

SPECIAL ARTICLE

Report of the NIH Task Force on Research Standards for Chronic Low Back Pain

Richard A. Deyo, MD, MPH,* Samuel F. Dworkin, DDS, PhD,[†] Dagmar Amtmann, PhD,[†] Gunnar Andersson, MD, PhD,[‡] David Borenstein, MD,[§] Eugene Carragee, MD,[¶] John Carrino, MD, MPH,** Roger Chou, MD,* Karon Cook, PhD,^{††} Anthony DeLitto, PT, PhD,^{‡‡} Christine Goertz, DC, PhD,^{§§} Partap Khalsa, DC, PhD,^{¶¶} John Loeser, MD,[†] Sean Mackey, MD, PhD,[¶] James Panagis, MD, MPH,^{***} James Rainville, MD,^{†††} Tor Tosteson, ScD,^{‡‡‡} Dennis Turk, PhD,[†] Michael Von Korff, ScD,^{§§§} and Debra K. Weiner, MD^{††}

*Oregon Health and Sciences University, Portland, Oregon; [†]University of Washington, ^{§§§}Group Health Research Institute, Seattle, Washington; [‡]Rush University Medical Center, Chicago, ^{¶¶}Northwestern University, Evanston, Illinois; [§]George Washington University, Washington, District of Columbia; ^{¶¶}Stanford University, Stanford, California; ^{**}Johns Hopkins University, Baltimore, ^{¶¶}National Center for Complementary and Alternative Medicine, ^{***}National Institute for Arthritis, Musculoskeletal and Skin Diseases, Bethesda, Maryland; ^{‡‡‡}VA Pittsburgh Healthcare System and University of Pittsburgh, Pittsburgh, Pennsylvania; ^{§§}Palmer College of Chiropractic, Davenport, Iowa; ^{†††}New England Baptist Hospital, Roxbury Crossing, Massachusetts; ^{‡‡}Dartmouth University, Hanover, New Hampshire, USA

Reprint requests to: Richard A. Deyo, MD, MPH, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Mail Code FM, Portland, OR 97239, USA. Tel: 503-494-1694; Fax: 503-494-2746; E-mail: deyo@ohsu.edu.

This study was supported by the National Center for Complementary and Alternative Medicine and the National Institute for Arthritis, Musculoskeletal, and Skin Diseases.

Acknowledgements: Published with permission from Elsevier, Inc. on the behalf of the American Pain Society. Original article first published in The Journal of Pain. 2014; 15(6): 569–585. Copyright is retained by the American Pain Society.

Abstract

Objective. Despite rapidly increasing intervention, functional disability due to chronic low back pain (cLBP) has increased in recent decades. We often cannot identify mechanisms to explain the major negative impact cLBP has on patients' lives. Such cLBP is often termed non-specific, and may be due to multiple biologic and behavioral etiologies. Researchers use varied inclusion criteria, definitions, baseline assessments, and outcome measures, which impede comparisons and consensus.

Design. Expert panel and preliminary evaluation of key recommendations.

Methods. The NIH Pain Consortium charged a Research Task Force (RTF) to draft standards for research on cLBP. The resulting multidisciplinary panel developed a 3-stage process, each with a 2-day meeting.

Results. The panel recommended using 2 questions to define cLBP; classifying cLBP by its impact (defined by pain intensity, pain interference, and physical function); use of a minimal data set to describe research subjects (drawing heavily on the PROMIS methodology); reporting “responder analyses” in addition to mean outcome scores; and suggestions for future research and dissemination. The Pain Consortium has approved the recommendations, which investigators should incorporate into NIH grant proposals.

Conclusion. The RTF believes these recommendations will advance the field, help to resolve controversies, and facilitate future research addressing the genomic, neurologic, and other mechanistic substrates of chronic low back pain. Greater consistency in reporting should facilitate comparisons among studies and the development of phenotypes. We expect the RTF recommendations will become a dynamic document, and undergo continual improvement.

Perspective. A task force was convened by the NIH Pain Consortium with the goal of developing research standards for chronic low back pain. The results included recommendations for definitions, a

minimum dataset, reporting outcomes, and future research. Greater consistency in reporting should facilitate comparisons among studies and the development of phenotypes.

Key Words. Low Back Pain; Chronic Low Back Pain; Research Standards; Minimum Dataset; NIH Task Force

Introduction

The Institute of Medicine recently estimated that chronic pain affects approximately 100 million adults in the United States, with an estimated annual cost of \$635 billion, including direct medical expenditures and loss of work productivity [1]. Activity-limiting low back pain (LBP), in particular, has a worldwide lifetime prevalence of approximately 39% and a similar annual prevalence of 38% [2]. The majority of people who have LBP experience recurrent episodes [3]. The use of all interventions for treating chronic LBP (cLBP) increased from 1995 to 2010, including surgical, pharmacologic, and nonpharmacologic approaches. Despite increased utilization, however, the prevalence of symptoms and expenditures has increased [4–6].

There is growing evidence that cLBP, like other chronic pain conditions, can progress beyond a symptomatic state to a complex condition [7] involving persistent anatomic and functional changes in the central nervous system [8–10], in addition to structural changes in the back (e.g., degenerative spinal changes, atrophy, or asymmetry of paraspinal muscles) [11–13]. Although some patients with cLBP may have clear pathoanatomic causes of pain, for many there is no clear association between pain and identifiable pathology of the spine or its associated soft tissues [14].

Many patients who undergo procedures intended to correct the putative causative pathoanatomy continue to have pain. Furthermore, we often cannot identify mechanisms to account for the substantial negative impact cLBP has on the lives of many patients [15]. Such cLBP is often termed nonspecific, idiopathic, mechanical, or due to instability, and may in fact be due to the contributions of different and multiple biologic and behavioral etiologies in different individuals [16].

Many classes of interventions have been developed and tested in adults with cLBP, including spine surgery, injections, medications, psychological interventions, manual therapies, exercise, nutritional supplements, and lifestyle change and self-management approaches [15,17–19]. Many of these have shown some clinical benefit, but few appear to consistently provide substantial, long-term reductions in pain with increased function [20–23].

A critical issue for advancing research on cLBP is comparing results from the many classes of interventions. In 2009 and 2010, the National Institutes of Health (NIH) Pain Consortium convened 2 workshops on LBP

research, inviting experts from the relevant scientific and clinical fields to provide research recommendations to NIH. These experts noted that clinical studies have used variable inclusion and exclusion criteria, case definitions for LBP chronicity or recurrence, baseline assessments, stratification criteria, and outcome measures. As a result, it is difficult to compare epidemiologic data and studies of similar or competing interventions, replicate findings, pool data from multiple studies, resolve conflicting conclusions, develop multidisciplinary consensus, or even achieve consensus within a discipline regarding interpretation of findings. Key recommendations from the workshops on how to advance cLBP research were to establish research standards for cLBP and to have the NIH facilitate this process.

In response, the NIH Pain Consortium established a Steering Committee for a Research Task Force (RTF) on Research Standards for cLBP. The Steering Committee was composed of representatives from the following NIH institutes/centers: National Center for Complementary and Alternative Medicine (NCCAM), National Institute on Aging, National Institute of Arthritis, Musculoskeletal, and Skin Diseases (NIAMS), National Institute of Child Health and Human Development, National Institute on Drug Abuse, National Institute of Dental and Craniofacial Research, National Institute of Neurological Disorders and Stroke (NINDS), and National Institute of Nursing Research. The Steering Committee developed goals for the RTF, identified the needed scientific and clinical expertise, selected 2 co-chairs, and invited 14 additional experts from outside NIH to join the RTF. The Steering Committee provided 2 representatives (Drs. Panagis and Khalsa) in ex-officio (i.e., nonvoting) capacity to the RTF.

The NIH Pain Consortium charged the RTF with developing a set of standards for clinical research on cLBP that would address the following:

- Consider the state of existing research relevant to the development of standards.
- Conduct a comprehensive review of existing case definitions, diagnostic criteria, and outcome measures that are relevant
- Develop a draft set of standards
- Engage the broader research community and representatives from relevant government agencies in developing these standards
- Chart a general plan for their incorporation into research studies and their future modification

This charge focused solely on developing standards for research and not for use in coding, billing, or general use in clinical settings.

