



Published in final edited form as:

Spine (Phila Pa 1976). 2011 December 15; 36(26): 2316–2323. doi:10.1097/BRS.0b013e31820c01a6.

Acute Low Back Pain and Primary Care: How to Define Recovery and Chronification?

Wolf E. Mehling, MD

University of California, San Francisco, CA

Viranjini Gopisetty, MD, MPH

University of California San Francisco

Michael Acree, PhD

University of California San Francisco

Alice Pressman, PhD

Kaiser Permanente Northern California Division of Research

Tim Carey, MD, MPH

University of North Carolina

Harley Goldberg, MD

Kaiser Permanente Northern California

Frederick Hecht, MD

University of California San Francisco

Andrew L Avins, MD, MPH

Kaiser Permanente Northern California Division of Research

Abstract

Study Design—Prospective cohort study

Objective—to establish outcome measures for recovery and chronic pain for studies with patients that present with recent-onset acute low back pain in primary care

Summary of Background Data—Among back pain researchers, no consensus exists about outcome definitions or how to identify primary-care patients as not-recovered from an episode of low back pain. Cut points for outcome scales have mostly been arbitrarily chosen. Theoretical models for establishing minimal important change (MIC) values in studies of patients with low back pain have been proposed and need to be applied to real data.

Methods—In a sample of 521 patients which presented with acute low back pain (<4 weeks) in primary care clinics and were followed for 6 months, scores for pain and disability were compared with ratings on a global perceived effect scale. Using multiple potential “gold standards” as anchors (reference standards), the receiver operating characteristics method was used to determine optimal cut points for different ways of defining non-recovery from acute low back pain.

Results—MIC values and upper limits for pain and disability scores as well as minimal important percent changes are presented for five different definitions of recovery. A previously

mehlingw@ocim.ucsf.edu .

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

suggested 30% change from baseline scores does not accurately discriminate between recovered and not recovered patients in patients presenting with acute low back pain in primary care.

Conclusions—Outcome definitions that combine ratings from perceived recovery scales with pain and disability measures provide the highest accuracy in discriminating recovered from non-recovered patients.

Keywords

Acute low back pain; primary care; outcome definitions; minimal important change; receiver operating characteristics

Introduction

Great efforts have been made in recent years to assess outcome measures and define minimally important clinical differences (MIC) when assessing the efficacy of treatments for low back pain (LBP)^{1–3}. Yet, despite long-existing expert recommendations,^{4, 5} no agreement exists regarding appropriate outcome criteria for defining recovery or chronic pain in patients who present with a new episode of acute LBP. This problem is particularly compelling in primary care, where few of these patients are on sick leave and using return-to-work as primary outcome is inappropriate.

Recent qualitative studies pointed out that patients' views of recovery are spread across multiple domains⁶, are highly individualized³ and do not fit any single standardized instrument used in prior prediction studies, such as pain scales or the Roland-Morris Disability Questionnaire (RM).^{7, 8} A low pain score does not clearly distinguish those viewing themselves as recovered from those who do not.³ Cut-offs vary widely ranging from 0–2 for pain and 2–4 for RM, are arbitrarily defined by median⁹ or quartile splits¹⁰, or percent changes.^{11–13} To address this measurement problem, numerous studies have (a) combined criteria of pain and function^{10, 14–19}, (b) used a symptom satisfaction scale²⁰, or (c) a global-perceived-effect (GPE) or recovery scale, commonly as a 7-point Likert scale^{20–23} and rarely as a dichotomous option.²⁴ One study used a 15-point Likert scale²⁵ that could be collapsed into a 7-point scale.¹³ Patients have expressed difficulties with self-classification into a binary judgment demanding options for ambiguous responses.³ However, a GPE Likert scale provides patient responses in a middle “gray” range¹² of the scale (“slightly improved”, “unchanged” or “slightly worsened”) and presents a challenge for measurement strategies that require binary classifiers with a defined cut point.^{21, 26} Binary classifiers are commonly used for both clinical decision making and prognosis studies, e.g. when assessing the odds of developing chronic pain with specific risk factors. The choice of a cut point for defining recovery versus chronic pain comes with a sensitivity-specificity tradeoff: if we place the cut point where we classify only a few patients with significant pain and/or disability as chronic LBP cases, then we may misclassify many patients with less pain and disability as recovered although they might self-classify as not recovered. For example, when de Vet et al.⁹ defined her reference standard for “important improvement” by using at least “slightly improved” as reference standard on the GPE scale, 35% of their hypothetical patients were viewed as misclassified. Beurskens et al.²¹ compared two different interpretations of recovery on the same scale with “slightly improved” either classified as recovered or non-recovered. Kamper et al.²⁷ used only “fully recovered” patients for their analyses ignoring “much improved” patients, which many might consider to be an overly stringent criterion for prognosis studies.

