Further Qualification of a Therapeutic Responder Index for Patients with Chronic Low Back Pain

CLAIRE BOMBARDIER, CHRIS J. EVANS, NATHANIEL KATZ, JACK MARDEKIAN, GERGANA ZLATEVA, and LEE S. SIMON

ABSTRACT. ***Objective.*** Previously, a preliminary patient responder index (RI) in chronic low back pain (CLBP) was developed and validated in 5 placebo-controlled clinical trials. The resulting RI was a > 30% improvement in CLBP and patient global assessment (PGA), and no worsening (< 20%) in the Roland Morris Disability Questionnaire (RMDQ) total score. Our objective was to provide further characterization of the preliminary RI in a trial with an active control.

Methods. Data from a 6-week randomized, double-blind study of celecoxib compared to tramadol hydrochloride was analyzed to determine differences by treatment group on the CLBP RI and its components, to compare the CLBP RI with each of its individual components, and to reanalyze the original cutoff points for the responder criteria.

Results. Of the celecoxib arm, 50.7%, and of the tramadol hydrochloride arm, 43.7% were classified as responders under the CLBP RI ($p = 0.043$). The PGA is the most important component in the RI (45% of the sample failed to reach the > 30% improvement criteria on the PGA compared to 34% on the low back pain visual analog scale and only 11% on the RMDQ. The agreement among the CLBP RI with each of its 3 components was largest for the PGA component (κ coefficient 0.849) and smallest for the RMDQ component (κ coefficient 0.207).

Conclusion. The RI appears to be particularly sensitive to the cutoff point used for improvement in the PGA component. Further testing of the index in trials with other active comparators is required to gain a fuller understanding of its performance. (J Rheumatol First Release Nov 1 2010; doi:10.3899/jrheum.091444)

Key Indexing Terms:
LOW BACK PAIN
NONSTEROIDAL ANTIINFLAMMATORY AGENT
RESPONDER INDEX
OPIOID

Low back pain (LBP) is a common condition that affects an estimated 70% to 80% of adults in the United States at some time in their lives[1]. It presents a significant health and economic burden to society. In the US, back pain is one of the most frequent reasons for hospitalization and physician visits, resulting in high medical care costs. The economic burden for patients with LBP in the US has been estimated at between $12.2 billion and $90.6 billion in direct medical costs and from $7.4 billion and $28.2 billion in indirect costs (i.e., lost productivity) annually[2]. Based on data from the Medical Expenditure Panel Survey, individuals with LBP have been found to have annual medical care costs $1320 greater ($3498 vs $2178) than those without LBP[3].

Chronic low back pain (CLBP), a subset of LBP in which pain is persistent for 3 months or more, is difficult to treat because its effect is multidimensional: it involves pain, limitation of normal activities, decreased health-related quality of life, the potential for the increased use of strong pain medications, depression, and sleep disturbances. Adults with LBP are nearly 3 times as likely to report fair or poor health than those without back pain (26% vs 9%, respectively), more than 4 times as likely to report arthritis-attributable activity limitations, 4 times as likely to be unable to work, twice as likely to report reduced sleep (< 6 hours per day), and 7 times more likely to report psychological distress[4]. The goal of CLBP treatment lies beyond pain control and includes the reduction of disability and the corresponding preservation of function[5]. It is therefore essential to assess physical functioning, the degree of limitation on activities of daily living, and overall well-being alongside pain intensity.

The selection and use of appropriate outcome measures in clinical trials of CLBP is difficult and is associated with several problems. First, researchers do not consistently use the same measures across clinical trials, which limits the

From the Institute for Work and Health, Toronto, Ontario, Canada; MAPI Values, and Analgesic Research, Boston, Massachusetts; Pfizer Inc., New York, New York; and SDG LLC, Cambridge, Massachusetts, USA.

Sponsored by Pfizer Inc. Drs. Bombardier, Katz, and Simon received a one-time fee in 2007 for advisory services to this study. Dr. Evans is an employee of MAPI Values, a research organization for this project, and was a paid consultant to Pfizer in connection with this research.

C. Bombardier, MD, Institute for Work and Health; C.J. Evans, PhD, MAPI Values; N. Katz, MD, Analgesic Research; J. Mardekian, PhD; G. Zlateva, PhD, Pfizer Inc.; L.S. Simon, MD, SDG LLC.