Methods

Creating the RTF

The Steering Committee selected 2 co-chairs with complementary leadership expertise. Dr. Deyo was chosen for his

Table 1 Task force members, affiliations and expertise

Members	Affiliations	Expertise
Co-Chairs		
Richard A. Deyo, MD, MPH	Departments of Family Medicine, Internal Medicine, and Public Health, Oregon Health and Science University and Clinical Investigator, Kaiser Center for Health Research	Primary care, health services research
Samuel F. Dworkin, DDS, PhD	Departments of Oral Medicine and Psychiatry and Behavioral Sciences, University of Washington	Temporomandibular joint disorders, chronic pain, clinical psychology
Task Force Members		
Gunnar Andersson, MD, PhD	Department of Orthopedic Surgery, Rush University Medical Center	Orthopedic spine surgery
David Borenstein, MD	Department of Rheumatology, George Washington University	Rheumatology
Eugene Carragee, MD	Department of Orthopaedic Surgery, Stanford University School of Medicine	Orthopedic spine surgery
John Carrino, MD, MPH	Department of Radiology, Johns Hopkins University School of Medicine	Musculoskeletal radiology
Roger Chou, MD	Departments of Medicine and of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University	General internal medicine, Systematic review
Anthony DeLitto, PT, PhD	Department of Physical Therapy, University of Pittsburgh School of Health Rehabilitation	Physical therapy
Christine Goertz, DC, PhD	Palmer College of Chiropractic	Chiropractic care, epidemiology
John Loeser, MD	Department of Neurological Surgery, University of Washington	Neurosurgery, pain management
Sean Mackey, MD, PhD	Department of Anesthesia, Stanford University School of Medicine	Pain management, functional brain imaging
James Rainville, MD	Physical Medicine and Rehabilitation, New England Baptist Hospital and Tufts University	Spine rehabilitation
Tor Tosteson, ScD	Department of Community and Family Medicine, Dartmouth University	Biostatistics
Dennis Turk, PhD	Department of Anesthesiology and Pain Medicine, University of Washington	Pain medicine, psychology
Michael Von Korff, ScD	Group Health Research Institute, Seattle	Epidemiology
Debra K. Weiner, MD	Geriatric Research, Education and Clinical Center, VA Pittsburgh Healthcare System and the Departments of Medicine, Psychiatry, and Anesthesiology; the Clinical and Translational Science Institute, University of Pittsburgh	Geriatric medicine, rheumatology

expertise in LBP research and Dr. Dworkin for his prior leadership in developing research diagnostic criteria for temporomandibular disorders (TMD), another set of chronic pain conditions. The co-chairs, in consultation with the Steering Committee, selected the RTF members for their needed scientific and clinical expertise (Table 1).

Work Plan

The RTF Evolved a 3-stage Work Plan, Each Involving a 2-day Meeting

Stage 1. The first meeting opened with remarks by the NIAMS and NCCAM directors, Stephen Katz, MD, PhD, and Josephine Briggs, MD, respectively. The directors emphasized the nature of chronic back pain as a highly

prevalent and costly public health challenge. They noted the existence of many stakeholders, including individuals with back pain, health care systems, clinicians, drug and device makers, regulatory agencies, and federal, state, and third-party payers. They emphasized the research—as opposed to clinical or administrative—focus of the task force.

Initial efforts of the RTF were directed at defining subsequent activities and products. At the initial and subsequent meetings, a consensus evolved on several important issues and strategies (Table 2).

The RTF noted that although the intended users of the proposed research standards would be investigators submitting grant applications to NIH, the standards would

Table 2 Key principles developed by the task force on research standards for chronic low back pain

1. The process should use an evidence-based approach that incorporates a biopsychosocial model of chronic pain.
2. Data should be useful for a wide range of conditions, including patients thought to have degenerative spinal disorders (e.g., herniated disc or lumbar stenosis) as well as those without identified pathoanatomy.
3. Patients with underlying systemic or highly specific diseases were not the target of the task force. Such conditions include cancer, spinal infections, fractures, and inflammatory spondylopathies, such as ankylosing spondylitis.
4. Patients with no identified pathoanatomy should not be assumed to have “psychological,” “psychosomatic,” “psychogenic,” or “somatoform” pain.
5. Given the current state of knowledge, stratifying chronic back pain by its impact is more feasible and potentially useful than attempting classification solely by pathoanatomy or pathophysiology. Impact will tentatively be defined in terms of pain intensity, interference with activities, and physical function.
6. A minimal uniform dataset should be reported in all studies of chronic back pain. This should be brief, so that investigators can supplement it with key measures for specific research questions.
7. The dataset should be relevant for population, observational, and interventional research.
8. The dataset should include both biomedical and psychosocial variables.
9. An investigator could substitute more detailed, precise, and well-validated measures for a particular domain, but should report data for each domain of the minimum dataset.
10. Additional “core” items would be recommended for specific study aims or populations, such as surgical trials or elderly populations.
11. A prognostic dimension for the classification of chronic low back pain would be desirable, but more evidence is needed before an explicit recommendation will be made.
12. Research standards should evolve, and the RTF will suggest a potential research agenda for refining the research standards.

be available to and encouraged for all researchers. The research standards could potentially allow cLBP phenotypes to be uncovered based on physical and psychosocial findings.

The RTF decided that it could not respond in detail to every component of the NIH Pain Consortium’s charge. For example, producing explicit evidence-based diagnostic criteria for conditions such as spinal stenosis, sciatica, or spine “instability” would be impossible given the available time and resources and the current lack of professional consensus. However, stratifying cLBP by its impact might have equally important descriptive and prognostic value and could supplement any pathophysiologic description.

Stage 2. The co-chairs conducted a series of e-mail surveys of RTF members. The surveys were based on item lists generated at the first RTF meeting and addressed key issues from the meeting. The following surveys and literature review efforts were conducted:

1. *Survey of candidate objective findings and medical history for a minimum dataset:* Members ranked the importance of potential baseline descriptors for patients with cLBP. These included items of medical history, comorbidity, physical examination, and laboratory and imaging tests.

2. *Survey of candidate self-report measures of behavior, mood, and symptoms:* Task force members were asked to rank the importance of measures of pain-related behavioral, emotional, and psychosocial domains influencing the expression of cLBP.

3. *Survey on feasibility of developing research diagnostic criteria for subsets of nonspecific chronic low back pain:* Part of the charge from the Pain Consortium was to consider developing a research diagnostic classification system based on pathophysiologic or etiologic features (i.e., criteria for subsets of nonspecific cLBP). This survey asked task force members to assess the feasibility of such an effort.

4. *Review of existing literature on back pain classification and prognosis:* The task force did not undertake a systematic literature review but considered previous work on back pain taxonomy [24–39], prognostic classification [19,40–71], pain and psychosocial measures [41,69,72–92], and outcome assessment [93–106]. These sources informed the deliberations and recommendations.

At the second RTF meeting, the most highly ranked candidate items for the minimum dataset based on survey responses were accepted with minimal disagreement or need for further discussion. Special attention was directed to the possible use of the Patient Reported Outcomes Measurement Information System (PROMIS) measures [93,95,102,103,105,106]. Progress was made toward defining cLBP and its impact. There was general agreement that developing pathophysiologic diagnostic criteria for subsets of nonspecific low back pain was unfeasible at present.

