A Method for Clinically Defining "Improvers" in Chronic Pain Studies

Robert W. Wassell, PhD
Senior Lecturer and Honorary Consultant
Department of Restorative Dentistry

M. Adel Moufti, PhD
Clinical Fellow
Department of Restorative Dentistry

John G. Meechan, PhD
Senior Lecturer and Honorary Consultant
Department of Oral and Maxillofacial Surgery

Ian N. Steen, PhD
Statistician and Principal Research Associate
Institute for Health and Society
University of Newcastle Upon Tyne

James G. Steele, PhD
Professor and Honorary Consultant
Department of Restorative Dentistry

The School of Dental Sciences
University of Newcastle Upon Tyne
United Kingdom

Correspondence to:
Dr Robert W. Wassell
The School of Dental Sciences
University of Newcastle Upon Tyne
Framlington Place
Newcastle upon Tyne
NE2 4BW
Fax: +0191 2226137
E-mail: r.w.wassell@ncl.ac.uk

Aims: *To test a measurement model based on clinicians' assessments of patient data that allows simple and confident clinical validation of any statistical or numerical technique designed to separate patients improving with treatment from those who are not, particularly for pain that shows large daily variation.* ***Methods:*** *Diaries using daily visual analog scales (VAS) of pain intensity were obtained from 39 patients treated for chronic temporomandibular disorders. Three experienced clinicians visually assessed 39 VAS/time graphs. Criteria indicating improvement (general trend, height and apparent frequency of graph spikes) evolved over 3 assessments. The third assessment defined improvers visually. Numeric analyses considered the difference between first and last months of treatment for mean, area under the curve (AUC), and maximum VAS scores. Thresholds of 40%, 50%, or 60% pain reduction defined improvement numerically. Aggregate sensitivity and specificity was compared with visual definition to find the optimal threshold.* ***Results:*** *Patients were defined visually as improvers, nonimprovers, and borderline cases. Interexaminer reliability for identifying improvers was good (k = 0.79). Mean VAS and AUC were highly correlated (r = 0.999). The optimal threshold of mean and maximum VAS relative to visual definition was 50% pain reduction. Cases defined as improvers by both mean and maximum agreed best with the visual definition (sensitivity 90%, specificity 84%).* ***Conclusion:*** *Visual assessment of VAS demonstrates distinct pain/time patterns that can validate numeric definition of complex pain recovery. No single numeric method can be guaranteed to give a clinically valid outcome.* J OROFAC PAIN 2008;22:30–40

Key words: chronic pain, pain diaries, pain measurement, stabilization splint, temporomandibular disorders, treatment outcomes, visual analog scales

Surprisingly, there is little in the literature to describe day-to-day variations in chronic pain suffered by patients. Measurements of acute pain have been standardized; however, this is not the case with chronic pain. Clinicians, especially those dealing with temporomandibular disorders (TMD) and other chronic musculoskeletal conditions, are familiar with patients whose pain fluctuates frequently and those whose pain remains at a more or less constant level. These differences are of interest academically as well as clinically, because they have the potential to affect the outcomes of studies comparing treatments, especially those with low numbers of participants; if pain is fluctuating, the intensity on the day of assessment may be high or low simply by chance.

Considerable attention has been paid to pain measurement and statistical analysis in intervention studies. At its simplest, pain may be measured in terms of perceived intensity by the use of verbal descriptors or numeric rating or visual analog scales (VASs).[1] Measures of pain relief are also described. If conducted retrospectively, however, these measures of pain intensity or pain relief suffer from the drawback that they rely on patient memory over relatively long periods[2] and do not always correlate well with contemporaneous measures of pain intensity.[3–5] Other methods include global rating scales and the measurement of physiologic responses. It is widely accepted that, as pain is a subjective experience, assessment of intensity or associated unpleasantness is best made by the patient; evaluations made by proxy are unreliable and often underestimate the patient's experience.[6]

Although intensity is often regarded as the primary outcome, especially in meta-analyses of pain relief trials,[7–9] other dimensions of pain may also be described, including its sensory and affective qualities,[10] total pain relief scores, number of patients with a percentage pain reduction (eg, 50%), analgesic consumption, and the need for rescue medication.[11,12] The International Association for the Study of Pain has published recommendations for core outcome domains[13] and outcome measures[14] to provide an overall understanding of patients' lived experience, psychosocial aspects, and journey through treatment. Nevertheless, pain intensity will continue to be an important treatment outcome.

