for treatment of breakthrough pain in cancer patients. *J Nat Cancer Int* 1998; 90:611-616.

- Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. *Pain* 2000; 88:287-294.
- Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measures on an 11-point numerical rating scale. *Pain* 2001; 94:149-158.
- Feine JS, Lavigne GJ, Thuan Dao TT, Morin C, Lund JP. Memories of chronic pain and perceptions of relief. *Pain* 1998; 137-141.
- Fernandez E, Turk DC. Demand characteristics underlying differential ratings of sensory versus affective components of pain. *J Behav Med* 1994; 17:375-390.
- Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. *JAMA* 1999; 282:1157-1162.
- Fishman B, Pasternak S, Wallenstein SL, Houde, RW, Holland JC, Foley KM. The Memorial Pain Assessment Card: A valid instrument for the evaluation of cancer pain. *Cancer* 1987; 60:1151-1158.
- Fowlow B, Price P, Fung T. Ambulation after sheath removal: a comparison of 6 and 8 hours of bedrest after sheath removal in patients following a PTCA procedure. *Heart Lung* 1995; 24:28-37.
- Fox EJ, Melzack R. Transcutaneous electrical stimulation and acupuncture: Comparison of treatment for low-back pain, *Pain* 1976; 2:141-148.
- Freeman K, Smyth C, Dallam L, Jackson B. Pain measurement scales: a comparison of the vision analogue and faces rating scales in measuring pressure ulcer pain. *J WOCN* 2001; 28:290-296.
- Frost S, Grossfeld S, Kirkley A, Litchfield B, Fowler P, Amendola A. The efficacy of femoral nerve block in pain reduction for outpatient hamstring anterior cruciate ligament reconstruction: a double-blind, prospective, randomized trial. *Arthroscopy* 2000; 16:243-248.
- Gagliese L, Melzack R. Age differences in the quality of chronic pain: a preliminary study. *Pain Res Manage* 1997; 2:157-162.
- Gallagher EJ, Bijur PE, Latimer C, Silver W. Reliability and validity of a visual analog scale for acute abdominal pain in the ED. *Am J Emerg Med* 2002; 20:287-290.
- Galer BS, Jensen MP. Development and preliminary validation of a pain measure specific to neuropathic pain: the Neuropathic Pain Scale. *Neurol* 1997; 48:332-338.
- Galer BS, Jensen MJ, Ma T, Davies PS, Rowbotham MC. The Lidocaine patch 5% effectively treatments all neuropathic pain qualities: results of a randomized, double-blind, vehicle-controlled, 3-week efficacy study with use of the Neuropathic Pain Scale. *Clin J Pain* 2002; 18:297-301.
- Gaston-Johansson F, Franco T, Zimmerman L. Pain and psychological distress in patients undergoing autologous bone marrow transplantation. *Oncol Nurs Forum* 1992; 19:41-48.
- Geddes D, Dones L, Hill E, Law K, Harper P, Spiro S, Tobias J, Souhami R. Quality of life during chemotherapy for small cell lung cancer: assessment and use of a daily diary card in a randomized trial. *Eur J Cancer* 1990; 26:484-492.
- Ger L, Ho S, Sun W, Wang M, Cleeland C. Validation of the Brief Pain Inventory in a Taiwanese population. *J Pain Sympt Manag* 1999; 8:316-322.

- Good M, Stiller C, Zauszniewski JA, Cranston Anderson G, Stanton-Hicks M, Grass JA. Sensation and distress of pain scales: reliability, validity, and sensitivity. *J Nurs Measur* 2001; 9:219-238.
- Gracely RH. Evaluation of multi-dimensional pain scales. *Pain* 1992; 48:297-300.
- Gracely RH, Dubner R, McGrath PA. Narcotic analgesia: Fentanyl reduces the intensity but not the unpleasantness of painful tooth pulp sensations. *Science* 1979; 203:1261-1263.
- Gracely RH, McGrath P, Dubner R. Ratio scales of sensory and affective verbal pain descriptors. *Pain* 1978a; 5:5-18.
- Gracely RH, McGrath P, Dubner R. Validity and sensitivity of ratio scales of sensory and affective verbal pain descriptors: Manipulation of affect by diazepam. *Pain* 1978b; 5:19-29.
- Graff-Radford SB, Shaw LR, Naliboff BN. Amitriptyline and fluphenazine in the treatment of postherpetic neuralgia. *Clin J Pain* 2000; 16:188-192.
- Graham C, Bond S, Gerkovich M. Use of the McGill Pain Questionnaire in the assessment of cancer pain: reliability and consistency. *Pain* 1980;8:377-387.
- Greenwald H, Bonica J, Bergner M. The prevalence of pain in four cancers. *Cancer* 1987; 60:2563-2569.
- Grond S, Radbruch L, Meuser T, Sabatowski R, Loick G, Lehmann KA. Assessment and treatment of neuropathic cancer pain following WHO guidelines. *Pain* 1999; 79:15-20.
- Grossman S, Sheidler V, McGuire D, Geer C, Santor D, Piantodosi S. A comparison of the Hopkins Pain Rating instrument with Standard Visual Analogue and Verbal Descriptor scales in patients with cancer pain. *J Pain Sympt Manag* 1992; 7:196-203.
- Haas M, Nyendo J, Aickin M. One year trend in pain and disability relief recall in acute pain and chronic ambulatory low back pain patients. *Pain* 2002; 95:83-91.
- Hall W. On "ratio scales of sensory and affective verbal pain descriptors." *Pain* 1981; 4:101-107.
- Hardy JD, Wolff HG, Goodell H. *Pain sensations and reactions*. Baltimore: Williams & Wilkins 1952.
- Hammerlid E, Bjordal K, Ahlner-Elmqvist M, Jannert M, Kaasa S, Sullivan M, Westin T. Prospective, longitudinal quality-of-life study of patients with Head and neck cancer: A feasibility study including the EORTC QLQ-C30. *Otolaring Head Neck Surg* 1997; 116:666-673.
- Harden RN, Carter TD, Gilman CS, Gross AJ, Peters JR. Ketorolac in acute headache management. *Headache* 1991; 31:463-464.
- Heft MW, Gracely RH, Dubner R. Nitrous oxide analgesia: A psychophysical evaluation using verbal descriptor scaling. *J Dent Res* 1982; 63:129-132.
- Holland J, Romano S, Heiligenstein J, Tepner R, Wilson M. A controlled trial of fluoxetine and desipramine in depressed women with advanced cancer. *Psycho-Oncol* 1998; 7:291-300.
- Hollen P, Gralla R, Kris M, Cox C. Quality of life during clinical trials: conceptual model for the Lung Cancer Symptom Scale (LCSS). *Sup Care Cancer* 1994; 2:213-222.
- Hollen P, Gralla R, Kris M, Cox C, Belani C, Grunberg S, Crawford J, Neidhard J. Measurement of quality of life in patients with lung cancer in multicenter trials of new therapies. *Cancer* 1994; 73:2087- 2098.