Address correspondence to Dr. G. Zlateva, Pfizer Inc., 235 East 42nd Street, New York, NY 10017, USA. E-mail: gergana.zlateva@pfizer.com

Accepted for publication September 10, 2010.

ability to compare results across studies. Second, outcome measures may focus narrowly on 1 aspect of treatment [e.g., the alleviation of the sensation of pain as measured on a numeric rating scale (NRS-Pain) or visual analog scale (VAS)] and ignore other important aspects such as the effect of CLBP on activities of daily living and sleep. Third, the interpretation of clinically important changes for some outcome measures has not been well researched and established.

A responder index (RI) is seen as a way to overcome some of these problems. An RI is a composite measure of face-valid and nonredundant clinical endpoints that usually measures different aspects of the disease manifestation. Response to treatment is measured by specific improvement criteria selected for the endpoints. These improvement criteria establish clinical efficacy and differentiate between placebo and active responses. To date, such improvement and response criteria have been developed and used in several different musculoskeletal disorders, including ankylosing spondylitis (Assessment of Spondyloarthritis International Society 20)[6], rheumatoid arthritis [American College of Rheumatology (ACR) 20][7], osteoarthritis (Outcome Measures in Rheumatology Clinical Trials/Osteoarthritis Research Society International)[8], and juvenile arthritis[9]. For example, the ACR20 requires > 20% improvement in swollen joint count, > 20% improvement in tender joint count; and > 20% improvement in 3 of the following 5 measures: patient global assessment (PGA), physician global assessment, patient pain (measured by VAS), Health Assessment Questionnaire (patient-assessed disability), and acute-phase reactant (C-reactive protein or erythrocyte sedimentation rate)[7].

Consistent with recommendations to standardize the selection and interpretation of outcome measures in chronic pain trials[10,11] and specifically LBP trials[12], we developed and validated a preliminary RI in CLBP in 5 clinical trials. The full results of that exercise are reported elsewhere[13]. The content of the initial RI was based on a review of the literature, discussions with patients with CLBP, and input from clinical experts. From these sources a list of candidate RI items was chosen. These items were subsequently tested in three 12-week placebo-controlled clinical trials of celecoxib therapy as compared to placebo in CLBP to identify a short list of candidate RI. The findings from these analyses were then validated in data available from two 12-week, placebo-controlled trials of valdecoxib as compared to placebo.

The resulting preliminary RI was > 30% improvement in LBP intensity as measured on a VAS and PGA, and no worsening (< 20%) in the Roland Morris Disability Questionnaire (RMDQ) total score. The LBP VAS is a 10-cm horizontal line (from no pain to extreme pain) that measures LBP severity. The PGA is a single question, "Considering all the ways your lower back pain affects you, how are you doing today?" and responses are recorded on a 5-point Likert-type response scale (from very good to very poor). The RMDQ[14,15] is a 24-item patient-completed measure designed to assess the degree of functional limitations in patients with LBP (e.g., getting dressed slowly because of back pain).

We describe the additional validation of this RI in CLBP. Our purpose is to provide further information on the performance of the CLBP RI in a prospective trial with an active control.

MATERIALS AND METHODS

Data from a multicenter, randomized, parallel-group, double-blind, double-dummy study at 56 centers in the US were analyzed posthoc[16]. The protocol was approved by the institutional review boards at each participating center, and written informed consent was obtained from each subject prior to study entry and before any study-related procedures were performed. The study was conducted in compliance with the Declaration of Helsinki and all International Conference on Harmonisation Good Clinical Practice guidelines.

Patients in this study were aged ≥ 18 years with a physician-confirmed diagnosis of CLBP. The primary location of back pain was between the 12th thoracic vertebra and the gluteal folds with or without radiation into the posterior thigh, classified as Category 1 or 2 according to the classification of the Quebec Task Force on Spinal Disorders. The duration of CLBP had to have been ≥ 3 months, requiring regular use of analgesics, and subjects had to have a moderate to severe LBP score of ≥ 4 on an NRS-Pain scale at baseline. Patients were randomized to receive either celecoxib 200 mg twice daily (bid) or tramadol hydrochloride (HCL) 50 mg 4 times daily (qid) for 6 weeks. Key exclusion criteria were CLBP of a neuropathic origin, history of rheumatoid arthritis, psoriasis, spondyloarthropathy, spinal stenosis, herniated disc for ≤ 2 years, fibromyalgia, and tumor or infection of the brain, spinal cord or peripheral nerves.

Use of nonselective nonsteroidal antiinflammatory drugs, cyclooxygenase (COX)-2 selective inhibitors (other than study medication), and other analgesics by any route was specifically excluded during the course of the study. Patients taking ≤ 325 mg of aspirin for nonanalgesic or arthritis reasons, at a stable dose for at least 30 days before the first dose of study medication, were allowed to continue their aspirin regimen for the duration of the study. Rescue medication was not allowed for CLBP during the study. Any medication and nondrug treatment that the patient took during the study other than study medication as specified in the protocol was recorded in the patient's medical record and on the case report forms.

