Pain 94 (2001) 149–158

PAIN

www.elsevier.com/locate/pain

Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale

John T. Farrar[a,*], James P. Young Jr.[b], Linda LaMoreaux[b], John L. Werth[b], R. Michael Poole[b]

[a]*Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Blockley Hall, Room 816, 423 Guardian Drive, Philadelphia, PA 19104, USA*

[b]*Pfizer Global Research and Development, Ann Arbor Laboratories, Ann Arbor, MI, USA*

Received 26 January 2001; received in revised form 16 April 2001; accepted 18 May 2001

Abstract

Pain intensity is frequently measured on an 11-point pain intensity numerical rating scale (PI-NRS), where 0 = no pain and 10 = worst possible pain. However, it is difficult to interpret the clinical importance of changes from baseline on this scale (such as a 1- or 2-point change). To date, there are no data driven estimates for clinically important differences in pain intensity scales used for chronic pain studies. We have estimated a clinically important difference on this scale by relating it to global assessments of change in multiple studies of chronic pain. Data on 2724 subjects from 10 recently completed placebo-controlled clinical trials of pregabalin in diabetic neuropathy, postherpetic neuralgia, chronic low back pain, fibromyalgia, and osteoarthritis were used. The studies had similar designs and measurement instruments, including the PI-NRS, collected in a daily diary, and the standard seven-point patient global impression of change (PGIC), collected at the endpoint. The changes in the PI-NRS from baseline to the endpoint were compared to the PGIC for each subject. Categories of 'much improved' and 'very much improved' were used as determinants of a clinically important difference and the relationship to the PI-NRS was explored using graphs, box plots, and sensitivity/specificity analyses. A consistent relationship between the change in PI-NRS and the PGIC was demonstrated regardless of study, disease type, age, sex, study result, or treatment group. On average, a reduction of approximately two points or a reduction of approximately 30% in the PI-NRS represented a clinically important difference. The relationship between percent change and the PGIC was also consistent regardless of baseline pain, while higher baseline scores required larger raw changes to represent a clinically important difference. The application of these results to future studies may provide a standard definition of clinically important improvement in clinical trials of chronic pain therapies. Use of a standard outcome across chronic pain studies would greatly enhance the comparability, validity, and clinical applicability of these studies.

Keywords: Analgesics; Human; Numeric rating scale; Pain measurement; Clinical trials; Treatment outcome

1. Introduction

The primary goal of any clinical trial is to evaluate the potential beneficial effect of a therapy. Statistically significant results are necessary but not sufficient to show a difference in effect between the two study groups. The clinical importance of the effect must also be demonstrated to make the results of the trial relevant to patient care. However, because of the subjective nature of pain, clinical importance is not always easy to determine (Farrar et al., 2001). Patients interpret measurement scales very differently when reporting pain and baseline scores can vary widely. This is especially true for scales that do not have any intrinsic meaning, such as the widely used 0–10 numeric rating scale (NRS) (Turk et al., 1993). To compensate for this variability, measures of improvement usually adjust for the individual's baseline by calculating raw change or percent change. Even so, without additional data it is difficult to evaluate the clinical importance of a numeric change, such as a 1- or 2-point decrease on a 0–10-point scale.

Since the NRS is a standard instrument in chronic pain studies, it has become important to define the level of change that best represents a clinically important improvement. To date, the criteria for this level of change have usually been determined based on face validity (Moore et al., 1996) or expert opinion (Goldsmith et al., 1993) since the data necessary for an analytic determination have been lacking. A recent publication defined these criteria based on data from a study of patients with acute pain (Farrar et al.,

* Corresponding author. Tel.: +1-215-898-5802; fax: +1-215-573-5315.
E-mail address: jfarrar@cceb.med.upenn.edu (J.T. Farrar).

0304-3959/01/$20.00
PII: S0304-3959(01)00349-9

2001), but we know of no comparable study of chronic pain therapy.

A series of ten recently completed studies of pregabalin treatment for chronic pain provided an appropriate set of data for such an analysis. The studies used a common design with identical pain measures and covered a wide range of indications in both neuropathic and non-neuropathic chronic pain, with a total enrollment of 2879 patients. In this paper, we present an analysis to determine the change in a pain intensity numeric rating scale (PI-NRS) that is most closely associated with improvement on the commonly used and validated measure of the patient's global impression of change (PGIC). This value should be useful in designing future studies and the determination of appropriate sample sizes. Such information also will facilitate the comparison of results across studies and help in determining the value of a therapy in clinical practice.

2. Methods

We examined the data for all patients enrolled in 10 double-blind, placebo-controlled, parallel, multi-center chronic pain studies that utilized the same study design and procedures. Outcome measures included a 0–10 PI-NRS as in Fig. 1A, and a PGIC scale, a 7-point categorical scale, as in Fig. 1B. Throughout each study, patients kept a daily diary in which they circled the number from 0 = 'no pain' to 10 = 'worst possible pain' that best described their pain over the preceding 24 h. The baseline score was computed as the mean of the seven diary entries prior to taking study medication and the endpoint score was the mean of the last seven diary entries while receiving study medication. At the endpoint of each study, patients completed the PGIC. In addition, the physician completed a clinical global impression of change (CGIC). The relationship between the PGIC and CGIC was examined using the Spearman rank correlation coefficient.

For each patient, we computed the raw change in the PI-NRS score by subtracting the baseline from the endpoint. Patients were stratified by the PGIC categories, and the mean raw change and percent change (raw change/baseline × 100) were calculated within each stratum by study. Since few patients chose 'very much worse' or 'much worse' on the PGIC, these categories were combined for our analysis. The same analyses were repeated separately comparing patients who received placebo to those who received active drug, males and females, patients in different age groups (18–49, 50–59, 60–69, 70 +), and patients with different baseline pain scores. In addition, descriptive statistics were calculated within each stratum for all patients combined.