- Hollen P, Gralla R, Kris M, Potanovich L. Quality of life assessment in individuals with lung cancer: Testing the Lung Cancer Symptom scale (LCSS). *Eur J Cancer* 1993;29 (Suppl. 1):S51-S58.
- Holzheimer A, McMillan SC, Weitzner M. Improving pain outcomes of hospice patients with cancer. *Oncol Nurs Forum* 1999;26:1499-1504.
- Hwang SS, Chang VT, Kasimis B. Cancer breakthrough pain characteristics and responses to treatment at a VA medical center. *Pain* 2003; 101:55-64.
- Jamison RN, Brown GK. Validation of hourly pain intensity profiles with chronic pain patients. *Pain* 1991; 45:123-128.
- Jamison RN, Sbrocco T, Parris WC. The influence of physical and psychological factors on accuracy of memory for pain in chronic pain patients. *Pain* 1989; 37:289-294.
- Jenkinson C, Carrol D, Egerton M, Frankland T, McQuay H., Nagle C. Comparison of the sensitivity to change of long and short form pain measures. *Qual Life Res* 1995; 4:353-357.
- Jensen MP, Chen C, Brugger AM. Postsurgical pain outcome assessment. *Pain* 2002; 99:101-109.
- Jensen, MP, Karoly P. Assessing the subjective experience of pain: What do the scale scores of the McGill Pain Questionnaire measure? Poster presented at the eighth annual scientific sessions of the Society of Behavioral Medicine, Washington, D.C. 1987.
- Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. *Pain* 1986; 27:117-126.
- Jensen MP, Karoly P, O'Riordan EF, Bland F Jr, Burns RS. The subjective experience of acute pain: An assessment of the utility of 10 indices. *Clin J Pain* 1989; 5:153-159.
- Joyce CRB, Zutshi DW, Hrubes V, Mason RM. Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur J Clin Pharm* 1975; 8:415-420.
- Jensen MP, McFarland C. Increasing the reliability and validity of pain intensity measurement in chronic pain patients. *Pain* 1993; 55:195-203.
- Jensen MP, Miller L, Fisher LD. Assessment of pain during medical procedures: a comparison of three scales. *Clin J Pain* 1998; 14:343-349.
- Jensen MP, Smith DG, Ehde DM, Robinson LR. Pain site and the effects of amputation pain: Further clarification of the meaning of mild, moderate, and severe pain. *Pain*, 2001; 91:317-322.
- Jensen MP, Turner LR, Turner JA, Romano JM. The use of multiple-item scales for pain intensity measurement in chronic pain patients. *Pain* 1996; 67:35-40.
- Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999; 83:157-162.
- Ingham J, Seidman A, Yao T-J, Lepore J, Portenoy R. An exploratory study of frequent pain measurement in a cancer clinical trial. *Qual Life Res* 1996; 5:503-507.
- Kaasa S, Bjordal K, Aaronson N, Moum T, Wist E, Hagen S, Kvikstad A. The EORTC Study Group on Quality of Life. The EORTC Core Quality of Life Questionnaire (QLQ-C30): Validity and reliability when analyzed with patients treated with palliative radiotherapy. *Eur J Cancer* 1995; 31:2260-2263.