Furthermore, the choice for how to divide a Likert scale might depend on the research question: efficacy studies interested in recovery may want to neglect the undecided and

move the cut point towards the recovery end of the scale, whereas prognostic studies interested, for example, in chronification of acute pain may move the cut point towards the opposite end of the scale. In addition, criteria for improvement from therapy for chronic pain are different from criteria for recovery from acute pain: a patient suffering from chronic LBP for several years might be content with a smaller improvement in pain and function than a patient with acute LBP, who generally experiences pain and function rapidly improved by more than 50% within a few weeks.²⁸

This paper presents analyses data from a prospective cohort study of patients seeking primary medical care for narrowly-defined acute LBP in the US (main results published separately). Its aim was to explore risk factors for chronic pain and to identify patients who might benefit from early intervention to prevent the progression to chronic pain.

For this study, we reasoned that patients who consider themselves “fully recovered” or “much improved” despite a minor degree of persistent pain and/or functional disability might be expected to not seek further medical services for their LBP and resume pre-episode activity levels. A low cut-off for pain or disability would count a large proportion of these patients as not recovered.³ Therefore, one option could be to use the criterion of at least “much improved” on the GPE scale at the 6-month follow-up as external criterion for recovery, as well as “slightly worse” and “much worse” as external criteria for non-recovery. For the more ambiguous criteria of “slightly improved” or “same” it was less clear how to force these patients into the dichotomy of recovered versus chronic. It has been shown that the way this is sorted according to different “gold standards” has a considerable effect on the sensitivity and specificity of disability measures^{21, 26}. Using a cohort of primary care patients with aLBP, we decided to (1) explore how self-reported global recovery relates to standard measures of pain and disability, (2) determine “optimal” cut-offs for discriminating between recovered and non-recovered patients, and (3) compare sensitivity and specificity of previously suggested “gold standards” with reference standards using *combined* outcome criteria. Although theory-driven face validity of integrated assessment strategies may be appreciated by researchers, they cannot be validated against an external criterion or purported “gold standard”.

Methods

Study Participants

Members of the largest health maintenance organization (HMO) in Northern California, seeking primary medical care for acute LBP were interviewed twice over the phone, at baseline and at six month follow-up. Acute LBP was defined as back pain between rib cage and buttocks of less than four weeks. Patients 18–70 years of age were included if they spoke English, had no prior LBP episode in the past year, no red flags (fever, cancer history, inflammatory/rheumatoid diseases), no history of spine surgery, no diagnosis of fibromyalgia, or current pregnancy. Patients with sciatica, defined as pain radiating below the knee, were not excluded unless they were scheduled for surgery at the time of the baseline interview. From February 2008 to March 2009, on the day following their clinic visit, consecutive patients were identified by a computer program from electronic medical records and invited by mail to participate in the study. The sample represented the socio-economic and ethnic diversity of the population of health-insured adults in Northern California seen in primary care for acute LBP.²⁹

Measures

We assessed pain scores for average pain, bothersomeness of pain^{20, 30}, least and worst pain (in past week) by 11-point numeric rating scales (NRS) and functional disability by RM at

both time points, and a GPE scale at follow-up²¹. We calculated absolute and percent changes. We collapsed answering options “much worsened” and “vastly worsened” on the GPE, thereby reducing the original 7-point Likert scale to 6 points. We explored additional criteria suggested by Jordan³¹, Ostelo¹ and Fritz¹³ assessing the proportion of patients that improved from their baseline parameters by 30% or 50%.

Analyses

We used the receiver operating characteristics (ROC) method with GPE as reference standard to assess (1) the minimally important change (MIC)^{1, 2} values for pain and disability perceived by patients as sufficient to self-classify as recovered, (2) upper limits for pain and functional disability compatible with perceived recovery²⁷, and (3) minimally important *percent* changes for pain and disability from baseline scores. We assessed the areas under the curves (AUC) as quantified measures for the overall ability of the scales to discriminate between patients who recovered and those who did not.^{32, 33} Similar to de Vet et al.² we determined cutoff scores that combined maximal sensitivity with optimal specificity for identifying non-recovered patients. Similar to Beurskens et al.²¹, in the absence of a gold standard, cutoffs for this sample were based on different GPE interpretations as reference standard: patients who were “slightly improved” at 6 months were either counted as recovered (Reference Standard 1) or non-recovered (Reference Standard 2).

We explored *combined* outcome criteria (Reference Standards 3–5): patients reporting to be at least “much improved” on the GPE scale were classified as recovered, and patients reporting to be “worse” were classified as non-recovered. Patients reporting to be “slightly improved” or “same” were classified as non-recovered if their scores at follow-up exceeded the upper limit of what all patients perceive as compatible with recovery as determined by ROC curves using Reference Standard 1. We conducted multiple analyses exploring which additional criterion would discriminate recovered from non-recovered with greatest sensitivity and specificity. Confidence intervals for MIC values were estimated by bootstrapping (1000 replications).