To better characterize the association between specific PI-NRS change scores and clinically important improvement, the sensitivity and specificity were calculated and receiver operating characteristic (ROC) curves were derived using logistic regression analyses (Hanley, 1989). For each analysis, clinical importance served as the dependent variable and either the raw or percent change scores served as the independent variable. Our a priori definition of clinical importance was the PGIC category of 'much improved' or better. However, since this definition is arbitrary, we also calculated the PI-NRS changes best associated with 'minimally improved' or better, and 'very much improved' alone.

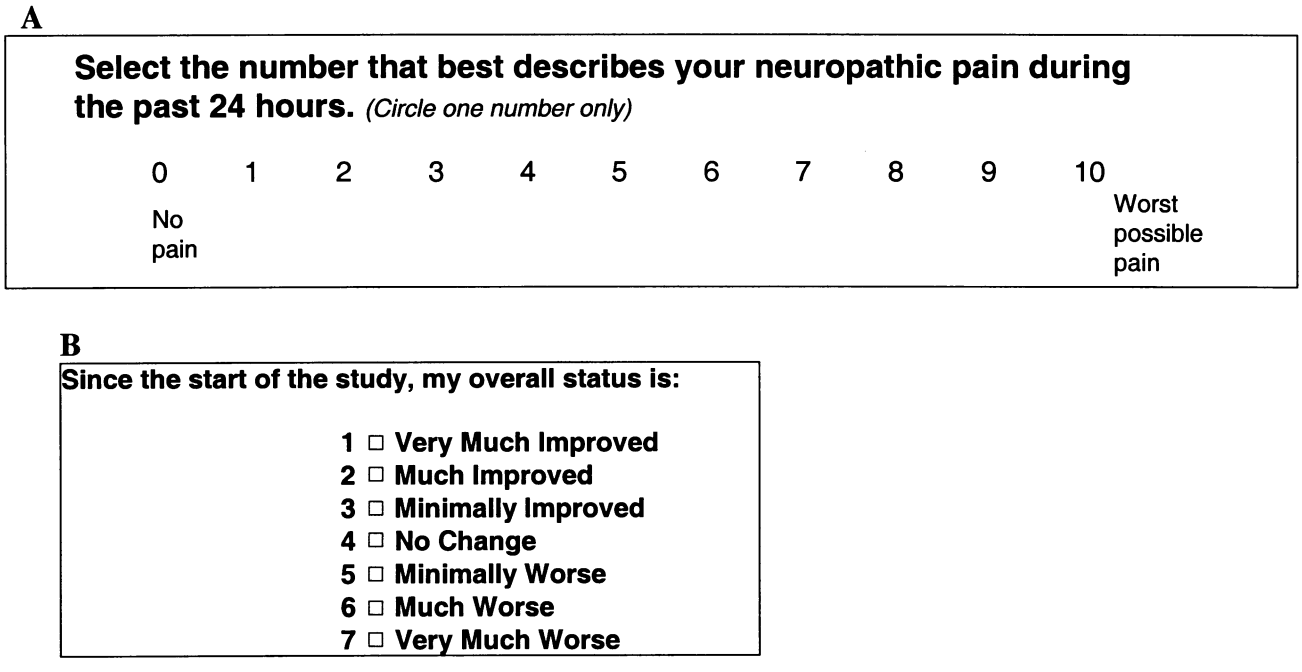

Fig. 1. (A) PI-NRS; Daily Pain Diary. (B) PGIC which was completed at study end.

ROC curves simultaneously describe the sensitivity and specificity of a predictive measure as different cutoff values are applied. In this case, the ROC curve describes the sensitivity and specificity of particular change (raw or percent change) values in predicting each definition of clinical importance. Values for the PI-NRS changes that were best associated with each of the categories were generated from the ROC curves, assuming equal importance of sensitivity and specificity. The value is defined by the intersection of a 45° tangent line with each ROC curve, which is mathematically equivalent to choosing the point at which sensitivity and specificity are the closest to being equal. The area under the ROC curve, reported as the *c*-statistic from the logistic regression, represents the total overall association between the PI-NRS and PGIC category used to construct the specific curve.

3. Results

The data are primarily summarized and presented in graphical format with specific values presented in tabular format. Since the PGIC and CGIC are so closely correlated (Spearman correlation = 0.87), and since the primary goal was to determine a clinically important difference from the patient's perspective, we present only the PGIC analyses. In the graphical displays, the mean values for the PI-NRS changes and percent changes within each stratum are connected by a line between each PGIC category in order to visualize the profile for each group analyzed. The distribution of the combined data are summarized and presented in box plot format.

3.1. The individual study model

A summary of the characteristics of the 10 studies is presented in Table 1. The studies ranged in duration from 5 to 12 weeks and each study enrolled between 146 and 529 subjects. Of the total of 2879 subjects, 2724 patients had scores recorded for baseline pain intensity, endpoint pain intensity and PGIC, and thus could be included in the analyses.

Patients from all treatment groups, including placebo, were combined for each of the 10 studies presented in Figs. 2 and 3. For all 10 studies, a similar relationship was observed between the change in PI-NRS and each PGIC category. Fig. 2 shows that, on average, decreases from baseline of two or more units were associated with the PGIC category of 'much improved', while decreases of at least four units generally corresponded to 'very much improved'. Fig. 3 shows the same relationship for the percent change in PI-NRS, with an association of changes of 30 and 50%, respectively. These associations, as measured by either raw or percent change, held across all ten studies regardless of differences in disease model, trial duration, and patient demographic characteristics. In addition, the relationship was consistent whether or not the drug was shown to be effective in a particular trial.

3.2. Active vs. placebo treatment

Fig. 4 displays the same data comparing active vs. placebo treatment groups, combined across all studies. The graph shows that placebo and active patients had nearly identical mean percent change scores associated with each level of the PGIC, indicating a consistent clinical interpretation of changes on the PI-NRS no matter which treatment they received.

3.3. Gender and age

Similar displays were constructed stratifying the data by gender (Fig. 5) and by age groups (18–49, 50–59, 60–69, 70 +) (Fig. 6). In both the cases, the relationship between PI-NRS percent change scores and PGIC category was nearly identical for each stratum.

