Minimal Clinically Important Difference. Low Back Pain: Outcome Measures

CLAIRE BOMBARDIER, JILL HAYDEN, and DORCAS E. BEATON

ABSTRACT. A proposed standard “core set” of outcome measures for low back pain includes 5 domains: back-specific function, generic health status, pain, work disability, and patient satisfaction. This paper focuses on the 2 recommended back-specific measures of function: the Roland-Morris Disability Questionnaire (RDQ) and the Oswestry Disability Index (ODI). We specifically address their ability to measure change. A systematic review of the literature identified a total of 78 and 71 (RDQ and ODI, respectively) articles as potentially relevant. Detailed tables are provided for each citation, with the type of back pain population studied, the type of change measured, the estimate of change, and the interval over which the change was studied. These tables should be used as a reference for sample size calculation. The responsiveness of the RDQ found in the literature ranges from 2 to 8 points on its 0 to 24 scale depending on what change is being measured. As a rough guide, Roland recommends that a change in 2–3 points on the RDQ should be considered the minimum clinically important change. Choosing any value larger than 5 in designing a clinical trial would risk underpowering the trial, since fewer patients are needed if a trial is designed on the basis of a large change score.

INTRODUCTION
Clinical studies of low back pain (LBP) vary widely in the type of outcomes used to assess the success of therapies. There is no consistency in the range of outcomes considered and even within a given outcome the measures selected vary considerably. For example, there are more than 10 different tools to measure back-specific function, and pain scales differ from one study to another. A standard set of outcome measures would make it easier to compare the magnitude of the treatment effects across studies. It would also encourage complete reporting of relevant outcomes so that investigators do not report on a single domain while ignoring others.

For all these reasons, an international group of investigators met and proposed a core set of 5 domains that should be used in all low back pain studies. These recommendations were published in Spine in 1998. Individual investigators may choose to add other measures in their studies, but the core set is intended to provide a common yardstick across all studies. The core set for clinical research is illustrated in Table 1. This core set has since been revisited for a forthcoming special issue of Spine and the same 5 domains are recommended with only slight differences in the specific instruments proposed for each domain. For example, the full Medical Outcome Survey Short Form-36 (SF-36) is now recommended instead of the SF-12 because of its superior measurement properties, particularly for detecting change in a clinical context.

In this paper, we focus on the 2 back-specific measures of function recommended in the “core-set,” the Roland-Morris Disability Questionnaire (RDQ) and the Oswestry Disability Index (ODI). They are the most commonly used measures of function in back pain research, but how well do they detect patients’ improvement and what is their smallest clinically relevant change? To answer these questions, we review the literature, identify the different approaches used to measure change, and document the magnitude of the change found. Finally, we classify our results using the framework proposed by Beaton elsewhere in this supplement.

METHODS
The Roland-Morris Disability Questionnaire (RDQ) measures 24 activity limitations due to back pain, while the Oswestry Disability Index (ODI) consists of 10 items assessing the level of pain interference with physical activities; the RDQ is mostly a measure of function while the ODI incorporates a measure of pain as well as physical function. In practice, the differences between these instruments are

Key Indexing Terms:
LOW BACK PAIN
OUTCOME ASSESSMENT

RANDOMIZED CLINICAL TRIALS
REPRODUCIBILITY OF RESULTS

Table 1. This core set has since been revisited for a forthcoming special issue of Spine and the same 5 domains are recommended with only slight differences in the specific instruments proposed for each domain. For example, the full Medical Outcome Survey Short Form-36 (SF-36) is now recommended instead of the SF-12 because of its superior measurement properties, particularly for detecting change in a clinical context.

In this paper, we focus on the 2 back-specific measures of function recommended in the “core-set,” the Roland-Morris Disability Questionnaire (RDQ) and the Oswestry Disability Index (ODI). They are the most commonly used measures of function in back pain research, but how well do they detect patients’ improvement and what is their smallest clinically relevant change? To answer these questions, we review the literature, identify the different approaches used to measure change, and document the magnitude of the change found. Finally, we classify our results using the framework proposed by Beaton elsewhere in this supplement.