The RTF also heard presentations of 2 related NIH efforts. The first was the NINDS effort to create “common data elements” for use by all NINDS-supported researchers. The second related to the NIH PROMIS effort, which

Table 3 Task force recommendations: Research standards for Chronic Low Back Pain (cLBP)

1. Defining the chronicity of cLBP: The RTF recommended 2 questions to achieve the definition of chronic (Questions 1 and 2 in Figure 1): 1) How long has back pain has been an ongoing problem for you? 2) How often has low back pain been an ongoing problem for you over the past 6 months? A response of greater than 3 months to question 1 and a response of "at least half the days in the past 6 months" to question 2 would define cLBP.
2. Stratify cLBP by impact: "Impact" was defined by pain intensity, pain interference with normal activities, and functional status. These items have major prognostic and discriminatory importance. Impact is calculated from 9 items of the 29-item PROMIS short form (marked with asterisks in Figure 1).
3. Report a Minimum dataset: A minimum dataset is recommended for describing subjects in all research on cLBP (Figure 1). Medical history included demographics, involvement in workers' compensation, work status, education, comorbidity, and previous treatment. Physical examination items were reserved for studies of invasive interventions or of older adults. MRI was recommended for studies of surgical interventions. Key self-report domains were pain intensity, pain interference, physical function, depression, sleep disturbance, and catastrophizing. The short form PROMIS measures were recommended as offering the best tradeoff of length with psychometric validity.
4. Outcome Measures: Many parts of the minimum dataset, such as PROMIS measures, are also appropriate as outcome measures. However, primary outcomes of clinical studies will vary, depending on study aims, and investigators are referred to earlier consensus documents on outcome measures. The RTF recommended reporting a "responder" analysis in addition to mean scores of outcome measures.
5. Research on the proposed standards: The RTF recommended new research to improve prognostic stratification of patients with chronic low back pain; refine and test composite outcome measures for increasing the clinical importance of study results; undertake patient stakeholder assessment of relevant outcomes; and further evaluate psychometric properties of the minimum dataset.
6. Dissemination: With adoption of recommendations by the NIH Pain Consortium, the RTF recommends dissemination to the broad research community. This would include publication of a report in multiple professional journals and presentations at professional meetings.

includes several psychometrically sound patient-reported outcomes measures directly relevant to the task force.

Stage 3. At the third meeting, the RTF agreed on a series of recommendations to be forwarded to the NIH Pain Consortium. These included a definition of cLBP and specific measures to stratify its impact. It also reached agreement on recommending specific domains and items to be integrated into a minimum dataset for research on cLBP. There followed a discussion of outcome measures and future research needs regarding the task force recommendations.

The task force also suggested strategies for obtaining feedback and support for its recommendations. These included consultation with the NIH Pain Consortium and relevant NIH institutes, other government agencies, and relevant journal editors. It would also include presentations at meetings of research and professional organizations.

Task Force Recommendations

The principles articulated in Table 2 led the task force to several specific recommendations that are summarized in Table 3. The rationales for these recommendations are discussed next. The first three recommendations refer to the questionnaire instrument in Figure 1.

Recommendation 1. Describe the Chronicity of Low Back Pain

The RTF recommended that "chronic low back pain" (cLBP) be defined as a back pain problem that has per-

sisted at least 3 months and has resulted in pain on at least half the days in the past 6 months. A human figure drawing would illustrate the region defined as the low back, indicating the space between the lower posterior margin of the rib cage and the horizontal gluteal fold (Figure 1).

The RTF considered definitions based on time with pain, days with pain, severity of pain, and varying durations of pain. Minimal durations of 3 months or 6 months were considered, as was the problem of intermittent symptoms.

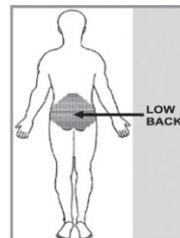
The RTF concluded that 2 questions should define chronicity (Questions 1 and 2 in Figure 1): 1) "How long has back pain has been an ongoing problem for you?" and 2) "How often has low back pain been an ongoing problem for you over the past 6 months?" A response of "greater than 3 months" to question 1 and a response of "at least half the days in the past 6 months" to question 2 would define cLBP. A patient with pain on at least half the days in the past 6 months would have accumulated at least 3 months' worth of pain days, and the Task Force concluded that this would be the recommended definition. It was decided that pain severity would not be included in the definition of cLBP.

Recommendation 2. Stratify Chronic Low Back Pain by Impact

The RTF overwhelmingly agreed that neither adequate data nor resources were available to offer a new pathoanatomic or pathophysiologic subclassification of cLBP that was clearly superior to those currently available.

Recommended Minimum Dataset

(PROMIS items marked with ¹; STarT Back or nearly identical items marked with ²; RTF Impact Classification items marked with *)



1. How long has low-back pain been an ongoing problem for you?
 - Less than 1 month
 - 1–3 months
 - 3–6 months
 - 6 months–1 year
 - 1–5 years
 - More than 5 years

2. How often has low-back pain been an ongoing problem for you over the past 6 months?
 - Every day or nearly every day in the past 6 months
 - At least half the days in the past 6 months
 - Less than half the days in the past 6 months

3. In the past 7 days, how would you rate your low-back pain on average?^{*1,2}

<input type="checkbox"/>									
1	2	3	4	5	6	7	8	9	10
No pain									Worst Imaginable pain

4. Has back pain spread down your leg(s) during the past 2 weeks?²
 - Yes
 - No
 - Not sure

5. During the past 4 weeks, how much have you been bothered by ...

	Not bothered at all	Bothered a little	Bothered a lot
• Stomach pain	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Pain in your arms, legs, or joints other than your spine or back	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Headaches	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Widespread pain or pain in most of your body	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6. Have you ever had a low-back operation?
 - Yes, one operation
 - Yes, more than one operation
 - No

7. If yes, when was your last back operation?
 - Less than 6 months ago
 - More than 6 months but less than 1 year ago
 - Between 1 and 2 years ago
 - More than 2 years ago

8. Did any of your back operations involve a spinal fusion? (also called an arthrodesis)
 - Yes
 - No
 - Not sure

PAIN INTERFERENCE

- | In the past 7 days... | Not at all | A little bit | Somewhat | Quite a bit | Very much |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 9. How much did pain interfere with your day-to-day activities? ^{*1} | <input type="checkbox"/> |
| 10. How much did pain interfere with work around the home? ^{*1} | <input type="checkbox"/> |

Figure 1 Recommended minimum dataset.

Standards for Chronic Low Back Pain

11. How much did pain interfere with your ability to participate in social activities? *1

12. How much did pain interfere with your household chores? *1

13. Have you used any of the following treatments for your back pain? (Check all that apply)

	Yes	No	Not sure
• Opioid painkillers (prescription medications such as Vicodin, Lortab, Norco, hydrocodone, codeine, Tylenol #3 or #4, fentanyl, Duragesic, MS Contin, Percocet, Tylox, OxyContin, oxycodone, methadone, tramadol, Ultram, Dilaudid)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
If you checked yes, are you currently using this medication?	<input type="checkbox"/>	<input type="checkbox"/>	
• Injections (such as epidural steroid injections, facet injections).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Exercise therapy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Psychological counseling, such as cognitive-behavioral therapy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The next two questions are for people who normally work outside the home.

14. I have been off work or unemployed for 1 month or more due to low-back pain.

Agree

Disagree

Does not apply

15. I receive or have applied for disability or workers' compensation benefits because I am unable to work due to low-back pain.

Agree

Disagree

Does not apply

PHYSICAL FUNCTION

Physical Function	Without any difficulty	With a little difficulty	With some difficulty	With much difficulty	Unable to do
16. Are you able to do chores such as vacuuming or yard work? *1	<input type="checkbox"/>				
17. Are you able to go up and down stairs at a normal pace? *1	<input type="checkbox"/>				
18. Are you able to go for a walk of at least 15 minutes? *1,2	<input type="checkbox"/>				
19. Are you able to run errands and shop? *1	<input type="checkbox"/>				

DEPRESSION

In the past 7 days...	Never	Rarely	Sometimes	Often	Always
20. I felt worthless ¹	<input type="checkbox"/>				
21. I felt helpless ¹	<input type="checkbox"/>				
22. I felt depressed ¹	<input type="checkbox"/>				
23. I felt hopeless ¹	<input type="checkbox"/>				

Figure 1 Continued

SLEEP DISTURBANCE

In the past 7 days...

	Very poor	Poor	Fair	Good	Very good
24. My sleep quality was ¹	<input type="checkbox"/>				

In the past 7 days...