The primary efficacy endpoint in the study was the proportion of subjects responding successfully to treatment, defined as ≥ 30% improvement from baseline on the NRS-Pain. A 2-stage analysis was used to test for noninferiority and superiority of celecoxib. Secondary efficacy endpoints included outcomes related to pain, functioning, overall impressions of health, tolerability, RI analysis, and safety assessments.

The objectives of this study were to analyze posthoc the performance of the CLBP RI in the following aspects: (1) to determine differences by treatment group on the preliminary CLBP RI in this population; (2) to compare the CLBP RI with each of its individual components; (3) to reanalyze the original cutoff points for the responder criteria; and (4) to examine the effect size of the RI.

The κ coefficient was used to measure the degree of reliability between the CLBP RI and the primary efficacy endpoint in the study in classifying responders. The κ coefficient is generally believed to be a more robust measure than simple percentage agreement since κ takes into account agreement occurring by chance. Possible values range from +1 (perfect agreement) to –1 (complete disagreement). A sensitivity analysis of the CLBP RI cutoff points was performed by varying each of its component cutoff values by ± 5%. The percentage of responders was computed for

each combination of cutoff values. The effect size for the change in LBP VAS, PGA, and RMDQ was computed using the mean difference divided by the common baseline standard deviation. The CLBP RI effect size was computed using the difference in proportions of responders divided by the square root of $p_{pooled} * (1 - p_{pooled})$, in which p_{pooled} is the combined proportion of CLBP RI responders. Primary analyses for the noninferiority assessment were performed on the evaluable population and repeated using the intent to treat (ITT) population. The ITT population includes randomized subjects who received at least 1 dose of study medication. All secondary analyses were done on the ITT population only. Dropouts were handled using last observation carried forward for secondary efficacy analyses. All analyses were conducted using SAS version 9.1 (SAS Institute, Cary, NC, USA).

RESULTS

A total of 1027 subjects were screened, of which 796 were randomized to study treatment: 404 to celecoxib 200 mg bid and 392 to tramadol HCL 50 mg qid (Figure 1). A higher percentage of subjects completed study treatment in the celecoxib group (85.6%) than in the tramadol HCL group (69.4%). In the celecoxib group, 58 subjects (14.4%) left the study: 22 (5.5%) for reasons related to study medication and 36 (9.0%) for reasons not related to study medication. In the tramadol HCL group, 119 subjects (30.6%) left the study: 71 (18.3%) for reasons related to study medication and 48