To estimate minimally important change (MIC) thresholds, we used the cut-point corresponding to the smallest residual sum of sensitivity and specificity, similar to the study by de Vet et al.² We used Stata11 software³⁴ with an additional module provided by R. Froud (London, UK).³⁵

Results

605 patients fulfilled eligibility criteria and were interviewed at baseline. This represents 25% of the 2,454 respondents to invitations mailed to 42,650 patients who were seen for any kind of LBP in clinics of the HMO during the twelve months of recruitment. 521 patients (86%) completed a 6-month follow-up interview. Table 1 shows mean self-ratings on the GPE scale and mean pain and disability scores for six response levels. At 6-month follow-up, 32% of patients reported to be “fully recovered”, 81% to be at least “much improved” and 91% at least “slightly improved”. If we classify patients reporting to be “slightly improved” as recovered (Reference Standard 1), 47 (9%) would be classified as non-recovered; if we classified the same patients as non-recovered (Reference Standard 2), we would classify 98 (19%) of all patients as non-recovered.

Table 2 shows the average percentage changes in pain and disability from baseline to 6-month follow-up for each GPE score. For the “completely recovered” and “much improved” GPE groups, pain and RM disability similarly improved on average by approximately 100% and 80%. The two GPE groups reported “slightly improved” or “same” improved by 30–

40% with mean RM change scores being identical for both groups (41%). The finding that patients with 30–40% improvement in pain or disability may report their follow-up situation as being the “same” illustrates the potential for misclassification if we use a single criterion of GPE, pain or disability for discriminating between recovery and chronic pain.

Table 3 shows the proportions of patients in each GPE category who improved by more than 50% or 30%, respectively, from baseline to six months. Though the proportions of patients who improved by either 30 or 50% in a parameter were quite similar within the subgroups at both ends of the GPE scale (“much improved” and “much worse”), these proportions clearly differed in the GPE scale’s middle range. In “slightly improved” patients, less than half of the patients reported a 50% reduction in pain or disability; in this GPE group the mean RM score was 8 (median = 7; see Table 1) which is above the reference standard score of ≥ 7 for chronic pain in several prior studies.^{9, 36} Consequently, half of these patients would fall into the chronic pain outcome group if we used a RM score of ≥ 7 or a 50% reduction in pain and function as reference. These findings question the accuracy of a dichotomous outcome using the GPE scale and classifying “slightly improved” patients as recovered². Although the number of self-reported “slightly worse” patients in our sample is too small ($N = 9$) to draw general conclusions, choosing a 30% improvement in the RM score as criterion for improvement would classify more than half of these as improved and therefore render this choice problematic. To reiterate, in general dichotomous classifications based on a single criterion may be problematic.

In which way do “completely recovered” patients differ from “much improved” patients? Almost half of the patients (117 of 253; 46%) reportedly not “completely recovered” but “much improved” were free of pain at 6 months, with a mean RM score of 1.8 ($SD \pm 2.5$) (data not presented in tables). In other words, the majority of “much improved” patients reported pain in the past week rated 1.8 for average intensity and 4.2 for worst pain. Generally, if at follow-up patients still had pain, worst pain in the last week was considerably higher than average pain (“slightly improved”: 5.6 vs. 3.3; “same”: 7.0 vs. 4.6). Worst pain in the past week, in addition to average pain intensity, may be a key aspect of GPE self-classification.

Tables 4 to 6 show reference standard-based cut-offs (and areas under the corresponding ROC curves) for: MIC values for pain and disability (Table 4), upper limits of pain and disability still compatible with self-reported recovery (Table 5), and minimally important percent changes for pain and disability from baseline (Table 6). Absolute values for MIC in pain or disability scores were expected to vary according to baseline scores; therefore we present separate results for patient subgroups with baseline scores either above (Table 4A) or below (Table 4B) the median. Table 7 shows confidence intervals estimated by bootstrapping to the results of Table 6.

Each table presents five rows of data for five different reference standards. For easy comparison, all reference standards are listed in a single legend in Tables 4 to 7. As de Vet et al. suggested², with Reference Standard 1 patients were counted as recovered, if they were “fully recovered”, “much improved” or “slightly improved”, whereas with Reference Standard 2 “slightly improved” patients were counted as non-recovered. Reference Standards 3, 4 and 5 add conditions to the patients self-classified as “slightly improved” or “same”. These patients were counted as recovered if they had pain of less than 3 out of 10 (NRS; Reference Standard 3), disability of less than 4 out of 24 on RM scale (Reference Standard 4) or fulfilled both conditions (Reference Standard 5). These cut-offs were taken from the assessment of the upper limits of these values for compatibility with self-reported recovery according to Reference Standard 1.