	Not at all	A little bit	Somewhat	Quite a bit	Very much
25. My sleep was refreshing ¹	<input type="checkbox"/>				

26. I had a problem with my sleep ¹	<input type="checkbox"/>				
--	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

27. I had difficulty falling asleep ¹	<input type="checkbox"/>				
--	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

28. It's not really safe for a person with my back problem to be physically active.²

- Agree
- Disagree

29. I feel that *my back pain is terrible and it's never going to get any better.*²

- Agree
- Disagree

30. Are you involved in a lawsuit or legal claim related to your back problem?

- Yes
- No
- Not sure

In the past year:

	Never	Rarely	Sometimes	Often
31. Have you been drunk or used drugs more than you meant to?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

32. Have you felt you wanted or needed to cut down on your drinking or drug use?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--	--------------------------	--------------------------	--------------------------	--------------------------

33. Age: _____ years (0–120)

34. Gender:

- Female
- Male
- Unknown
- Unspecified

35. Ethnicity: (*"X" ONLY the one with which you MOST CLOSELY identify*)

- Hispanic or Latino
- Not Hispanic or Latino
- Unknown
- Not Reported

36. Race: (*"X" those with which you identify*)

- American Indian or Alaska Native
- Asian
- Black or African-American
- Native Hawaiian or Other Pacific Islander
- White
- Unknown
- Not Reported

Figure 1 Continued

37. Employment Status:

- Working now
- Looking for work, unemployed
- Sick leave or maternity leave
- Disabled due to back pain, permanently or temporarily
- Disabled for reasons other than back pain
- Student
- Temporarily laid off
- Retired
- Keeping house
- Other, Specify: _____
- Unknown

38. Education Level: (select the highest level attained)

- No high school diploma
- High school graduate or GED
- Some college, no degree
- Occupational/technical/vocational program
- Associate degree: academic program
- Bachelor's degree
- Master's degree (e.g., M.A., M.S., M.Eng., M.Ed., M.B.A.)
- Professional school degree (e.g., M.D., D.D.S., D.V.M., J.D.)
- Doctoral degree (e.g., Ph.D., Ed.D.)
- Unknown

39. How would you describe your cigarette smoking?

- Never smoked
- Current smoker
- Used to smoke, but have now quit

- 40. Height:** ____ inches centimeters measured self-reported
- Weight:** ____ pounds kilograms measured self-reported

Figure 1 Continued

Rather, the RTF recommended stratification of cLBP by the personal impact of low back pain. "Impact" was defined as a combination of pain intensity, pain interference with normal activities, and functional status, using 9 items of the 29-item PROMIS short form (marked with asterisks in Figure 1). These items have substantial research support to validate their discriminatory and prognostic importance [19,40–71,107].

This stratification of cLBP by impact would be appropriate whether or not there appears to be contributory degenerative pathoanatomy. Even when pathoanatomic conditions are thought to contribute to symptoms and dysfunction, they often coexist and overlap, and sometimes fail to respond to specific interventions. Thus, the stratification of impact seems to be a useful addition to but not a substitute for pathoanatomic, physiologic, or symptomatic classification.

After considerable discussion about formal prognostic scales for stratification, such as the Subgroups for Targeted Treatment (STarT) Back instrument [52], the RTF decided that there remained substantial uncertainty about generalizability to subspecialty patients and older adults.

Thus, the RTF recommended further research in this area and included several items from the STarT Back instrument in the minimum dataset, but chose not to require them for stratification purposes.

The recommended RTF Impact Stratification approach uses the raw PROMIS scores with the usual scoring of the Physical Function items reversed. Thus, for each item in the Impact Stratification, a score of 1 is least severe, and 5 most severe. The exception is the single item on pain intensity, which ranges from 0 (least severe) to 10 (most severe). Thus, scores on the 9 PROMIS-based items yielding Impact Stratification range from 8 (least impact) to 50 (greatest impact). Items in Figure 1 with an asterisk comprise the Impact Stratification score.

Because the proposed impact score is a novel combination of 3 constructs (pain intensity, interference, and function), the RTF undertook a preliminary assessment of its validity and performance with the assistance of PROMIS investigators. The validation used existing PROMIS data from a group of patients with LBP, with or without leg pain, who underwent epidural steroid injections. This analysis was covered by an existing institutional review board

approval from the University of Washington. Given the intervention, an improvement in average functional scores was expected.

The sample included 218 patients with a mean age of 54 years; 56% were females. There were 41% employed full or part time, 22% retired, and 12% receiving disability compensation, with the remainder being homemakers, students, or unemployed. The racial mix included 87% white, 3.8% African American, 4% American Indian, and 5% Asian or Pacific Islander. There were 46% with a college or more advanced degree, and 5% with less than a high school diploma.

The dataset included legacy measures of back pain-related physical function: the Roland and Morris Disability Questionnaire and the Oswestry Disability Index (collected at baseline only). The RTF Impact Stratification showed strong correlations with legacy measures. Furthermore, score changes on the RTF Impact Stratification correlated more strongly with patient satisfaction at follow-up than did change on the Roland-Morris score (Table 4).

In this rather severely affected sample, baseline RTF Impact scores were almost equally distributed among mild, moderate, and severe impacts. Although the cutoffs used in Table 4 for mild, moderate, and severe scores were deemed potentially useful by the RTF, they are relatively arbitrary. Simply reporting actual scores is recommended, along with any categorization that investigators may choose.

As expected, scores on the Impact Stratification measure for this sample improved over time. Measures of effect size and standardized response mean for the 170 patients available for 3-month follow-up suggested

that the RTF Impact Stratification was more responsive than the Roland-Morris Disability Questionnaire (Table 3).

The task force found the results encouraging but acknowledged that the analyses reported reflect only an initial assessment. As suggested in the recommendations below for future research, further assessment of the reliability, validity, and clinical utility of this stratification strategy is a high priority.

Recommendation 3. Report a Minimum Dataset

A minimum dataset is recommended for describing individuals participating in all research studies on cLBP (Figure 1); the minimum dataset includes items of demographics, medical history, and self-report of symptoms and function.

Medical History, Physical Examination, Diagnostic Testing

In the survey of RTF members regarding items for a minimum dataset, the most highly ranked items of medical history and examination included demographics, involvement in workers' compensation or legal claims, work status, education, various measures of comorbidity, and previous treatment history. For many of these measures, the RTF adopted the format of the Common Data Elements system implemented by the NINDS (<http://www.commondataelements.ninds.nih.gov>).

The key comorbid conditions were judged to be smoking status, obesity, substance abuse, and widespread pain symptoms. The two-item conjoint scale (TICS) was judged

Table 4 Performance of the research task force impact stratification among 218 subjects undergoing epidural steroid injections

Construct Validation	Measurement Tool	
Correlation of RTF impact stratification with legacy measures of physical function, baseline (Spearman R)	Oswestry Disability Index (.806)	Roland-Morris Disability Questionnaire (.661)
Correlation of score changes with patient satisfaction with treatment at follow-up, scored 1–4 (Spearman R, absolute value)	Change, Roland-Morris Disability Questionnaire (.148)	Change, RTF impact stratification score (.251)
Distribution of RTF Impact Stratification Scores	Baseline (N = 218), n (%)	Follow-up (N = 170), n (%)
RTF impact stratification score 8–27 (mild)	66 (30%)	107 (63%)
RTF impact stratification score 38–34 (moderate)	73 (34%)	31 (18%)
RTF impact stratification score ≥ 35 (severe)	79 (36%)	32 (19%)
Mean RTF impact stratification score (SD)	32 (8.3)	25 (9.7)
Responsiveness	Effect Size (Change/Baseline SD)	Standardized Response Mean (Change/SD of Change)
RTF impact stratification	.69	.75
Roland-Morris Disability Questionnaire	.39	.41

Note. Three-month follow-up was available for 170 of these subjects (78%).

to be an adequate and suitably brief screen for substance abuse [18]. The key items of treatment history were thought to be history of surgical interventions and use of opioid analgesics.

Measures from the physical examination ranked lower than items of medical history. However, the most highly ranked of these were straight leg raising for patients with leg pain; hip internal rotation as a screen for hip arthritis (a potential cause of LBP); and lower extremity strength. There was general agreement that such physical examination items could be reserved for studies of invasive interventions (straight leg raising and lower extremity strength) or of older adults (hip examination). Thus, for example, physical examination measures would not be required of all epidemiologic studies.