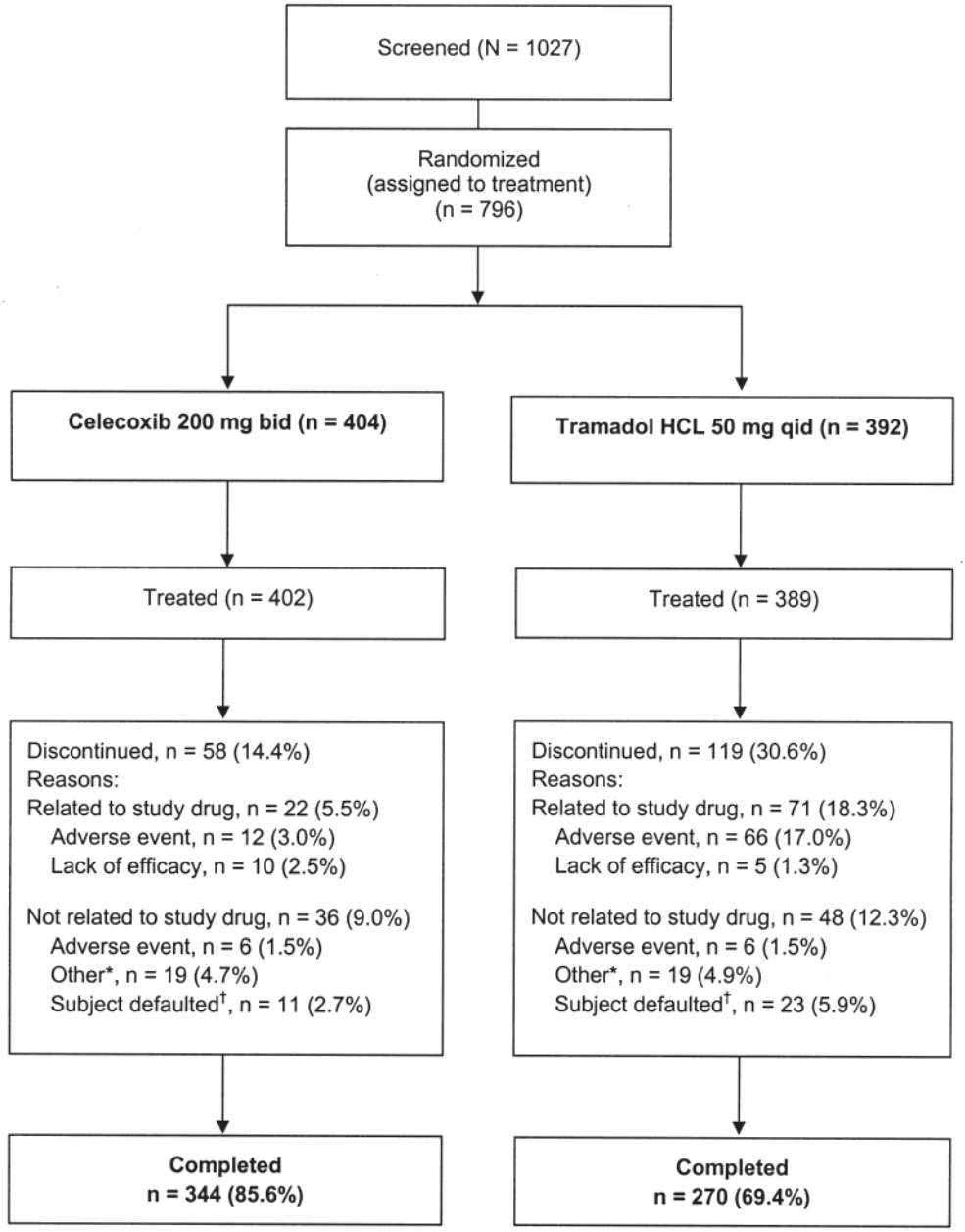

Figure 1. Patient disposition. *Includes subjects who did not meet entrance criteria, protocol violations, withdrawal because of pregnancy, or other specified. †Includes subjects lost to followup and subjects no longer willing to participate. HCL: hydrochloride.

(12.3%) for reasons not related to study medication. A review of patient source records and case report forms noted 8 celecoxib and 21 tramadol patients used analgesic therapies during the course of the study. These patients were appropriately included in the ITT analysis but the statistical analysis was not controlled for these interventions.

Demographic and clinical characteristics are reported in Table 1. The mean age of patients was 49 years in the celecoxib arm and 48 years in the tramadol arm. The majority of patients were women (58% and 59% in the celecoxib and tramadol arms, respectively) and white (67% and 61%, respectively). On average, patients experienced LBP for 419 weeks and 365 weeks in the celecoxib and tramadol arms, respectively. The majority of patients included in the trial characterized their pain as severe (Table 1).

The results of the individual components of the RI are reported in Figure 2 (A-C). On the CLBP VAS there was a significant difference in mean change from baseline to 6 weeks (–34.6 in celecoxib vs –30.4 for tramadol; p = 0.008). Although the numeric improvements were greater for celecoxib compared with tramadol from baseline to Week 6, for PGA and the RMDQ there were no statistically significant differences on either measure between arms. Under the CLBP RI (p = 0.043) criteria, 50.7% of subjects in the celecoxib arm and 43.7% in the tramadol arm were classified as responders (p = 0.043; Figure 3).

The PGA is the most important component in the RI: 45% of the sample failed to reach the > 30% improvement criteria on the PGA compared with 34% on the LBP VAS and only 11% on the RMDQ. The agreement among the CLBP RI with each of its 3 components was largest for the PGA component (κ coefficient = 0.849; near-perfect agreement[17]) and smallest for the RMDQ component (κ coefficient = 0.207; weak agreement). The κ coefficient (0.625) showed substantial agreement with the VAS component (Table 2).

Table 1. Demographic characteristics.

Characteristics	Celecoxib 200 mg bid, n = 402	Tramadol HCl 50 mg qid, n = 389	p
Age, yrs			
Mean (SD)	49.1 (14.8)	47.9 (14.5)	
Range	18–88	18–83	0.314
Sex, n (%)			
Male	170 (42.3)	159 (40.9)	
Female	232 (57.7)	230 (59.1)	0.8384
Race, n (%)			
White	271 (67.4)	236 (60.7)	
Black	70 (17.4)	82 (21.2)	
Asian	8 (2.0)	12 (3.1)	0.066
Other	53 (13.2)	59 (15.2)	
Primary diagnosis			
Back pain, n (%)	402 (100)	389 (100)	
Duration since first diagnosis (wks)			0.0729
Mean	418.9	364.7	
Range	2.1–2896.1	13.7–2988.9	
Unspecified (n)	1	0	
Severity of LBP (11-point NRS), mean	6.76	6.80	0.8300

LBP: low back pain; NRS: numeric rating scale.