In recent years, analysis of pain measurements in clinical trials has started to move away from consideration of mean pain changes in treatment and comparison groups, which can be misleading, toward comparison of the proportions of patients in such groups who have shown a clinically significant improvement.[12,15] Such a distinction is used in calculating the "number needed to treat," which is often quoted in meta-analyses.[16] The goal here is to define a clinically significant improvement in relation to the condition being treated. A pragmatic approach used by many investigators has been to use a 50% reduction in pain from the start of treatment.[7–9,16–20] Others have used a reduction in pain of 25% and 75%, but they have failed to provide a clear rationale.[12] A scientific basis for a percentage improvement that is clinically meaningful (40% median pain reduction) has been provided for rheumatoid arthritis.[21] This figure was based on consensus assessment by personnel interested in management of this condition, who considered other clinical parameters in their assessment of patients. Another approach was used with complex regional pain syndrome type 1. In this case, patients' global perceptions of successful and unsuccessful treatment were related to a reduction in VAS pain scores.[22] In this study a pain reduction of at least 50% and an absolute reduction of 3 cm on a 10-cm VAS were accurate in predicting successful treatment.

The concept of a clinically significant improvement over time is clearly at the heart of any statistical approach to evaluating the efficacy of an intervention. This is constrained, however, when the condition manifests fluctuating levels of pain on a day-to-day basis. Measurements made only at the start and end of the trial may not be representative of the patient's overall response to treatment. Attempts to compensate for fluctuating pain levels include asking the patient to provide, in addition to current pain levels, an estimate of the range and mean level of pain experienced over a certain time, which is termed the "reference period." These reference periods may vary between studies. In addition, patients' memories of pain can be unreliable, as they may be influenced by the current pain intensity.[2,23]

Clearly, if pain levels are monitored frequently, then such variations may be taken into account during analysis. In trials of acute pain management, intensity is often measured using a pain diary,[24] which allows multiple samples to be taken. This approach is rarely used in chronic pain trials, but the method has the potential to provide much useful information,[25,26] not only for clinical trial analysis, but for other purposes as well.

In summary, there are problems with recording measurements only at the start and end of treatment for studies of chronic conditions where pain levels fluctuate. Pain-diary data may be useful to record pain intensity, a primary outcome in clinical trials. Furthermore, there is now a movement in chronic pain studies to consider the proportion of patients showing improvement in the test and control groups. However, it is necessary to determine how to define "improvers." So far, statistical approaches alone have been unable to provide *clinically realistic thresholds* of pain reduction, which may vary for different diseases. There have been 3 approaches to deciding appropriate thresholds: *(1)* an arbitrary approach, *(2)* an approach based on patients' global rating of improvement, and *(3)* an approach based on clinicians' assessment of clinical data. However, patients' global ratings can be distorted by memory effects, and there is little in the way of systematic research supporting clinicians' assessments. The aim of this study was to test a

