CrossMark

Twin Peaks? No Evidence of Bimodal Distribution of Outcomes in Clinical Trials of Nonsurgical Interventions for Spinal Pain: An Exploratory Analysis

Neil E. O'Connell,* Steven J. Kamper,† Matthew L. Stevens,† and Qiang Li†

*Health Economics Research Group, Institute of Environment, Health and Societies, Department of Clinical Sciences, Brunel University London, Uxbridge, United Kingdom.
†The George Institute for Global Health, Sydney Medical School, University of Sydney, Sydney, Australia.

Abstract: The presence of bimodal outcome distributions has been used as a justification for conducting responder analyses, in addition to, or in place of analyses of the mean between-group difference, in clinical trials and systematic reviews of interventions for pain. The aim of this study was to investigate the distribution of participants' pain outcomes for evidence of bimodal distribution. We sourced data on participant outcomes from a convenience sample of 10 trials of nonsurgical interventions (exercise, manual therapy, medication) for spinal pain. We assessed normality using the Shapiro-Wilk test. When the Shapiro-Wilk test suggested non-normality we inspected distribution plots visually and attempted to classify them. To test whether responder analyses detected a meaningful number of additional patients experiencing substantial improvements we also calculated the risk difference and number needed to treat to benefit. We found no compelling evidence suggesting that outcomes were bimodally distributed for any of the intervention groups. Responder analysis would not meaningfully alter our interpretation of these data compared with the mean between group difference. Our findings suggest that bimodal distribution of outcomes should not be assumed in interventions for spinal pain and do not support the automatic prioritization of responder analysis over the between group difference in the evaluation of treatment effectiveness for pain.

Perspective: Secondary analysis of clinical trials of nonsurgical interventions for spinal pain found no evidence for bimodally distributed outcomes. The findings do not support the automatic prioritization of responder analyses over the average between group difference in the evaluation of treatment effectiveness for spinal pain.

Key words: Pain, clinical trials, systematic reviews, effectiveness.

Randomized controlled trials (RCTs), and systematic reviews of RCTs are widely considered the most robust method to evaluate the efficacy and effectiveness of clinical interventions. In pain treatment a critical outcome is pain intensity, which is commonly measured using a self-reported visual analog scale (VAS) or numeric rating scale. In RCTs with outcomes measured on these scales, the only direct estimate of the effects caused by the intervention of interest is the average between group difference in pain score after the intervention. For example, if a hypothetical RCT, comparing an electrotherapy intervention with a placebo control, reported a postintervention between group mean difference of 1 point on a 0 to 10 pain intensity VAS, then we would interpret that the intervention specifically improved pain by 1 point, on average.

A range of examples from trials of pharmacologic therapies for pain show that that the pattern of participant outcome is often bimodally distributed.[21,23,24] Simply put, of the patients who receive the active intervention some experience substantial improvement (and are often referred to as "responders"), some have minimal to no change in their outcome ("nonresponders"), and very few experience intermediate (moderate) effects.

This bimodality of outcomes is proposed to reflect that these interventions deliver large treatment effects for a minority of patients and that this large benefit to those individuals is essentially washed out in the between-group average difference, because of the large proportion of nonresponders, leading to underestimation of the effectiveness of the intervention. In this instance, the average treatment effect, estimated according to the between group difference, may not reflect the experience of many or indeed any participants who received the intervention. On that basis it has been proposed that in trials for chronic pain, the mean between group difference lacks utility as a measure of effectiveness.[22,24,25]

One proposed solution to this problem is to conduct a 'responder analysis,'[9,10,22,24,25] which compares the proportion achieving a predetermined clinically important improvement from baseline in the treatment and control groups. It is proposed that this type of analysis better quantifies individual participant responses to treatment, enables the calculation of easily interpreted measures such as the number needed to treat to benefit (NNTB)[9,10,22,24,25] and thereby provides more meaningful and interpretable estimates of treatment effectiveness than the between group mean difference. Such analyses can use thresholds for important change derived from patients. For example, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) consensus group recommendations[5] use benchmarks on the basis of studies that compared pain scores with global impression of change in patients with neuropathic pain,[7] arthropathies,[27] and pain after spinal cord injury or amputation.[13]

In responder analysis of clinical trials, the between group difference in the proportion of participants who experience a good outcome reflects the net increase in the proportion of patients who achieve the desired outcome because of the intervention. However, responder analyses have attracted criticism over certain conceptual and practical limitations. The term "responder analysis" encourages the conflation of patient outcomes with treatment effects.[28] 'Responders' are identified by within person change from baseline. That change (or outcome) may be because of natural recovery, nonspecific treatment effects, and/or regression to the mean as well as, or instead of, the effects of the intervention and it is not possible to isolate treatment effects from these other causes of within person change.[14] As such, the label "responder" simply identifies a participant whose outcome improved, regardless of whether improvement was because of the treatment or not. In fact, treatment may be responsible for very little, or none of that improvement in outcome. It is also possible that some individuals who actually responded to the intervention might not be counted as responders. For example, if the natural history of a person during the treatment period would have been significant worsening, yet with treatment their condition remains stable, they will be counted as nonresponders despite receiving significant benefit from the intervention. Similarly, a lack of symptom improvement and significant worsening of symptoms are both counted simply as nonresponse. Finally, the dichotomization of outcomes in

responder analyses greatly reduces the precision of estimates of effect, necessitating larger sample sizes to detect significant differences.[28]

Fig 1 illustrates this issue. Data from the treatment arm of a hypothetical RCT are bimodally distributed with a small but distinct group of "responders" achieving substantial (≥50%) pain relief. Because a larger group achieves little or no relief the average reduction in pain in this group is small. In Fig 1B the control group data are added and are unimodally distributed. In this example the between group average difference in pain relief is small because of the large number in the treatment group who experienced little pain relief. In this case, a treatment that is actually effective for a certain group of people might be considered not meaningfully better than control, on the basis of a small mean between group difference. In this instance a case can be made that a responder analysis is more informative than the average between group difference in pain scores.