- Keefe FJ, Affleck G, Lefebvre JC, Starr K, Caldwell DS, Tennen H. Pain coping strategies and coping efficacy in rheumatoid arthritis: a daily process analysis. *Pain* 1997; 69:35-42.
- Keefe FJ, Schapira B, Williams RB, Brown C, Surwit RS. EMG-assisted relaxation training in the management of chronic low back pain. *Am J Clin Biofeedback* 1981; 4: 93-103.
- King RB. Topical aspirin in chloroform and the relief of pain due to herpes zoster and postherpetic neuralgia. *Arch Neurol* 1993; 50:1046-1053.
- Klepstad P, Loge JH, Borchgrevink PC, Mendoza TR, Cleeland CS, Kaasa S. The Norwegian Brief pain Inventory Questionnaire: translation and validation in cancer pain patients. *J Pain Sympt Manage* 2002; 24:517-535.
- Klepstad P, Borchgrevink PC, Kaasa S. Effects on cancer patients' health-related quality of life after the start of morphine therapy. *J Pain Sympt Manage* 2000; 20:19-26.
- Kremer EF, Atkinson JH, Jr, Ignelzi RJ. Measurement of pain: Patient preference does not confound measurement. *Pain* 1981; 10:241-248.
- Kremer EF, Atkinson JH, Jr, Ignelzi RJ. Pain measurement: the affective dimensional measure of the McGill Pain Questionnaire with a cancer pain population. *Pain* 1982; 12:153-163.
- Kucuk O, Fisher E, Moinpour CM, Coleman D, Hussain MHA, Sartor AO, Chatta GS, Lowe BA, Eisenberger MA, Crawford ED. Phase II trial of bicalutamide in patients with advanced prostate cancer in whom conventional hormonal therapy failed: a southwest oncology group study (SWOG 9235). *Urology* 2001; 58:53-58.
- Leavitt F, Garron DC. Validity of a back pain classification scale for detecting psychological disturbance as measured by the MMPI. *J Clin Psych* 1980; 36:186-189.
- Leksowski K. Thoracoscopic splanchnicectomy for control of intractable pain due to advanced pancreatic cancer. *Surg Endosc* 2001; 15:129-131.
- Levine FM, De Simone LL. The effects of experimenter gender on pain report in male and female subjects. *Pain* 1991; 44:69-72.
- Lin CC. Applying the Am Pain Society QA standards to evaluate the quality of pain management among surgical, oncology, and hospice inpatients in Taiwan. *Pain* 2000; 87:43-49.
- Lingard EA, Wright EA, Sledge CB, Kinemax Outcomes Group. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Joint Surg* 2001; 83:1149-1156.
- Linton SJ. Memory for chronic pain intensity: correlates of accuracy. *Percept Motor Skills* 1991; 72:1091-1095.
- Linton SJ, Göttestam KG. A clinical comparison of two pain scales: correlation, remembering chronic pain, and a measure of compliance. *Pain* 1983; 17:57-65.
- Littman GS, Walker BR, Schneider BE. Reassessment of verbal and visual analog ratings in analgesic studies. *Clin Pharm Therap* 1985; 38:16-23.
- Love A, Leboeuf C, Crisp TC. Chiropractic chronic low back pain sufferers and self-report assessment methods. Part I. A reliability study of the visual analogue scale, the pain drawing and the McGill Pain Questionnaire. *J Manipulative Physiol Ther* 1989; 12:21-25.

- Lundeberg T, Lund I, Dahlin L, Borg E, Gustafsson C, Sandin L, Rosén A, Kowalski J, Eriksson SV. Reliability and responsiveness of three different pain assessments. *J Rehabil Med* 2001; 33:279-283.
- Magnusson T, List T, Helkimo M. Self-assessment of pain and discomfort in patients with temporomandibular disorders: a comparison of five different scales with respect to their precision and sensitivity as well as their capacity to register memory of pain and discomfort. *J Oral Rehabil* 1995; 22:549-556.
- Manfredi PL, Chandler S, Pigazzi A, Payne R. Outcome of cancer pain consultations. *Cancer* 2000; 89:920-924.
- Masedo AI, Esteve R. Some empirical evidence regarding the validity of the Spanish version of the McGill pain questionnaire (MPQ-SV). *Pain* 2000; 85:451-456.
- Masson EA, Hunt L, Gem JM, Boulton AJM. A novel approach to the diagnosis and assessment of sympathetic diabetic neuropathy. *Pain* 1989; 38:25-28.
- Maunsell E, Allard P, Dorval M, Labbe J. A brief pain diary for ambulatory patients with advanced cancer. *Cancer* 2000; 88:2387-2397.
- McMillan S, Tittle M, Hagan S, Laughlin J: Management of pain and pain-related symptoms in hospitalized veterans with cancer. *Cancer Nursing* 23:327-336, 2000
- McMillan S, Williams F, Chatfield R, Camp D: A Validity and Reliability Study of Two Tools for Assessing and Managing Cancer Pain. *Oncol Nurs For* 15:735-741, 1998
- Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975; 1:277-299.
- Melzack R., Terrence C, Fromm G, Amsel R. Trigeminal neuralgia and atypical facial pain: Use of the McGill Pain Questionnaire for discrimination and diagnosis. *Pain* 1986; 27:297-302.
- Melzack R. The short-form McGill Pain Questionnaire. *Pain* 1987; 30:191-197.
- Meuser T, Pietruck C, Radbruch L, Stute P, Lehmann KA, Grond S. Symptoms during cancer pain treatment following WHO-guidelines: a longitudinal follow-up study of symptom prevalence, severity and etiology. *Pain* 2001; 93:247-257.
- Molenaar S, Sprangers MA, Rutgers EJ, Luiten EJ, Mulder J, Bossuyt PM, van Everdingen JJ, Oosterveld P, de Haes HC. Decision support for patients with early-stage breast cancer: effects of an interactive breast cancer CDROM on treatment decision, satisfaction, and quality of life. *J Clin Oncol* 2001; 19:1676-1687.
- Mongini F, Italiano M. TMJ disorders and myogenic facial pain: a discriminative analysis using the McGill Pain Questionnaire. *Pain* 2001; 91:323-330.
- Moore MJ, Osoba D, Murphy K, Tannock IF, Armitage A, Findlay B, Coppin C, Neville A, Venner P, Wilson J. Use of palliative end points to evaluate the effects of mitoxantrone and low-dose prednisone in patients with hormonally resistant prostate cancer. *J Clin Oncol* 1994; 12:689-694.
- Mystakidou K, Mendoza T, Tsilika E, Befon S, Parpa E, Bellos G, Vlahos L, Cleeland C. Greek Brief Pain Inventory: validation and utility in cancer pain. *Oncology* 2001; 60:35-42.