No laboratory or imaging tests were highly ranked, because of the widely recognized weak association between degenerative spine changes on imaging and patient symptoms or function [14]. However, magnetic resonance imaging (MRI) was considered the most valuable of potential tests, and there was agreement that it should be required in studies of surgical interventions.

Self-report of Functional Status, Psychosocial Factors, and Mood Disturbance

With regard to other self-report measures, there was discussion first about the domains to be included, then the potential sources of items, and then the desirable number of items. The key domains were judged to be physical function, depression, sleep disturbance, and catastrophizing. The task force felt that these constructs were important for a wide range of patients with chronic back pain, with or without specific pathoanatomic diagnoses. For parsimony, other important constructs such as anxiety, fatigue, and satisfaction with social role were considered but not included in the minimum dataset.

Although the minimum dataset in Figure 1 is recommended for inclusion in all NIH-funded research on cLBP and is available for use by all researchers, the RTF did not in any way intend to constrain the scope of investigators' proposed scientific inquiries. On the contrary, the RTF believes that the minimum dataset represents a major advance toward standardization of research reporting by asking researchers to include, at a minimum, a set of items that evidence supports as critical to scientifically advancing our understanding of cLBP.

After considering several potential instruments for assessing these domains, the RTF concluded that the short-form PROMIS measures [108] offered the best trade-off of length with psychometric validity for a minimum dataset. Therefore, it recommended use of the relevant scales from the 29-item PROMIS short-form, which includes 4 items for each domain. Investigators and patient samples with access to computer adaptive testing could use the entire PROMIS item bank to measure the domains included on

the PROMIS 29 Profile version 1.0, an acceptable or even preferable alternative [109].

There was agreement that it would be acceptable if investigators preferred well-validated, lengthier legacy measures of these domains. For example, if investigators wanted more extensive legacy measures of physical function, they might substitute the Oswestry or Roland-Morris disability scale for the PROMIS physical function items. If they wanted legacy measures of depression, they might substitute the Patient Health Questionnaire (PHQ-9) [57] or Beck Depression Inventory [72]. In Figure 1, we have labeled the PROMIS constructs to facilitate such substitution if desired, though investigators may wish to remove the labels when using the dataset. If such substitutions are made, all the other recommended domains should still be assessed.

Investigators may find it useful to consult the website PROsetta Stone, supported by NCI-funded investigators at Northwestern University (www.prosetta.org) [110]. This website provides a "cross-walk" between scores on the PROMIS measures and scores on several "legacy" measures, such as the Brief Pain Inventory [41], the Center for Epidemiologic Studies Depression Scale (CES-D) [83], the PHQ-9 [80], and the Short Form-36 [90]. The resulting proposed minimum dataset is presented in Figure 1. PROMIS items are identified with a superscript 1, and STarT Back items (or very similar items) are identified with a superscript 2.

The RTF was able to obtain institutional review board approval at Stanford University (RTF member Sean Mackay, principal investigator) to conduct an internet survey of back pain patients using the RTF recommended version of the Minimum Dataset. This cross-sectional sample was distinct from the patients described above for validity testing, who underwent intervention and follow-up. There were 221 participants recruited from the San Francisco Bay Area using high-visibility ads. Participants had a mean age of 46.2 years (range, 19–81), with 53% female subjects. Participants included 72% whites, 17% Asians, 7% African Americans, and 3.8% each of American Indians and Pacific Islanders. There were 52% with at least a bachelor's degree and only a single participant with no high school diploma. Thirty-nine percent were employed, 5% were retired, and 16% described themselves as disabled. Thirty-eight percent described leg pain in addition to back pain, and the mean pain intensity (on a 0–10 scale) was 5.5. In this sample, the median time-to-completion was 7 minutes, and 75% of subjects completed the questionnaire in less than 10 minutes.

Proposed Supplemental Data for Specific Situations

For studies of invasive therapies such as spine surgery, the RTF recommended that physical examination and imaging data be added to the minimum dataset. Straight leg raising, lower extremity reflexes, and lower extremity strength as indicators of radiculopathy were

recommended as a minimum physical examination. Lumbar MRI was recommended in such studies as the minimal imaging evaluation.

In older adults, there is increased likelihood of hip osteoarthritis contributing to low back pain. Thus, for studies of adults mainly over age 65, the task force recommended testing internal hip rotation to help screen for potential osteoarthritis. A screen for cognitive function also may be important in such studies, as dementia may impair the validity of assessments or of consent for research.

In studies focused on behavioral or mood correlates of cLBP, the RTF recommended that investigators be free to incorporate additional measures. These might include, for example, assessment of emotional status, physical function and pain behaviors, substance abuse, interpersonal violence, or quality of life relevant to specific study interests. Such measures should have published reliability, validity, and responsiveness data at least equal to those of the minimum dataset's PROMIS short-form items. These additional measures should have population-based normative data to be included when relevant. The IMMPACT statement can be recommended as a starting point for selection of desired supplemental measures [99].

Recommendation 4. Outcome Measures

Investigators are referred to earlier consensus documents on outcome measures [99,111,112]. However, the RTF recommends reporting a “responder” analysis in addition to mean scores of outcome measures.

The RTF recognized that many parts of the baseline minimum dataset, such as the PROMIS measures, were highly appropriate as outcome measures, remembering that the initial focus of the NIH PROMIS effort was on patient-reported outcomes. It was also recognized that the primary outcomes of clinical studies would vary, depending on study aims. For example, some might focus on pain relief, whereas others might focus on return to work, physical function, mood, or need for subsequent therapy. Thus, the RTF did not make a recommendation regarding a minimum outcome dataset beyond recommending consideration of the minimum dataset for standardized recording of both baseline assessment and outcomes evaluation. Investigators are referred to earlier consensus statements on outcome measures for studying chronic pain in general or back pain in particular [99,111,112].

Reporting of Outcomes

An important discussion centered on reporting of outcomes. There was general agreement that for (at least theoretically) continuous measures, such as pain or function, in addition to mean scores and score changes, the proportion of participants achieving certain thresholds also should be reported. For example, the proportion of participants achieving a prespecified minimum clinically important change might be reported. Investigators have

proposed minimally important differences in PROMIS short forms, at least in the context of cancer therapy [113]. Calculating the percentage of study participants who achieve such landmarks is referred to by the U.S. Food and Drug Administration (FDA) as a “responder” analysis [114].

For example, other expert panels have suggested that a 30% improvement in pain or function might be a clinically important difference, and recommended reporting the proportion of participants with this degree of improvement [115]. Statistical analysts have suggested potential problems with the use of percentage changes [116], but the approach has clinical appeal. One might alternatively specify a certain number of points as the relevant change, or the percentage of participants reaching some threshold pain level (e.g., pain score below 3 out of 10).

An attractive option to the RTF was reporting the “cumulative distribution function” of responses for the treatment and control groups. This is a continuous plot of the proportion of patients at each scale score who experience change at that level or better. This amounts to calculating the percentage of responders at each value of the outcome score. This approach acknowledges the lack of consensus on the approach for establishing a responder threshold and provides information for any given threshold [114].

Composite Outcome Measures

The RTF also discussed the potential for use of composite outcome measures. One member noted that it is common in studies of osteoarthritis to require improvement in pain score *and* functional status *and* global self-assessment before judging treatment successful. Similar combinations have been proposed for evaluating back pain [17,117].

Composite measures are often required in FDA trials for drug or device approval. For example, “success” in trials of artificial disc replacement required functional improvement of 15 points on the Oswestry scale, improvement in quality of life on the Short Form-36, proper radiographic placement, and absence of new neurologic deficits or revision surgery [118]. Such composites offer the potential advantage of defining success in terms that are clearly clinically important, and not merely statistically significant.

However, the RTF concluded that with the paucity of data on performance of such composite measures for low back pain, it could not make a recommendation about composite outcome measures. Instead, this was recommended as an important topic for future research.

Time Frames for Outcome Measures

The RTF chose not to make specific recommendations for timing of outcome assessments because appropriate timing would vary depending on an intervention. For some treatments (e.g., analgesics or spinal manipulation), the goal may be short-term relief. For others, such as surgery,

the goal more often is long-term relief. For studying patients with chronic pain, longer-term follow-up (e.g., at least 6–12 months) is generally preferred.