Examples of bimodal outcome distributions have been reported in selected examples of pharmacological interventions for pain[21,23,24] but it has not yet been established whether this pattern is commonly present across a wide range of interventions. As such, the case for the wider replacement of the between group difference with responder analysis in trials and reviews of interventions for pain has not been established. One finding that would support such a case would be the observation of bimodal outcome distributions (expressed as percentage change from baseline) in groups receiving various different types of interventions. This case would be further supported if the equivalent outcomes from placebo, waiting list, or minimal care control groups did not also show a bimodal outcome distribution, because evidence of bimodal outcomes is not in itself evidence of bimodal effects. When outcomes are normally distributed in all groups or similarly distributed in intervention and control groups, it is likely that the average between group effect remains the most representative and informative estimate of treatment effectiveness.

Aim

The aim of our exploratory study was to investigate the distribution of participants' pain outcomes using data from a selection of trials of nonsurgical interventions for spinal pain.

Methods

This was a secondary analysis of data from recent trials of different interventions for spinal pain. We sourced data on participant outcomes from a convenience sample of 10 trials of nonsurgical interventions for low back and neck pain conducted by researchers at the George Institute for Global Health that measured pain intensity as a primary outcome. Our analysis included all relevant trials conducted by this group over a 9-year period.

For each trial, we calculated the percentage change in pain intensity score from baseline to the first follow-up point after intervention for each individual participant
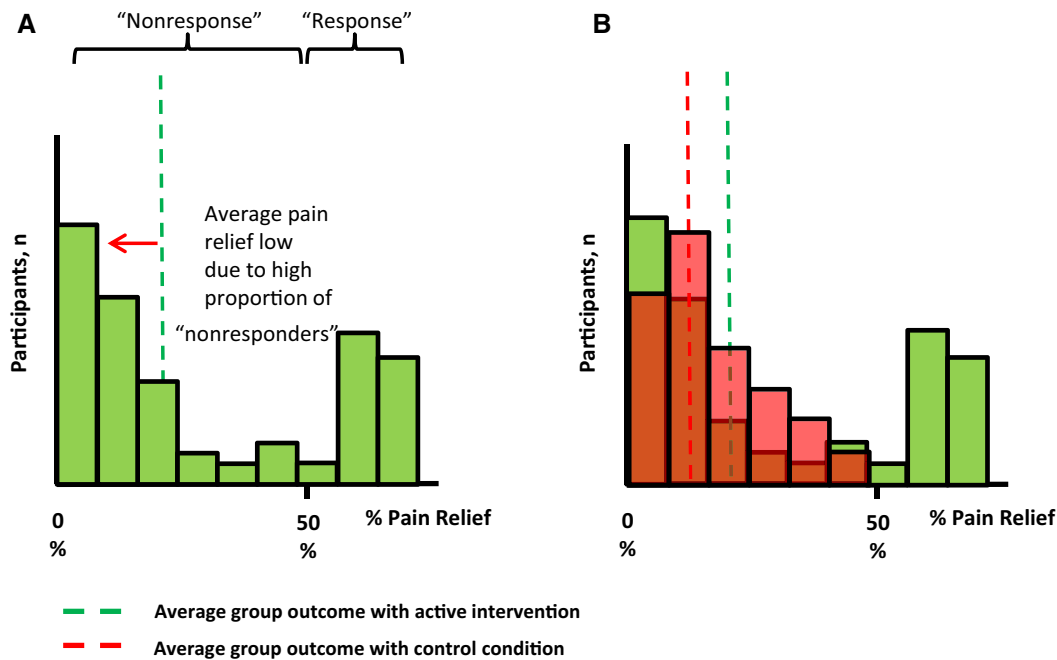
Figure 1. Impact of bimodally distributed outcomes. Hypothetical data from an active intervention arm of an RCT. A small proportion of participants experience substantial (≥50%) pain relief in the intervention whereas most experience little improvement. Because most participants experience little or no change the estimate of average pain relief experienced is modest. The control group data are superimposed (red). Outcomes are unimodally distributed in this group. The difference in average pain relief between groups is small despite the presence of more "responders" in the active group.

in the trial. We then aggregated these to create a distribution of change scores (outcomes) for each arm of the trial.

We divided the samples according to duration of pain (acute and chronic) and pooled data according to the type of intervention. Before combining data from different trials we checked, using visual inspection, that baseline pain scores were similarly distributed. When this was the case we combined data from different trials in the following groups, established a priori:

1. Exercise interventions,
2. Manual therapy interventions,
3. Medications, and
4. Placebo/sham control, minimal intervention conditions.

Data from all comparison/control groups (placebo/sham, wait list, minimal care) were combined. For dropouts, missing data were imputed using the baseline observation carried forward method.

Tests for normality of the outcome distribution were conducted using the Shapiro-Wilk test. When these tests suggested non-normality we inspected the distribution plots visually and attempted to classify them. When separate distribution peaks suggestive of groups with distinct distributions were observed we planned to assign a cut point ("knot"), decided post hoc upon observation of the data and consider whether distributions on each side of the knot followed known parametric distributions, for example, normal, binomial, or Poisson distribution. However, we did not observe such distinct peaks in our data so this was not conducted. We conducted a sensitivity analysis to examine the effect of imputing missing data by reanalyzing the data using an available case approach. No adjustment was made for baseline levels of pain intensity.