- Mystakidou K, Parpa E, Tsilika E, Kalaidopoulou O, Georgali S, Galanos A, Vlahos L. Greek McGill pain questionnaire: validation and utility in cancer patients. *J Pain Symptom Manage* 2002; 24:379-387.
- Naeser MA, Hahn K-AK, Lieberman BE. Carpal tunnel syndrome pain treated with low-level laser and microamperes transcutaneous electric nerve stimulation: a controlled study. *Arch Phys Med Rehabil* 2002; 83:978-988.
- Nikolajsen L., Hansen CL, Nielsen J, Keller J, Arendt-Nielsen L, Jensen TS. The effect of ketamine on phantom pain: A central neuropathic disorder maintained by peripheral input. *Pain* 1996; 67:69-77.
- Norvell K, Gaston-Johansson F, Fridh G. Remembrance of labor pain: how valid are retrospective pain measurements? *Pain* 1987; 31:77-86.
- Nunnally JC. *Psychometric Theory*, 2nd Edition. New York, McGraw-Hill, 1978.
- Ohnhaus JE, Adler R. Methodological problems in the measurement of pain: A comparison between the verbal rating scale and the visual analogue scale. *Pain* 1975; 1:379-384.
- Owen JE, Klapow JC, Casebeer L. Evaluating the relationship between pain presentation and health-related quality of life in outpatients with metastatic or recurrent neoplastic disease. *Qual Life Res* 2000; 9:855-863.
- Padilla GV, Presant C, Grant MM, Metter G, Lipsett J, Heide F. Quality of life index for patients with cancer. *Res Nurs Health* 1983; 6:117-126.
- Paice J, Cohen F. Validity of a verbally administered numeric rating scale to measure cancer pain intensity. *Cancer Nurs* 1997; 20:88-93.
- Palangio M, Northfelt DW, Portenoy RK, Brookoff D, Doyle RT, Dornseif BE, Damask M. Dose conversion and titration with a novel once-daily, OROS osmotic technology, extended-release hydromorphone formulation in the treatment of chronic malignant or nonmalignant pain. *J Pain Symptom Manage* 2002; 23:355-368.
- Perry F, Heller PH, Levine JD. Differing correlations between pain measures in syndromes with or without explicable organic pathology. *Pain* 1988; 34:185-189.
- Perry F, Heller PH, Levine JD. A possible indicator of functional pain: Poor pain scale correlation. *Pain* 1991; 46:191-193.
- Philip BK. Parametric statistics for evaluation of the visual analog scale. *Anesthesia Analgesia* 1990; 71:710.
- Plesh O, Curtis D, Levine J, McCall WD, Jr. Amitriptyline treatment of chronic pain in patients with temporomandibular disorders. *J Oral Rehabil* 2000; 27:834-841.
- Portenoy R, Hanen NA. Breakthrough pain: definition, prevalence, and characteristics. *Pain* 1990; 41:273-281.
- Portenoy RK, Payne R, Coluzzi P, Raschko JW, Lyss A, Busch MA, Frigerio V, Ingham J, Loseth DB, Nordbrock E, Rhiner M. Oral transmucosal fentanyl citrate (OTFC) for the treatment of breakthrough pain in cancer patients: a controlled dose titration study. *Pain* 1999; 79:303-312.

- Poulos AR, Gertz MA, Pankratz VS, Post-White J. Pain, mood disturbance, and quality of life in patients with multiple myeloma. *Oncol Nurs Forum* 2001; 28:1163-1171.
- Pozehl B, Barnason S, Zimmerman L, Nieveen J, Crutchfield J. Pain in the postoperative coronary artery bypass graft patient. *Clin Nurs Res* 1995; 4:208-222.
- Price DD, Harkins, SW, Baker C. Sensory-affective relationship among different types of clinical and experimental pain. *Pain* 1987; 28:297-307.
- Price, DD. Roles of psychophysics, neuroscience, and experimental analysis in the study of pain. In L. Kruger & J. K. Liebeskind (Eds.), *Advances in Pain Research and Therapy*, Vol. 6. New York: Raven Press 1984, pp.341-355
- Price DD, Barber J. An analysis of factors that contribute to the efficacy of hypnotic analgesia. *J Abnormal Psych* 1987; 96:46-51.
- Price DD, Barrell JJ, Gracely RH. A psychophysical analysis of experiential factors that selectively influence the affective dimension of pain. *Pain* 1980; 8:137-149.
- Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain* 1994; 56:217-226.
- Price DD, Harkins SW. Combined use of experimental pain and visual analogue scales in providing standardized measurement of clinical pain. *The Clinical Journal of Pain* 1987; 3:1-8.
- Price DD, Harkins SW, Baker C. Sensory-affective relationships among different types of clinical and experimental pain. *Pain* 1987; 28:297-307.
- Price DD, Harkins SW, Rafii A, Price C. A simultaneous comparison of fentanyl's analgesic effects on experimental and clinical pain. *Pain* 1986; 24:197-203.
- Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 1983; 17:45-56.
- Price DD, Von der Gruen A, Miller J, Rafii A, Price C. A psychophysical analysis of morphine analgesia. *Pain* 1985; 22:261-269.
- Price DD, Riley JL, III, Wade JB. Psychophysical approaches to measurement of the dimensions and stages of pain. In: Turk DC, Melzack R. editors. *Handbook of Pain Assessment*, 2nd Edition. New York: Guildford Press, 2001, pp. 53-75.
- Puntillo KA, Neighbor ML. Two methods of assessing pain intensity in English-speaking and Spanish-speaking emergency department patients. *J Emergency Nurs* 1997; 23:597-601.
- Putzke JD, Richards JS, Hicken BL, Ness TJ, Kezar L, DeVivo M. Pain classification following spinal cord injury: the utility of verbal descriptors. *Spinal Cord* 2002; 40:118-127.
- Radbruch L, Sabatowski R, Loick G, Jonen-Thielemann I, Kasper M, Gondek B, Lehmann KA, Thielemann I. Cognitive impairment and its influence on pain and symptom assessment in a palliative care unit: development of a Minimal Documentation System. *Palliat Med* 2000; 14:266-276.
- Ramer L, Richardson JL, Zichi Cohen M, Bedney C, Lackerdas Danley K, Judge EA. Multimeasure pain assessment in an ethnically diverse group of patients with cancer. *J Trans Nurs* 1999; 10:94-101.