Adverse Events

Reporting of adverse events was recognized as an important outcome measure. Because the likely adverse events vary enormously with the nature of an intervention, the RTF did not make recommendations for reporting specific adverse events. There was general agreement that for most intervention studies, it would be desirable to specify certain adverse events in advance and measure them prospectively, along with open-ended reporting of unanticipated events.

Recommendation 5. Research on the Proposed Standards

The RTF recommended new research to improve prognostic stratification of patients with cLBP; refine and test composite outcome measures for increasing the clinical importance of study results; undertake patient stakeholder assessment of relevant outcomes; and further evaluate psychometric properties of the minimum dataset.

Because the measures in the minimum dataset will often not comprise the sole measures used in a study, their widespread use will not only provide researchers a standardized set of data but also provide accumulating evidence for (or against) the reliability, validity, and clinical utility of the RTF recommendations. The potential for such an iterative approach to reevaluate scientific measures of chronic pain was successfully modeled in developing research diagnostic criteria for TMD. An iterative scientific process has successfully evolved the next generation of evidence-based measures for diagnosing and classifying the most common subtypes of TMD, including physical, behavioral, and psychosocial domains [119].

Beyond viewing the present set of recommendations as appropriate topics for future research, the RTF identified several related knowledge gaps that limit our ability to define and classify critical domains and variables. These were seen as important topics for which further research should be encouraged.

Prognosis

Improving prognostic stratification of patients with cLBP is important clinically to help guide the nature and intensity of therapy, and important for researchers to adjust for confounding and to improve comparability among studies. Recent work such as the STarT Back project from the United Kingdom has made important advances in this regard [49–52], and others have systematically reviewed risk factors for the emergence of chronic back pain [120]. However, the generalizability of such studies to interventions and populations outside primary care remains uncertain. Other approaches may be important for specific populations or for predicting specific treatment outcomes.

Additional work in this area might improve the ability to characterize clinically important subgroups of patients with cLBP and improve our “impact stratification.”

Composite Outcome Measures

An ongoing frustration has been the seeming lack of progress in reducing back-related disability at a population level. In part, this may be a result of claiming treatment efficacy based on statistically significant but clinically trivial results. More work is needed to understand how certain outcome scores are associated with major events, such as return to work. Composite outcome measures, such as requiring simultaneous improvement in pain, function, and global self-assessment, may move us closer to important outcomes. However, more data are needed to determine the performance of such measures in terms of validity, reliability, responsiveness, and prognostic value.

Patient Stakeholder Assessment

Little work has addressed the outcomes judged most important by patients with cLBP. Such outcomes may vary with demographic features and diagnosis.

Psychometric Properties of the Proposed Minimum Dataset

Extensive effort has been made to validate the PROMIS measures [93,95,102,103,105,106,121–123], but there is modest information on their performance specifically in the context of cLBP. One recent study suggested excellent performance of the PROMIS physical function item bank among patients with back and neck problems [124]. Further data on the precision of the domains is important (e.g., the optimal number of items), as are data on responsiveness to change and sensitivity to small differences. Creating a “cross-walk” of scores with legacy measures, such as the Oswestry and Roland-Morris disability questionnaires, is also important.

Recommendation 6. Dissemination of the Report of the NIH Task Force on Research Standards for Chronic Low Back Pain

With adoption of recommendations by the NIH Pain Consortium, the RTF recommends dissemination to the broad research community, including publication of a report in multiple professional journals and presentations at professional meetings.

The NIH Pain Consortium has accepted the RTF report (to view the full NIH-approved RTF report on Standards for Research on Chronic Low Back Pain, see painconsortium.nih.gov). The consortium is recommending that all NIH institutes and centers require grant applications proposing clinical studies of cLBP to use the research standards set forth in the RTF report. Similarly, NIH encourages all other agencies that fund research on cLBP to consider incorporating these research standards for their respective awardees or investigators, as

appropriate. The RTF proposed to disseminate these recommendations in professional journals and presentations at scientific meetings.

Discussion

Consistent with its charge from NIH, the RTF strove to recommend standards for conducting research into the complex, intertwined factors that influence the onset, natural history, and clinical course of cLBP. This remains one of the most important and costly of all public health conditions affecting the U.S. population. As adopted by NIH, these recommendations have the potential to standardize methods for identifying cLBP research cases, describing research subjects, and comparing published reports.

The new research standards should improve the comparability of research studies on cLBP, facilitate pooling data from multiple studies (e.g., for meta-analyses), and improve the ability to define phenotypes among patients with low back pain. These standards will allow comparable core summary statistics to be included in all published reports without interfering with collection of specific measures needed to address specific research questions.

After extended review and discussion, the RTF concluded that at the current state of scientific evidence on cLBP, it was not realistic to create operationally defined research diagnostic criteria for subsets of cLBP. While creation of research diagnostic criteria has proven beneficial to research for some other conditions (e.g., TMD [119] and Alzheimer's disease [125]), the multifactorial nature of most cases of cLBP decreased enthusiasm for attempting to do so in this condition. However, creation of an impact stratification and a uniform minimum dataset will achieve many of the same goals.

In summary, the RTF has recommended a definition of cLBP and has proposed classifying it in terms of its impact, in addition to any presumed pathoanatomic diagnosis. Impact is conceived as a combination of pain intensity, interference with activities, and physical function. The RTF also has recommended a uniform minimum dataset, with recommendations for medical history, physical examination, diagnostic tests, and self-report measures of physical function, depression, and sleep disturbance, in addition to pain intensity and interference. Finally, recommendations have been made for reporting patient outcomes, further research, and dissemination of the recommendations.

Any effort to standardize research methods is only a starting point for further testing and refinement. The final recommendations were seen as a first step toward creating standards for research in cLBP. We anticipate that further validation, refinement, and possible extension of these recommendations will require years and the efforts of many investigators. Nonetheless, the RTF believes these recommendations can advance the field, help to resolve controversies, and facilitate future research addressing the

prevalence and incidence and genomic, neurologic, and other mechanistic substrates of cLBP. Furthermore, it can help reveal the biologic-behavioral interfaces that confront our present-day understanding of cLBP and its evidence-based management.

It is anticipated that the RTF recommendations will become a dynamic document, and that the proposals are likely to undergo continual improvement. The proposed research agenda should facilitate this evolution.

References

- 1 Pizzo PA, Clark NM, Carter-Pokras O, et al. Institutes of Medicine: Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. Washington, DC: The National Academies Press; 2011.
- 2 Hoy D, Bain C, Williams G, et al. A systematic review of the global prevalence of low back pain. *Arthritis Rheum* 2012;64:2028–37.
- 3 Hoy D, Brooks P, Blyth F, Buchbinder R. The Epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 2010;24:769–81.
- 4 Deyo RA, Mirza SK, Turner JA, Martin BI. Overtreating chronic back pain: Time to back off? *J Am Board Fam Med* 2009;22:62–8.
- 5 Kenan K, Mack K, Paulozzi L. Trends in prescriptions for oxycodone and other commonly used opioids in the United States, 2000–2010. *Open Med* 2012;6:e41–7.
- 6 Rajae SS, Bae HW, Kanim LE, Delamarter RB. Spinal fusion in the United States: Analysis of trends from 1998 to 2008. *Spine (Phila Pa 1976)* 2012;37:67–76.
- 7 Tracey I, Bushnell MC. How neuroimaging studies have challenged us to rethink: Is chronic pain a disease? *J Pain* 2009;10:1113–20.
- 8 Baliki MN, Petre B, Torbey S, et al. Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat Neurosci* 2012;15:1117–9.
- 9 Rodriguez-Raecke R, Niemeier A, Ihle K, Ruether W, May A. Structural brain changes in chronic pain reflect probably neither damage nor atrophy. *PLoS ONE* 2013;8:e54475.
- 10 Seminowicz DA, Wideman TH, Naso L, et al. Effective treatment of chronic low back pain in humans reverses abnormal brain anatomy and function. *J Neurosci* 2011;31:7540–50.
- 11 Barker KL, Shamley DR, Jackson D. Changes in the cross-sectional area of multifidus and psoas in patients with unilateral back pain: The relationship to