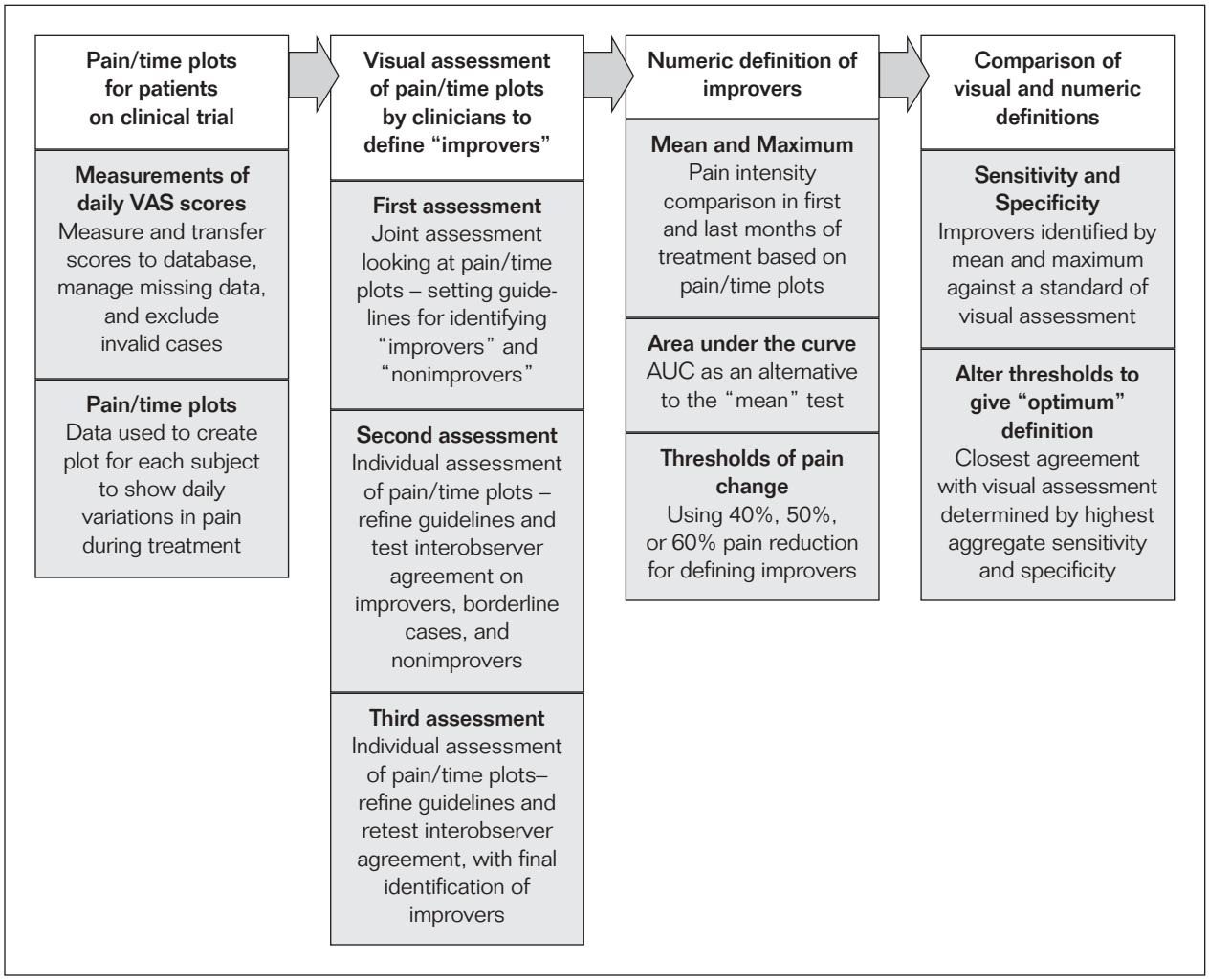

Fig 1 Outline of the method used to define improvers.

measurement model based on clinicians' assessments of patient data that allows simple and confident clinical validation of any statistical or numeric technique designed to separate patients improving with treatment from those who are not, particularly for pain that shows large daily variation. Although reported here for the management of TMD, the approach is straightforward and could be applied to any condition where there may be a risk of daily variation in the recovery period.

Materials and Methods

An overview of the stages used in this study is shown in Fig 1.

Pain/Time Plots for Patients on Clinical Trial

Data were obtained from a prospective trial of TMD management. The original study tested the effectiveness of stabilization splints in 72 patients who completed an initial treatment in 3 to 5 months.[27] Treatment outcomes were assessed using daily diaries of pain intensity. In addition, other clinical assessments were made at reviews scheduled at 3-week intervals. Each diary page contained a 100-mm VAS anchored at each end by the term "no pain" at one end and "unbearable pain" at the other. Patients were asked at the end of each day to give an estimate of the worst pain experienced for that day. In addition, the number and type of analgesics used were recorded. Completed diaries were collected at each review, and VASs were measured. Scores were rounded to the nearest millimeter and inserted into an Excel spreadsheet. Patients were

Table 1 Diagnoses of Patients Participating in the Original Study According to IHS Criteria[28]

Subdiagnosis	Frequency	%
Muscle disorders		
None	1	2.6
Reflex splinting	14	35.9
Myofascial pain	24	61.5
Total	39	100.0
Joint disorders		
None	23	59.0
Disc displacement with reduction	15	38.4
Capsulitis	1	2.6
Total	39	100.0

excluded for 2 reasons: *(1)* incorrect completion (5 patients) or *(2)* incomplete pages resulting in loss of more than 10% data overall or more than 25% missing data in any consecutive 30 days (28 patients). These levels were set to ensure sufficient data for visual assessment of the resulting pain/time graphs.

The diaries of 39 patients were considered valid. This represented 54% of patients (35 female and 4 male) completing treatment. The average length of follow-up for these patients was 127 days, and the average number of missed days across the sample was only 1.5 days (range, 0 to 15 days). The mean age was 35 years (range, 19 to 65 years). Sixteen patients were in the treatment splint group (stabilization splints), 13 were controls (fitted with nonoccluding splints), and 10 were crossed-over from the control group to the stabilization splint group. Detailed TMD diagnoses according to International Headache Society (IHS) criteria[28] for the 39 patients are shown in Table 1.

Individual pain/time graphs were drawn for these 39 patients. To identify patients showing improvement (improvers) and validate the process, the graphs underwent a sequence of visual and numeric analyses.

Visual Assessment of Pain/Time Plots to Define Improvers

The graphs were examined by 3 consultant clinicians experienced in the management of TMD. Two were specialists in restorative dentistry, and the third was a specialist in oral and maxillofacial surgery. The process to identify definitive criteria was developed over 3 evaluation sessions. To reduce memory effects, these assessments were separated by at least 2 months. The refinements established at each session are described here.

On the first occasion, the examiners consulted with one another to provide a mutually agreed-upon categorization of each patient's pain/time graphs as follows (Fig 2):

- **Improvers:** Graphs demonstrating a clear trend in pain reduction
- **Nonimprovers:** Graphs with no trend in pain reduction, with either major fluctatuations, giving a saw-tooth appearance, or with pain remaining constant
- **Borderline cases:** Graphs showing a possible trend toward pain reduction but accompanied by several spikes of pain representing days where pain intensity was much higher than usual.