Using Reference Standards 3, 4 or 5 with *combined* criteria, 70 (13%), 82 (16%) or 67 (13%) patients, respectively, would be classified as having chronic LBP. In our sample of patients with *acute* LBP, perceived recovery required percent changes from baseline pain and disability to be well above 50%. As expected, absolute values for MICs were dramatically higher for patients with higher baseline scores than for those with lower baseline values.

In addition to average pain in the past week, we assessed bothersomeness of pain, a parameter used in numerous previous LBP studies^{5, 20, 30, 36–39}. All of our analyses showed virtually identical results for both pain measures (data not presented). Regarding the parameter's ability to discriminate between recovery and non-recovery, bothersomeness of pain in the past week was not superior to average pain in the past week (*p*-values for comparing AUCs were between 0.12 and 0.76). As expected, integrating pain or disability or both into the classification criteria for recovery or non-recovery improves the discriminative ability. Among the combination criteria, the discriminatory accuracy appears to be strongest with the inclusion of either pain into the GPE scale, or both pain and disability conjoined.

Discussion

De Vet et al. presented methods for establishing MIC values on multi-item questionnaires for studies of LBP and used a hypothetical sample of 500 patients for correlating the hypothetical responses on the GPE scale as reference standard with a hypothetical multi-item scale.² Their theoretical model described a situation identical to the one we explored in our study. The results of the current study put flesh on that theoretical skeleton by providing data for 521 patients.

For the reference scale, we used an identical Reference Standard 1: patients self-reporting as at least “slightly improved” were classified as “importantly improved”. The hypothetical questionnaire for physical functioning consisted of a continuous scale scoring from 0–50. In our study we used the RM Disability Questionnaire, a validated scale from 0–24. If we were to translate the resulting hypothetical MIC value on De Vet's 51-point scale into an MIC on the RM scale, we could expect a change score of 5.0 (95% CI: 2.7–6.8) as MIC value. Using identical methods, our acute LBP sample showed a higher proportion of “importantly improved” patients (91% versus 80%) with an MIC of 11.6 (95% CI: 8.5–14.7) on the RM scale (range 0–24).

Similar to Beurskens et al.²¹, the current study provides and compares MIC values for multiple hypothetical “gold standards” with Reference Standards 1 and 2 being identical to those used by Beurskens. However, the results are quite different from prior studies, as the population samples differ considerably with respect to symptom duration (limited to 4 weeks in our study), which is a well-known key factor for the prognosis of LBP.⁴⁰ At the 6-month follow-up in the current study, 32% of patients reported to be “fully recovered”. This is different from the 8% reported by Kamper et al.²⁷ and may reflect the differences in the participants' duration of LBP or in months of follow-up. Kamper et al. presented data for a sub-sample of 239 patients with acute LBP; however, this group was only followed for 3 weeks, and only “fully recovered” patients were analyzed. Beurskens et al.²¹ excluded patients with LBP of less than 6 weeks; Demoulin et al.²¹ of less than 3 months. In the samples examined by Hill et al.⁹ and Dunn et al.¹⁹, 75 and 83% of participants had LBP for more than 4 weeks (up to 3 years). In the study by Fritz et al.¹³ 24% reported symptoms of more than 3 months, and 68% had a history of prior LBP with an unknown time interval to the current episode. Finally, the population from which we draw these data is not easily comparable to the US sample of patients with any duration of LBP, in which Von Korff et al.¹⁸ developed and validated a graded definition of chronic pain using a 6-month recall time

frame. Similarly in the replication study in the UK, 81% of participants suffered from LBP for at least 3 months.¹⁹

A limitation of this study is that we only interviewed patients who responded to our invitation letter. Therefore, this inception cohort is a small portion of all the patients seen for any type of LBP in that HMO setting during the time of enrolment. We do not have comprehensive information for the patients who did not respond to our invitation. We know, however, that 1) our patient sample was similar in key characteristics (age, sex, ethnicity, education, income) to the insured patients of that HMO according to membership surveys,²⁹ 2) respondents were slightly older and slightly more likely female than non-respondents, which is common for respondents in membership surveys of this HMO.²⁹

Conclusion

In the absence of a real gold standard for the definition of recovery from acute LBP or for its chronification, data from a primary care cohort of patients with acute LBP are provided to inform the discussion of (a) MIC values and (b) upper limits for pain and functional disability associated with perceived recovery at follow-up. Although we explored multiple cut points and reference standards, the previously suggested outcome of a 30% decrease in pain or RM scores^{1, 31} did not discriminate between recovered and non-recovered patients in a sample of patients strictly limited to acute LBP of up to four weeks.

For studies of acute LBP that require a bivariate outcome criterion for recovery versus non-recovery, we presented “optimal cut-offs” for standard measures of pain and disability and assessed their discriminatory capability. Our data suggest a cut-off of <3 for pain and <4 for the RM scores as upper limits of recovery at follow-up. Our data also suggest values for MICs and minimally important percent changes compatible with perceived recovery from acute LBP. If we were using minimally important *percent* changes as outcome measures, which are less vulnerable to baseline differences, these appear to provide good discriminatory accuracy at change scores generally above 50%.

