

Outcome Measures for Low Back Pain Research: A Proposal for Standardized Use

Deyo, Richard A. MD, MPH*; Battie, Michele PhD, PT†; Beurskens, A. J. H. M. PhD, PT‡; Bombardier, Claire MD§; Croft, Peter MD ; Koes, Bart PhD#; Malmivaara, Antti MD, PhD††; Roland, Martin MD‡‡; Von Korff, Michael PhD§§; Waddell, Gordon DSc, MD

From the *Center for Cost and Outcomes Research, the Department of Medicine, and the Department of Health Services, University of Washington, Seattle, Washington; the †University of Alberta, Edmonton, Alberta, Canada; ‡The Department of Epidemiology, Maastricht University, Maastricht, The Netherlands; §The Institute for Work and Health, Toronto, Ontario, Canada; #Keele University School of Postgraduate Medicine, Stoke-on-Trent, United Kingdom; the #EMGO Institute, Vrije University, Amsterdam, The Netherlands; the ††Finnish Institute of Occupational Health, Helsinki, Finland; the ‡‡National Primary Care Research and Development Centre, University of Manchester, Manchester, United Kingdom; §§Group Health Cooperative, Seattle, Washington; and The Glasgow Nuffield Hospital, Glasgow, Scotland.

Supported in part by a grant (HS-08194) from the Agency for Health Care Policy and Research, Washington DC.

Presented in part at the Second International Forum for Primary Care Research on Low Back Pain, The Hague, The Netherlands, May 30-31, 1997.

Acknowledgment date: August 14, 1997.

First revision date: January 30, 1998.

Acceptance date: February 11, 1998.

Device status category: 1.

Address reprint requests to: Richard A. Deyo, MD, MPH; Division of General Internal Medicine; University of Washington; Box 356429; Seattle, WA 98195; E-mail: deyo@u.washington.edu.

Abstract [TOP](#)

Study Design. An international group of back pain researchers considered recommendations for standardized measures in clinical outcomes research in patients with back pain.

Objectives. To promote more standardization of outcome measurement in clinical trials and other types of outcomes research, including meta-analyses, cost-effectiveness analyses, and multicenter studies.

Summary of Background Data. Better standardization of outcome measurement would facilitate comparison of results among studies, and more complete reporting of relevant outcomes. Because back pain is rarely fatal or completely cured, outcome assessment is complex and involves multiple dimensions. These include symptoms, function, general well-being, work disability, and satisfaction with care.

Methods. The panel considered several factors in recommending a standard battery of outcome measures. These included reliability, validity, responsiveness, and practicality of the measures. In addition, compatibility with widely used and promoted batteries such as the American Academy of Orthopaedic Surgeons Lumbar Cluster were considered to minimize the need for changes when these instruments are used.

Results. First, a six-item set was proposed, which is sufficiently brief that it could be used in routine care settings for quality improvement and for research purposes. An expanded outcome set, which would provide more precise measurement for research purposes, includes measures of severity and frequency of symptoms, either the Roland or the Oswestry Disability Scale, either the SF-12 or the EuroQol measure of general health status, a question about satisfaction with symptoms, three types of "disability days," and an optional single item on overall satisfaction with medical care.

Conclusion. Standardized measurement of outcomes would facilitate scientific advances in clinical care. A short, 6-item questionnaire and a somewhat expanded, more precise battery of questionnaires can be recommended. Although many considerations support such recommendations, more data on responsiveness and the minimally important change in scores are needed for most of the instruments.