3.4. Baseline pain

A comparison of the analysis of the PI-NRS change score, stratified by the baseline pain level is presented in Fig. 7 (raw change) and Fig. 8 (percent change). As would be expected, higher baseline pain scores necessitated larger raw change scores to achieve the same level of PGIC,

Table 1
Study and patient characteristics

Study number	Indication	Length (weeks)	Number of patients	Baseline pain mean (SD)	Age in years mean (SD)	Percent women
1	Diabetic peripheral neuropathy	6	246	6.7 (1.5)	57.1 (9.6)	39
2	Diabetic peripheral neuropathy	5	337	6.4 (1.4)	60.0 (10.5)	40
3	Postherpetic neuralgia	5	255	6.6 (1.6)	71.7 (9.2)	49
4	Osteoarthritis	12	296	6.4 (1.3)	60.9 (9.3)	62
5	Chronic low back pain	7	253	6.2 (1.4)	50.6 (13.5)	51
6	Chronic low back pain	8	406	6.5 (1.4)	51.0 (13.5)	52
7	Fibromyalgia	8	529	7.0 (1.3)	48.6 (10.6)	91
8	Postherpetic neuralgia	8	173	6.3 (1.5)	71.5 (10.9)	54
9	Diabetic peripheral neuropathy	8	146	6.3 (1.6)	59.6 (11.4)	44
10	Postherpetic neuralgia	8	238	6.8 (1.6)	72.2 (10.2)	55

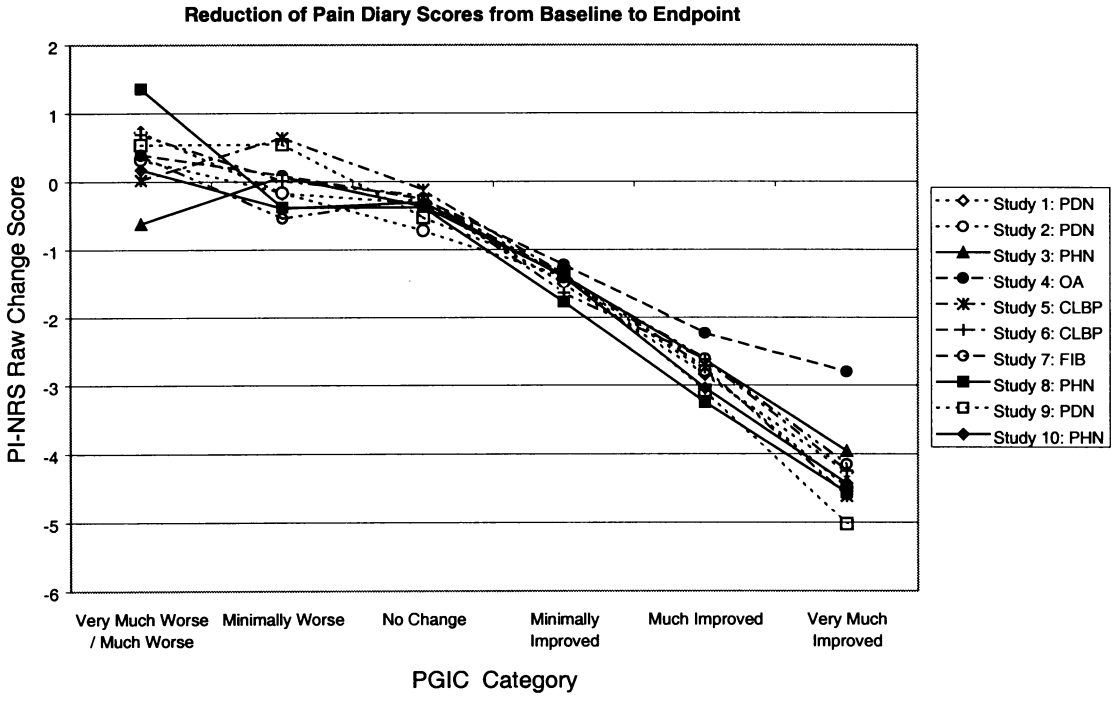

Fig. 2. Stratification by study – raw change (baseline to endpoint) in PI-NRS compared to PGIC assessment recorded at endpoint. PDN, peripheral diabetic neuropathy; PHN, postherpetic neuralgia; OA, osteoarthritis of the hip or knee; CLBP, chronic low back pain and FIB, fibromyalgia. Note: 'very much worse' and 'much worse' are combined because of low numbers in these groups.