INTRODUCTION
Clinical studies of low back pain (LBP) vary widely in the type of outcomes used to assess the success of therapies. There is no consistency in the range of outcomes considered and even within a given outcome the measures selected vary considerably. For example, there are more than 10 different tools to measure back-specific function, and pain scales differ from one study to another. A standard set of outcome measures would make it easier to compare the magnitude of the treatment effects across studies. It would also encourage complete reporting of relevant outcomes so that investigators do not report on a single domain while ignoring others.

For all these reasons, an international group of investigators met and proposed a core set of 5 domains that should be used in all low back pain studies. These recommendations were published in Spine in 1998. Individual investigators may choose to add other measures in their studies, but the core set is intended to provide a common yardstick across all studies. The core set for clinical research is illustrated in Table 1. This core set has since been revisited for a forthcoming special issue of Spine and the same 5 domains are recommended with only slight differences in the specific instruments proposed for each domain. For example, the full Medical Outcome Survey Short Form-36 (SF-36) is now recommended instead of the SF-12 because of its superior measurement properties, particularly for detecting change in a clinical context.

In this paper, we focus on the 2 back-specific measures of function recommended in the “core-set,” the Roland-Morris Disability Questionnaire (RDQ) and the Oswestry Disability Index (ODI). They are the most commonly used measures of function in back pain research, but how well do they detect patients’ improvement and what is their smallest clinically relevant change? To answer these questions, we review the literature, identify the different approaches used to measure change, and document the magnitude of the change found. Finally, we classify our results using the framework proposed by Beaton elsewhere in this supplement.

METHODS
The Roland-Morris Disability Questionnaire (RDQ) measures 24 activity limitations due to back pain, while the Oswestry Disability Index (ODI) consists of 10 items assessing the level of pain interference with physical activities; the RDQ is mostly a measure of function while the ODI incorporates a measure of pain as well as physical function. In practice, the differences between these instruments are

Key Indexing Terms:
LOW BACK PAIN
OUTCOME ASSESSMENT

RANDOMIZED CLINICAL TRIALS
REPRODUCIBILITY OF RESULTS

Table 1. This core set has since been revisited for a forthcoming special issue of Spine and the same 5 domains are recommended with only slight differences in the specific instruments proposed for each domain. For example, the full Medical Outcome Survey Short Form-36 (SF-36) is now recommended instead of the SF-12 because of its superior measurement properties, particularly for detecting change in a clinical context.

In this paper, we focus on the 2 back-specific measures of function recommended in the “core-set,” the Roland-Morris Disability Questionnaire (RDQ) and the Oswestry Disability Index (ODI). They are the most commonly used measures of function in back pain research, but how well do they detect patients’ improvement and what is their smallest clinically relevant change? To answer these questions, we review the literature, identify the different approaches used to measure change, and document the magnitude of the change found. Finally, we classify our results using the framework proposed by Beaton elsewhere in this supplement.

METHODS
The Roland-Morris Disability Questionnaire (RDQ) measures 24 activity limitations due to back pain, while the Oswestry Disability Index (ODI) consists of 10 items assessing the level of pain interference with physical activities; the RDQ is mostly a measure of function while the ODI incorporates a measure of pain as well as physical function. In practice, the differences between these instruments are
not great. They are both easy to use, reliable, and valid. Some floor and ceiling effects have been described; the ODI may be better at measuring higher disability levels, while the RDQ may perform better in general populations, where lower disability levels are found.

To assess their ability to measure change we undertook electronic and manual citation searches of the original publications of Roland and Morris (1983)\textsuperscript{2b} and Fairbank, et al (1980)\textsuperscript{3}. The electronic search covered the time period 1993 to March 2000; it identified 205 and 334 articles for the RDQ and ODI, respectively. Manual citation searches for the years prior to 1993, back to the year of publication of the instrument, identified a further 36 and 47 articles resulting in a total of 241 and 381 references for the RDQ and ODI, respectively. These articles were assessed by title and abstract (if available) to determine eligibility. A total of 78 and 71 (RDQ and ODI, respectively) articles were considered potentially relevant to responsiveness and were retrieved for closer examination. (Note that all articles identified from the manual search were retrieved since electronic titles and abstracts were unavailable for screening.)