- Rasmussen JL. Analysis of Likert-Scale data: A reinterpretation of Gregoire and Driver. *Psychol Bull* 1989; 105:167-170.
- Rathmell, AJ, Ash DV, Howes M, Nicholls J. Assessing quality of life in patients treated for advanced head and neck cancer. *Clin Oncol* 1991; 3:10-16.
- 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.
- Rice ASC, Maton S. Postherpetic Neuralgia Study Group. Gabapentin in postherpetic neuralgia: a randomized, double-blind, placebo controlled study. *Pain* 2001; 94:215-224.
- Roach KE, Brown MD, Dunigan KM, Kusek CL, Walas M. Test-retest reliability of patient reports of low back pain. *JOSPT* 1997; 26:253-259.
- Rogers SN, Humphris, G, Lowe D, Brown JS, Vaughan JD. The impact of surgery for oral cancer on quality of life as measured by the Medical Outcome Short Form 36. *Oral Oncol* 1998; 34:171-179.
- Rogers SN, Lowe D, Brown JS, Vaughan JD. A comparison between the University of Washington Head and Neck Disease-Specific Measure and the Medical Short Form 36, EORTC QOQ-C33 and EORTC Head and Neck 35. *Oral Oncol* 1998;34:361-372.
- Rofe Y, Algom D. Accuracy of remembering post delivery pain. *Perceptual Motor Skills* 1985;60:99-105.
- Rowbotham M, Harden N, Stacey B, Bernstein P, Magnus-Miller L. Gabapentin for the treatment of postherpetic neuralgia. *JAMA* 1998;280:1837-1842.
- Rybstein-Blinchik E. Effects of different cognitive strategies on chronic pain experience. *J Behav Med* 1979;2:93-101.
- Salovey P, Smith, AF, Turk DC, Jobe JB, Willis GB. The accuracy of memory for pain: not so bad most of the time. *APS J* 1993; 2:184-191.
- Sandouk P, Serrie A, Urtizbera M, Debray M, Got P, Scherrmann JM. Morphine pharmacokinetics and pain assessment in patients with terminal cancer. *Clin Pharmacol Ther* 1991; 49:442-448.
- Sandblom G, Carlsson P, Sigsjo P, Varenhorst E. Pain and health-related quality of life in a geographically defined population of men with prostate cancer. *Br J Cancer* 2001; 85:497-503.
- Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the functional living index-cancer: Development and validation. *J Clin Oncol* 1984; 2:472-483.
- Schwartz L, Engel JM, Jensen MP. Pain in persons with cerebral palsy. *Arch Phys Med Rehab* 1999;80:1243-1246.
- Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain* 1995; 61:277-284.
- Serrao JM, Marks RL, Morley SJ, Goodchild CS. Intrathecal midazolam for the treatment of chronic mechanical low back pain: a controlled comparison with epidural steroid in a pilot study. *Pain* 1992; 48:5-12.