In a secondary analysis we tested the proposition that small average between group treatment effects in our data set might hide the fact that a meaningful number of additional patients experienced substantial improvements above what they would have received in the comparison condition. Using the IMMPACT group recommend thresholds of 30% and 50% improvement from baseline[5] as markers of moderately and substantially important change, respectively, we calculated the risk difference (absolute risk increase or decrease) and number needed to treat for all trials that reported an average between group difference of <10 mm or 1 point on a 100-mm 10-point pain scale. This threshold was chosen on the basis of the recent Outcome Measures in Rheumatology (OMERACT12) consensus group recommendations for a threshold to define a minimal clinically important effect.[3]

Results

Description of Included Studies

Data were available from 10 RCTs published between 2007 and 2015 that included a total of 2,926 participants and were conducted by the George Institute for Global Health and the University of Sydney. In acute pain conditions we included data from 3 RCTs[12,17,32] and in chronic pain conditions we included 6 RCTs.[4,8,11,16,20,29] One study[26] recruited participants with subacute low back pain, with pain of 6 to 12 weeks duration; we included this study in the chronic category but conducted a sensitivity analysis to see whether its exclusion from this group affected our results.

Table 1. Included Trials According to Intervention Group

ACUTE/CHRONIC	INTERVENTION	TRIALS INCLUDED	POOLED N FOR GROUP
Acute	Placebo	Williams et al[32]	606
		Hancock et al[12]	
	Exercise	Machado et al[17]	74
	Manual therapy	Hancock et al[12]	60
	Medications	Williams et al[32]	1,154
		Hancock et al[12]	
Chronic	Placebo	Costa et al[4]	311
		Hall et al[11]	
		Michaleff et al[20]	
	Exercise	Hall et al[11]	641
		Costa et al[4]	
		Macedo et al[16]	
		Pengel et al[26]	
		Ferreira et al[8]	
		Michaleff et al[20]	
		Stewart et al[29]	
	Manual therapy	Ferreira et al[8]	80

Interventions included exercise interventions (Tai Chi, motor control exercise, McKenzie spinal exercises, multimodal exercise, graded activity), manual therapy (spinal manipulative therapy), medication (paracetamol 665 mg modified-release regular dose plus placebo tablets "as needed," or 665 mg modified-release regular dose and 500 mg paracetamol immediate-release tablets "as needed," diclofenac 50 mg twice daily). Placebo/minimal care control interventions included placebo tablets, a single advice session and booklet, and detuned electrotherapy. Two trials included participants with whiplash-associated disorder[20,29] and 8 included participants with low back pain.[4,8,11,12,16,17,26,32] Table 1 outlines the included trials and interventions contributing to the analysis and Table 2 summarizes the characteristics of participants in the included trials.

Distributions of Outcomes

Acute

Participant outcomes were not normally distributed for any of the groups (Shapiro-Wilk test for all groups $P < .05$; Table 3). No evidence of bimodal of distribution was observed for any of the groups (Fig 2). Distributions generally showed a unimodal distribution with a strong positive skew reflecting that most participants in these trials experienced complete or substantial improvement in their pain, whichever group there were assigned to, including placebo and minimal care groups. Participants receiving exercise interventions showed a broader and more variable distribution.

Chronic

Participant outcomes were not normally distributed for any of the groups (Shapiro-Wilk test for all groups $P < .05$; Table 3). No clear evidence of bimodal of distribution was observed. The placebo/minimal care and exercise group distributions were unimodal with a positive skew (Fig 3) with the peak of the distribution reflecting that most participants experienced little change from baseline. The manual therapy distribution showed a broader and more random distribution but without strong evidence to suggest bimodality. Reanalyzing with data from the study of Pengel et al[26] removed did not alter the results.

Sensitivity Analysis

Reanalysis of all groups using an available case approach instead of imputing missing data using baseline observation carried forward did not reveal evidence of bimodal distribution for any of the intervention classes.

Secondary Analysis

We calculated the risk difference and NNTB for comparisons that showed small (<1 point change on a VAS) between group mean differences. These results are shown in Tables 4 and 5. In our sample, no comparisons with between group differences <1 showed statistical significance at the $P < .05$ level. Similarly, the NNTB were frequently high (>10) with 95% confidence intervals (CIs) that crossed the line of no effect for all comparisons.

Discussion

In this sample of RCTs of interventions for spinal pain we found no compelling evidence suggesting that outcomes are bimodally distributed for the interventions studied. The distribution of outcomes was different

Table 2. Participant Characteristics for Included Studies

TRIAL	NUMBER OF PARTICIPANTS, N	MEAN AGE (SD), Y	FEMALE SEX, N (%)	PAIN CONDITION	MEAN DURATION OF SYMPTOMS (SD)	MEAN PAIN INTENSITY AT BASELINE (SD)	OUTCOMES ASSESSED AT
Costa et al[4]	154	53.7 (12.8)	93 (60.4)	Chronic LBP	331.5 (392.4) Wk	6.7 (2.0)	8 Wk
Ferreira et al[8]	240	53.7 (14.9)	165 (68.8)	Chronic LBP	103.1 (120.7) Mo	6.3 (2.0)	8 Wk
Hall et al[11]	160	43.4 (13.3)	119 (74.4)	Chronic LBP	n/a	4.4 (2.0)	8 Wk
Hancock et al[12]	239	40.7 (15.6)	105 (43.9)	Acute LBP	9.1 (9.3) D	6.5 (1.7)	4 Wk
Macedo et al[16]	172	49.2 (15.0)	102 (59.3)	Chronic LBP	87.7 (105.0) Mo	6.0 (2.1)	8 Wk
Machado et al[17]	148	46.6 (14.7)	74 (50.0)	Acute LBP	n/a	6.7 (1.9)	3 Wk
Michaleff et al[20]	172	42.9 (12.4)	110 (64.0)	Chronic WAD	21.5 (16.8) Mo	5.1 (2.0)	14 Wk
Pengel et al[26]	259	49.9 (15.8)	124 (47.9)	Subacute LBP	n/a	5.4 (2.0)	6 Wk
Stewart et al[29]	134	43.3 (14.7)	89 (66.4)	Chronic WAD	9.0 (2.4) Mo	5.3 (2.0)	6 Wk
Williams et al[32]	1,643	45.0 (15.8)	766 (46.6)	Acute LBP	9.8 (10.0) D	6.3 (1.9)	4 Wk

Abbreviations: LBP, low back pain; WAD, whiplash associated disorder; n/a, not available.
NOTE. Acute defined as <6 weeks, subacute as 6 to 12 weeks, and chronic as >12 weeks.