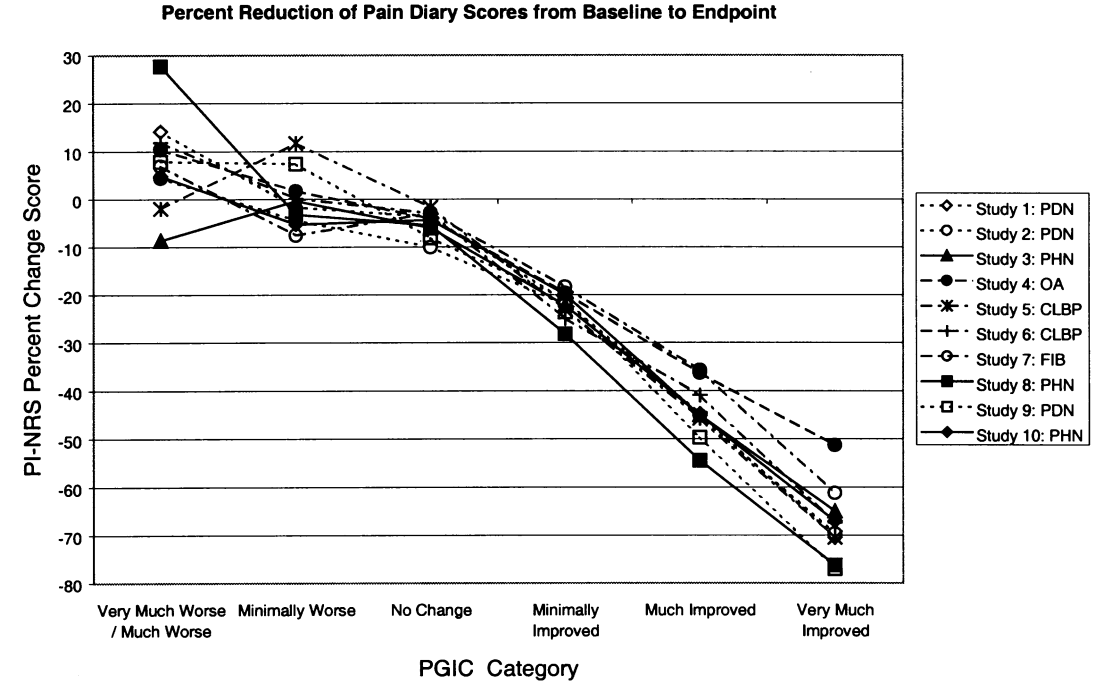

Fig. 3. Stratification by study – percent change (baseline to endpoint) in PI-NRS compared to PGIC assessment recorded at endpoint. PDN, peripheral diabetic neuropathy; PHN, postherpetic neuralgia; OA, osteoarthritis of the hip or knee; CLBP, chronic low back pain and Fib, fibromyalgia. Note: 'very much worse' and 'much worse' are combined because of low numbers in these groups.

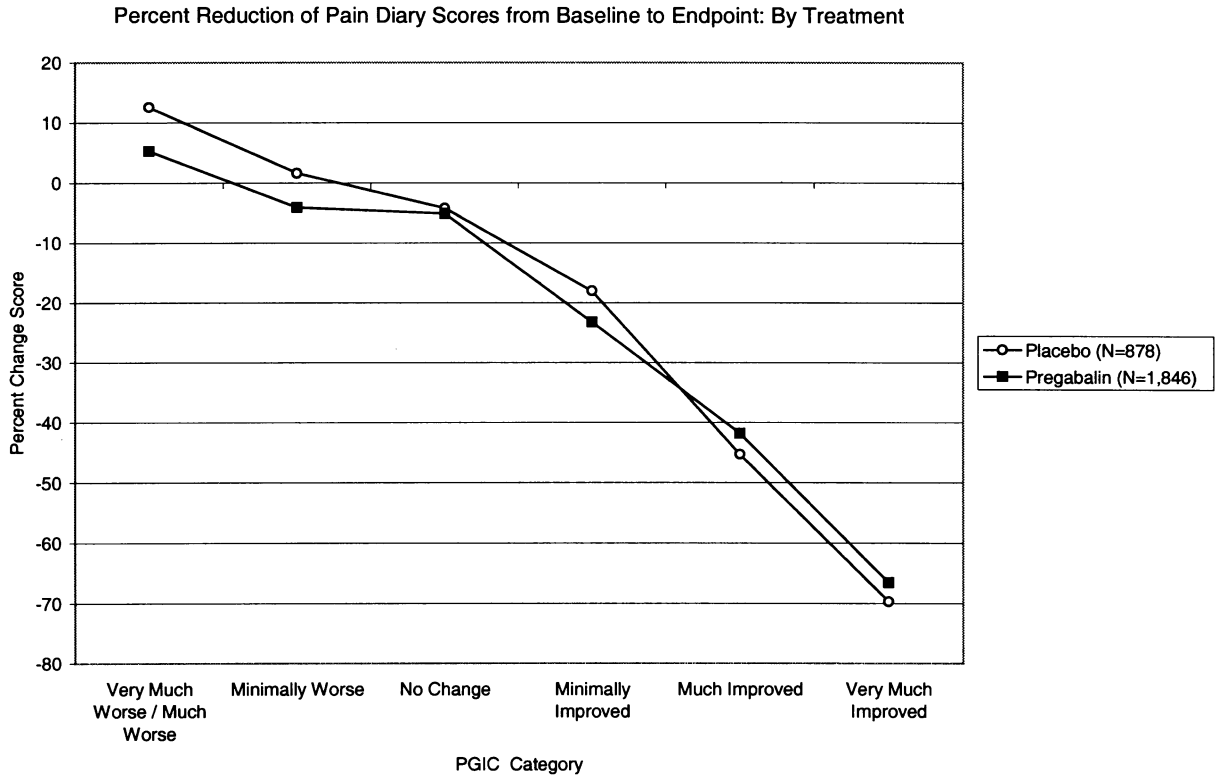

Fig. 4. Active treatment group (pregabalin) compared to the placebo group. Percent change (baseline to endpoint) in PI-NRS pain score vs. PGIC assessment at endpoint.