The measurement of patient outcomes in clinical studies of low back pain has been vexing for many investigators. Traditionally, in an effort to achieve objectivity, physiologic measures such as range of motion and muscle strength were widely used. However, in many cases, such measures are only weakly associated with outcomes that are more relevant to patients and to society, such as symptom relief, daily functioning, and work status.⁷ In recent years, social science methods and clinical expertise have been fused in the creation of a series of questionnaire measures that seek to capture this broader range of relevant outcomes.^{7,8} However, there is little standardization of use of these instruments, and comparisons among studies are therefore difficult or impossible. New questionnaires seem constantly to emerge. As a result, there is little shared understanding of what certain results mean, what their clinical relevance may be, or how the patient populations and results of different studies may compare. [Table 1](#)

A multinational group of investigators met as part of an international program on primary care research on back pain held in The Hague, The Netherlands, in May 1997. The group considered the question of whether a relatively standardized set of outcome measures could be recommended, based on published studies and the group's assessment of the important domains. This report summarizes the discussion and provides initial recommendations for investigators in this field.

A standardized "core" set of questions and questionnaires would have many advantages for clinicians and investigators. Such a core would facilitate many types of comparison and pooling of data, while leaving investigators free to augment the core with a wide variety of measures of their own choice. Thus, the effort to develop a standardized set of instruments is not intended to force investigators into a straitjacket, but to provide a common yardstick that is appropriate for use in many types of studies. For individual investigations, it may be important to augment this core with measures of specific clinical effects or to experiment with new measures of the constructs included in the standardized core. Thus, it should be anticipated that the "core" measures may change somewhat with time but that change may be gradual enough to maintain some standardization within the field.

It seems clear that the traditional surgical outcome measure of a single rating scale (excellent, good, fair, and poor) is no longer sufficient. Howe and Frymoyer¹³ demonstrated that different definitions of these terms could result in widely differing conclusions about the success of a given surgical procedure. Furthermore, such a rating scale fails to indicate, for example, how many patients are employed at the beginning of a study and how many are employed at the end of a study. It also combines multiple dimensions of outcome, such as symptoms, function, and work disability, which may be relatively unrelated in their clinical trajectory after a treatment intervention. For example, it has been shown that among patients undergoing discectomy for sciatica, surgery offers substantial advantages in symptom and functional outcomes at 1 year, but that return to work is equivalent for surgical and nonsurgical treatments.¹ Thus, a standardized set of instruments should measure various outcomes in different dimensions and keep them relatively distinct.

Advantages of Standardized Outcome Measurement [TOP](#)

A standardized set of clinical outcome measures would make it easier to compare the results of clinical studies of similar treatments. Presently, this is almost impossible, even when outcome ratings are superficially similar (such as the excellent-good-fair-poor scale). Even when studies are concordant in demonstrating benefit for a given treatment, differences in outcome measures may make it difficult to assess the relative magnitude of treatment effects among various studies. This may make it difficult or impossible to pool the results of multiple studies in the form of a meta-analysis. Such pooling of results may be important when there are multiple small studies in which results show a treatment benefit, but the individual studies are too small to show statistical significance; when clinical subgroup analyses may be important, but studies must be combined to achieve adequate numbers of patients; and when a single assessment of the magnitude of

treatment effects is sought for purposes of cost-effectiveness analyses. Because cost-effectiveness analysis is always a ratio of cost to treatment effectiveness, the denominator must be a valid estimate of the magnitude of treatment effects, along with confidence intervals. If the outcome measures can be integrated with an assessment of patient preferences or weighting of the outcomes (utilities) the effectiveness and cost-effectiveness measures can be transferred into a standardized metric that permits comparison with many other treatments for other medical conditions. This may be important for resource allocation purposes.

The availability of a standardized outcome measurement set should also improve the quality of the medical literature. It would encourage more complete reporting of relevant outcomes, so that investigators do not simply report a single dimension of outcome while ignoring others. Having a common set of measures that are widely used would also encourage development of cooperative multicenter studies, which offers the prospect of large, rapid, and generalizable efficacy and effectiveness studies. If various centers were familiar with a core set of measures, had the instruments readily available, and were familiar with their administration and manipulation, the design and conduct of such cooperative studies would be facilitated. Finally, having some standardized measures of outcome would simplify the process of designing and reviewing research proposals, manuscripts, and published studies. [Table 2](#)

Dimensions of Outcome [TOP](#)

Some have previously argued for measuring the outcomes of low back pain in terms of symptoms, functional status, overall well-being, and work disability.¹⁴ Valid measures of all of these constructs are available, and such dimensions have been incorporated into the outcome instruments of the American Academy of Orthopaedic Surgeons (AAOS) and the North American Spine Society (NASS). Thus, there seems to be a growing consensus that these are appropriate dimensions of outcome for patients with back pain.