The detailed results of our literature search are presented in Appendix 1 for the RDQ (3 articles could not be retrieved) and in Appendix 2 for the ODI (retrieval is continuing for this measure). These appendices offer a summary of this literature to date. For each citation, we extracted the type of back pain population studied, the type of change measured, the estimate of change, and the interval over which the change was studied. We classify the type of change measured using the framework proposed by Beaton\textsuperscript{4a}: does the change refer to individuals or groups of individuals, does it describe a difference between individuals or within individuals or both, and what type of change is being studied?: minimum potentially detectable, minimum actually detectable beyond error, change observed in a population, or change observed in those estimated to have an important difference.

**RESULTS**

**Changes in individuals.** Changes in individuals were measured less frequently in the literature than changes in groups. Stratford, \textit{et al}\textsuperscript{4b} estimate that the minimally detectable change on the RDQ is 5 points on the 0–14 point scale, with less confidence of detecting changes at the extremes of the scale (Figure 1, number 1). Dionne, \textit{et al}\textsuperscript{5} estimate that a 10% change is the best discriminator between those who return to work and those who do not, or a change of 2 to 3 points on the scale (Figure 1, number 2). Stratford, \textit{et al}\textsuperscript{6} and Riddle, \textit{et al}\textsuperscript{7} calculate that a 5 point change is an important difference since it is the change found in patients where both the patient and the clinician believe an important change has occurred and treatment goals have been achieved (Figure 1, number 3). The approach most often used for deciding on an important change in individuals is to contrast the distributions of changes in the RDQ and in the ODI in individuals who meet a given criterion with the distribution of changes in individuals who do not meet that criterion. For example, Deyo and Centor\textsuperscript{8} compared the distributions of changes in RDQ scores in those who resumed full activity to those who did not.

### Table 1. A proposed “core set” of instruments for clinical researchers. Reproduced with permission from \textit{Spine}\textsuperscript{1}.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Specific Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain symptoms</td>
<td>Bothersomeness or severity and frequency of low back and leg pain (sciatica)</td>
</tr>
<tr>
<td>Back related function</td>
<td>Roland-Morris Disability Questionnaire (or adaptations)</td>
</tr>
<tr>
<td></td>
<td>Oswestry Disability Index (or adaptations)</td>
</tr>
<tr>
<td>Generic well being</td>
<td>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?”</td>
</tr>
<tr>
<td>Disability (social role)</td>
<td>Days of work absenteeism, cut down activities, bed rest</td>
</tr>
<tr>
<td>Satisfaction with care</td>
<td>Single question on overall satisfaction (optional)</td>
</tr>
</tbody>
</table>

### Table 2. Comparison of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>No. of Items (response options)</th>
<th>Score (best to worse)</th>
<th>Time to complete, min</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roland-Morris</td>
<td>24 (yes/no)</td>
<td>0–24</td>
<td>5</td>
<td>Physical activities, housework, mobility, dressing, getting help, appetite, irritability, pain</td>
</tr>
<tr>
<td>Oswestry</td>
<td>10 (6 levels)</td>
<td>0–100</td>
<td>5</td>
<td>Pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, traveling,</td>
</tr>
</tbody>
</table>

These articles were assessed by title and abstract (if available) to determine eligibility. A total of 78 and 71 (RDQ and ODI, respectively) articles were considered potentially relevant to responsiveness and were retrieved for closer examination. (Note that all articles identified from the manual search were retrieved since electronic titles and abstracts were unavailable for screening.)

The detailed results of our literature search are presented in Appendix 1 for the RDQ (3 articles could not be retrieved) and in Appendix 2 for the ODI (retrieval is continuing for this measure). These appendices offer a summary of this literature to date. For each citation, we extracted the type of back pain population studied, the type of change measured, the estimate of change, and the interval over which the change was studied. We classify the type of change measured using the framework proposed by Beaton\textsuperscript{4a}: does the change refer to individuals or groups of individuals, does it describe a difference between individuals or within individuals or both, and what type of change is being studied?: minimum potentially detectable, minimum actually detectable beyond error, change observed in a population, or change observed in those estimated to have an important difference.