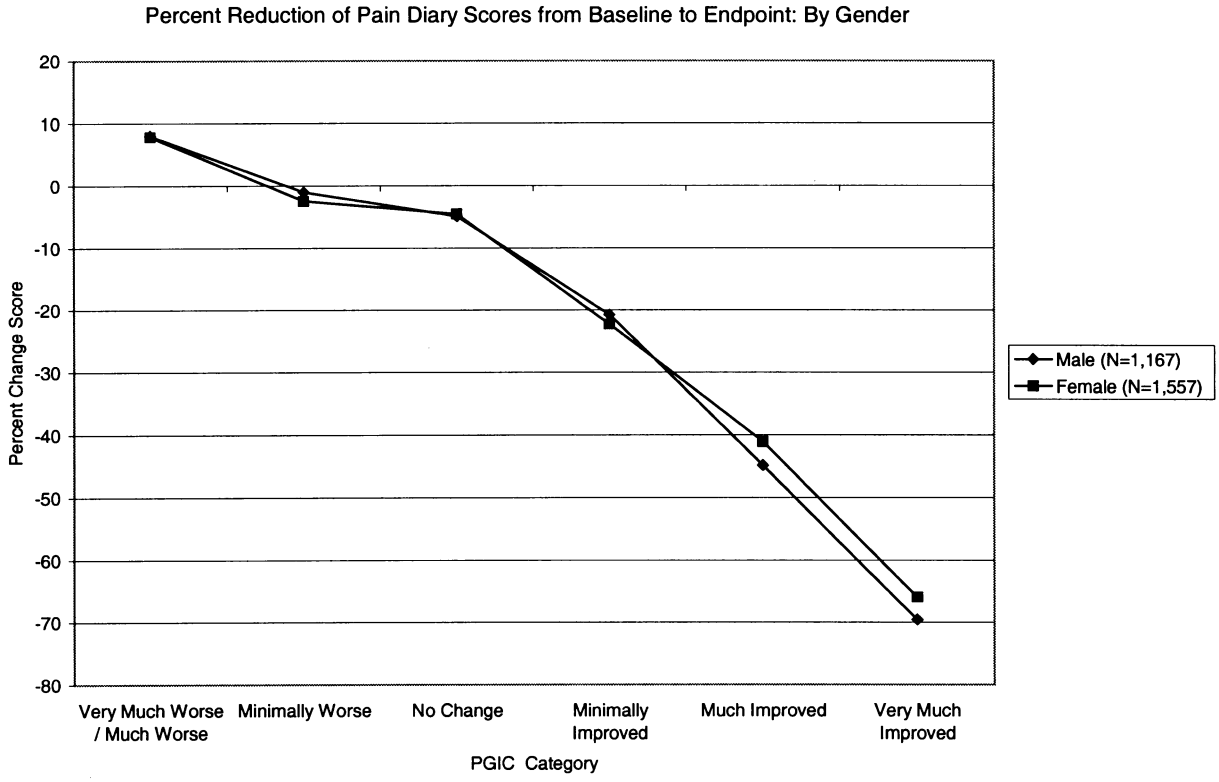

Fig. 5. Comparison by gender. Percent change (baseline to endpoint) in PI-NRS pain score vs. PGIC assessment at endpoint.

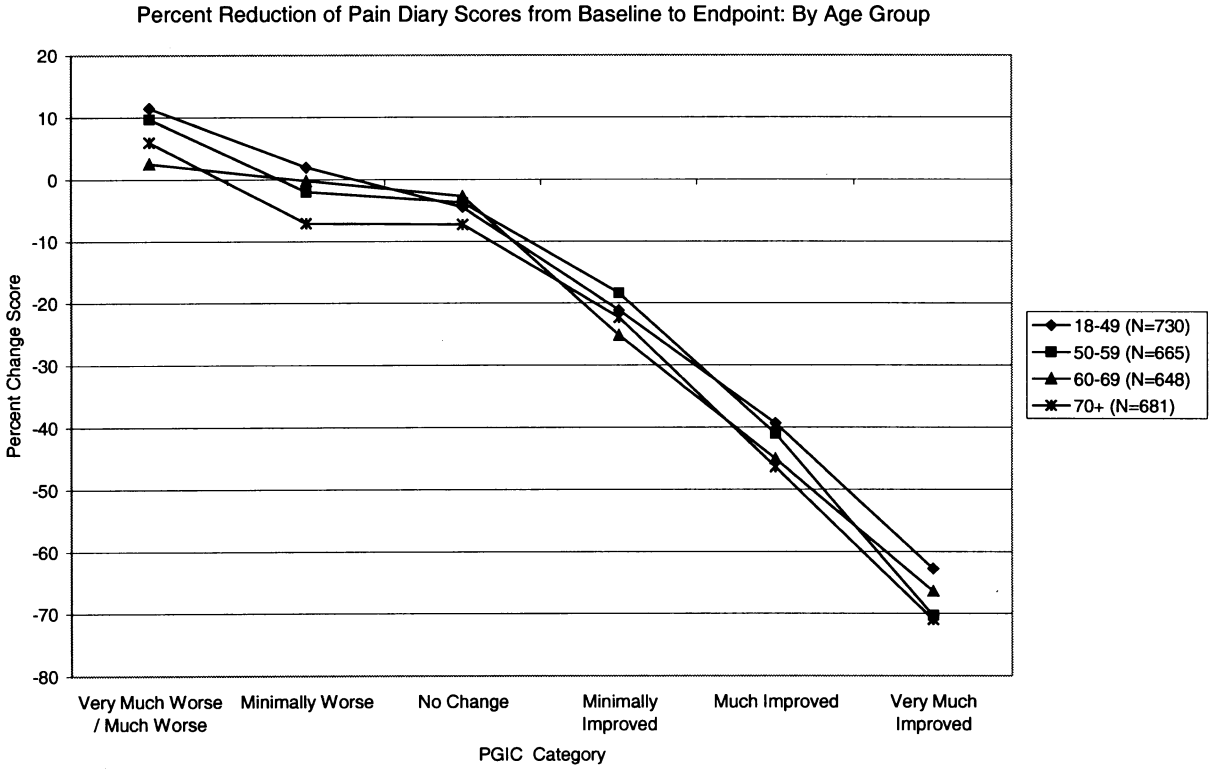

Fig. 6. Comparison by age group. Percent change (baseline to endpoint) in PI-NRS pain score vs. PGIC assessment at endpoint.