There is substantial evidence that symptoms and functional status change fairly readily among patients with acute low back pain and that these aspects of outcome can be influenced by various treatments. In contrast, results in some studies show that affecting work disability or employment status may be very difficult.¹ Although work status has the advantage that it is easily assessed, objective, and highly relevant, it is usually a function of many factors, among which medical intervention may be only minor. For example, physical job demands, job satisfaction, relationships with fellow employees, supervisor ratings, income, regional job availability, closeness to retirement age, or the availability of another breadwinner in the family all influence the likelihood of return to work. Thus, although return to work is an outcome of great social and personal importance, it may also be one that is less responsive to clinical treatment than symptoms or daily functioning. In considering a core set of outcome measures, it seems appropriate to consider outcomes that are highly responsive to treatment and those that are of major social importance, even though some outcome measures do not have both traits.

For many purposes, physiologic outcomes such as neurologic function, range of motion, or muscle strength are important outcomes to measure as well. However, the wide range of methods for measuring such physiologic functions and the extensive literature on the reliability and validity of such measures is beyond the scope of this article. Furthermore, there are many examples in which measures such as electromyogram activity, spine mobility, or straight leg raising results were poorly associated with pain relief, functional status, or use of health care resources.⁷ Thus, the current group chose to focus on questionnaire measures that could be easily administered in multiple settings and by a wide variety of personnel.

Finally, health care use may be considered an important outcome domain, because hospitalization, use of imaging, surgery, and health care costs are important considerations in studying therapy. However, these are measures of the process of care, rather than outcomes. They do not necessarily reflect health status and are not easily measured with a brief, uniform set of items. Again, measures of cost or health care use could easily be added by individual investigators, depending on the goals of their research.

Factors to Consider in Recommending a Standard Battery of Outcome Measures [TOP](#)

Many factors could be considered in selecting questions or questionnaires for standardized use. The recommendations made here were drawn up primarily in consideration of the demonstrated validity of instruments, their responsiveness to change with time, and their practicality. For validity, the major concern was whether the items selected had evidence of construct validity—that is, associations with different but related measures in a predictable direction and of moderate magnitude. Construct validation is necessary, because for symptoms, function, and well-being, there is no widely accepted gold standard for measurement.

Responsiveness refers to the ability to detect true changes in patient status beyond the random variability that is expected on repeat measurement of any sort. All measures of responsiveness are in some way measures of "signal" to "noise." The signal is the true change in patient status and the noise is the random variability of measures. Several indicators of responsiveness have been proposed, including effect size, the standardized response mean, and variations of these statistics. Some investigators have also examined score changes as they relate to other external measures of improvement or deterioration, using either correlation analysis or receiver operating characteristic (ROC) curves.⁹ For an instrument designed to measure change in longitudinal studies, such measures of responsiveness may be more important than cross-sectional reliability, in which the signal-to-noise ratio represents differences between patients (rather than longitudinal change for a single patient) related to cross-sectional score variability.

The major concern regarding practicality was the length of the questionnaires. To be practical, they should be as brief as possible, minimizing response burden and the costs of data collection and management. Furthermore, the core data set should be brief if investigators are to feel free to add other measures to the battery.