The results of the RI are driven by the improvement criterion definition for the PGA. The highest and lowest responder rates occur with changes to the PGA improvement criterion. If the improvement criterion on the PGA is increased from 30% (the recommended level) to 35%, reflecting a higher hurdle for patients to achieve in terms of clinical benefit, the percentage of responders drops to 21.9% (Table 3A). If the PGA criterion is decreased to 25%, then the percentage of responders increases to 51.2% (Table 3A). A change to the improvement criterion for the RMDQ (moving it either up or down 5%) has only a negligible effect on the percentage of responders (Table 3B, 3C).

Analysis of the individual components of the RI and the total index reveal low effect sizes[18]. In the original study[13] used to develop the preliminary RI, the effect size for the RI was slightly higher (0.19 and 0.23 in the original trials) compared with this study (0.14; Table 4).

DISCUSSION

The proposed RI performed well in this clinical trial: it was able to differentiate the effects of the treatments on patients based on outcome measures they find to be clinically relevant and important in their everyday lives. Moreover, although the observed effect sizes were small, the criteria used to demonstrate improvement or no worsening in the index incorporate the concept of clinically meaningful change (> 30% for the pain VAS and the PGA and < 20% in the RMDQ). Decreases in pain intensity on a VAS of around 30% have been reported to be above a minimal amount of change considered to be important to patients with chronic pain[19,20]. On the RMDQ a 5-point change (20%) is considered the minimal amount of change that must occur before it is noted by patients according to anchor-based methods for calculating the minimal important difference, and improvements around 30% are considered important based on distribution-based calculations of the minimal important difference. The > 30% improvement criteria for the PGA in this study exceeds the > 20% improvement originally recommended as part of a responder criteria for a PGA in osteoarthritis[21]. Further, the RI aligns well with recommendations from leading clinical experts and a consensus panel on pain measurement[8,12].

That the PGA is the major driver of the RI is not surprising, as some consider this simple, single-item measure a valid method of determining the patient's overall impression of a treatment: pain relief, effect on functioning, and tolerability.