However, as qualitative studies have previously suggested^{3, 41}, our data confirmed that single parameters such as pain or disability do not easily translate into perceived recovery. We found large AUC values in ROC curves when we used reference standards with combined outcome criteria (GPE scale with the addition of pain or disability scales for patients that self-classify as neither much improved nor worse). Combined outcomes showed improved discriminatory ability between recovered and chronic pain patients and may be considered as alternative to single parameter outcomes. Our results suggest that for studies with acute LBP patients, a combination of the GPE with pain scores may be used for the middle group of patients that self-classify as neither much improved nor worse.

There is a need to define recovery and non-recovery from acute low back pain in primary care. In 521 patients with acute low back pain followed over 6 months, perceived recovery and pain and disability scores are used to establish minimal important change scores using the receiver operating characteristics method.

References

1. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine*. Jan 1; 2008 33(1):90–94. [PubMed: 18165753]
2. de Vet HC, Terluin B, Knol DL, et al. Three ways to quantify uncertainty in individually applied “minimally important change” values. *J Clin Epidemiol*. Jun 18.2009

3. Hush JM, Refshauge K, Sullivan G, De Souza L, Maher CG, McAuley JH. Recovery: what does this mean to patients with low back pain? *Arthritis Rheum.* Jan 15; 2009 61(1):124–131. [PubMed: 19116958]
4. Deyo R. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil.* 1988; 69:1044–1053. [PubMed: 2975164]
5. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine.* Sep 15; 1998 23(18):2003–2013. [PubMed: 9779535]
6. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain.* Feb; 2008 9(2):105–121. [PubMed: 18055266]
7. Burton AK, Tillotson KM, Main CJ, Hollis S. Psychosocial predictors of outcome in acute and subchronic low back trouble. *Spine.* Mar 15; 1995 20(6):722–728. [PubMed: 7604349]
8. Grotle M, Brox JI, Veierod MB, Glomsrod B, Lonn JH, Vollestad NK. Clinical course and prognostic factors in acute low back pain: patients consulting primary care for the first time. *Spine.* Apr 15; 2005 30(8):976–982. [PubMed: 15834343]
9. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum.* May 15; 2008 59(5):632–641. [PubMed: 18438893]
10. Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. *Pain.* Oct; 2005 117(3):304–313. [PubMed: 16153772]
11. Kent PM, Keating JL. Can we predict poor recovery from recent-onset nonspecific low back pain? A systematic review. *Man Ther.* Feb; 2008 13(1):12–28. [PubMed: 17658288]
12. Hilfiker R, Bachmann LM, Heitz CA, Lorenz T, Joronen H, Klipstein A. Value of predictive instruments to determine persisting restriction of function in patients with subacute non-specific low back pain. Systematic review. *Eur Spine J.* Aug 15.2007
13. Fritz JM, Hebert J, Koppenhaver S, Parent E. Beyond minimally important change: defining a successful outcome of physical therapy for patients with low back pain. *Spine (Phila Pa 1976).* Dec 1; 2009 34(25):2803–2809. [PubMed: 19910868]
14. Croft PR, Macfarlane GJ, Papageorgiou AC, Thomas E, Silman AJ. Outcome of low back pain in general practice: a prospective study. *Bmj.* May 2; 1998 316(7141):1356–1359. [PubMed: 9563990]
15. Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson MI, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. *Bmj.* Jun 19; 1999 318(7199):1662–1667. [PubMed: 10373170]
16. Coste J, Lefrancois G, Guillemin F, Pouchot J. Prognosis and quality of life in patients with acute low back pain: insights from a comprehensive inception cohort study. *Arthritis Rheum.* Apr 15; 2004 51(2):168–176. [PubMed: 15077256]
17. Jones GT, Johnson RE, Wiles NJ, et al. Predicting persistent disabling low back pain in general practice: a prospective cohort study. *Br J Gen Pract.* May; 2006 56(526):334–341. [PubMed: 16638248]
18. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain.* Aug; 1992 50(2):133–149. [PubMed: 1408309]
19. Dunn KM, Croft PR, Main CJ, Von Korff M. A prognostic approach to defining chronic pain: replication in a UK primary care low back pain population. *Pain.* Mar; 2008 135(1–2):48–54. [PubMed: 17570585]
20. Cherkin DC, Deyo RA, Street JH, Barlow W. Predicting poor outcomes for back pain seen in primary care using patients' own criteria. *Spine.* Dec 15; 1996 21(24):2900–2907. [PubMed: 9112715]
21. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain.* Apr; 1996 65(1):71–76. [PubMed: 8826492]
22. Smeets RJ, Vlaeyen JW, Hidding A, et al. Active rehabilitation for chronic low back pain: Cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial [ISRCTN22714229]. *BMC Musculoskelet Disord.* Jan 20.2006 7(1):5. [PubMed: 16426449]