A final consideration was the compatibility of questions with widely promoted instruments or batteries, such as the Lumbar Cluster developed by the AAOS, the NASS Lumbar Spine Outcome Assessment Instrument, or the typology of patient experience (TyPE) originally developed by InterStudy (later reverting to the Health Outcomes Institute, Bloomington, Minnesota, now a part of Stratis Health, Bloomington). Thus, it would be desirable for the core items to be included in commonly used instruments so that practitioners or health care systems that have already implemented one of these batteries would not have to make changes (or could make minor changes) to include the core battery.

A Very Parsimonious Six-Item Core Set [TOP](#)

A proposed core set of just six questions was developed ([Table 3](#) and [Appendix I](#)) that would be practical for use in a wide variety of settings, including routine clinical care, quality improvement efforts, and more formal research. It is so brief that even lengthy supplemental measures (e.g., physiologic measures, detailed questionnaires) could be added and still be practical, retaining a core set of multidimensional measures that would assure comparability with results in a wide variety of other studies. Most investigators would want to collect other demographic and clinical information for baseline description, and this set focuses strictly on outcomes of care. However, many investigators would also want to measure each of these constructs at a baseline time, or before some intervention. This set has the advantage of measuring several dimensions of outcome, each with a single item. Even this short set of measures would be a substantial improvement compared with simply measuring pain severity, or excellent-good-fair-poor outcomes.

The image shows a table with multiple columns and rows, which is mostly illegible due to low resolution. It appears to be a structured list of items, possibly the six questions mentioned in the text, with associated details like response options or scales.

Table 3. A Proposed "Core Set" of Six Questions, Practical for Routine Clinical Use, Quality Improvement, and as a Component of More Formal Research*

Each of the questions in this short set has been studied and validated elsewhere and is already incorporated into other widely promulgated questionnaires. For example, the questions on bothersomeness of back and leg pain, function, and well-being are all included in the Lumbar Cluster of the AAOS and the NASS outcome questionnaire.⁵ However, some of the questions (e.g., those concerning the "bother" of symptoms) have not been translated into languages other than English or tested outside the United States.

The first item in this core set is a measure of pain symptoms that have been divided into low back symptoms and leg symptoms. For this purpose, the current suggestion is that pain severity be measured in one of two ways at the discretion of the investigator. First, a set of questions could be used to determine how bothersome pain symptoms have been, as was reported in the Maine Lumbar Spine Study and is incorporated in the AAOS, NASS, and TyPE instruments. This question is shown in [Appendix I](#). A 1-week time frame for these symptoms is suggested because it allows the patient to integrate recent experience of a long enough interval to be meaningful, but short enough that memory is not likely to be a problem and short-term improvements are apparent.⁵ The severity of low back and leg pain could also be measured using conventional visual analog scales, a type of measure more familiar to some investigators. Guyatt et al¹² demonstrated nearly identical statistical measures of responsiveness for 7-point Likert Scales and Visual Analog Scales. However, they found that the Likert format with verbal descriptors (such as the "bothersome" question) was easier to understand for patients and required less instruction. This is likely to be especially important with elderly patients, such as those with spinal stenosis.

Although some investigators may wish to ask much more extensive pain questions, such as the most severe pain, the duration of pain, or other aspects of pain symptomatology, the current investigators recommend these items on pain severity for the past week as a minimum set of core questions. The questions on "bothersomeness" of pain symptoms have been validated in patients with sciatica by demonstrating highly significant associations with measures of functional status, absence from work, reflex changes, straight leg raising, and use of opioid analgesics.^{1,14} Similarly, visual analog pain scales have enjoyed wide use, and similar evidence of construct validity has accrued.