On the second occasion, the 3 examiners made their assessments individually to determine interexaminer agreement. After individual assessment, cases where there was not agreement were discussed, and allocation was resolved. Two more guidelines were introduced for use in a third assessment. These guidelines directed examiners to consider the apparent frequency of the pain spikes and reduction in the apparent area under the graph line with time.

The third assessment was again performed individually, and interexaminer agreement assessed. Based on these final results, patients were defined visually as improvers,[20] nonimprovers,[10] or borderline cases[9] according to agreement among at least 2 examiners.

Numeric Definition of Improvers

From the visual examination of the pain data, it was decided to represent mathematically for each patient:

1. An overall trend in pain reduction, calculated first as the percentage change in mean pain and

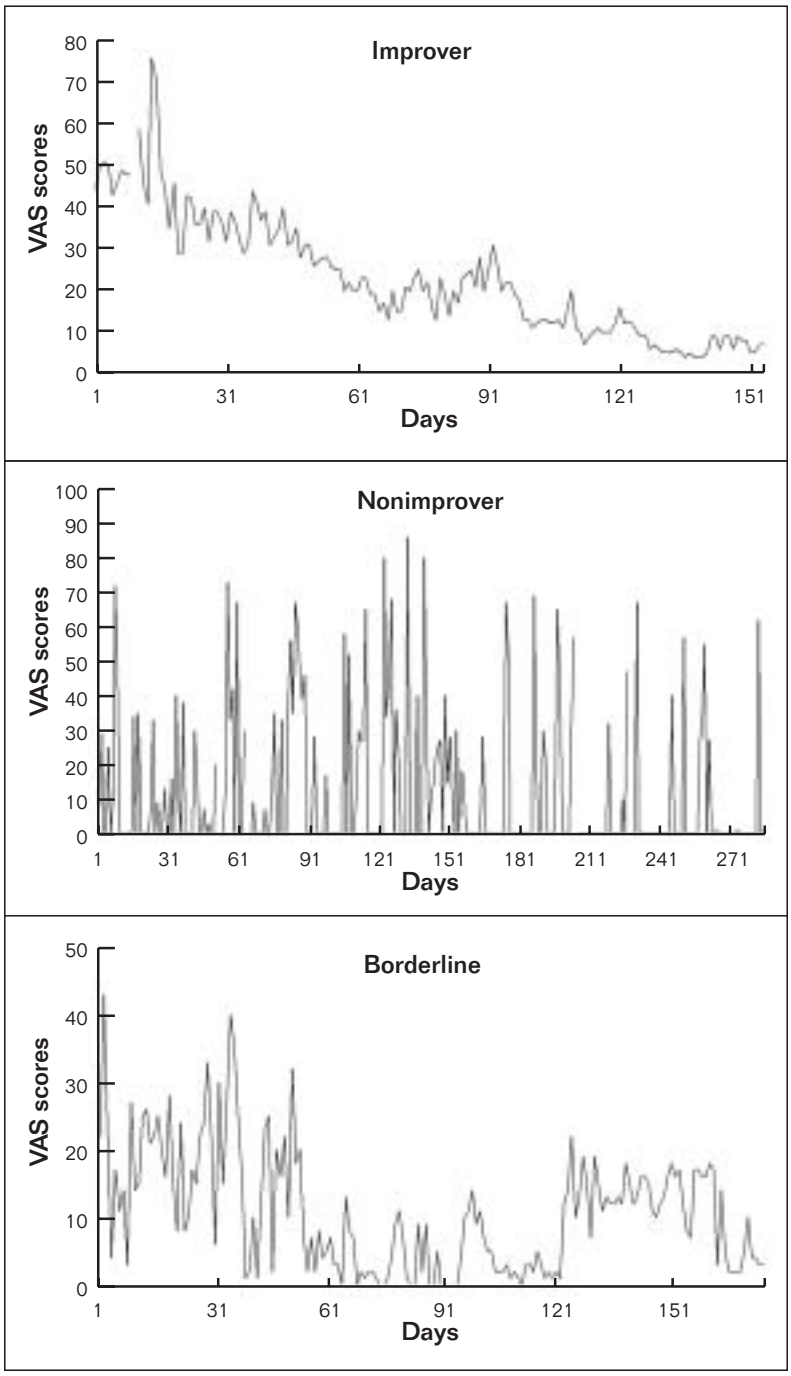

Fig 2 Pain/time graphs for patients classified by the 3 assessors as improver, nonimprovers, or borderline.

second as the change in area under the curve (AUC) between the first and last months (30 days) of treatment.

2. The change in the "spikes" of pain seen in the pain/time graphs (Fig 2), represented by the percentage difference in the maximum pain between the 2 sampling periods. The highest spike in the graph constituted the maximum pain.

The correlation between mean pain and AUC was assessed. The investigators believed that if the 2 approaches gave the same result it would be easier for this and future studies to adopt the first approach.