- Seymour RA. The use of pain scales in assessing the efficacy of analgesics in post-operative dental pain. *Eur J Clin Pharm* 1982; 23:441-444.
- Seymour RA, Charlton JE, Phillips ME. An evaluation of dental pain using visual analogue scales and the McGill Pain Questionnaire. *J Oral Maxillofac Surg* 1983; 41:643-648.
- Seymour RA, Simpson JM, Charlton JE, Phillips ME. An evaluation of length and end-phrase of visual analogue scales in dental pain. *Pain* 1985; 21:177-185.
- Shacham S, Dar R, Cleeland CS. Relationship of mood state to the severity of clinical pain. *Pain* 1984; 18:187-197.
- Shannon MM, Ryan MA, D'Agostino N, Brescia FJ. Assessment of pain in advanced cancer patients. *J Pain Sympt Manag* 1995; 10:274-278.
- Shiffmann R, Kopp JB, Austin HA, Sabnis S, Moore DF, Weibel T, Balow JE, Brady RO. Enzyme replacement therapy in Fabry disease: a randomized controlled trial. *JAMA* 2001; 285:2743-2749.
- Singer AJ, Kowalska A, Thode HC. Ability of patients to accurately recall the severity of acute painful events. *Acad Emerg Med* 2001; 8:292-295.
- Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, Chambers WA, Cairns Smith W. The chronic pain grade questionnaire: validation and reliability in postal research. *Pain* 1997; 71:141-147.
- Smith WB, Safer MA. Effects of present pain level on recall of chronic pain and medication use. *Pain* 1993; 55:355-361.
- Smith EL, Hann DM, Ahles TA, Furstenberg CT, Mitchell TA, Meyer L, Maurer LH, Rigas J, Hammond S. Dyspnea, anxiety, body consciousness, and quality of life in patients with lung cancer. *J Pain Symptom Manage* 2001; 21:323-329.
- Smith WB, Gracely RH, Safer MA. The meaning of pain: cancer patients' rating and recall of pain intensity and affect. *Pain* 1998; 78:123-129.
- Sneeuw KCA, Aaronson NK, Osoba D, Miller MJ, Hsu M-A, Yung A, Brada M, Newlands ES. The use of significant others as proxy raters of the quality of life of patients with brain cancer. *Med Care* 1997; 35:490-506.
- Sneeuw KCA, Aaronson NK, Spranger MAG, Detmar SB, Wever LDV, Schornagel JH. Evaluating the quality of life of cancer patients: assessments by patients, significant others, physicians and nurses. *Br J Cancer* 1999; 81:87-94.
- Soh G, Ang, HG. Comparison of two pain rating scale among Chinese cancer patients. *Chin Med J* 1997; 105:953-956.
- Sriwatanakul K, Kelvie W, Lasagna L, Calimlim JF, Weis OF, Mehta G. Studies with different types of visual analog scales for measurement of pain. *Clin Pharmacol Ther* 1983; 34:234-239.
- Stahmer SA, Shofer FS, Marino A, Shepherd S, Abbuhl S. Do quantitative changes in pain intensity correlate with pain relief and satisfaction. *Acad Emerg Med* 1998; 5:851-857.
- Stambough JE, Sarajian C. Analgesic efficacy of zomepirac sodium in patients with pain due to cancer. *J Clin Pharm* 1981; 21:501-507.

- Steiner TJ, Lange R, Mvoelker. Aspirin in episodic tension-type headache: placebo-controlled dose ranging comparison with paracetamol. *Cephalgia* 2003; 23:59-66.
- Stelian J, Gil I, Habot B, Rosenthal M, Abramovici I, Kutok N, Khahil A. Improvement of pain and disability in elderly patients with degenerative osteoarthritis of the knee treated with narrow-band light therapy. *J Am Geriatr Soc* 1992; 40:23-26.
- Stenn PG, Mothersill KJ, Brooke RI. Biofeedback and a cognitive behavioral approach to treatment of myofascial pain dysfunction syndrome. *Behavior Therapy* 1979; 10:29-36.
- Stockler MR, Osoba D, Goodwin P, Corey P, Tannock IF. Responsiveness to change in health-related quality of life in a randomized clinical trial: a comparison of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) with analogous scales from the EORTC QLQ-C30 and a trial specific module. *J Clin Epidemiol* 1998; 51:137-145.
- Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *BMJ* 2002; 324:1193-1194.
- Sze FK, Chung TK, Wong E, Lam KK, Lo R, Woo J. Pain in Chinese cancer patients under palliative care. *Pall Med* 1998; 12:271-277.
- Talmi YP, Waller A, Bercovici M, Horowitz Z, Pfeffer MR, Adunski A, Kronenberg J. Pain experienced by patients with terminal head and neck carcinoma. *Cancer* 1997; 80:1117-1123.
- Tannock I, Gospodarowicz M, Meakin W, Panzarella T, Stewart L, Rider W. Treatment of metastatic prostatic cancer with low-dose prednisone. Evaluation of pain and quality of life as pragmatic indices of response. *J Clin Oncol* 1989; 7:590-597.
- Tannock IF, Osoba D, Stockler MR, Ernst D, Neville AJ, Moore MJ, Armitage GR, Wilson JJ, Venner PM, Coppin CML, Murphy KC. Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: A Canadian randomized trial with palliative end points. *J Clin Oncol* 1996; 14:1756-1764.
- Tasmuth T, Estlanderb A-M, Kalso E. Effect of present pain and mood on the memory of past postoperative pain in women treated surgically for breast cancer. *Pain* 1996; 68:343-347.
- Tesfaye S, Watt J, Benbow SJ, Pang KA, Miles J, MacFarlane IA. Electrical spinal-cord stimulation for painful diabetic peripheral neuropathy. *Lancet* 1996; 348:1698-1701.
- Thie NMR, Prasad NG, Major PW. Evaluation of glucosamine sulfate compared to ibuprofen for the treatment of temporomandibular joint osteoarthritis: a randomized double blind controlled 3 month clinical trial. *J Rheum* 2001; 28:1347-1355.
- Thomas V, Heath M, Rose D, Flory P. Psychological characteristics and the effectiveness of patient-controlled analgesia. *Brit J Anaesthesia* 1995; 74:271-276.
- Trotti A, Johnson DJ, Gwede C, Casey L, Sauder B, Cantor A, Pearlman J. Development of a head and neck companion module for the quality of life-radiation therapy instrument (QOL-RTI). *Int J Rad Oncol -Biology-Physics* 1998; 42:257-261.
- Turk DC, Rudy TE, Salovey P. The McGill Pain Questionnaire reconsidered: Confirming the factor structure and examining appropriate uses. *Pain* 1985; 21:385-397.