**RESULTS**

**Changes in individuals.** Changes in individuals were measured less frequently in the literature than changes in groups. Stratford, \textit{et al}\textsuperscript{4b} estimate that the minimally detectable change on the RDQ is 5 points on the 0–14 point scale, with less confidence of detecting changes at the extremes of the scale (Figure 1, number 1). Dionne, \textit{et al}\textsuperscript{5} estimate that a 10% change is the best discriminator between those who return to work and those who do not, or a change of 2 to 3 points on the scale (Figure 1, number 2). Stratford, \textit{et al}\textsuperscript{6} and Riddle, \textit{et al}\textsuperscript{7} calculate that a 5 point change is an important difference since it is the change found in patients where both the patient and the clinician believe an important change has occurred and treatment goals have been achieved (Figure 1, number 3). The approach most often used for deciding on an important change in individuals is to contrast the distributions of changes in the RDQ and in the ODI in individuals who meet a given criterion with the distribution of changes in individuals who do not meet that criterion. For example, Deyo and Centor\textsuperscript{8} compared the distributions of changes in RDQ scores in those who resumed full activity to those who did not.
not resume full activity. Beurskens, et al contrasted the RDQ change scores in those who improved using a global assessment of improvement versus those who did not, while Stratford, et al used as a criterion improvements according to both the clinician and the patients’ global rating. All these authors used a well known method called the receiver operand curve (ROC), which results in various cutoff points for changes, each associated with a given sensitivity and specificity.

Changes in groups. There are many studies looking at changes in groups of individuals. The most common ones are randomized trials where the treatment effects for the 2 back-specific measures, the RDQ or the ODI, can be calculated. The treatment effect is the mean change in the treatment group contrasted with the mean change in the comparison group (Figure 2, number 4); more specifically, it is the difference in observed changes between the 2 cohorts.

Appendices 1 and 2 provide examples of treatment effects in randomized trials. Another type of change is illustrated in Figure 2, number 5; it is the change observed in a single cohort of patients. The magnitude of this change will vary depending on the patient population (acute vs chronic back pain), on the length of followup, or on the types of interventions used during the followup period (surgery vs analgesics). Several examples of cohorts are included in the appendices. A third type of change shown in Figure 2, number 6, is the change observed in those estimated to have changed. Patrick, et al found a change of 8.7 for those with improved sciatica, improved quality of life (8.4). 7. Mean change in those patients deemed to have had an important improvement: 7.2–7.6: achieving treatment goals, clinician/patient rating.

DISCUSSION
The most common reason for using patient based outcome measures is to assess patients’ response to treatment. Is the patient better? This paper reviewed the types and magnitude of changes found in 2 measures of back-specific function: the Roland-Morris Disability Questionnaire and the
Oswestry Disability Index. Both are part of the recommended "core set" of measures for use in back pain research. For a comprehensive review of other measurement properties of these 2 scales the reader can refer to a recent review by Roland and Fairbank.

We found that there is no set answer to the question of what is an important change. The magnitude of change varies depending on the approach used to measure change. Some authors label as responsiveness any change greater than the measurement error of the instrument, others use the change observed in a cohort of patients, while others restrict it to the change found in patients who said they improved. It is not surprising, then, that the responsiveness of the RDQ found in the literature will range from 2 to 8 points on its 0 to 24 scale depending on what change is being measured. Since the decision about what is a clinically important change is key for sample size calculations, the reader should carefully apply Beaton’s guidelines and refer to the tables in the appendices of this paper to find a match between the responsiveness reported in the literature and the study you are planning. As a rough guide Roland recommends that a change in 2 to 3 points on the RDQ should be considered the minimum clinically important change. Choosing any value larger than 5 in designing a clinical trial would risk underpowering the trial, since fewer patients are needed if a trial is designed on the basis of a large change score.

Appendix 1. Roland-Morris Disability Questionnaire.