23. Ferreira ML, Ferreira PH, Latimer J, et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. *Pain*. Jan 22.2007
24. Heneweer H, Aufdemkampe G, van Tulder MW, Kiers H, Stappaerts KH, Vanhees L. Psychosocial variables in patients with (sub)acute low back pain: an inception cohort in primary care physical therapy in The Netherlands. *Spine*. Mar 1; 2007 32(5):586–592. [PubMed: 17334295]
25. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. Dec; 1989 10(4):407–415. [PubMed: 2691207]
26. Demoulin C, Ostelo R, Knottnerus JA, Smeets RJ. What factors influence the measurement properties of the Roland-Morris disability questionnaire? *Eur J Pain*. May 12.2009
27. Kamper SJ, Maher CG, Herbert RD, Hancock MJ, Hush JM, Smeets RJ. How little pain and disability do patients with low back pain have to experience to feel that they have recovered? *Eur Spine J*. Mar 13.2010
28. Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *Bmj*. Aug 9.2003 327(7410):323. [PubMed: 12907487]
29. Gordon, NP. How Does the Adult Kaiser Permanente Membership in Northern California Compare with the Larger Community?. 2006. Unpublished Work, summary at http://www.dor.kaiser.org/external/uploadedFiles/content/research/mhs/Other_Reports/mhs_project_trends-1993-2005_notes.pdf
30. Dunn KM, Croft PR. Classification of low back pain in primary care: using “bothersomeness” to identify the most severe cases. *Spine*. Aug 15; 2005 30(16):1887–1892. [PubMed: 16103861]
31. 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*. Jan; 2006 59(1): 45–52. [PubMed: 16360560]
32. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL. Performance of screening and diagnostic tests. Application of receiver operating characteristic analysis. *Arch Gen Psychiatry*. Jun; 1987 44(6):550–555. [PubMed: 3579501]
33. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. Apr; 1993 39(4):561–577. [PubMed: 8472349]
34. Stata11. *Statistics/Data Analysis*. Stata Corporation; College Station, TX: 2010.
35. Froud, R. *Statistical Software Components S457052*. Boston College Department of Economics; 2009. ROCMIC: Stata module to estimate minimally important change (MIC) thresholds for continuous clinical outcome measures using ROC curves. www.robertfroud.info/software.html
36. Hill JC, Dunn KM, Main CJ, Hay EM. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain*. Jan; 2009 14(1):83–89. [PubMed: 19223271]
37. Sherman KJ, Cherkin DC, Erro J, Miglioretti DL, Deyo RA. Comparing yoga, exercise, and a self-care book for chronic low back pain: a randomized, controlled trial. *Ann Intern Med*. Dec 20; 2005 143(12):849–856. [PubMed: 16365466]
38. Eisenberg DM, Post DE, Davis RB, et al. Addition of choice of complementary therapies to usual care for acute low back pain: a randomized controlled trial. *Spine*. Jan 15; 2007 32(2):151–158. [PubMed: 17224808]
39. Jarvik JG, Hollingworth W, Martin B, et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *Jama*. Jun 4; 2003 289(21):2810–2818. [PubMed: 12783911]
40. Dunn KM, Croft PR. The importance of symptom duration in determining prognosis. *Pain*. Mar; 2006 121(1–2):126–132. [PubMed: 16472916]
41. Turk DC, Dworkin RH, Revicki D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. *Pain*. Jul 15; 2008 137(2):276–285. [PubMed: 17937976]

Table 1

Mean Values and Standard Deviation (SD) for Pain and Disability at Follow-up as well as percent change from baseline by Level of Global Perceived Effect

	N (%)	average pain		worst pain		RM	
		mean (SD)	median	mean (SD)	median	mean (SD)	median
total at baseline	521 (100)	5.4 (1.8)		8.5 (1.5)		16 (5)	
total at follow-up		1.2 (2.0)		2.1 (3.0)		4 (5)	
completely recovered	169 (32)	0 (0)	0	0 (0)	0	0 (1)	0
much improved	254 (49)	1.0 (1.6)	0	2.3 (2.7)	1.5	4 (4)	2
slightly improved	51 (10)	3.0 (3.0)	3	5.1 (2.5)	5	8 (5)	7
same	31 (6)	4.0 (2.6)	5	6.1 (3.0)	7	9 (6)	8
slightly worse	9 (2)	4.0 (1.2)	4	6.1 (1.4)	6	10 (4)	11
much worse	7 (1)	6.6 (1.8)	7	8.6 (1.6)	9	15 (7)	16

SD: Standard Deviation

Average pain: mean and median values for average pain in past week [range 0–10]

Worst pain: mean and median values for worst pain in past week [range 0–10]

RM: Roland Morris functional disability score [range 0–24]