- pain and disability. *Spine (Phila Pa 1976)* 2004;29:E515–9.
- 12 Battie MC, Niemelainen R, Gibbons LE, Dhillon S. Is level- and side-specific multifidus asymmetry a marker for lumbar disc pathology? *Spine J* 2012;12:932–9.
 - 13 Beneck GJ, Kulig K. Multifidus atrophy is localized and bilateral in active persons with chronic unilateral low back pain. *Arch Phys Med Rehabil* 2012; 93:300–6.
 - 14 Chou R, Deyo RA, Jarvik JG. Appropriate use of lumbar imaging for evaluation of low back pain. *Radiol Clin North Am* 2012;50:569–85.
 - 15 Bunzli S, Watkins R, Smith A, Schutze R, O’Sullivan P. Lives on hold: A qualitative synthesis exploring the experience of chronic low-back pain. *Clin J Pain* 2013;29:907–16.
 - 16 Negrini S, Zaina F. The chimera of low back pain etiology: A clinical rehabilitation perspective. *Am J Phys Med Rehabil* 2013;92:93–7.
 - 17 Bombardier C, Evans CJ, Katz N, et al. Further qualification of a therapeutic responder index for patients with chronic low back pain. *J Rheumatol* 2011; 38:362–9.
 - 18 Brown RL, Leonard T, Saunders LA, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. *J Am Board Fam Pract* 2001;14:95–106.
 - 19 Bruyere O, Demoulin M, Brereton C, et al. Translation validation of a new back pain screening questionnaire (the STarT Back Screening Tool) in French. *Arch Public Health* 2012;70:12.
 - 20 Chou R, Atlas SJ, Stanos SP, Rosenquist RW. Non-surgical interventional therapies for low back pain: A review of the evidence for an American Pain Society clinical practice guideline. *Spine (Phila Pa 1976)* 2009;34:1078–93.
 - 21 Chou R, Huffman LH. Medications for acute and chronic low back pain: A review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:505–14.
 - 22 Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: A review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:492–504.
 - 23 Chou R, Loeser JD, Owens DK, et al. Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: An evidence-based clinical practice guideline from the American Pain Society. *Spine (Phila Pa 1976)* 2009;34:1066–77.
 - 24 Abraham I, Killackey-Jones B. Lack of evidence-based research for idiopathic low back pain: The importance of a specific diagnosis. *Arch Intern Med* 2002;162:1442–4. discussion 1447.
 - 25 Apeldoorn AT, Bosmans JE, Ostelo RW, de Vet HC, van Tulder MW. Cost-effectiveness of a classification-based system for sub-acute and chronic low back pain. *Eur Spine J* 2012;21:1290–300.
 - 26 Bogduk N. On the definitions and physiology of back pain, referred pain, and radicular pain. *Pain* 2009;147:17–9.
 - 27 Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. *Ann Intern Med* 2004;141:920–8.
 - 28 Croft P, Dunn KM, Von Korff M. Chronic pain syndromes: You can’t have one without another. *Pain* 2007;131:237–8.
 - 29 Deyo RA. Diagnostic evaluation of LBP: Reaching a specific diagnosis is often impossible. *Arch Intern Med* 2002;162:1444–7. discussion 1447–1448.
 - 30 Fairbank J, Gwilym SE, France JC, et al. The role of classification of chronic low back pain. *Spine (Phila Pa 1976)* 2011;36:S19–42.
 - 31 Genevay S, Atlas SJ, Katz JN. Variation in eligibility criteria from studies of radiculopathy due to a herniated disc and of neurogenic claudication due to lumbar spinal stenosis: A structured literature review. *Spine (Phila Pa 1976)* 2010;35:803–11.
 - 32 Hall H, McIntosh G, Boyle C. Effectiveness of a low back pain classification system. *Spine J* 2009;9:648–57.
 - 33 Konstantinou K, Hider SL, Jordan JL, et al. The impact of low back-related leg pain on outcomes as compared with low back pain alone: A systematic review of the literature. *Clin J Pain* 2013;29:644–54.
 - 34 Loisel P, Vachon B, Lemaire J, et al. Discriminative and predictive validity assessment of the Quebec task force classification. *Spine (Phila Pa 1976)* 2002;27:851–7.
 - 35 Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. *JAMA* 2008;299:656–64.

- 36 Shah RV. Spine pain classification: The problem. *Spine (Phila Pa 1976)* 2012;37:1853–5.
- 37 Spitzer WO, Abenhaim L, Dupuis M, et al. Quebec task force on spinal disorders. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. Chapter 3: Diagnosis of the problem (the problem of diagnosis). *Spine (Phila Pa 1976)* 1987;12:S16–21.
- 38 Stanton TR, Latimer J, Maher CG, Hancock MJ. A modified Delphi approach to standardize low back pain recurrence terminology. *Eur Spine J* 2011;20:744–52.
- 39 Waddell G. Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain. *Spine (Phila Pa 1976)* 1987;12:632–44. 1987.
- 40 Beneciuk JM, Bishop MD, Fritz JM, et al. The STarT back screening tool and individual psychological measures: Evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Phys Ther* 2013;93:321–33.
- 41 Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994;23:129–38.
- 42 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* 2008;135:48–54.
- 43 Dunn KM, Von Korff M, Croft P. Defining chronic pain by prognosis. In: Hasenbring MI, Rusu AC, Turk DC, eds. *From Acute to Chronic Back Pain: Risk Factors, Mechanisms, and Clinical Implications*. New York, NY: Oxford University Press; 2012:21–40.
- 44 Edelen MO, Saliba D. Correspondence of verbal descriptor and numeric rating scales for pain intensity: An item response theory calibration. *J Gerontol A Biol Sci Med Sci* 2010;65:778–85.
- 45 Friedman BW, Mulvey L, Davitt M, et al. Predicting 7-day and 3-month functional outcomes after an ED visit for acute nontraumatic low back pain. *Am J Emerg Med* 2012;30:1852–9.
- 46 Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Phys Ther* 2011;91:722–32.
- 47 Gerbershagen HJ, Rothaug J, Kalkman CJ, Meissner W. Determination of moderate-to-severe postoperative pain on the numeric rating scale: A cut-off point analysis applying four different methods. *Br J Anaesth* 2011;107:619–26.
- 48 Gusi N, del Pozo-Cruz B, Olivares PR, Hernandez-Mocholi M, Hill JC. The Spanish version of the “STarT Back Screening Tool” (SBST) in different subgroups. *Aten Primaria* 2011;43:356–61.
- 49 Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008;59:632–41.
- 50 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* 2010;14:83–9.
- 51 Hill JC, Vohora K, Dunn KM, Main CJ, Hay EM. Comparing the STarT back screening tool’s subgroup allocation of individual patients with that of independent clinical experts. *Clin J Pain* 2010;26:783–7.
- 52 Hill JC, Whitehurst DG, Lewis M, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. *Lancet* 2011;378:1560–71.
- 53 Jensen MP, Smith DG, Ehde DM, Robinsin LR. Pain site and the effects of amputation pain: Further clarification of the meaning of mild, moderate, and severe pain. *Pain* 2001;91:317–22.
- 54 Jones KR, Vojir CP, Hutt E, Fink R. Determining mild, moderate, and severe pain equivalency across pain-intensity tools in nursing home residents. *J Rehabil Res Dev* 2007;44:305–14.
- 55 Kapstad H, Hanestad BR, Langeland N, Rustoen T, Stavem K. Cutpoints for mild, moderate and severe pain in patients with osteoarthritis of the hip or knee ready for joint replacement surgery. *BMC Musculoskelet Disord* 2008;9:55.
- 56 Kongsted A, Johannesen E, Leboeuf-Yde C. Feasibility of the STarT back screening tool in chiropractic clinics: A cross-sectional study of patients with low back pain. *Chiropr Man Therap* 2011;19:10.
- 57 Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. *J Gen Intern Med* 2009;24:733–8.
- 58 Mallen CD, Peat G, Thomas E, Dunn KM, Croft PR. Prognostic factors for musculoskeletal pain in primary care: A systematic review. *Br J Gen Pract* 2007;57:655–61.