**Table 3.** Results of Tests of Normality According to Intervention Group (Shapiro-Wilk Test)

ACUTE/CHRONIC	INTERVENTION GROUP	STATISTIC	DF	P
Acute	Placebo	.691	605	<.01
	Exercise	.895	74	<.01
	Manual therapy	.660	59	<.01
	Medications	.715	1,154	<.01
Chronic	Placebo	.736	307	<.01
	Exercise	.776	636	<.01
	Manual therapy	.961	80	.016

between chronic and acute conditions, with acute conditions predictably showing a positively skewed, but essentially unimodal, distribution, indicating that most participants experienced substantial pain relief. This pattern was the same regardless of the receipt of any particular treatment, or control intervention. For chronic conditions outcome distributions did not meet the threshold for normality as per the Shapiro-Wilk test, but were more evenly spread from mild worsening, to little or no change, to mild or moderate improvement.

Recommendations to prioritize or at least coreport responder analyses over the mean between group differ-

ence have been made with particular regard to the evaluation of drug therapies for pain.[3,5,22,24,25] Specific examples of bimodal distribution of outcomes have been presented for: nonsteroidal anti-inflammatory drugs (etoricoxib and naproxen) in acute postoperative pain, osteoarthritis pain, chronic low back pain, and ankylosing spondylitis[23]; and duloxetine in osteoarthritis, fibromyalgia, chronic low back pain, and painful diabetic neuropathy.[21] In the data presented in this report we found no evidence for bimodality of outcomes when we pooled data from trials of paracetamol and diclofenac (n = 1,114) for acute low back pain. In the primary analyses of both of those trials no significant between group difference was observed between the drug and placebo groups, suggesting a lack of efficacy, which may explain the lack of bimodal distribution. There was also no evidence of bimodality when we investigated each drug individually or for nonpharmacological interventions. These findings suggest that bimodal distribution of outcomes for pain interventions cannot necessarily be assumed.