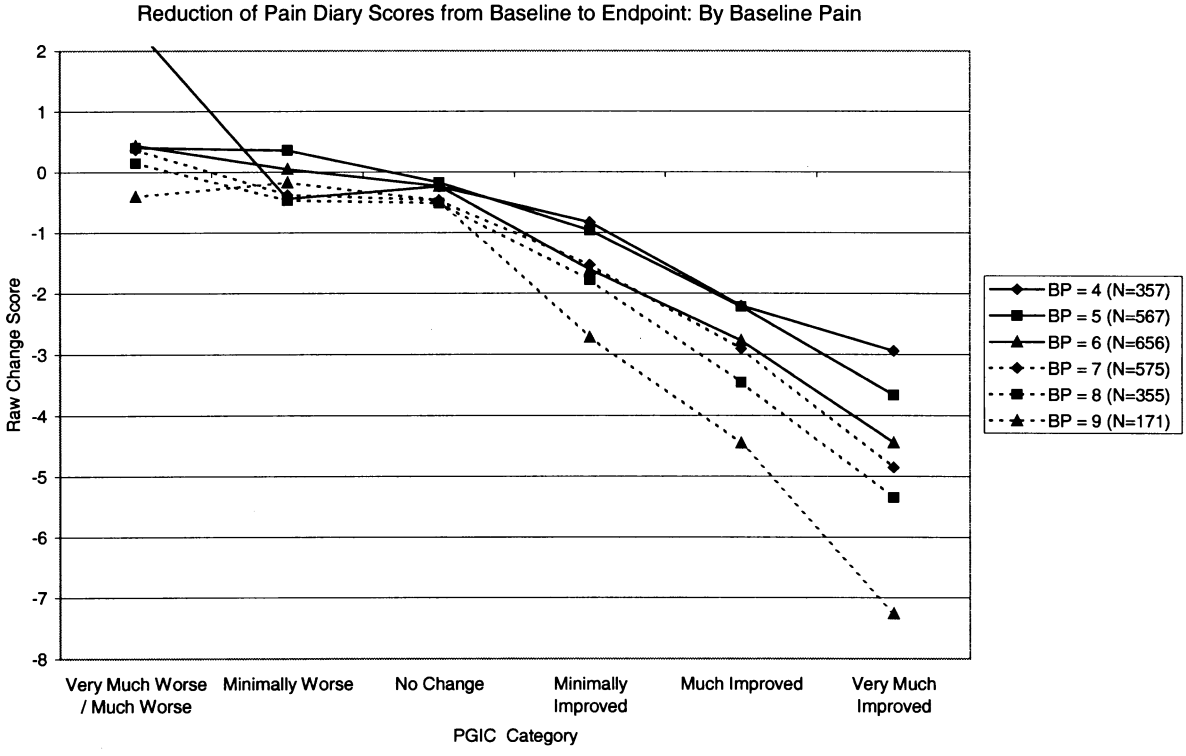

Fig. 7. Comparison by baseline pain. Raw change (baseline to endpoint) in PI-NRS pain score vs. PGIC assessment at endpoint.

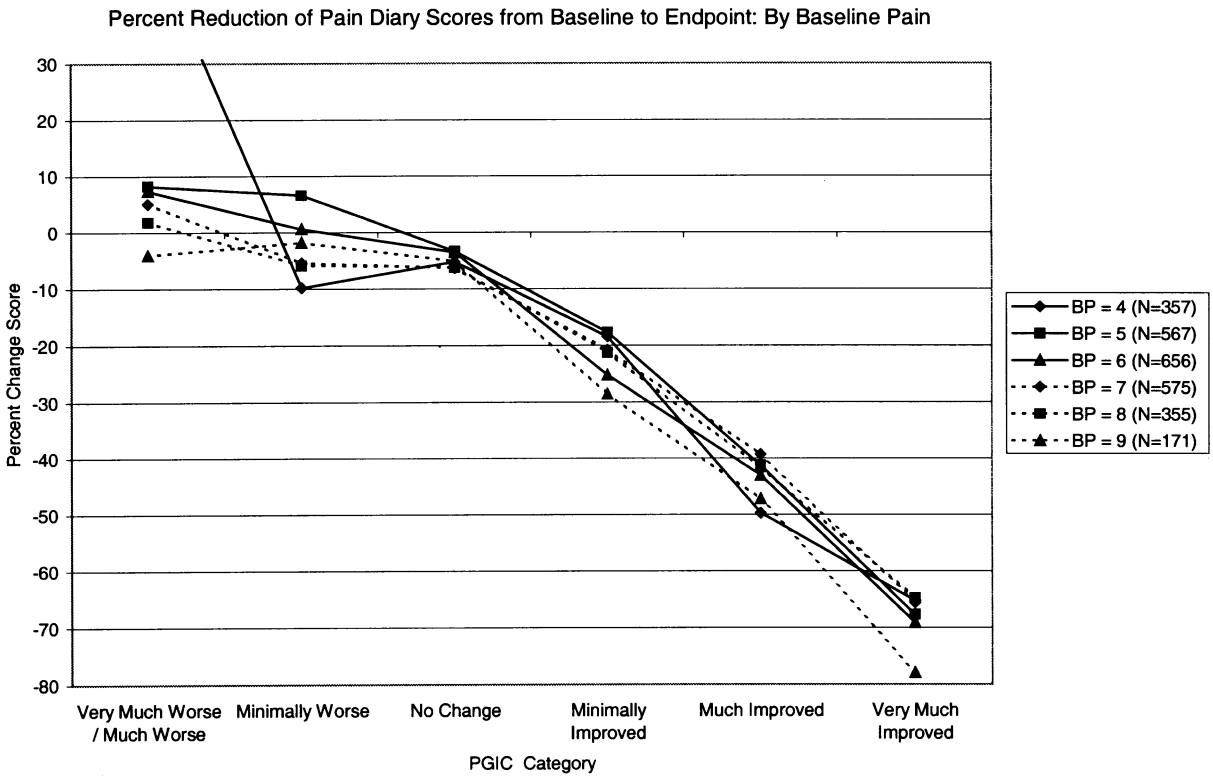

Fig. 8. Comparison by baseline pain. Percent change (baseline to endpoint) in PI-NRS pain score vs. PGIC assessment at endpoint.

while this was not true for percent change from baseline, which adjusts for the baseline pain score.

3.5. All data combined

Given the consistent clinical interpretation of a given PI-NRS change across these different strata, data from all patients were combined for subsequent analyses. The box plot in Fig. 9 shows the full distribution of the PI-NRS change scores for each PGIC category. This figure illustrates that almost all patients who considered themselves 'much improved' or 'very much improved' had at least some decrease in the NRS score and most had a decrease of two points or greater.

3.6. Receiver operating characteristics

Table 2 provides specific values generated from the ROC analyses for both raw change and percent change in the PI-NRS score best associated with several definitions of clinically important improvement (i.e. 'minimally improved' or better, 'much improved' or better, and 'very much improved' only). The areas under the ROC curves for the PI-NRS change and percent change are nearly identical for each definition of improvement. A raw change of −1.74 and a percent change of −27.9% were best associated with our definition of clinically important improvement. Table 3 shows the sensitivity and specificity for a range of convenient PI-NRS change and percent change scores associated with this same definition of clinically important improvement.