Regarding function, the parsimonious set includes a single question about how much pain interfered with normal work. This question is derived from the widely used Short Form 36 (SF-36)¹⁶ and its derivative shorter version, the SF-12, developed by Ware and colleagues.¹⁷ It is also incorporated into the Lumbar Cluster of the AAOS and was a part of the Maine Lumbar Spine Study.¹

The current proposal is that a single item be used to measure overall well-being by asking, "If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?" This item was used by the low back pain Patient Outcome Research Team at the University of Washington, and was also used by the initial Patient Outcome Research Teams for studies in coronary heart disease, prostate disease, and cataract disease that were funded by the Agency for Health Care Policy and Research. This item was also adopted as part of the Lumbar Cluster of the AAOS and was incorporated into the Maine Lumbar Spine Study. The performance of this particular question has been studied in relation to Roland Disability Scores, symptom bother-someness, and the presence of symptoms and depression.⁴ In the Maine Lumbar Spine Study, it was highly responsive to differences in treatment outcomes.¹⁴

Two items on disability related to social role are included in the core set. Both are derived from the U. S. National Health Interview Survey, and have been validated by identifying significant associations with compensation status, use of opioid analgesics, abnormal lower extremity strength, symptom frequency and severity, and changes in symptoms.¹⁴ The first item asks on how many days patients have had to cut down on their normal activities, an item that is applicable even for retired or nonworking people. The second question asks about days lost from usual work or school because of leg or back pain. Although this item is not as responsive to changes occurring after therapy as direct measures of symptoms or functional status, it reflects significant differences in treatment outcomes for surgical *versus* nonsurgical patients in the Maine Lumbar Spine Study.^{1,14} For the working-age population, specific time of absence from work was thought to be a key outcome measure. This question is also a part of the Low Back Pain TyPE questionnaire.

Finally, the parsimonious core set includes an optional question regarding patient satisfaction with care. Although this is not a measure of health outcome, it is an important concern in many types of interventions and in quality improvement applications. As ineffective treatments are used less frequently and as efforts are made to reduce unnecessary imaging, it may be important to measure and maintain patient satisfaction with care.

Proposed Core Instruments for Clinical Researchers [TOP](#)

For many clinical research purposes, greater precision in measurement than is achievable with a single question is desirable. Thus, a somewhat expanded set of core instruments is recommended for investigators who have sufficient resources to collect and analyze such data. It is still a brief set of instruments, designed with the intent that other specific measures could be added at an investigator's discretion. This modestly expanded set of outcome measures is listed in [Table 4](#). Some of the actual instruments are included as [Appendices II-IV](#).

The same dimensions of outcome used in the parsimonious six-item set are preserved. The same measures of pain severity or bothersomeness are recommended as are used in the parsimonious core set. This includes separate measures of back pain severity and leg pain severity. In this expanded set, however, the determination of symptom frequency is also recommended.

With regard to functional status, the use of either the Roland and Morris Disability Scale¹⁵ (or its adaptations¹⁴) or the Oswestry Disability Questionnaire¹¹ (and its adaptations) is recommended. These are among the most widely used and well validated of functional status questionnaires, and both would be highly acceptable. The Roland and Morris Scale, which was derived from the longer Sickness Impact Profile,^{6,15} is well suited to administration by telephone, which may be important when seeking high follow-up rates at low cost. The Oswestry Scale would be tedious at best to administer by telephone. The Roland and Morris Disability Scale may be most useful in primary care settings, or in any situation in which the anticipated level of dysfunction at the end of a trial is small. The Oswestry Disability Questionnaire may be most useful in specialty care settings or in situations in which the disability level is likely to remain relatively high throughout a trial (e.g., chronic severe low back pain).² A modest adaptation of the Roland and Morris Disability Scale has been studied and validated with the intent of producing a more responsive instrument, although head-to-head comparisons are unavailable.¹⁴ The original version is reproduced in [Appendix II](#). This instrument is now available in eight non-English languages (French, Dutch, Flemish, Spanish, Romanian, Italian, Portuguese, Polish, and Czech). These translations are available from Dr. Sandra Sinclair (e-mail address: ssinclair@wh.on.ca) or from the MAPI Research Institute (27 Rue de la Villette, 69003 Lyon, France). In addition, a modest adaptation of the Oswestry Disability Questionnaire has been developed and validated for use in the AOS and NASS outcome questionnaires⁵ ([Appendix III](#)).