Table 2 Comparison of the No. of Patients Defined Visually as Improvers, Borderline Improvers, or Nonimprovers

	First assessment	Second assessment	Third assessment
Improvers	20	18	20
Borderline improvers	5	7	9
Nonimprovers	14	14	10

To define improvers, various thresholds of remaining pain were used for mean pain and maximum pain. Remaining pain was calculated using the formula "pain in last month/pain in first month × 100%," and thresholds of improvement were defined at ≤40%, ≤50%, or ≤60% of pain remaining. Furthermore, improvers were also defined as those who showed both mean and maximum pain reduction at each of these thresholds. The improvers derived from these numeric approaches were compared with those defined visually (ie, by the examiners' analyses).

Statistical Analyses

Visual Assessments. Interexaminer reliability in the second and third visual assessments was determined using Cohen's kappa (Landis-Koch extension for 3 or more examiners[29]). This test gives 3 kappa values representing the comparison of each group against the amalgam of the other 2 (eg, improvers versus the amalgam of borderline cases and nonimprovers). In other words, each kappa value is effectively a 2-group comparison.

Numeric Definitions. The correlation coefficient *(r)* was calculated for the relationship between mean pain and AUC. To allow calculation of AUC within the statistical package (SPSS 12.0.1 for Windows), missing values were imputed using expectation maximization algorithms.[30] This method is considered superior to traditional methods of imputation, such as mean substitution and multiple regression, which tend to reduce artificially the variability in the data.[30]

Comparison of Numeric and Visual Definitions. For this comparison sensitivity and specificity calculations were made using the clinician-agreed improvers and nonimprovers as the gold standard. The numerical definition of improvers and nonimprovers was made using *mean* pain, *maximum* pain, and the combination of *mean and maximum* pain. To determine the optimal threshold of each numeric definition (≤40%, ≤50%, or ≤60% remaining pain), the highest aggregate sensitivity and specificity in relation to the visual definition was identified.

Results

Visual Assessment of Pain/Time Plots to Define Improvers

The initial assessment, based on a mutually agreed-upon categorization by the 3 examiners, identified 20 improvers, 14 nonimprovers (of whom only 1 had pain remaining at a constant high level), and 5 borderline patients. In comparison, the results of the second and third assessments, which were carried out individually by each examiner, revealed a trend toward an increase in borderline cases and a decrease in nonimprovers; the number of improvers remained relatively constant (Table 2).

In the second assessment, there was total agreement between examiners in 24 of 39 cases (61%), and partial agreement (2 of 3 examiners agreeing) for the remaining 15 cases. In the third assessment, total agreement increased to 29 cases (74%), with partial agreement for 10 cases.

Agreements (kappa) between examiners during the second and third visual assessments are shown in Table 3. Best agreement for improvers, borderline cases, and nonimprovers was found in the third visual assessment; accordingly, the results of the third assessment provided the visual definition of definite improvers (kappa = 0.79 for improvers versus borderline cases and nonimprovers combined).

A similar proportion of improvers, borderline cases, and nonimprovers, defined visually, occurred in each TMD treatment group (Table 4); within the control and stabilization splint groups, the percentage of improvers was similar, at 53% and 56%, respectively.

Table 3 Kappa Tests for Agreement Between Assessors After the Second and Third Assessments

	Improvers	Borderline improvers	Nonimprovers
2nd assessment	0.76	0.26	0.69
3rd assessment	0.79	0.44	0.79

Within assessments each kappa value represents the comparison of that group against the combination of the other 2 groups.

Table 4 Cross-tabulation of TMD Treatment Group with Visually Defined Treatment Outcome Showing Similar Proportions of Improvers in the Control, Stabilization, and Crossover Groups

Trial group	Improvers	Borderline	Nonimprovers	Total
Control	7	2	4	13
Stabilization	9	5	2	16
Crossover*	4	2	4	10
Total	20	9	10	39

*Patients crossed over from control group to stabilization-splint group if not improving.[27]

Table 5 Sensitivity and Specificity of Numeric Definitions Against the Visual Definition, with Different Thresholds of Remaining Pain

	Percentage pain reduction*		
	≤ 40%	≤ 50%	≤ 60%
Mean			
Sensitivity (%)	85	95	95
Specificity (%)	74	68	58
Aggregate	159	163	153
Maximum			
Sensitivity (%)	80	90	95
Specificity (%)	74	68	63
Aggregate	154	158	158
Combined			
Sensitivity (%)	75	90	95
Specificity (%)	84	84	74
Aggregate	159	174	169

*Percentage pain reduction required for a patient to be categorized as an improver.

Numeric Definition of Improvers

Unlike the visual definition, the numerical definitions did not allow for the classification of borderline cases, so patients were classified as improvers or nonimprovers. The AUC test identified exactly the same patients as improvers and nonimprovers as did the mean pain test. The correlation coefficient *r* for AUC against mean pain was 0.999. As a result, AUC was discarded in favor of mean pain in the subsequent analyses.