Table 2

Percent Change from Baseline (95% Confidence Intervals (CI)) by Level of Global Perceived Effect (GPE)

	N (%)	average pain –% (CI)	worst pain –% (CI)	RM –% (CI)
total	521 (100)	77% (73.3 – 80.8)	74% (70.3 – 76.7)	76% (72.4 – 78.9)
completely recovered	169 (32)	100%	100	97% (95.7 – 97.9)
much improved	254 (49)	81% (76.2 – 85.3)	73% (69.2 – 77.0)	78% (74.3 – 80.8)
slightly improved	51 (10)	41% (27.1 – 55.3)	37% (26.7 – 48.3)	41% (24.6 – 56.8)
same	31 (6)	27% (8.8 – 45.8)	29% (15.7 – 42.2)	41% (30.5 – 51.2)
slightly worse	9 (2)	13% (–23.5 – 50.3)	18% (4.3 – 31.5)	0% (–81.0 – 80.8)
much worse	7 (1)	–54% (–137 – 30.2)	–20% (–74.0 – 34.9)	2% (–36.4 – 40.2)

–%: mean percentage of improvement/change from baseline [range –100 to +100]; negative value indicates worsening of the parameter.

Table 3

Proportion of patients for each level of GPE which have improved in a parameter by 50 / 30%

	N	average	worst	RM
total	519	.82 / .87	.77 / .85	.84 / .90
completely recovered	169	1	1	1
much improved	253	.88 / .92	.80 / .88	.89 / .94
slightly improved	50	.46 / .66	.37 / .61	.47 / .75
same	31	.29 / .39	.23 / .45	.42 / .61
slightly worse	9	.22 / .33	.11 / .33	.33 / .56
much worse	7	0 / 0	0 / 0	.14 / .14

Legend: see Table 1

Table 4Minimal Important *Change* from Baseline to Perceived Recovery (AUC*; N = 521) for

	Average Pain Past Week	Worst Pain Past Week	RM Score
Reference Standard 1 ¹	2 (.82; .74 – .89)	5 (.88; .83 – .93)	12 (.83; .76 – .89)
Reference Standard 2 ²	3 (.79; .74 – .85)	6 (.86; .82 – .90)	11 (.82; .77 – .87)
Reference Standard 3 ³	3 (.91; .87 – .94)	5 (.91; .89 – .94)	11 (.86; .82 – .91)
Reference Standard 4 ⁴	3 (.82; .76 – .87)	5 (.89; .86 – .93)	11 (.87; .84 – .91)
Reference Standard 5 ⁵	3 (.91; .88 – .95)	5 (.92; .89 – .94)	11 (.87; .83 – .91)

* AUC = Area under the ROC curve.

¹Reference Standard 1: at least “slightly improved” = recovered;

²Reference Standard 2: at least “much improved” = recovered, “slightly improved” = non-recovered;

³Reference Standard 3: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain at follow-up ≤ 2 .

⁴Reference Standard 4: at least “much improved” = recovered, “slightly improved” and “same” = recovered if Roland Morris at follow-up ≤ 3

⁵Reference Standard 5: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain ≤ 2 and RM ≤ 3 at follow-up.

Table 4A

Same as Table 4 but only in patients with above median baseline scores

	Average Pain Past Week	Worst Pain Past Week	RM Score
Reference Standard 1 ¹	5 (.84; .75 – .92)	5 (.88; .83 – .94)	14 (.85; .76 – .94)
Reference Standard 2 ²	5 (.82; .76 – .89)	6 (.86; .81 – .91)	14 (.85; .79 – .92)
Reference Standard 3 ³	5 (.95; .92 – .97)	7 (.91; .88 – .95)	14 (.93; .89 – .97)
Reference Standard 4 ⁴	5 (.86; .80 – .92)	6 (.89; .85 – .93)	14 (.94; .91 – .97)
Reference Standard 5 ⁵	5 (.95; .93 – .98)	5 (.92; .89 – .95)	14 (.94; .90 – .97)

¹Reference Standard 1: at least “slightly improved” = recovered;

²Reference Standard 2: at least “much improved” = recovered, “slightly improved” = non-recovered;

³Reference Standard 3: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain at follow-up ≤ 2 .

⁴Reference Standard 4: at least “much improved” = recovered, “slightly improved” and “same” = recovered if Roland Morris at follow-up ≤ 3

⁵Reference Standard 5: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain ≤ 2 and RM ≤ 3 at follow-up.