With regard to generic well-being, inclusion of either the SF-12 (adapted from the SF-36¹⁶) or the EuroQoL ([Appendix IV](#)) is tentatively recommended.^{10,17} Both are measures of general health that are not disease specific and that provide an assessment of a patient's overall health status. The SF-12 has been incorporated into the Lumbar Cluster of the AOS and is widely used in a variety of settings within the United States. Translations are available in several languages. The EuroQoL instrument has the advantage that it produces a preference-weighted score that is necessary for use in formal decision analysis and cost-effectiveness analysis. That is, it measures a patient's "utility" for various health states in the manner that is theoretically necessary for these forms of analysis. Unfortunately, there is little evidence regarding the responsiveness of either the SF-12 or the EuroQoL instrument, although the larger SF-36 has been found to be responsive.^{3,14} In general, such generic measures have proved to be less responsive than disease-specific questionnaires. In addition, the use of the single item from the parsimonious set that asks "if you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?" is recommended.

Regarding disability from social role, the questions about days of work absenteeism and days in which the patient had to cut down on normal activities are recommended, but days spent in bed for at least half a day

should also be determined. Again, these questions have been adapted from the National Health Interview Survey, validated for use in patients with back problems, and shown to discriminate between surgical and nonsurgical outcomes in the Maine Lumbar Spine Study.^{1,14} Finally, the optional inclusion of a single item on overall patient satisfaction with care is again recommended.

The SF-12 includes questions related to mental health and depression, but the battery proposed here does not include a more extensive or formal measure of depression or mental health. Depending on the goal of a particular study, such measures might be added to this battery, but in many studies of back problems, measures of mental health are not as responsive to treatment effects as measures of physical health.¹

Interpreting Changes or Differences in Outcome Measures [TOP](#)

In some cases there are sufficient data to make recommendations for the smallest clinically relevant change or difference in outcome measures. For example, data from the Maine Lumbar Spine Study show that the minimum change that is of clinical importance in Roland Scale Score is between 2 and 3 points and in the SF-36 Physical Function subscale, it is approximately 7 points.¹⁴ Effect sizes for many of the outcome measures listed here have been published among a cohort of patients with sciatica who demonstrated general improvements in 3 months.¹⁴ These effect sizes could be used for purposes of calculating sample sizes for clinical trials. Nonetheless, there is a general need for more data on the responsiveness and the minimal changes of clinical importance in these outcome measures.

In summary, there are several scientific advantages to the use of a standardized set of measures for most clinical trials of treatment for low back pain. There are wellvalidated measures available that capture the many dimensions of outcome that are important for most patients with back pain, including symptoms, daily functioning, well-being, work disability, and satisfaction. The current group of investigators has prepared a core set of six items for extensive use and a modestly larger set of instruments that could be widely used by investigators with sufficient resources. There can be no illusion that either of these standardized sets will be sufficient for all research studies, and it is anticipated that most investigators will want to add specific and more detailed measures of outcomes that are particularly important for their individual settings or particular forms of treatment. In addition, it is hoped that the battery of questionnaires proposed here will undergo further evaluation, and it is anticipated that it will be continually refined and updated. To the extent that research on back pain incorporates this standardized set, however, it will facilitate comparability among studies, formal pooling of data, and an increase in large, multicenter studies.