- Turner JA. Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low back pain. *J Cons Clin Psych* 1982; 50:757-765.
- Tursky B, Jamner LD, Friedman R. The pain perception profile: a psychophysiological approach to the assessment of pain report. *Behav Ther* 1982; 13:376-394.
- Uki J, Mendoza T, Cleeland CS, Nakamura Y, Takeda F: A brief cancer pain assessment tool in Japanese: The utility of the Japanese Brief Pain Inventory – BPI-J. *J Pain Sympt Manag* 16:364-373, 1998
- Urban BJ, Keefe FJ, France RD. A study of psychophysical scaling in chronic pain patients. *Pain* 1984;20:157-168.
- Valdix SW, Puntillo KA. Pain, pain relief and accuracy of their recall after cardiac surgery. *Progress Cardiovascular Nurs* 1995; 10:3-11.
- Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain* 1992; 50:133-149.
- Wallenstein SL. The VAS relief scale and other analgesic measures. *Adv Pain Res Ther* 1991; 18:97-103.
- Wallenstein SL, Heidrich GIII, Kaiko R, Houde RW. Clinical evaluation of mild analgesics: The measurement of clinical pain. *Brit J Clin Pharm* 1980;10:319S-327S.
- Walsh TD, Leber B. Measurement of chronic pain: Visual analog scales and McGill Melzack pain questionnaire compared. *Adv Pain Res Ther* 1983;5:897-899.
- Wang XS, Cleeland CS, Mendoza TR, Engstrom MC, Liu S, Xu G, Hao X, Wand Y, Ren XC. The effect of pain severity on health-related quality of life. *Cancer* 1999; 86:1848-1855.
- Wilkie DJ, Kampbell J, Cutshall S, Halabisky H, Harmon H, Johnson LP, Weinacht L, Rake-Marona M. Effects of massage on pain intensity, analgesics and quality of life in patients with cancer pain: a pilot study of a randomized clinical trial conducted within hospice care delivery. *Hosp J* 2000;15:31-53.
- Wilkie DJ, Keefe FJ, Dodd MJ, Copp LA. Behavior of patients with lung cancer: description and associations with oncologic and pain variables. *Pain* 1992; 51:231-240.
- Wilkie D, Lovejoy N, Dodd M, Tesler M: Cancer pain intensity measurement: concurrent validity of three tools – Finger dynamometer, Pain Intensity Number Scale, Visual Analogue Scale. *Hosp J* 1990; 6:1-13.
- Williams AC de C, Oakley Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. *Pain* 2000;85:457-463.
- Zalon M.L. Comparison of pain measures in surgical patients. *J Nurs Measur* 1999; 7:135-152.
- Zeppetella G: An assessment of the safety, efficacy, and acceptability of intranasal fentanyl citrate in the management of cancer-related breakthrough pain: a pilot study. *J Pain Symptom Manage* 2000; 20:253-258.

Table 1. The Strengths and Weaknesses of Three Measures of Pain Intensity

Scale	Strengths	Weaknesses
Visual Analogue Scale	<ul style="list-style-type: none"> -Many ("infinite") response categories. -Average (group) scores can be treated as ratio data. -Good evidence for validity. 	<ul style="list-style-type: none"> -Extra step in scoring the paper-and-pencil version can take more time and adds an additional source of error. -Some people, especially older people, have difficulty using VASs.
Numerical Rating Scale	<ul style="list-style-type: none"> -Easy to administer. -Many response categories if NRS-101 chosen; adequate Number of response categories if 0 – 10 NRS is chosen. - Easy to score. -Good evidence for validity. -Compliance with measurement task is high. 	<ul style="list-style-type: none"> -Average (group) scores cannot necessarily be treated as ratio data.
Verbal Rating Scale	<ul style="list-style-type: none"> -Easy to administer. -Easy to score. -Good evidence for validity. -Compliance with measurement task is high. -May approximate ratio scaling if CMM methods (or scores developed from CMM methods) are used. 	<ul style="list-style-type: none"> -Can be difficult for persons with limited vocabulary. -Relatively few response categories compared to the VAS or NRS. -If scored using the ranking method, the scores do not necessarily have ratio qualities. -People forced to choose one word, even if no word on the scale adequately describes their pain intensity.

Table 2. Summary of the Primary Research Findings on Measures of Pain Relief

-
- Relatively little research has compared VAS, NRS, and VRS pain relief measures to each other.
 - Pain relief measures are sensitive (sometimes more so than pain intensity measures) to the effects of pain treatments.
 - Measures of pain relief are statistically distinct from measures of pain intensity:
 - Pain relief is sometimes endorsed even when pain changes little or worsens.
 - Perceived pain relief is more strongly associated than change in pain intensity with treatment satisfaction.
-

Table 3. Summary of the Primary Research Findings on Measures of the Temporal Aspects of Pain

-
- Dimensions
 - Variability
 - Frequency
 - Duration
 - Pattern
 - “Breakthrough” pain
 - Time to analgesia onset/time to meaningful pain relief
 - Temporal pain qualities are distinct from pain intensity.
 - Temporal pain qualities may predict patient function over and above effects of pain intensity.
 - Measures of temporal qualities are under-utilized in pain clinical trials.
-

Table 4. Summary of the Primary Research Findings on Measures of Pain Qualities (Including Pain Affect)

-
- Measures of pain quality and affect used relatively infrequently in pain clinical trials.
 - Evidence supports the validity of the MPQ and SF-MPQ as outcome measures.
 - But both MPQ and SF-MPQ appear to be less sensitive than measures of pain intensity to changes in pain.
 - Use of MPQ and SF-MPQ scale scores obscured the specific qualities of pain.
 - MPQ is probably not practical in most clinical trials, but SF-MPQ appears to be.
 - More research is needed to examine the psychometric qualities of the SF-MPQ and other possible pain quality measures.
-

Table 5. Pain Assessment Recommendations

Pain intensity

- In most trials, a measure of pain intensity is the appropriate primary outcome dimension.
- 0 – 10 NRS-I appears to have the most strengths and fewest weaknesses of pain intensity measures.
- VRS-4-I (none, mild, moderate, severe) may be a useful secondary measure.