In our secondary analysis we sought to see if small (<1 point on a 0–10 VAS) between group differences in pain intensity corresponded to potentially a favorable NNTB. The proposed value of responder analysis may lie with interventions for which a between group difference
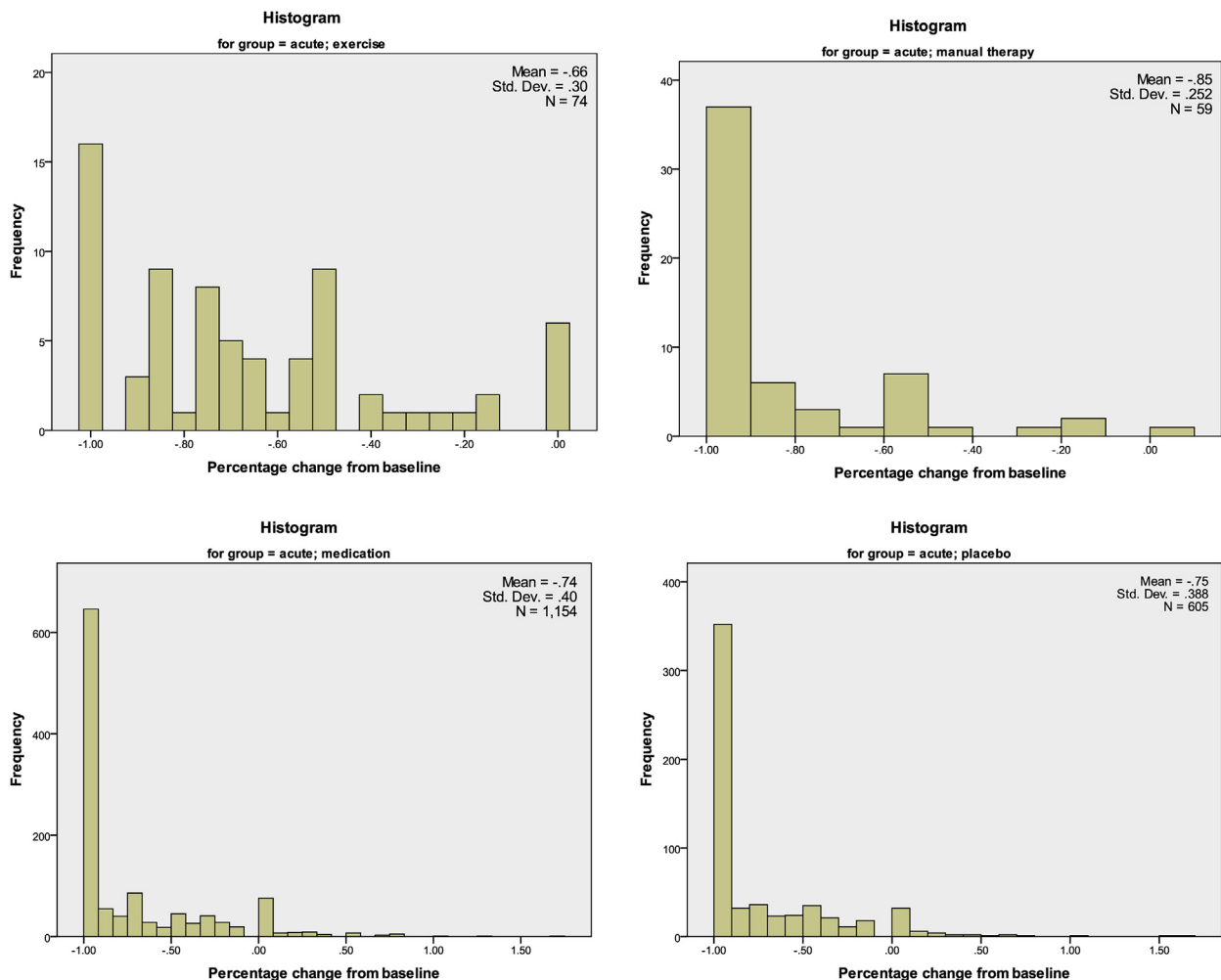
Figure 2. Distribution of participant outcomes in acute pain conditions; −1.0 = 100% pain relief. Please see this figure in color at http://www.jpain.org/.
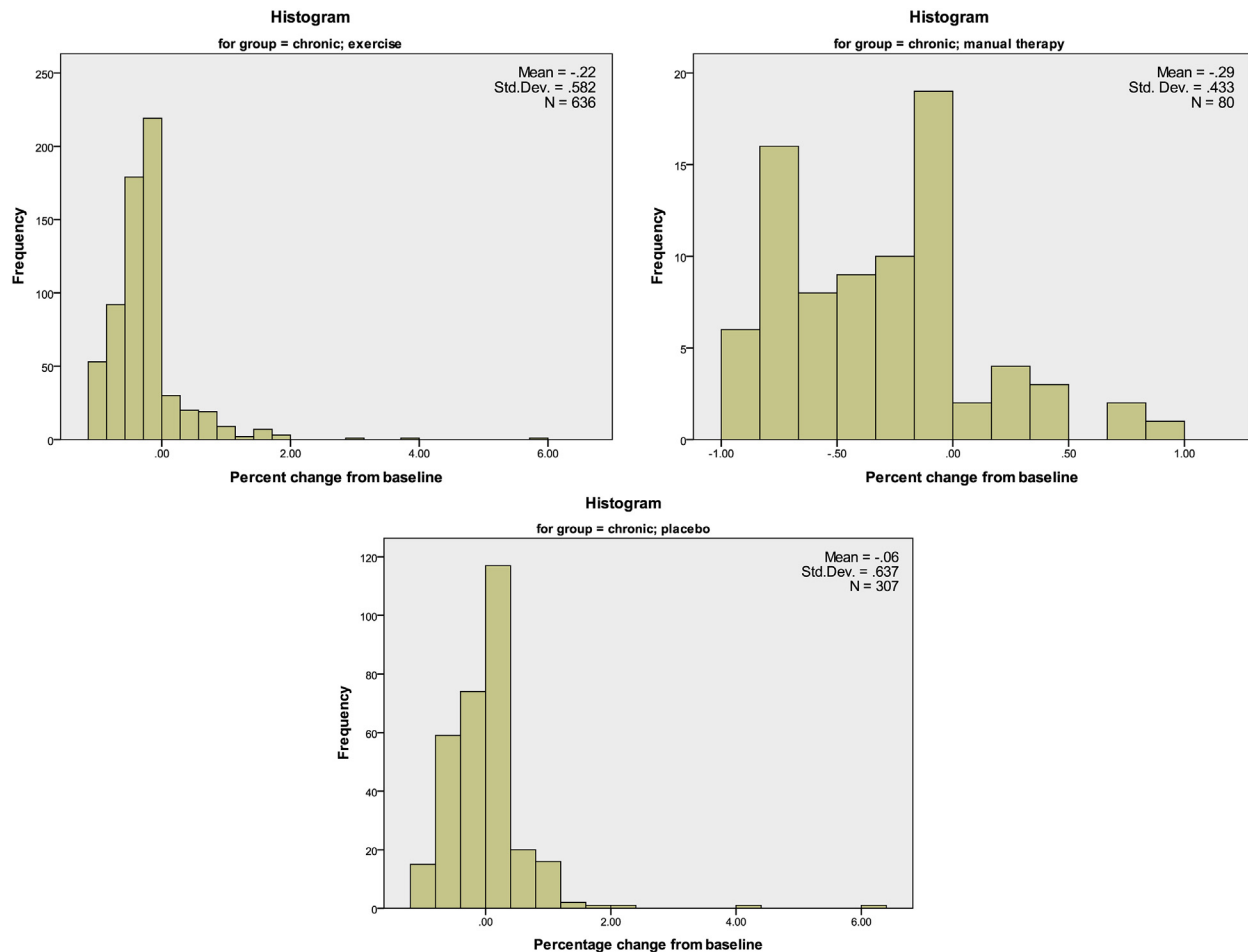
Figure 3. Distribution of participant outcomes in chronic pain conditions; −1.0 = 100% pain relief. Please see this figure in color at http://www.jpain.org/.