Comparison of Visual and Numeric Definitions

Table 5 shows the results for the sensitivity and specificity tests comparing the visual and numeric definitions (for mean, maximum, and combined mean and maximum pain tests) carried out with different thresholds of percentage remaining pain to discriminate between improvers and nonimprovers. Sensitivity describes the percentage of the clinically defined (ie, visually assessed) improvers who would also be defined as improvers using each of the numeric thresholds, with visual assessment as the gold standard. The specificity describes the percentage of patients clinically defined as not being improvers (borderline and nonimprovers) who would also be defined as not being improvers using the numerical definition, again using the visual assessment as the gold standard. Clearly a high sensitivity and specificity are essential for a numeric definition to be clinically valid. The optimal threshold, showing the highest aggregate sensitivity and specificity, was 50%, although this was less clear-

cut for the maximum pain test. The combined pain tests (where subjects had to meet both mean and maximum thresholds) gave the highest sensitivity and specificity (90% and 84%, respectively), especially at the 50% threshold.

Discussion

Evidence on which to base treatment judgments for chronic pain conditions largely depends on well-run clinical trials. As discussed in the introduction, there is an important difference between an approach comparing mean VAS scores of patient and control groups over time and an approach labeling individual patients as improvers or nonimprovers to see what proportion of each group has responded positively. In this context the judgment about what constitutes improvement is critical but, often, completely arbitrary. Does a "50% mean improvement" have any clinical significance for either the clinician or the patient? In some circumstances a 20% reduction in mean daily pain score may be an acceptable therapeutic achievement; in others, a 60% improvement may not be enough to justify treatment. This study deals with this anomaly by describing a method for ensuring that any numerical definition of "improvement" has some clinical validity.

The authors are not advocating simply using clinical judgment as an outcome in clinical trials but suggest applying mutually agreed-upon clinician judgments to a measurement model to make sure that any statistical definition of success has some clinical meaning, rather than being arbitrary. As noted by others, many researchers would be reluctant to label a patient as "improved" based solely on the patient's own perception without some judgment by a clinician.[31] Clearly then, a rational means of applying clinicians' judgment is required. The method described is rigorous, clinician-led, and easy to undertake.

The ability of clinicians to agree independently is a prerequisite, so this was tested before the numerically defined rules of success were tested against these decisions. In this case, there was good agreement at a 50% improvement in scores. Had 50% not been appropriate, the numeric definition could have been changed to find something that gave better agreement with the clinical judgment.

In discussing the results, 2 caveats must be emphasized in relation to TMD. Firstly, the assessment here relates only to the recording of pain intensity. It does not incorporate other aspects of TMD that make the clinician's overall assessment more complex. However, the method provides data that can inform discussions between patient and clinician, which may help determine progress. Secondly, TMD consists of a variety of subdiagnoses, the most common of which are myofascial pain and disc displacement with reduction. These conditions are well represented in the present sample. For subdiagnoses occurring less frequently (eg, disc displacement without reduction and arthritis), the pain/time characteristics may well be different. The Research Diagnostic Criteria for Temporomandibular Disorders[32] would have been preferable to the criteria of the IHS,[28] but at the time of planning the original clinical study, these had not been universally accepted. The methods used to standardize the IHS diagnoses are explained elsewhere.[27] The range of diagnoses in the present study can be considered typical of patients attending general dental practice who have signs and symptoms of TMD requiring treatment.

Visual Assessment of Pain/Time Plots to Define Improvers

The clinicians found it surprisingly easy to reach a consensus decision for each case. They were also surprised by the amount of day-to-day variation reported for some patients. The visual assessment was refined to ascribe suitable criteria and optimize agreement between observers (as determined by kappa tests). With the individual assessments, there was remarkable consistency in identifying improvers, with some minor disagreement in separating borderline and nonimprovers. It is not surprising that the classification between borderline and nonimprovers was indistinct, as there was considerable variation in daily pain scores. The good agreement between clinicians concerning the improver category is the important finding.

A factor in support of the validity of the visual definition was the similarity in percentage of improvers seen in the control (53%) and stabilization splint (56%) groups. During the clinical trial there was no significant difference between these 2 groups for any of the clinical outcome measures.[27]

Numeric Definition of Improvers

The role of the numeric analysis was to find an approach that was best supported by the visual definition in terms of balanced sensitivity and specificity, potentially making it more applicable for use in a trial. A similar approach to determine "adequate pain relief" has been described for an acute pain trial in cancer patients,[33] but does not seem to have been used in chronic pain trials.

There is much to commend the use of mean pain in preference to AUC measurements. There was almost perfect correlation between mean pain and AUC (r = 0.999) and both provided exactly the same discrimination of improvers at the 50% level of improvement. As AUC is simply not available as part of many statistical and graphics programs, analysis of mean pain would be advantageous in centers where statistical support is limited.