4. Discussion

In the data generated from 10 different studies of pregabalin for the treatment of chronic pain, we have demonstrated a close association between changes on the PI-NRS and the PGIC. The association is highly consistent over multiple trials regardless of the disease causing the chronic pain, treatment administered (drug or placebo), trial outcome (positive or negative), or the patient factors of age or gender. The consistency of these results suggests that widely different patient populations interpret changes in the PI-NRS similarly. The diversity of the patients that make up our study population provides strong support for the external validity of these results. The similarity in study designs makes it less likely that external factors are responsible for our findings.

The choice of percent change vs. raw change score, calculated from the PI-NRS, as the best indicator of the level of improvement is controversial (Price et al., 1983). In our studies, nearly equivalent associations were found between PGIC and both the raw change and percent change, as demonstrated by the equivalent area under the ROC. A possible explanation for this finding is that the baseline pain entry requirement of four on the 0–10 PI-NRS resulted

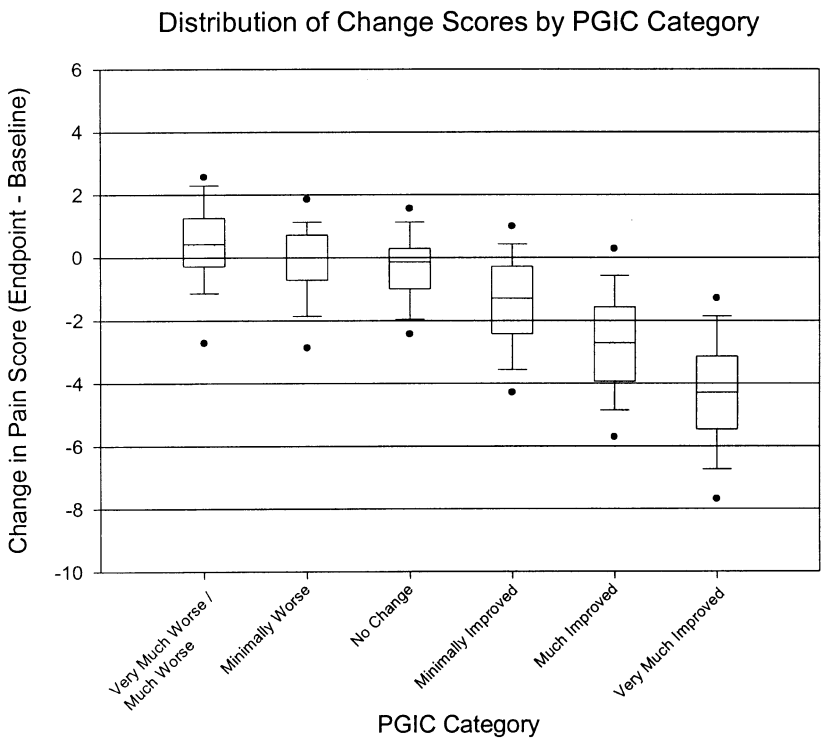

Fig. 9. This box-plot displays the median, 25th, and 75th percentiles (solid line box); the 10th and 90th percentiles (whiskers); the 5th and 95th percentiles (dots) as well as the mean value (dotted line) for the change in the PI-NRS within each PGIC category for all patients.

in a mean baseline pain score of 6.6 for this population. Thus, in our study population, a 30% decrease is generally equivalent to a raw reduction of two points on average. In studies with more variability in baseline pain, such as studies with no minimum baseline pain requirement, the relationship between percent change and PGIC will be more consistent than the relationship between raw change and PGIC. In such studies, clinical relevance should be defined in terms of percent change.

We have shown that a PI-NRS change score of −1.74 and a percent change score of −27.9% were best associated with our a priori definition of clinically important improvement, namely the PGIC category of 'much improved' or better. However, when using PI-NRS change scores in a clinical setting, integer values are more convenient for clinicians and patients. Table 3 shows that a reduction from baseline of two points or 30% on the PI-NRS provides essentially the same sensitivity and specificity as the more exact values. These values should be appropriate for use in the interpretation of clinical study results for clinical care and for the design and analysis of future clinical trials of chronic pain therapy.

These values are consistent with recently published findings generated by different methods. In one recent study of acute breakthrough cancer pain, a change score of −2.0 and a percentage change of −33% on the PI-NRS were shown to be associated with the clinically important outcome defined as a patient's need to take additional medication to treat pain (Farrar et al., 2001). Using a different paradigm to evaluate chronic pain data from arthritis patients, the Outcome

Table 2
ROC analyses: model statistics at tangent

PI-NRS score type	Model	Area under the curve	Change	Sensitivity (%)	Specificity (%)	Percent correct (%)
Raw change	Very much improved	0.873	−2.76	79.2	80.1	80.0
Raw change	Much or very much improved	0.853	−1.74	77.0	78.6	78.0
Raw change	Minimally, much or very much improved	0.832	−1.0	77.9	75.3	76.8
Percent change	Very much improved	0.890	−46.51	81.5	81.5	81.5
Percent change	Much or very much improved	0.859	−27.9	78.4	78.4	78.4
Percent change	Minimally, much or very much improved	0.832	−14.5	76.8	76.8	76.8

Table 3
Model statistics for selected change scores

Raw change from baseline	Sensitivity	Specificity	Percent change from baseline	Sensitivity	Specificity
0	95.32	31.80	−10	92.56	55.36
−1	89.90	61.64	−20	85.65	69.83
−2	74.92	80.60	−30	76.51	80.09
−3	56.43	91.81	−50	54.94	92.54
−4	35.92	96.92	−75	21.47	98.88

Measurement in Rheumatoid Arthritis Clinical Trials (OMERACT) study group determined the change in patient reported arthritis specific scores compared to a set of written vignettes describing the change in the patient's condition. Experts were asked to comment on whether or not they thought the patient vignettes described important improvement (Goldsmith et al., 1993). In that study, a 36% change in pain score was shown to be best associated with the expert's opinion that the improvement had been clinically important.