compared with the comparison condition is present (for example as indicated by a threshold of statistical significance) but where that difference is small. The absence of statistical significance in the available comparisons in our analysis is reflected in CIs for the risk difference and number needed to treat that all cross the line of no effect. In these data, responder analysis would not meaningfully alter our interpretation of the data compared with the average between-group difference. The lack of bimodality of outcomes in our data may be interpreted as reflecting comparisons of treatments that lack even marginal effectiveness. The 2 comparisons in our data set that showed as statistically significant between group differences (Tai Chi vs waiting list control for chronic low back pain, mean difference 1.2 (95% CI, .6–1.8).[11] Advice with exercise versus placebo for subacute low back pain, mean difference 1.56 (95% CI, .7–2.4)[26] had mean between group differences of >1 point on a pain scale which exceeds the OMERACT 12 group's recent recommended threshold for a minimal clinically important effect. For those comparisons, outcomes in the intervention group did not show a bimodal distribution. Post hoc analysis of these comparisons reveals relatively a favorable NNTB for a 50% reduction in pain of 4 (95% CI, 2–12) for Tai Chi versus usual care and 4 (95% CI, 2–26) for advice and exercise versus placebo. In these examples a

'responder' analysis, although offering an alternative way to present the data, does not change our interpretation of the results compared with a traditional means analysis. Because of the absence of bimodal distribution in these data one interpretation may be that the between group difference may be representative of the average true treatment effect but may have been adequate to tip a larger number of participants over the threshold from "non-responder" to "responder" in the treatment group. In this way generally modest treatment effects may give the false impression of large treatment "responses."

It is clear that the mean between group difference is a somewhat blunt tool with limitations. It tells us little about the individual experience of participants in clinical trials or that factors that drive them. In certain circumstances (ie, in which strong evidence of bimodality is present), responder analysis may be useful although it is crucial that the thresholds used to establish response have validity and are established a priori. Despite the problems outlined previously in establishing clinical "response," it is arguable that responder analyses offer a simple and understandable estimate of what is likely to happen to a patient given a treatment in a clinical setting and therefore useful to inform clinician and patient decision-making. However, it remains important to recognize its limitations as an estimate of

Table 4. Results of Responder Analyses for Effect Sizes <1 Point on a 0 to 10 Pain Scale (Acute Pain)

Trial	Comparison	Between Group Difference (95% CI)*	Risk Difference 30% Change in Pain (95% CI)*	NNTB/NNTH (95% CI)†	Risk Difference 50% Change in Pain (95% CI)*	NNT 50% Change in Pain (95% CI)†
Hancock et al[12]	NSAID versus SMT	−.71 (−1.51 to −.08)	−.03 (−.17 to .11)	NNTH 31 (NNTH 6 to NNTB 9)	−.07 (−.12 to .28)	NNTH 13 (NNTH 8 to NNTB 4)
	NSAID versus NSAID with SMT	−.47 (−1.28 to .08)	−.03 (−.17 to .11)	NNTH 30 (NNTH 6 to NNTB 9)	−.03 (−.20 to .13)	NNTH 30 (NNTH 5 to NNTB 8)
	NSAID versus placebo	−.17 (−.95 to .62)	−.03 (−.17 to .11)	NNTH 30 (NNTH 6 to NNTB 9)	−.03 (−.20 to .13)	NNTH 30 (NNTH 5 to NNTB 8)
	SMT versus NSAID with SMT	.25 (−.53 to 1.02)	−.001 (−.13 to .13)	NNTH 885 (NNTH 8 to NNTB 8)	.03 (−.11 to .17)	NNTB 31 (NNTH 6 to NNTB 9)
	SMT versus placebo	.55 (−.20 to 1.29)	−.001 (−.13 to .13)	NNTH 885 (NNTH 8 to NNTB 8)	.03 (−.11 to .17)	NNTB 31 (NNTH 6 to NNTB 9)
	NSAID with SMT versus placebo	.30 (−.46 to 1.06)	0 (−.13 to .13)	NNT ∞ (NNTH 8 to NNTB 8)	0 (−.15 to .15)	NNT ∞ (NNTH 7 to NNTB 7)
Machado et al[17]	GP (control) versus McKenzie	−.54 (−1.41 to .33)	−.09 (−.27 to .08)	NNTH 11 (NNTH 6 to NNTB 9)	−.16 (−.36 to .04)	NNTH 6 (NNTH 3 to NNTB 28)
Williams et al[32]	Placebo versus Paracetamol (regular dose + as needed)	.17 (−.21 to .46)	.02 (.04 to .08)	NNTB 41 (NNTH 27 to NNTB 12)	.05 (−.02 to .12)	NNTB 21 (NNTH 43 to NNTB 8)
	Placebo versus Paracetamol (regular dose)	−.07 (−.40 to .27)	−.007 (−.05 to .06)	NNTB 128 (NNTH 19 to NNTB 15)	.01 (−.06 to .08)	NNTB 91 (NNTH 18 to NNTB 13)
	Paracetamol (regular dose + as needed) versus Paracetamol (regular dose)	−.19 (−.53 to .15)	−.016 (−.08 to .04)	NNTH 61 (NNTH 13 to NNTB 22)	−.04 (−.11 to .03)	NNTH 28 (NNTH 9 to NNTB 29)

Abbreviations: NNTH, number needed to treat to harm; NNT, number needed to treat; NSAID, nonsteroidal anti-inflammatory drug; SMT, spinal manipulative therapy; GP, general practitioner management.
*Negative values indicate the comparator intervention was more beneficial.
†Because lower values of NNT represent larger effect sizes, no effect represents infinity; see Altman.[1]

Table 5. Results of Responder Analyses for Effect Sizes <1 Point on a 0 to 10 Pain Scale (Chronic Pain)