References [TOP](#)

1. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, II: One year outcomes of surgical and non-surgical management of sciatica. *Spine* 1996;21:1777-86.
[\[Context Link\]](#)
2. Baker CD, Pynsent PB, Fairbank JCT. The Oswestry Disability Index revisited: Its reliability, repeatability, and validity, and a comparison with the St. Thomas's Disability Index. In: Roland MO, Jenner JR, eds. *Back Pain: New Approaches to Education and Rehabilitation*. Manchester, UK: Manchester University Press, 1989:174-86.
[\[Context Link\]](#)
3. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997;50:79-93.
[\[CrossRef\]](#) [\[Context Link\]](#)
4. Cherkin DC, Deyo RA, Street JH, Barlow W. Predicting poor outcomes for back pain seen in primary care using patient's own criteria. *Spine* 1996;21:2900-7.
[\[Fulltext Link\]](#) [\[CrossRef\]](#) [\[Context Link\]](#)
5. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American Spine Society Lumbar Spine Outcome Assessment Instrument: Reliability and validity tests. *Spine* 1996;21:741-9.
[\[Context Link\]](#)
6. Deyo RA. Comparative validity of the Sickness Impact Profile and Shorter Scales for Functional Assessment in low back pain. *Spine* 1986;11:951-4.
[\[Medline Link\]](#) [\[Context Link\]](#)

7. Deyo RA. Measuring the functional status of patients with low back pain. Arch Phys Med Rehabil 1988;69:1044-53.
[\[Medline Link\]](#) [\[Context Link\]](#)
8. Deyo RA, Andersson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. Spine 1994;19(Suppl):S2032-6.
[\[Context Link\]](#)
9. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. J Chronic Dis 1986;39:897-906.
[\[Medline Link\]](#) [\[CrossRef\]](#) [\[Context Link\]](#)
10. EuroQoL Group. EuroQoL: A new facility for the measurement of health-related quality-of-life. Health Policy 1990;16:199-208.
[\[Medline Link\]](#) [\[CrossRef\]](#) [\[Context Link\]](#)
11. Fairbank JCT, Davies JB, Mbaot JC, O'Brien JT. The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy 1980;66:271-3.
[\[Medline Link\]](#) [\[Context Link\]](#)
12. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987;40:1129-233.
[\[Medline Link\]](#) [\[CrossRef\]](#) [\[Context Link\]](#)
13. Howe J, Frymoyer JW. Effects of questionnaire design on determination of end results in lumbar spine surgery. Spine 1985;10:804-5.
[\[Medline Link\]](#) [\[Context Link\]](#)
14. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health related quality of life in patients with sciatica. Spine 1995;20:1899-909.
[\[Context Link\]](#)
15. Roland M, Morris R. A study of the natural history of back pain. I: Development of a reliable and sensitive measure of disability in low back pain. Spine 1983;8:141-4.
[\[Medline Link\]](#) [\[Context Link\]](#)
16. Ware JE, Sherbourne C. The MOS 36-item short-form survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473-83.
[\[Context Link\]](#)
17. Ware JE, Kosinski, Keller SD. A 12-item short form health survey. Med Care 1996;34:220-33.
[\[Context Link\]](#)

Appendix I: [TOP](#)

Proposed Core Outcome Measures [TOP](#)

[\(Figure\)](#) [\[Context Link\]](#)

Appendix II: [TOP](#)

Roland and Morris Disability Questionnaire (with instructions) [TOP](#)

When your back hurts, you may find it difficult to do some of the things you normally do.

This list contains some sentences that people have used to describe themselves when they have back pain. When you read them, you may find that some stand out because they describe you *today*. As you read the list, think of yourself today. When you read a sentence that describes you today, put a tick against it. If the sentence does not describe you, then leave the space blank and go on to the next one. Remember, only tick the sentence if you are sure that it describes you today.

[\(Figure\)](#) [\[Context Link\]](#)

Appendix III: [TOP](#)

Modified version of the Oswestry Disability Questionnaire used in the AAOS Lumbar Cluster [TOP](#)

[\(Figure\)](#) [\[Context Link\]](#)

Appendix IV: [TOP](#)

The EuroQoL instrument for utility-weighted health status (useful for cost-effectiveness analysis) [TOP](#)

[\(Figure\)](#) [\[Context Link\]](#)