<table>
<thead>
<tr>
<th>Author</th>
<th>Patient Population</th>
<th>Which comparison</th>
<th>Who is focus</th>
<th>Type of change</th>
<th>Estimate of change (24)</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanier &amp; Stockton, 1988(17)</td>
<td>LBP &lt;28 days; presentation to GP</td>
<td>Over time</td>
<td>Group</td>
<td>Observed</td>
<td>Mean: 6-week: 9.1</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Deyo &amp; Cenior, 1986(8) (24 items of RMDQ from SIP)</td>
<td>LBP; presentation to hospital walk-in clinic; n=120</td>
<td>Over time</td>
<td>A. Group B. Individual</td>
<td>Observed: 1. Natural history of LBP 2. Resumed full activities 3. Patient and examiner: improved</td>
<td>A. Mean (sd) 1. 3.0 (5.2) 2. 3.8 (5.3) 3. 4.4 (5.4) B. ROC area (se) 2. 0.72 (.047) 3. 0.67 (.065)</td>
<td>3-week</td>
</tr>
<tr>
<td>Beurskens et al., 1996(9)</td>
<td>Non-specific LBP &gt;6 weeks; n=81; treatment = traction.</td>
<td>Over time</td>
<td>A. Group B. Individual</td>
<td>Estimated: Global perceived effect</td>
<td>A. Mean (sd): 7.8 (3.9) Effect size: 2.02 B. ROC area: 0.93 (change 2.5-5 points)</td>
<td>5 weeks</td>
</tr>
<tr>
<td>Patrick et al., 1995(19) (**Modified 23-item)</td>
<td>Sciatica; presenting to orthopedic and neurologic surgeon offices; surgical vs. non-surgical treatment; n=427</td>
<td>Over time</td>
<td>Group</td>
<td>Observed: 1. Treatment effect Estimated: 2. Improved leg pain (patient) 3. Improved quality of life (patient)</td>
<td>A. Mean: 1. 7.2 2. 8.7 3. 8.4 Effect size: 2.1.6 3. 1.6</td>
<td>3 months</td>
</tr>
<tr>
<td>Stratford et al., 1996(4)</td>
<td>Musculoskeletal LBP; referred to outpatient physiotherapy; n=60; Physiotherapy treatment</td>
<td>Over time</td>
<td>Individual</td>
<td>Minimum detectable change (CSEM)</td>
<td>Mean (95% CI): 5 (4.5) *cannot confidently detect improvement in patients with initial scores &lt;4, or deterioration if &gt;20.</td>
<td>4 to 6 weeks</td>
</tr>
<tr>
<td>Constant et al., 1995(13)</td>
<td>Chronic LBP (&gt;1 year); GP referred; n=62; Spa therapy treatment</td>
<td>Between group, within person</td>
<td>Group</td>
<td>Observed (Treatment, control groups)</td>
<td>Mean (sd) 3 week: Tx - 3.5 (4.4) C - 0.1 (2.7) 6 month: Tx - 5.1 (4.4) C - 0.9 (3.4)</td>
<td>3 weeks, 6 months</td>
</tr>
<tr>
<td>Author</td>
<td>Patient Population</td>
<td>Which comparison</td>
<td>Who is focus</td>
<td>Type of change</td>
<td>Estimate of change (24)</td>
<td>Interval</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>-----------------------------------------</td>
<td>----------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Burton et al., 1999(11)</td>
<td>Non-specific LBP; new episode of acute or recurrent (&lt;3 months); primary care (GP, osteopath) practices; n=162; Educational booklet</td>
<td>Over time</td>
<td>Group</td>
<td>Observed: Treatment effect</td>
<td>No significant differences between groups; NB. Study assessed RR of “important” change in two groups as: a. &gt;4 point improvement in FABphys b. &gt;3 point improvement in RMDQ</td>
<td>2 week, 3 month, 1 year</td>
</tr>
<tr>
<td>Sinclair et al., 1997(20)</td>
<td>Soft tissue injuries inception cohort; WCB claimants; n (LBP)=885; Community clinic program</td>
<td>Between group, within person</td>
<td>Group</td>
<td>Observed: Treatment effect</td>
<td>No significant differences in rate of change between groups</td>
<td>4, 10, 16, 52 weeks</td>
</tr>
<tr>
<td>Leclaire et al., 1996(15)</td>