Trial	Comparison	Between-Group Difference (95% CI)*	Risk Difference 30% Change in Pain (95% CI)*	NNTB/NNTH 30% Change in Pain (95% CI)†	Risk Difference 50% Change in Pain (95% CI)*	NNTB/NNTH 50% Change in Pain (95% CI)†
Costa et al[4]	Motor control versus placebo	.81 (−.11 to 1.72)	.14 (−.07 to .36)	NNTB 7 (NNTH 13 to NNTB 3)	.08 (−.12 to .28)	NNTB 13 (NNTH 8 to NNTB 4)
Ferreira et al[8]	General ex versus specific ex	−.68 (−1.48 to .13)	−.16 (−.37 to .05)	NNTH 6 (NNTH 3 to NNTB 19)	−.13 (−.33 to .07)	NNTH 8 (NNTH 3 to NNTB 13)
	General ex versus SMT	−.55 (−1.37 to .27)	−.11 (−.33 to .10)	NNTH 9 (NNTH 3 to NNTB 10)	−.13 (−.33 to .07)	NNTH 8 (NNTH 3 to NNTB 13)
	Specific ex versus SMT	.13 (−.70 to .95)	.05 (−.17 to .27)	NNTB 20 (NNTH 6 to NNTB 4)	0 (−.21 to .21)	NNT ∞ (NNTH 5 to NNTB 5)
Macedo et al[16]	Graded activity versus motor control	.12 (−.73 to .96)	.04 (−.17 to .25)	NNTB 20 (NNTH 6 to NNTB 4)	.004 (−.20 to .21)	NNTB 247 (NNTH 5 to NNTB 5)
Michaleff et al[20]	Ex versus control	.12 (−.59 to .82)	0 (−.21 to .21)	NNT ∞ (NNTH 5 to NNTB 5)	.01 (−.18 to .20)	NNTB 85 (NNTH 6 to NNTB 5)
Pengel et al[26]	Advice with ex versus advice	.74 (−.20 to 1.69)	.12 (−.11 to .34)	NNTB 9 (NNTH 9 to NNTB 3)	.10 (−.14 to .35)	NNTB 10 (NNTH 7 to NNTB 3)
	Advice with ex versus ex	.80 (−.04 to 1.64)	.125 (−.10 to .35)	NNTB 8 (NNTH 10 to NNTB 3)	.05 (−.20 to .29)	NNTB 21 (NNTH 5 to NNTB 3)
	Advice versus ex	.52 (−.81 to .91)	.01 (−.23 to .25)	NNTB 103 (NNTH 4 to NNTB 4)	−.05 (−.3 to .19)	NNTH 18 (NNTH 3 to NNTB 5)
	Advice versus placebo	.81 (−.05 to 1.69)	.21 (−.03 to .44)	NNTB 5 (NNTH 34 to NNTB 2)	.17 (−.07 to .40)	NNTB 6 (NNTH 15 to NNTB 2)
	Ex versus placebo	.77 (.003 to 1.53)	.20 (−.04 to .43)	NNTB 5 (NNTH 26 to NNTB 2)	.22 (−.01 to .46)	NNTB 5 (NNTH 102 to NNTB 2)

Abbreviations: NNTH, number needed to treat to harm; ex, exercise.
*Negative values indicate the comparator intervention was more beneficial.
†Because lower values of NNT represent larger effect sizes, no effect represents infinity; see Altman.[1]

efficacy. When outcomes are not bimodally distributed it is arguable as to whether responder analysis performs that task any better.

As well as aiming to offer a more meaningful interpretation of bimodally distributed outcomes, responder approaches have been recommended for identifying potentially important characteristics of responders that may act as effect modifiers.[2] However as discussed previously, because "response" includes the influences of natural history, regression to the mean, and the nonspecific effects of interventions, such approaches conflate outcome with treatment response and risk identifying general prognostic indicators as opposed to actual treatment effect modifiers. There are alternative methods for identifying patients who are more likely to respond to a particular treatment that minimize that risk. These include exploring treatment effect modifiers using interaction analyses[15,30] or by exploring proposed mechanisms of action of different treatments using mediation analyses. These studies can also inform clinical decision-making although still require judgement regarding the size and likely clinical importance of any observed effects.[18,31]

Other approaches to interpreting patient-reported outcomes are emerging. For example, cumulative distribution functions have been proposed, in which the full distribution of participants' outcomes in each group is plotted. This approach visualizes the difference between groups in the proportion of participants with a specific outcome across the full range of possible outcomes allowing for the difference between groups to be assessed at any threshold for defining "response."[6,19] Such an approach mitigates the risks associated with relying on a single threshold, although potentially allows increased freedom for selective interpretation.

An important limitation of our study is that it is restricted to a convenience sample of data from a relatively small number of trials of interventions for spinal pain. We may have lacked adequate power to reliably detect bimodality. Further analysis with data from a larger set of trials for a broader range of interventions and painful conditions would be useful. In addition, we have focused on pain as a sole outcome and it is possible that other important outcomes, for example, disability or quality of life, may have shown different results.

Conclusions

In a convenience sample of clinical trials of a variety of conservative interventions for spinal pain we did not find evidence that clinical outcomes were bimodally distributed. These findings suggest that bimodal distribution of outcomes should not be assumed in interventions for spinal pain and do not support the automatic prioritization of responder analysis over the between group difference in the evaluation of treatment effectiveness for pain. In light of this, a pragmatic approach for trials might be to prospectively plan for either analysis approach, contingent on the observed distribution of outcomes. We note, however, that this approach is likely to have practical implications for study design and conduct, because the required sample size for a responder analysis may be much larger than that for a between group difference. For systematic reviews of trials of interventions for pain, when information regarding the distribution of outcomes is likely to be missing in most cases and when individual patient data are often not accessible, the decision is likely to be driven by what is reported in the included studies. There is a case for including both types of analysis.