Pain relief

- Should be strongly considered as a secondary outcome measure in pain clinical trials.
- No strong evidence to support one type of pain relief measures (VAS, NRS, VRS) over the others, although concerns raised about VAS-I may encourage investigators to select a NRS (e.g., 0 = none; 10 = complete) or VRS (e.g., none, a little, some, a lot, complete relief) over a VAS for this purpose.

Temporal aspect of pain

- Temporal aspects of pain should be strongly considered as a secondary outcome dimensions in pain clinical trials.
- Temporal dimension selected should be consistent with the expected effects of treatment:
 - Time to analgesia onset/Time to meaningful pain relief for fast-acting analgesics is appropriate.
 - Presence/absence, intensity, and frequency of breakthrough pain for BP treatments.
 - Frequency of pain for treatment of intermittent pain problems (e.g., headache).

Qualitative aspects of pain (including affective quality[ies])

- Should be considered as secondary outcome measure(s) in pain clinical trials.
- SF-MPQ appears the most useful measure of pain qualities. SF-MPQ construct validity might be improved by adding descriptors, and by increasing response levels (e.g., from 4 to 11). The SF-MPQ's strongest asset (ability to detect changes in specific pain qualities) has been under-utilized in clinical trials.
- Single-item measures (VAS-A, NRS-A, VRS-A) may provide a useful summary measure of pain affect. VRS-A may be more effective than VAS-A or NRS-A for helping subjects distinguish between pain affect and intensity.

Table 6. Issues in Pain Assessment in Clinical Trials

How often and for how long should pain be measured?

At a minimum, pain needs to be assessed both before and after the treatment conditions in a clinical trial. Ideally, more assessment points should be included, extended to beyond the point at which the experimental intervention is thought to be effective, to allow for comparisons between treatment conditions concerning the pattern of effects of the interventions on pain.

When measuring pain intensity, should study participants only be asked to rate their current pain, perhaps on multiple occasions, or can recalled pain (e.g., worst, least, and average over a specified period of time) can trusted?

The evidence suggests that recall measures are not specifically accurate, but are valid (i.e., they reflect) measures of previous pain. They can therefore be used as treatment outcome variables, eliminating the need for repeated (e.g., daily diary) measures in situations where “average” or “usual” pain is the primary outcome dimension (e.g., most chronic pain studies). However, no studies have compared the relative sensitivity of single ratings of previous pain versus diary averages; it is possible that recall ratings may be less sensitive in some situations.

To what extent should unsupervised (e.g., at-home) pain diaries be used, and what measures, if any, do investigators need to take to ensure that these are completed as instructed?

If diary data are needed, the veracity of unsupervised data collection can be called into question; the findings from such data should be considered preliminary and not conclusive.

Can single-item measures be used exclusively, or are there any situations when composite measures of pain would be appropriate?

The evidence indicates that single-item measures are adequately valid and reliable for most situations; composite measures may increase the reliability and validity of pain assessment a little, on average, but perhaps not enough to warrant a requirement or recommendation that they always be used. Future research is needed to replicate this conclusion, which is based on a relatively few number of studies.

To what extent can, or should, rescue dose requests be used as outcome measures in pain clinical trials?

The incidence of rescue dose requests should be strongly considered as one of the secondary outcome measures when appropriate; and the use of such measures is probably appropriate in nearly all analgesic clinical trials.

Table continues

Table 6. Issues in Pain Assessment in Clinical Trials (continued)

How should the fact that many patients with pain problems experience multiple pain complaints be taken into account when assessing pain in clinical trials?

The extent to which subjects with multiple pain problems provide questionable responses to single-item pain measures, and the impact of this on the findings of clinical trials, is unclear. This potential problem and confound needs to be examined further among pain populations with a high incidence of multiple pain problems. In the meantime, investigators would do well to consider assessing pain in multiple sites at each assessment point at each assessment point.

Should (or is it practical for) there be standardization in the format of and endpoints for pain intensity and affect measures in clinical trials?

It may be time to consider making specific recommendations for standardized pain measures in clinical trials.

Table 7. Pain Assessment Research Recommendations

Pain intensity

- Individual versus composite measures.
- Relative sensitivity of actual usual pain versus recalled usual pain intensity.

Pain relief

- What contributes to a pain relief score in addition to change in pain?
 - Hope engendered by treatment?
 - Area under the curve (SPID)?
 - Change in pain qualities?
 - Other?

Temporal aspect of pain

- Are recall measures of temporal aspects reliable and valid?
- Additional brief and psychometrically sound measures of temporal components should be developed

Qualitative aspects of pain (including affective quality[ies])

- Are single-item measures adequate?
- Relative sensitivity of single-item versus multiple-item scales.
- Can SF-MPQ be improved?
 - More efficient – drop descriptors rarely used.
 - More content validity – add descriptors frequently used.
 - More sensitive – increase number of levels.