Keywords:

back pain; clinical research; disability; health status; outcomes

Improve comparability of results among clinical studies
Improve comparability of baseline patient characteristics among clinical studies
Facilitate meta-analysis (pooling of results from multiple studies)
Facilitate cost-effectiveness analysis by creating an accepted metric for effectiveness (would also facilitate comparison of cost-effectiveness with treatments for other medical conditions)
Encourage more complete reporting of relevant outcomes
Facilitate conduct of multicenter studies
Facilitate design and review of manuscripts, publications, research proposals
Avoid “reinventing the wheel”

Table 1. Advantages of a Standardized Set of Outcome Measures

Breadth of coverage
 Demonstrated validity and reproducibility
 Demonstrated responsiveness
 Practicality (brevity and low cost)
 Compatibility with widely promoted instruments or batteries (e.g., SF-36, AAOS, NASS)
 Importance to patient
 Importance to society

AAOS = American Academy of Orthopaedic Surgeons; NASS = North American Spine Society; SF-36 = Short Form, 36 items (from Medical Outcomes Study¹⁶).

Table 2. Factors to Consider in Choosing "Standardized" Outcome Measures

Domain	Specific Question	Source, References	Where Used
Pain symptoms	During the past week, how bothersome have the following symptoms been? a) low back pain b) leg pain (sciatica)	From back pain PORT; Patrick et al, ¹⁴ Atlas et al ¹	MLSS AAOS NASS TyPE
or			
Function	Conventional visual analog pain scales During the past week, how much did pain interfere with your normal work (including both work outside the home and housework?	SF-36 and SF-12 ^{16,17}	AAOS MLSS TyPE
Well-being	If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?	From back Pain PORT; Cherkin et al ⁴	AAOS MLSS
Disability	During the past 4 weeks, about how many days did you cut down on the things you usually do for more than half of the day because of back pain or leg pain (sciatica)?	Adapted from Questions in NHIS; Patrick et al ¹⁴	MLSS TyPE
Disability (social role)	During the past 4 weeks, how many days did low back pain or leg pain (sciatica) keep you from going to work or school?		
Satisfaction with care	Over the course of treatment for your low back pain or leg pain (sciatica), how would you rate your overall medical care? (optional)		TyPE

* The exact questions and response options are listed in Appendix I.

PORT = patient outcome research team; MLSS = Maine Lumbar Spine Study; AAOS = American Academy of Orthopaedic Surgeons; NASS = North American Spine Society; NHIS = National Health Interview Survey; TyPE = typology of patient experience; SF = Short Form.

Table 3. A Proposed "Core Set" of Six Questions, Practical for Routine Clinical Use, Quality Improvement, and as a Component of More Formal Research*

Domain	Specific Instrument	Source References
Pain symptoms	Bothersomeness or severity and frequency of low back pain and leg pain (sciatica)	From LBP TyPE, used in AAOS and NASS instruments. Patrick et al, ¹⁴ Atlas et al, ¹ Daltroy et al ⁵
Back-related function	Roland and Morris Disability Scale (or adaptations)	Roland et al, ¹⁵ Fairbank et al, ¹¹ Daltroy et al, ⁵ Patrick et al, ¹⁴ Modified Oswestry used in NASS and AAOS
	or	
	Oswestry Disability questionnaire (or adaptations)	
Generic well-being	SF-12 or EuroQoL; also, "If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?"	From SF-36, ¹⁶ Ware et al, ¹⁷ Patrick et al, ¹⁴ Used in AAOS, EuroQoL Group
Disability (social role)	Days of work absenteeism, cut down activities, bed rest	"Cut down days" from NHIS, Patrick et al ¹⁴
Satisfaction with care	Single question on overall satisfaction (optional)	TyPE

NHIS = National Health Interview Survey; TyPE = typology of patient experience. LBP = low back pain; AAOS = American Academy of Orthopaedic Surgeons; NASS = North American Spine Society; SF-36 = Short Form, 36 items.

Table 4. A Proposed "Core Set" of Instruments for Clinical Researchers