References

1. Altman DG: Confidence intervals for the number needed to treat. BMJ 317:1309-1312, 1998

2. Bouhassira D, Attal N: Translational neuropathic pain research: A clinical perspective. Neuroscience 338:27-35, 2016

3. Busse JW, Bartlett SJ, Dougados M, Johnston BC, Guyatt GH, Kirwan JR, Kwoh K, Maxwell LJ, Moore A, Singh JA, Stevens R, Strand V, Suarez-Almazor ME, Tugwell P, Wells GA: Optimal strategies for reporting pain in clinical trials and systematic reviews: recommendations from an OMERACT 12 workshop. J Rheumatol 42:1962-1970, 2015

4. Costa LO, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, McAuley JH, Jennings MD: Motor control exercise for chronic low back pain: A randomized placebo-controlled trial. Phys Ther 89:1275-1286, 2009

5. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S: Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 9:105-121, 2008

6. Farrar JT, Dworkin RH, Max MB: Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: Making clinical trial data more understandable. J Pain Symptom Manage 31:369-377, 2006

7. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM: Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 94:149-158, 2001

8. Ferreira ML, Ferreira PH, Latimer J, Herbert RD, Hodges PW, Jennings MD, Maher CG, Refshauge KM: Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. Pain 131:31-37, 2007

9. Froud R, Eldridge S, Lall R, Underwood M: Estimating the number needed to treat from continuous outcomes in randomised controlled trials: Methodological challenges and worked example using data from the UK Back Pain Exercise and Manipulation (BEAM) trial. BMC Med Res Methodol 9:35, 2009

10. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS: Interpreting treatment effects in randomised trials. BMJ 316:690-693, 1998

11. Hall AM, Maher CG, Lam P, Ferreira M, Latimer J: Tai chi exercise for treatment of pain and disability in people with persistent low back pain: A randomized controlled trial. Arthritis Care Res (Hoboken) 63:1576-1583, 2011

12. Hancock MJ, Maher CG, Latimer J, McLachlan AJ, Cooper CW, Day RO, Spindler MF, McAuley JH: Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: A randomised controlled trial. Lancet 370:1638-1643, 2007

13. Hanley MA, Jensen MP, Ehde DM, Robinson LR, Cardenas DD, Turner JA, Smith DG: Clinically significant changes in pain intensity ratings in persons with spinal cord injury or amputation. Clin J Pain 22:25-31, 2006

14. Jafari N, Hearne J, Churilov L: Why caution is recommended with post-hoc individual patient matching for estimation of treatment effect in parallel-group randomized controlled trials: The case of acute stroke trials. Stat Med 32:4467-4481, 2013

15. Kamper SJ, Maher CG, Hancock MJ, Koes BW, Croft PR, Hay E: Treatment-based subgroups of low back pain: A guide to appraisal of research studies and a summary of current evidence. Best Pract Res Clin Rheumatol 24:181-191, 2010

16. Macedo LG, Latimer J, Maher CG, Hodges PW, McAuley JH, Nicholas MK, Tonkin L, Stanton CJ, Stanton TR, Stafford R: Effect of motor control exercises versus graded activity in patients with chronic nonspecific low back pain: A randomized controlled trial. Phys Ther 92:363-377, 2012

17. Machado LA, Maher CG, Herbert RD, Clare H, McAuley JH: The effectiveness of the McKenzie method in addition to first-line care for acute low back pain: A randomized controlled trial. BMC Med 8:10, 2010

18. Mansell G, Kamper SJ, Kent P: Why and how back pain interventions work: What can we do to find out? Best Pract Res Clin Rheum 27:685-697, 2013

19. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD: Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res 11:163-169, 2011

20. Michaleff ZA, Maher CG, Lin CW, Rebbeck T, Jull G, Latimer J, Connelly L, Sterling M: Comprehensive physiotherapy exercise programme or advice for chronic whiplash (PROMISE): A pragmatic randomised controlled trial. Lancet 384:133-141, 2014

21. Moore RA, Cai N, Skljarevski V, Tölle TR: Duloxetine use in chronic painful conditions–individual patient data responder analysis. Eur J Pain 18:67-75, 2014

22. Moore A, Derry S, Eccleston C, Kalso E: Expect analgesic failure; pursue analgesic success. BMJ 346:f2690, 2013

23. Moore RA, Derry S, Simon LS, Emery P: Nonsteroidal anti-inflammatory drugs, gastroprotection, and benefit-risk. Pain Pract 14:378-395, 2014

24. Moore RA, Derry S, Wiffen PJ: Challenges in design and interpretation of chronic pain trials. Br J Anaesth 111:38-45, 2013

25. Moore RA, Eccleston C, Derry S, Wiffen P, Bell RF, Straube S, McQuay H, ACTINPAIN Writing Group of the IASP Special Interest Group on Systematic Reviews in Pain Relief, Cochrane Pain, Palliative and Supportive Care Systematic Review Group Editors: "Evidence" in chronic pain-establishing best practice in the reporting of systematic reviews. Pain 150:386-389, 2010

26. Pengel LH, Refshauge KM, Maher CG, Nicholas MK, Herbert RD, McNair P: Physiotherapist-directed exercise, advice, or both for subacute low back pain: A randomized trial. Ann Intern Med 146:787-796, 2007

27. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W: Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 8:283-291, 2004

28. Senn S: Individual therapy: New dawn or false dawn. Drug Inf J 35:1479-1494, 2001

29. Stewart MJ, Maher CG, Refshauge KM, Herbert RD, Bogduk N, Nicholas M: Randomized controlled trial of exercise for chronic whiplash-associated disorders. Pain 128:59-68, 2007

30. Sun X, Briel M, Walter SD, Guyatt GH: Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 340:850-854, 2010

31. Van der Weele TJ: A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology 24:224-232, 2013

32. Williams CM, Maher CG, Latimer J, McLachlan AJ, Hancock MJ, Day RO, Lin CW: Efficacy of paracetamol for acute low-back pain: A double-blind, randomised controlled trial. Lancet 384:1586-1596, 2014