Our criteria are less stringent than the previous 'gold standard' criterion based on face validity (a 50% reduction in pain from baseline) (Cooper et al., 1976; Turk et al., 1993; McQuay et al., 1995; Follett, 1999; Seres, 1999). In our data, a 50% reduction in pain from baseline was associated with the highest degree of improvement on the PGIC ('very much improved'). Other studies have addressed the determination of clinically important improvement, but the underlying assumptions of the analyses, the methodologies used, and in some cases the scales being compared were different and have led to differing results (Cook et al., 1993; Jason et al., 1995; Vanspall et al., 1996; Todd et al., 1996; Bucher et al., 1997; Guyatt et al., 1998). In contrast to other potential data sets, these data offer substantial advantages for this type of analysis. Specifically, our data are derived from multiple studies across five disease conditions in clinical trial settings conducted in multiple countries with essentially identical measurement instruments and methodology. This common study design removes a substantial degree of the unknown variability that may exist between unrelated studies while incorporating a relatively broad range of known and testable variables. Thus, the relationship between the change in the pain intensity PI-NRS and the categories of the PGIC is less likely to have alternative explanations.

However, there are limitations that should be considered in interpreting these results. Although a number of diseases are represented, these results may not generalize to all chronic pain syndromes. Caution should be exercised when applying these findings to studies with periods of observation longer than 12 weeks. In addition, other components of a patient's response to pain are known to influence their perception of overall improvement. We have not tried to examine these other components, but the high degree of correlation between change in pain score and the PGIC in several pain states strongly supports the concept that pain intensity is a major component of the patient's global response.

There may also be concerns that the PGIC may not always be an appropriate measure, especially in the assessment of pain conditions with significant psychiatric overlay (Just et al., 1999). However, the PGIC has been well validated and has been extensively used by pain researchers as a standard outcome and for comparison to other outcome measures (Goldsmith et al., 1993; Jenkinson et al., 1994; Juniper et al., 1994; Buchbinder et al., 1995). For example, it has been used as a comparison standard for the analyses of several scales by the OMERACT study group for the purpose of establishing standardized criteria for clinical trials in arthritis (Buchbinder et al., 1995). The PGIC was also used as the reference measure in a recent study comparing three functional assessment scales for chronic low back pain (Beurskens et al., 1996).

Unlike the numeric PI-NRS, the PGIC is specifically tied to the conceptual framework of overall improvement. While improvement remains a subjective concept, the concept of 'much improved' or better as an indicator of clinically important benefit is conceptually reasonable and clinically relevant. It is thought to assess the patient's overall status since starting the study, integrating the effect of treatment, side effects, and patient expectations. This integration is seen by some as a concern, since more than one factor can affect this outcome. However, in our data there was a high correlation between the CGIC and the PGIC, which lends additional credibility to the validity of this measure and the comparisons made in this study. Even if CGIC were deemed to be a more appropriate criterion, our analysis as presented here would remain applicable.

In conclusion, we have presented an extensive analysis that determined a data derived value of the change in PI-NRS that best represents a clinically important improvement. The application of these results to future studies may provide a standard definition of clinically important improvement in clinical trials of chronic pain therapies. The definition proposed by this analysis could, for example, be used to calculate the number needed to treat (NNT) enabling comparisons to other studies (Cook and Sackett, 1995; Bucher et al., 1997). Use of a standard outcome across chronic pain studies would greatly enhance the comparability, validity and clinical applicability of these studies. Even if there is disagreement about our definition of clinically important improvement, these data provide information

about the range of PI-NRS values associated with various degrees of global improvement and will facilitate the evaluation of treatments for chronic pain.

References

Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996;65:71–76.

Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 1995;38:1568–1580.

Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;50:683–691.

Cook DJ, Guyatt GH, Davis C, Willan A, McIlroy W. A diagnostic and therapeutic N-of-1 randomized trial. Can J Psychiatry 1993;38:251–254.

Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. Br Med J 1995;310:452–454.

Cooper SA, Beaver WT. A model to evaluate mild analgesics in oral surgery outpatients. Clin Pharmacol Ther 1976;20:241–250.

Farrar JT, Portenoy RK, Berlin JA, Kinman J, Strom BL. Defining the clinically important difference in pain outcome measures. Pain 2000;88:287–294.

Follett KA. The fallacy of using a solitary outcome measure as the standard for satisfactory pain treatment outcome. Pain Forum 1999;8:189–191.

Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. J Rheumatol 1993;20:561–565.

Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. Br Med J 1998;316:690–693.

Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging 1989;29:307–335.

Jason N, Slater MA, Atkinson JH. The descriptor differential scale of pain intensity: an evaluation of item and scale properties. Pain 1995;61:251–260.

Jenkinson C, Peto V, Coulter A. Measuring change over time: a comparison of results from a global single item of health status and the multidimensional SF-36 health status survey questionnaire in patients presenting with menorrhagia. Qual Life Res 1994;3:317–321.

Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994;47:81–87.

Just N, Ciccone DS, Bandilla EB, Wu W. Global impressions versus validated measures of treatment effectiveness in patients with chronic nonmalignant pain. Rehabil Psychol 1999;44:194–207.

McQuay H, Carroll D, Jadad AR, Wiffen P, Moore A. Anticonvulsant drugs for management of pain: a systematic review. Br Med J 1995;311:1047–1052.

Moore A, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics. Pain 1996;66:229–237.

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.

Seres JL. The fallacy of using 50% pain relief as the standard for satisfactory pain treatment outcome. Pain Forum 1999;8:183–188.

Todd KH, Funk KG, Funk JP, Bonacci R. Clinical significance of reported changes in pain severity. Ann Emerg Med 1996;27:485–489.

Turk DC, Rudy TE, Sorkin BA. Neglected topics in chronic pain treatment outcome studies: determination of success. Pain 1993;53:3–16.

Vanspall B, Richardson J, Moreland J. Variables for assessing shoulder pain following stroke: a modified Delphi survey. Can J Rehabil 1996;10:25–34.