For this study we were interested in the possibility of

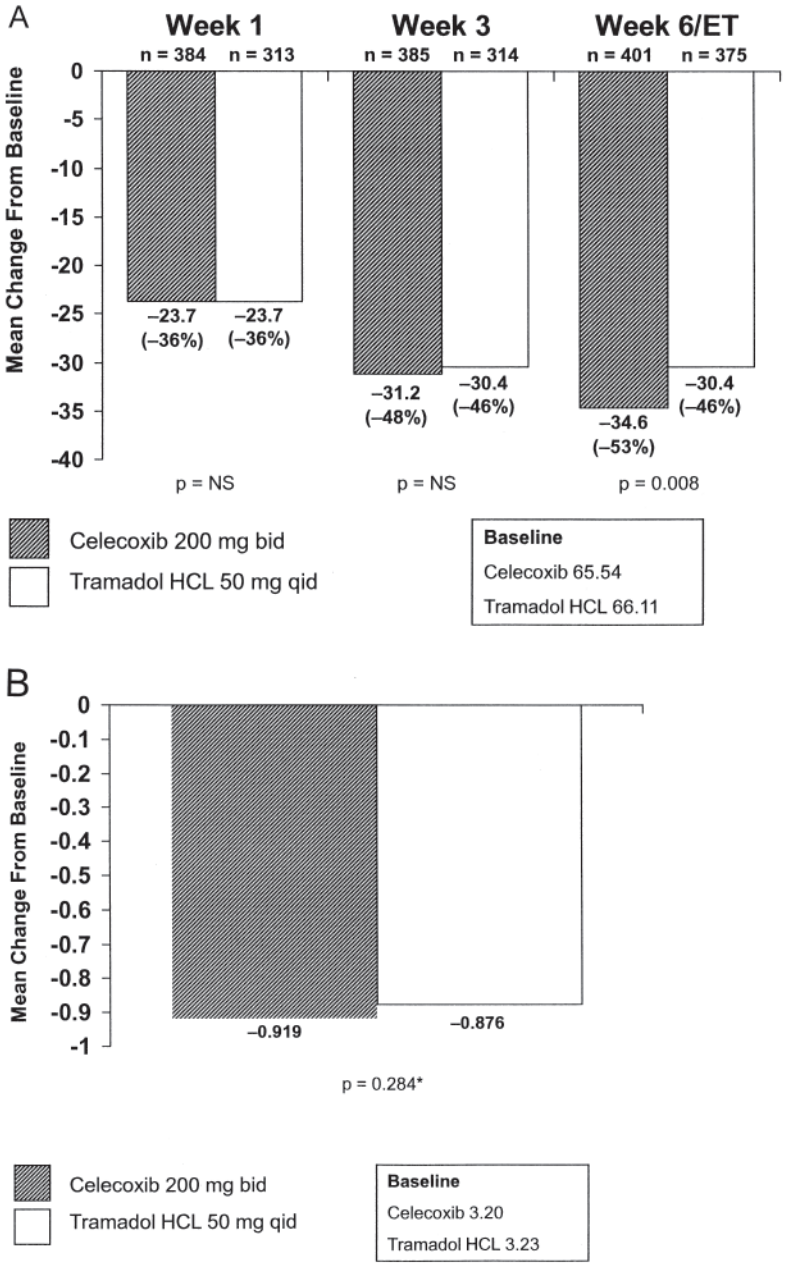

Figure 2. A. Analysis of severity of low back pain as measured by VAS. B. Change in patient global assessment of disease activity at Week 6. *Celecoxib vs tramadol HCL, based on a general linear model with a change from baseline as dependent variable and factors for treatment and center and baseline score as a covariate. C. (overleaf) Roland-Morris Disability Questionnaire scores. HCL: hydrochloride; ET: early termination; NS: not significant.

improving the preliminary RI by substituting the short-form version of the RMDQ (18 items) with the longer-form version used in this study (24 items). There is some preliminary evidence that the content validity of the short version is better than the long version; therefore it would be anticipated that the measure may perform better in a clinical trial[22]. When we scored the long form excluding the items that are not included in the short form, only 21 of the 791 patients (2.7%) would be flagged differently for the RMDQ component of the CLBP RI using the RMDQ 18 versus the RMDQ 24. More importantly, only 6 of the 791 patients (0.8%) would have a different CLBP RI classification if based on the RMDQ 18 instead of the RMDQ 24. Therefore, there is little advantage to using the short version of the RMDQ in the RI except for reduction of burden to the study subject.

We also considered whether it might be useful to use an

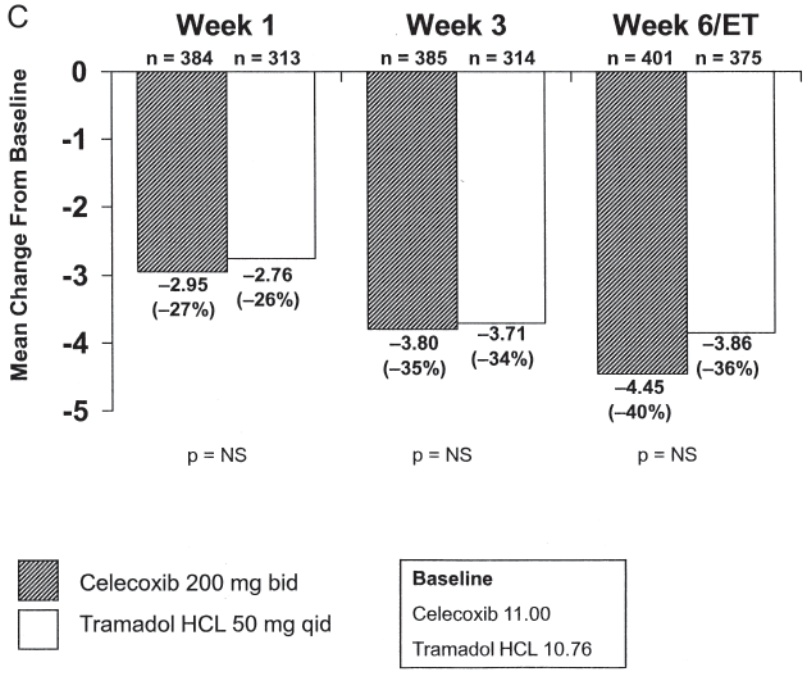

Figure 2C.