The first and last months of treatment were chosen for the numerical definition, as they appeared to provide representative samples of daily pain measurements for these TMD patients, most of whom were treated over a 3-month period. In other conditions where treatment times are shorter, it may be appropriate to select a shorter period of time over which to sample measurements for numeric definition. More frequent pain measurements may also be considered depending on the variability in pain intensity during each day. Decisions of this type would need to be made on a trial-by-trial basis, with the frequency of pain sampling decided pretrial. Appropriate sampling periods can be determined after scrutinization of the pain intensity/time graphs.

Regression analysis[34] has been used in the study of pain. In the present study, however, the characteristics of the pain intensity/time graphs of improvers often showed a much faster rate of improvement at the beginning of the trial than at the end. This differential rate of improvement would have made linear regression inappropriate, as it assumes a constant rate of progression. Furthermore, a significant improvement on a regression line may not have any clinical relevance.

Other, more sophisticated statistical tests, such as cluster analysis, approaches based on multivariate analysis, and growth curve analysis (which samples all rather than part of the data)[35] may ultimately be helpful in determining thresholds for improvers but will still rely on clinical validation.

Comparison of Visual and Numeric Definitions

As an individual numerical test, mean pain set at a 50% threshold gave the best comparison to the clinical definition. Nevertheless, mean pain does not reflect the spikes of pain, which represented days when the patient's discomfort was markedly more severe. These spikes were considered in the present study by the evaluation of changes in maximum pain between the first and last months of treatment. The best comparison with the visual definition was obtained when these 2 tests were combined with a 50% threshold. Once other diseases have been scrutinized for their pain/time characteristics, it may be possible to come to a consensus as to the best methods of numerically classifying improvers and nonimprovers. These numeric methods could then be recommended for general use.

Limitations

This study was limited by the lack of true pretreatment data as well as by a lack of valid pain diaries. Ideally, in any pain study a baseline evaluation of pain should take place before any intervention. Unless waiting lists are used constructively for this purpose, however, there are ethical difficulties in withholding treatment to facilitate measurement. In this study, pain diaries were issued after the initial consultation, which occurred 2 weeks prior to starting splint treatment. Although it is tempting to label this period "preintervention," in reality the initial interaction between patient and clinician may have influenced pain perception, so these were not truly pretreatment measurements.

In this study, 54% of the 72 patients who finished treatment had adequately completed diaries. This provided sufficient data for the purpose of this methodologic study; however, it would have been a disappointing result if the original trial[27] had relied entirely on the diaries as the sole outcome measure. Much better compliance with pain diaries has been reported by LeResche et al[26] and de Wit et al.[25] LeResche et al paid patients to participate and also telephoned them as a reminder to complete the diaries. The de Wit study was carried out in cancer patients, where the majority found the process of completing the diary helped them to cope with the pain. Clearly, if pain diaries are to be adopted as a principal outcome for TMD and other chronic pain trials, considerable efforts will be needed to ensure adequate compliance, especially if long-term follow-up is envisaged. Indeed, the lack of reporting of patient compliance and the impact of missing data have been raised as significant issues affecting the quality of many clinical trials of pain management.[14] Electronic pain diaries coupled with suitable encouragement to complete them may help with these problems.[36]

Future Studies

Assuming improved compliance with pain diary completion, a number of interesting developments may result from this work. For example, it would be helpful to explore the correlation between the described method of clinical categorizaton of improvers, borderline cases, and nonimprovers

with the often-used patients' perception of global improvement.[14] If it can be shown that start-of-treatment and end-of-treatment measurements are representative of the overall change shown with pain diaries in certain chronic pain conditions, it may be possible to adapt these to give valid results. For example, Jensen et al[37] have shown that when chronic pain patients completed hourly pain diaries for 6 to 14 days the average pain over those periods correlated well with the patient's estimate of their "least pain" over a 2-week period. Also, Bolton[38] has shown that back pain patients were able to reliably estimate their average pain intensity in comparison to daily pain intensity measurements over a 1-week reference period. An important principle in extending this study will be standardization of the reference period.

The identification of improvers and nonimprovers is not only good practice for clinical trial analysis, it may also be usefully employed in creating characteristic pain/time responses for patients being treated for different subdiagnoses of TMD and perhaps also other diseases. This information would help clinicians and their patients understand the time needed on average before a chronic pain condition responsive to treatment showed a clinically significant improvement rather than rely on opinion or guesswork.

Acknowledgments

The authors would like to thank the British Dental Association Research Foundation for sponsoring the pain diary project through the Shirley Glasstone Hughes Memorial Prize. They are also grateful to the International Association for the Study of Pain for sponsoring Dr M. Adel Moufti to attend the 11th World Congress on Pain, Sydney, Australia, where this work was presented.