<td>Work related LBP (&lt;3 months); Private psychiatric outpatient clinic attendants; n=168; physiotherapy treatment or physio + back school</td>
<td>Over time</td>
<td>Groups</td>
<td>Observed: Treatment effect</td>
<td>Difference between groups (95% CI): Post-treatment 6.54 (1.59, 11.49) No significant differences at 6, 12 months</td>
<td>Post-treatment, 6, 12 months</td>
</tr>
<tr>
<td>Cherkin et al., 1998(12)</td>
<td>LBP &gt;7 days following primary care presentation; n=323; physiotherapy, chiropractic, educational booklet</td>
<td>Over time</td>
<td>Group</td>
<td>Observed: a. booklet b. chiropractic c. physiotherapy</td>
<td>Mean: 4 weeks a. 6.8 b. 8.4 c. 8.1 12 weeks a. 7.4 b. 9 c. 8.1</td>
<td>4, 12 weeks</td>
</tr>
<tr>
<td>Constant et al., 1998(14)</td>
<td>Chronic LBP (&gt;1 year); GP referred; n=224; Spa therapy treatment added to usual drug treatment</td>
<td>Between group, within person</td>
<td>Group</td>
<td>Observed: a. drug b. drug + spa</td>
<td>Mean (ad): 3 weeks a. 0.6 (3.4) b. 3.4 (3.4) 3 months a. 1.1 (3.1) b. 4.0 (4.1)</td>
<td>3 weeks, 3 months</td>
</tr>
<tr>
<td>Dionne et al., 1999(5)</td>
<td>LBP; GP visit, HMO enrollees; n=720; standard treatment</td>
<td>Over time</td>
<td>Individual</td>
<td>Estimated: Work status improvement at 2 years (employed vs. unemployed)</td>
<td>ROC analysis: 10% decrease maximum discrimination (sen=84, spec=54)</td>
<td>2 years</td>
</tr>
<tr>
<td>Stratford et al., 1994(10)</td>
<td>Mechanical LBP; referred by GP to physiotherapy; n=88;</td>
<td>Over time</td>
<td>A. Group B. Individual</td>
<td>Observed: 1. Treatment effect 2. Average of clinician and patient global rating, importance of change (&gt;5 on 15 point scale)</td>
<td>Mean (sd): 1. 4.7 (5.0) Correlation with criterion change: 2. 0.60 ROC analysis: 2. ROC area=0.79</td>
<td>4 to 6 weeks</td>
</tr>
<tr>
<td>Author</td>
<td>Patient Population</td>
<td>Which comparison</td>
<td>Who is focus</td>
<td>Type of change</td>
<td>Estimate of change (24)</td>
<td>Interval</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
</tr>
</tbody>
</table>
| Stratford et al., 1998(6) | Mechanical LBP less than 6 weeks duration; referred by GP to physiotherapy; n=226; | Over time        | Individual   | Important: Average of clinician and patient global rating of important change (>5 on 15 point scale) | Mean: 5  
ROC area: 0.84  
Initial scores: 0-8, 5-12, 9-16, 13-20, 17-24  
Estimates of important change: 2, 4, 5, 8, 8, respectively. | 3 to 6 weeks                  |
| Kopec et al.,1995(21)  | LBP patients seeking care from various treatment centers (physio, physiatry, orthopedic, rheumatology, pain clinics); n=242; | Over time        | Group        | Estimated: (Time 2: 2 to 4 days): (self-rated; back pain “better, worse, or about the same") (Time 3: 6 months): (any positive change in ADL on 15 point scale) | Time 1 to 2:  
Effect size = 0.59; N&S sensitivity coeff = 0.20;  
Time 1 to 3:  
Effect size = 1.01; Correlation = 0.47; N&S sensitivity coeff = 0.35 | Time1: baseline  
Time 2: 2 to 4 days (within 2 weeks)  
Time 3: 6 months (within 80 days) |
| Riddle et al., 1998(7) | Mechanical LBP less than 6 weeks duration; referred by GP to physiotherapy; n=143; | Over time        | Individual   | Important: Achievement of treatment goals | Mean: 5  
ROC area: 0.68  
Initial scores: 0-8, 5-12, 9-16, 13-20, 17-24  
Estimates of important change: 3, 5, 8, 11, 13, respectively. | Initial visit to discharge     |