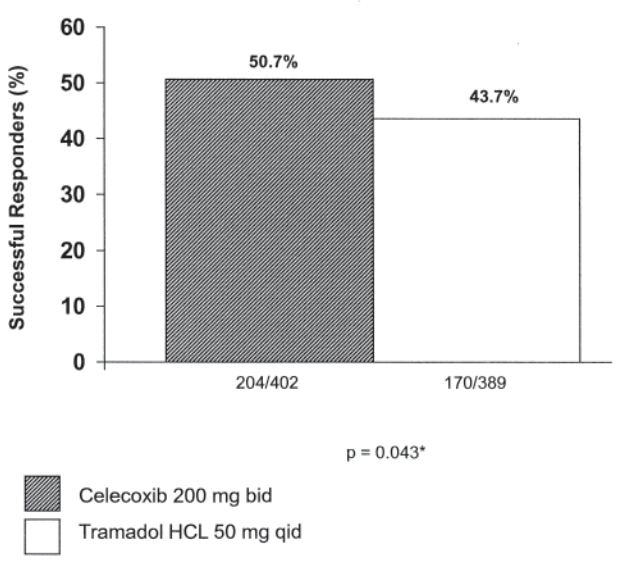

Figure 3. Results on the chronic low back pain responder index at Week 6. *Celecoxib vs tramadol hydrochloride (HCL), based on the Cochran-Mantel-Haenszel test, stratified by center.

NRS rather than a VAS in the CLBP RI as patients have shown a preference for using an NRS over a VAS[23], and therefore there might be fewer missing data on an NRS. Among the 3805 observed cases generated by the 791 patients across visits, there were only 3 cases in which the VAS was completed and the NRS was not completed. Moreover, there were only 2 cases where the NRS was completed and the VAS was not completed.

It may be useful to alter the threshold cutoff point for the RMDQ. In our study, about 700 of 800 subjects met the RMDQ threshold. It is hypothesized, therefore, that making this cutoff more stringent may improve the performance of the index.

The effect sizes found in our study are low for the individual components of the index and for the total RI. This indicates a low magnitude of treatment effect; however, given the current available treatment options for CLBP, a moderate to large effect size would have been a surprise. A systematic literature review of effect sizes of nonsurgical treatments for CLBP identified small treatment effect sizes, with treatment effect being lowest in nonspecific LBP[24]. Acupuncture, behavioral therapy, exercise therapy, and nonsteroidal antiinflammatory drugs (standardized mean differences 0.61, 0.57, and 0.52, respectively; RR 0.61) had almost equivalent effect sizes. The small difference between interventions and placebo as measured by a pain endpoint supports further the need for multidomain criteria that may enhance discrimination between responders and nonresponders. Overall, the RI has greater construct validity than the individual components alone and it is able to differentiate between 2 active treatments as well as an active treatment versus placebo. Further, the RI discriminates better compared with using the RMDQ alone.

One weakness of our study is that it is purely quantitative. We did no further testing of the content validity of the CLBP RI in patients and have essentially started from the point that the components of the RI are correct. In the orig-

Table 2. Components of the responder index.

	PGA > 30% Improvement		LBP VAS ≥ 30% Improvement		RMDQ < 20% Worsening		
	No	Yes	No	Yes	No	Yes	Total (%)
Nonresponder	357	60	266	151	90	327	417 (52.7)
Responder	0	374	0	374	0	374	374 (47.3)
Total (%)	357 (45.1)	434 (54.9)	266 (33.6)	525 (66.4)	90 (11.4)	701 (88.6)	791
κ coefficient	0.849		0.625		0.207		

PGA: patient global assessment; LBP: low back pain; VAS: visual analog scale; RMDQ: Roland-Morris Disability Questionnaire.

Table 3A. Sensitivity analysis of chronic low back pain responder index cutoff points (keeping the RMDQ at 20%.

LBP VAS Cutoff	PGA Cutoff	RMDQ Cutoff	Responders, %
25	25	20	52.0
25	30	20	47.7
25	35	20	21.9
30	25	20	51.2
30	30	20	47.3
30	35	20	21.9
35	25	20	49.4
35	30	20	46.3
35	35	20	21.7

LBP: low back pain; VAS: visual analog scale; PGA: patient global assessment; RMDQ: Roland-Morris Disability Questionnaire.

Table 3B. Sensitivity analysis of chronic low back pain responder index cutoff points (decreasing the RMDQ cutoff by 5%).

LBP VAS Cutoff	PGA Cutoff	RMDQ Cutoff	Responders, %
25	25	15	51.8
25	30	15	47.7
25	35	15	21.9
30	25	15	51.1
30	30	15	47.3
30	35	15	21.9
35	25	15	49.3
35	30	15	46.3
35	35	15	21.7

RMDQ: Roland-Morris Disability Questionnaire; LBP: low back pain; VAS: visual analog scale; PGA: patient global assessment.

Table 3C. Sensitivity analysis of chronic low back pain responder index cutoff points (increasing the RMDQ cutoff by 5%).