Table 4B

Same as Table 4 but only in patients with below median baseline scores

	Average Pain Past Week	Worst Pain Past Week	RM Score
Reference Standard 1 ¹	2 (.88; .76 – 1.00)	2 (.89; .83 – .96)	6 (.83; .75 – .90)
Reference Standard 2 ²	2 (.83; .73 – .93)	5 (.91; .87 – .96)	11 (.84; .78 – .89)
Reference Standard 3 ³	2 (.96; .93 – .99)	4 (.94; .91 – .97)	10 (.84; .77 – .90)
Reference Standard 4 ⁴	2 (.86; .74 – .97)	5 (.92; .88 – .96)	11 (.84; .79 – .90)
Reference Standard 5 ⁵	2 (.96; .93 – .99)	4 (.94; .90 – .97)	10 (.85; .79 – .91)

¹Reference Standard 1: at least “slightly improved” = recovered;

²Reference Standard 2: at least “much improved” = recovered, “slightly improved” = non-recovered;

³Reference Standard 3: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain at follow-up ≤ 2 .

⁴Reference Standard 4: at least “much improved” = recovered, “slightly improved” and “same” = recovered if Roland Morris at follow-up ≤ 3

⁵Reference Standard 5: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain ≤ 2 and RM ≤ 3 at follow-up.

Table 5

Upper Limits for Pain and Disability Compatible with Perceived Recovery (AUC*) for

	Average Pain Past Week	Worst Pain Past Week	RM Score
Reference Standard 1 ¹	2 (.86; .80 – .92)	5 (.87; .81 – .93)	3 (.87; .83 – .91)
Reference Standard 2 ²	2 (.84; .80 – .89)	3 (.87; .83 – .91)	4 (.86; .82 – .90)
Reference Standard 3 ³	3 (.97; .95 – .98)	4 (.93; .91 – .95)	4 (.91; .88 – .94)
Reference Standard 4 ⁴	3 (.88; .83 – .92)	4 (.90; .86 – .93)	4 (.92; .90 – .94)
Reference Standard 5 ⁵	3 (.96; .95 – .98)	4 (.93; .91 – .95)	4 (.92; .90 – .94)

* AUC = Area under the ROC curve.

¹Reference Standard 1: at least “slightly improved” = recovered;

²Reference Standard 2: at least “much improved” = recovered, “slightly improved” = non-recovered;

³Reference Standard 3: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain at follow-up ≤ 2 .

⁴Reference Standard 4: at least “much improved” = recovered, “slightly improved” and “same” = recovered if Roland Morris at follow-up ≤ 3

⁵Reference Standard 5: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain ≤ 2 and RM ≤ 3 at follow-up.

Table 6

Minimum Percent Change from Baseline to Perceived Recovery (AUC*) for

	Average Pain Past Week	Worst Pain Past Week	RM score
Reference Standard 1 ¹	58% (.86; .79 – .92)	51% (.87; .82 – .93)	78% (.88; .85 – .92)
Reference Standard 2 ²	58% (.84; .89 – .89)	58% (.87; .83 – .91)	78% (.88; .84 – .92)
Reference Standard 3 ³	64% (.96; .94 – .98)	58% (.93; .91 – .95)	68% (.92; .90 – .95)
Reference Standard 4 ⁴	58% (.87; .82 – .92)	58% (.90; .86 – .93)	78% (.93; .91 – .96)
Reference Standard 5 ⁵	58% (.96; .94 – .97)	45% (.93; .91 – .95)	68% (.93; .91 – .95)

* AUC = Area under the ROC curve.

¹Reference Standard 1: at least “slightly improved” = recovered;

²Reference Standard 2: at least “much improved” = recovered, “slightly improved” = non-recovered;

³Reference Standard 3: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain at follow-up ≤ 2 .

⁴Reference Standard 4: at least “much improved” = recovered, “slightly improved” and “same” = recovered if Roland Morris at follow-up ≤ 3

⁵Reference Standard 5: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain ≤ 2 and RM ≤ 3 at follow-up.

Table 7

Minimum Percent Change from Baseline to Perceived Recovery (Confidence Intervals by bootstrapping) for

	Average Pain Past Week	Worst Pain Past Week	RM score
Reference Standard 1 ¹	58% (33–83%)	51% (35–67%)	78% (60–96%)
Reference Standard 2 ²	58% (38–78%)	58% (46–70%)	78% (66–91%)
Reference Standard 3 ³	64% (54–73%)	58% (40–76%)	68% (54–81%)
Reference Standard 4 ⁴	58% (44–72%)	58% (48–69%)	78% (67–89%)
Reference Standard 5 ⁵	58% (49–67%)	45% (30–61%)	68% (54–81%)

¹Reference Standard 1: at least “slightly improved” = recovered;

²Reference Standard 2: at least “much improved” = recovered, “slightly improved” = non-recovered;

³Reference Standard 3: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain at follow-up ≤ 2 .

⁴Reference Standard 4: at least “much improved” = recovered, “slightly improved” and “same” = recovered if Roland Morris at follow-up ≤ 3

⁵Reference Standard 5: at least “much improved” = recovered, “slightly improved” and “same” = recovered if average pain ≤ 2 and RM ≤ 3 at follow-up.