References

1. Huskisson EC. Measurement of pain. Lancet 1974;2: 1127–1131.
2. Eich E, Reeves JL, Jaeger B, Graff-Radford SB. Memory for pain: Relation between past and present pain intensity. Pain 1985;23:375–380.
3. 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.
4. Jensen MP, Chen C, Brugger AM. Postsurgical pain outcome assessment. Pain 2002;99:101–109.
5. Feine JS, Lavigne GJ, Dao TT, Morin C, Lund JP. Memories of chronic pain and perceptions of relief. Pain 1998;77:137–141.
6. Grootendorst PV, Feeny DH, Furlong W. Does it matter whom and how you ask? Inter- and intra-rater agreement in the Ontario Health Survey. J Clin Epidemiol 1997;50:127–135.
7. Mason L, Moore RA, Edwards JE, Derry S, McQuay HJ. Topical NSAIDs for chronic musculoskeletal pain: Systematic review and meta-analysis. BMC Musculoskelet Disord 2004;5:28.
8. Mason L, Moore RA, Derry S, Edwards JE, McQuay HJ. Systematic review of topical capsaicin for the treatment of chronic pain. BMJ 2004;328:991.
9. Mason L, Moore RA, Edwards JE, McQuay HJ, Derry S, Wiffen PJ. Systematic review of efficacy of topical rubefacients containing salicylates for the treatment of acute and chronic pain. BMJ 2004;328:995.
10. Katz J, Melzack R. Measurement of pain. Surg Clin North Am 1999;79:231–252.
11. Harden RN, Weinland SR, Remble TA, et al. Medication Quantification Scale Version III: Update in Medication Classes and Revised Detriment Weights by Survey of American Pain Society Physicians. J Pain 2005;6:364–371.
12. Farrar JT. What is clinically meaningful: outcome measures in pain clinical trials. Clin J Pain 2000;16:S106–S112.
13. Turk DC, Dworkin RH, Allen RR, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106:337–345.
14. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.
15. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain 2000;88:287–294.
16. McQuay HJ. Systematic Reviews Pain 2005—An Updated Review: Refresher Course Syllabus. Seattle: IASP Press; 2005:259–264.
17. McQuay HJ, Tramer M, Nye BA, Carroll D, Wiffen PJ, Moore RA. A systematic review of antidepressants in neuropathic pain. Pain 1996;68:217–227.
18. Guignard S, Job-Deslandre C, Sayag-Boukris V, Kahan A. Pamidronate treatment in SAPHO syndrome. Joint Bone Spine 2002;69:392–396.
19. Green S, Buchbinder R, Barnsley L, et al. Acupuncture for lateral elbow pain. Cochrane Database Syst Rev 2002:CD003527.
20. Freedman BA, Cohen SP, Kuklo TR, Lehman RA, Larkin P, Giuliani JR. Intradiscal electrothermal therapy (IDET) for chronic low back pain in active-duty soldiers: 2-year follow-up. Spine J 2003;3:502–509.
21. 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.
22. Forouzanfar T, Weber WE, Kemler M, van Kleef M. What is a meaningful pain reduction in patients with complex regional pain syndrome type 1? Clin J Pain 2003;19:281–285.
23. Matera D, Morelli M, La Grua M, Sassu B, Santagostino G, Prioreschi G. Memory distortion during acute and chronic pain recalling. Minerva Anestesiol 2003;69:775–780, 780–783.
24. Ong KS, Seymour RA. Pain measurement in humans. Surgeon 2004;2:15–27.

25. de Wit R, van Dam F, Hanneman M, et al. Evaluation of the use of a pain diary in chronic cancer pain patients at home. Pain 1999;79:89–99.
26. LeResche L, Mancl L, Sherman JJ, Gandara B, Dworkin SF. Changes in temporomandibular pain and other symptoms across the menstrual cycle. Pain 2003;106:253–261.
27. Wassell RW, Adams N, Kelly PJ. Treatment of temporomandibular disorders by stabilising splints in general dental practice: Results after initial treatment. Br Dent J 2004;197:35–41.
28. McNeill C. Craniomandibular Disorders: Guidelines for Evaluation, Diagnosis and Management. American Academy of Craniomandibular Disorders. Chicago: Quintessence, 1990:39–60.
29. Landis JR, Koch GG. A one-way components of variance model for categorical data. Biometrics 1977;33:671–679.
30. Schafer JL. Monographs on Statistics and Applied Probability: Analysis of Incomplete Multivariate Data. Book No. 72. London: Chapman & Hall, 1997.
31. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. Controlled Clin Trials 1991;12:142S–158S.
32. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord 1992;6:301–355.
33. Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: A validation study. J Pain Symptom Manage 2003;25:406–411.
34. Caraceni A, Brunelli C, Martini C, Zecca E, De Conno F. Cancer pain assessment in clinical trials. A review of the literature (1999–2002). J Pain Symptom Manage 2005;29:507–519.
35. Duncan TE, Duncan SC. An introduction to latent growth curve modeling. Behavior Therapy 2004;35:333–363.
36. Gendreau M, Hufford MR, Stone AA. Measuring clinical pain in chronic widespread pain: Selected methodological issues. Best Pract Res Clin Rheumatol 2003;17:575–592.
37. 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.
38. Bolton JE. Accuracy of recall of usual pain intensity in back pain patients. Pain 1999;83:533–539.