- 59 Martell BA, O'Connor PG, Kerns RD, et al. Systematic review: Opioid treatment for chronic back pain: Prevalence, efficacy, and association with addiction. *Ann Intern Med* 2007;146:116–27.
- 60 Morso L, Albert H, Kent P, Manniche C, Hill J. Translation and discriminative validation of the STarT Back Screening Tool into Danish. *Eur Spine J* 2011;20:2166–73.
- 61 Muller S, Thomas E, Dunn KM, Mallen CD. A prognostic approach to defining chronic pain across a range of musculoskeletal pain sites. *Clin J Pain* 2013;29:411–6.
- 62 Salovey P, Sieber W, Smith AF, et al. Reporting Chronic Pain episodes on Health Surveys. Vital and Health Statistics Series. Hyattsville, MD: National Technical Information Service; 1992. NTIS Issue Number 9303.
- 63 Stewart WF, Lipton RB, Simon D, Liberman J, Von Korff M. Validity of an illness severity measure for headache in a population sample of migraine sufferers. *Pain* 1999;79:291–301.
- 64 Thomas E, Dunn KM, Mallen C, Peat G. A prognostic approach to defining chronic pain: Application to knee pain in older adults. *Pain* 2008;139:389–97.
- 65 Turner JA, Shortreed SM, Saunders KW, et al. Optimizing prediction of back pain outcomes. *Pain* 2013;154:1391–401.
- 66 Von Korff M. Assessment of chronic pain in epidemiological and health services research: Empirical cases and new directions. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment*. New York, NY: Guilford Press; 2001:455–73.
- 67 Von Korff M, Dunn KM. Chronic pain reconsidered. *Pain* 2008;138:267–76.
- 68 Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. *Pain* 2005;117:304–13.
- 69 Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain* 1992;50:133–49.
- 70 Von Korff M, Shortreed SM, Saunders KW, et al. Comparison of back pain prognostic risk stratification item sets. *J Pain* 2014;15:81–9.
- 71 Wideman TH, Hill JC, Main CJ, et al. Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: The whole is greater than the sum of its parts. *Pain* 2012;153:2182–91.
- 72 Beck AT, Steer RA, Ball R, Ranieri W. Comparison of beck depression inventories—IA and—II in psychiatric outpatients. *J Pers Assess* 1996;67:588–97.
- 73 Dworkin RH, Turk DC, Revicki DA, et al. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). *Pain* 2009;144:35–42.
- 74 Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
- 75 Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: Validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997;36:551–9.
- 76 Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment*. New York, NY: Guilford Press; 2001:15–34.
- 77 Keefe FJ, Williams DA, Smith SJ. Assessment of pain behaviors. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment*. New York, NY: Guilford Press; 2001:170–90.
- 78 Kerns RD, Haythornthwaite J, Rosenberg R, et al. The Pain Behavior Check List (PBCL): Factor structure and psychometric properties. *J Behav Med* 1991;14:155–67.
- 79 Kori S, Miller R, Todd D. Kinesiophobia: A new view of chronic pain behavior. *Pain Management* 1990;3:35–43.
- 80 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
- 81 Kroenke K, Spitzer RL, Williams JB. The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;64:258–66.
- 82 Margolis RB, Tait RC, Krause SJ. A rating system for use with patient pain drawings. *Pain* 1986;24:57–65.
- 83 Radloff LS. The CES-D scale: A self-report depression scale for research in general populations. *Appl Psych Meas* 1977;1:385–401.
- 84 Roland M, Morris R. A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 1983;8:141–4.
- 85 Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: Relationship to patient characteristics and current adjustment. *Pain* 1983;17:33–44.

Deyo et al.

- 86 Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med* 2006;166:1092–7.
- 87 Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: Development and validation. *Psychol Assess* 1995;7:524–32.
- 88 Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157–68.
- 89 Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
- 90 Ware JE Jr, Sherbourne CD. The MOS 36-item short- form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- 91 Weathers FW, Ford J. Psychometric properties of the PTSD Checklist (PCL–C, PCL–S, PCL–M, PCL–PR). In: Stamm BH, ed. *Measurement of Stress, Trauma, and Adaptation*. Lutherville, MD: Sidran Foundation & Press; 1996:250–2.
- 92 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- 93 Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain* 2010;150:173–82.
- 94 Atlas SJ, Deyo RA, van den Ancker M, et al. The Maine-Seattle back questionnaire: A 12-item disability questionnaire for evaluating patients with lumbar sciatica or stenosis: Results of a derivation and validation cohort analysis. *Spine* 2003;28:1869–76.
- 95 Buysse DJ, Yu L, Moul DE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep* 2010;33:781–92.
- 96 Cherkin DC, Deyo RA, Street JH, Barlow W. Predicting poor outcomes for back pain seen in primary care using patients' own criteria. *Spine (Phila Pa 1976)* 1996;21:2900–7.
- 97 Cook KF, Choi SW, Crane PK, et al. Letting the CAT out of the bag: Comparing computer adaptive tests and an 11-item short form of the Roland-Morris Disability Questionnaire. *Spine* 2008;33:1378–83.
- 98 Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613–9.
- 99 Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113:9–19.
- 100 Dworkin RH, Turk DC, McDermott MP, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain* 2009;146:238–44.
- 101 Froud R, Eldridge S, Kovacs F, et al. Reporting outcomes of back pain trials: A modified Delphi study. *Eur J Pain* 2011;15:1068–74.
- 102 Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: A patient-reported outcomes measurement information system initiative. *J Clin Oncol* 2007;25:5106–12.
- 103 Hahn EA, Devellis RF, Bode RK, et al. Measuring social health in the patient-reported outcomes measurement information system (PROMIS): Item bank development and testing. *Qual Life Res* 2010;19:1035–44.
- 104 Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine (Phila Pa 1976)* 1995;20:1899–908. discussion 1909.
- 105 Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): Depression, anxiety, and anger. *Assessment* 2011;18:263–83.
- 106 Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol* 2008;61:17–33.
- 107 Field J, Newell D. Relationship between STaRT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropr Man Therap* 2012;20:17.
- 108 Cella D, Gershon R, Bass M, Rothrock N. Accessing PROMIS Short Form Instruments. 2007. Available at: <http://www.assessmentcenter.net/PromisForms.aspx> (accessed February 2014).
- 109 Cella D, Riley W, Stone A, et al. The Patient- Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63:1179–94.

- 110 Cella D. PROsetta Stone: Linking Patient-Reported Outcome Measures. Available at: <http://www.prosettastone.org> (accessed February 2014).
- 111 Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: Summary and general recommendations. *Spine (Phila Pa 1976)* 2000;25:3100–3.
- 112 Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine (Phila Pa 1976)* 1998;23:2003–13.
- 113 Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol* 2011;64:507–16.
- 114 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* 2011;11:163–9.
- 115 Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001;94:149–58.
- 116 Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Med Res Methodol* 2001;1:6.
- 117 Simon LS, Evans C, Katz N, et al. Preliminary development of a responder index for chronic low back pain. *J Rheumatol* 2007;34:1386–91.
- 118 Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)* 2007;32:1155–62. discussion 1163.
- 119 Schiffman E, Ohrbach R, Truelove E, et al. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: Recommendations of the International RDC/ TMD Consortium Network* and Orofacial Pain Special Interest Groupdagger. *J Oral Facial Pain Headache* 2014; 28:6–27.
- 120 Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? *JAMA* 2010;303:1295–302.
- 121 Askew RL, Kim J, Chung H, et al. Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. *Qual Life Res* 2013;22:2769–76.
- 122 Kim J, Chung H, Amtmann D, Revicki DA, Cook KF. Measurement invariance of the PROMIS pain interference item bank across community and clinical samples. *Qual Life Res* 2013;22:501–7.
- 123 Revicki DA, Cook KF, Amtmann D, et al. Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. *Qual Life Res* 2014;23:245–55.
- 124 Hung M, Hon SD, Franklin JD, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine (Phila Pa 1976)* 2014;39:158–63.
- 125 Sarazin M, de Souza LC, Lehericy S, Dubois B. Clinical and research diagnostic criteria for Alzheimer's disease. *Neuroimaging Clin N Am* 2012;22:23–32. viii.