LBP VAS Cutoff	PGA Cutoff	RMDQ Cutoff	Responders, %
25	25	25	52.2
25	30	25	47.8
25	35	25	22.0
30	25	25	51.5
30	30	25	47.4
30	35	25	22.0
35	25	25	49.7
35	30	25	46.4
35	35	25	21.9

RMDQ: Roland-Morris Disability Questionnaire; LBP: low back pain; VAS: visual analog scale; PGA: patient global assessment.

Table 4. Comparison of chronic low back pain (CLBP) responder index (RI) effect size.

Variable	Effect Size in Current Study	Pooled Effect Size in Initial CLBP RI Studies
LBP VAS	–0.24	–0.25
PGA	–0.24	–0.20
RMDQ	–0.12	–0.08
Responder index	0.14	0.19 (in COX-A-244) 0.23 (in COX-245)

LBP: low back pain; VAS: visual analog scale; PGA: patient global assessment; RMDQ: Roland-Morris Disability Questionnaire.

inal study, the components of the index were determined based on a review of the literature, 3 CLBP focus groups, and clinical opinion, so there is a clear rationale for the index; however, there can be no assurances that if additional patients or clinicians were interviewed regarding the correct makeup of the index, alternative content would be derived. Further, even if the content of the index is largely correct, a constraint is imposed because of the availability of existing measures that may be used to measure the content. For instance, the RMDQ is used to measure functioning and daily activities. It is possible that there are other measures, as yet to be developed, that would more appropriately tap into those constructs and prove more responsive in a CLBP RI.

A further weakness of our study is that the validation of the RI has been limited to trials of COX-2 selective inhibitors (i.e., celecoxib and valdecoxib) in LBP of nociceptive origin. Although our trial offers the advantage of an active comparator, it is necessary to determine how this index performs in trials of other analgesics (e.g., opioids) and LBP of a different etiology (e.g., neuropathic). There

could be a risk also of concluding effectiveness in treatments where the therapeutic effect on individual endpoints is minimal, thus warranting assessment of outcomes based on clinical relevance.

Significantly more patients improved with celecoxib compared with tramadol HCL based on the CLBP RI proposed in the initial investigation. The RI appears to be particularly sensitive to the cutoff point used for improvement in the PGA component. Changes in the improvement criterion for this component determine, to a large extent, the percentage of patients classified as responders. Further testing of the CLBP RI in clinical trials of other agents is necessary to confirm its validity in this population.

ACKNOWLEDGMENT

L. Prevost of Parexel provided support for formatting the manuscript.

REFERENCES

1. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine 1995;20:1899-909.
2. Dagenis S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8-20.
3. Luo X, Pietrobon X, Sun SX, Liu GG, Hey L. Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. Spine 2004;29:79-86.
4. US Department of Health and Human Services, National Center for Health Statistics: National Health Interview Survey (NHIS), 2007. [Internet. Accessed Sept 13, 2010.] Available from: http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm
5. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9-19.
6. Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing Spondylitis Assessment Group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876-86.
7. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
8. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389-99.
9. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202-9.
10. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106:337-45.
11. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105-21.
12. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine 1998;23:203-13.
13. Simon LS, Evans C, Katz N, Bombardier C, West C, Robbins J, et al. Preliminary development of a responder index for chronic low back pain. J Rheumatol 2007;34:1386-91.
14. Roland MO, Morris RW. A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low back pain. Spine 1983;8:141-4.
15. Roland MO, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 2000;25:3115-24.
16. O'Donnell JB, Ekman E, Spalding WM, Bhadra P, McCabe D, Berger MF. The effectiveness of a weak opioid medication versus a cyclo-oxygenase-2 (COX-2) selective non-steroidal anti-inflammatory drug in treating flare-up of chronic low-back pain: results from two randomized, double-blind, 6-week studies. J Int Med Res 2009;37:1789-802.
17. Landis J, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
18. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
19. Dworkin R, Turk D, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105-21.
20. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability questionnaire for low back pain. J Clin Epidemiol 2006;59:45-52.
21. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389-99.
22. Horowicz-Mehler N, Evans CJ, Abetz L, Copley-Merriman K. The content validity of clinician derived patient reported outcomes (PRO) measures: the Roland Morris Disability Questionnaire [abstract]. Value Health 2005;8:245.
23. Gagliese L, Weizblit N, Ellis W, Chan VW. The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. Pain 2005;117:412-20.
24. Keller A, Hayden J, Bombardier C, van Tulder M. Effect sizes of non-surgical treatments of non-specific low-back pain. Eur Spine J 2007;16:1776-88.