**Appendix 2. Oswestry Disability Index.**

<table>
<thead>
<tr>
<th>Author</th>
<th>Patient Population</th>
<th>Which comparison</th>
<th>Who is focus</th>
<th>Type of change</th>
<th>Estimate of change (%)</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morrison et al., 1988(18)</td>
<td>LBP patients admitted to community hospital treatment program; n=120</td>
<td>Over time</td>
<td>Group</td>
<td>Observed</td>
<td>Mean change 6.0</td>
<td>Post-treatment, 1 year</td>
</tr>
</tbody>
</table>
| Taylor et al., 1999(22) | LBP and sciatica patients; recruited from orthopedic clinic in tertiary care hospital; n=318 | Over time        | Group        | Observed:  
1. Treatment effect  
Estimated (patient’s perceived global rating):  
2. Better  
3. Worse  
4. No change | Median change:  
1. 8.0  
Mean change:  
2. 18.4  
3. -8.6  
4. 2.1  
Effect size (Cohen):  
2. 1.1  
3. 0.4 | physio/ injection - 2 months; surgery - 6, 12, 24 months |
| Beurskens et al., 1996(9) | Non-specific LBP >6 weeks; n=81; treatment = traction. | Over time        | A. Group  
B. Individual | Estimated:  
Patient global perceived effect | A. Mean:  
11.9  
Effect size:  
0.80  
B. ROC area:  
0.76 (change 4-6 points) | 5 weeks |

---

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2001. All rights reserved.
<table>
<thead>
<tr>
<th>Author</th>
<th>Patient Population</th>
<th>Which comparison</th>
<th>Who is focus</th>
<th>Type of change</th>
<th>Estimate of change (%)</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malmivaara et al., 1995(16)</td>
<td>Acute non-specific LBP; employees; n=186; treatment=bed rest, exercises, control</td>
<td>Between group, within person</td>
<td>Group</td>
<td>Observed difference: a. bed rest b. exercises</td>
<td>Mean (95% CI): 3 weeks a. 3.9 (-0.2, 8.0) b. 6.6 (2.0, 11.1) 12 weeks a. 3.8 (0.1, 7.5) b. 2.6 (-1.6, 6.7)</td>
<td>3, 12 weeks</td>
</tr>
<tr>
<td>Leclaire et al., 1996(15)</td>
<td>Work related LBP &lt;3 months; Private; psychiatric outpatient clinic attendants; n=168; physiotherapy treatment or physio + back school</td>
<td>Over time</td>
<td>Groups</td>
<td>Observed: Treatment effect</td>
<td>Difference between groups (95% CI): Post-treatment 4.49 (37, 81.2) 40 significant differences at 6, 12 months</td>
<td>Post-treatment, 6, 12 months</td>
</tr>
<tr>
<td>Stratford et al., 1994(10)</td>
<td>Mechanical LBP; referred by GP to physiotherapy; n=88</td>
<td>Over time</td>
<td>A. Group B. Individual</td>
<td>Observed: 1. Treatment effect Important: 2. Average of clinician and patient global rating, importance of change (&gt;5 on 15 point scale)</td>
<td>Mean (sd): 1. 16.1 (15.9) Correlation with criterion change: 2. 0.57 ROC analysis: 2. ROC area=0.78</td>
<td>4 to 6 weeks</td>
</tr>
<tr>
<td>Kopec et al., 1995(21)</td>
<td>LBP patients seeking care from various treatment centers (physio, physiary, orthopedic, rheumatology, pain clinics); n=142</td>
<td>Over time</td>
<td>Group</td>
<td>Estimated (Time 2: 2 to 4 days): (self-rated: back pain &quot;better, worse, or about the same&quot;) (Time 3: 6 months): (any positive change in ADL on 15 point scale)</td>
<td>Time 1 to 2: Effect size = 0.50; N&amp;S sensitivity coeff = 0.15; Time 1 to 3: Effect size = 0.61; Correlation = 0.35; N&amp;S sensitivity coeff = 0.07</td>
<td>Time1: baseline Time 2: 2 to 4 days (within 2 weeks), Time 3: 6 months (within 80 days)</td>
</tr>
</tbody>
</table>

REFERENCES
12. Cherkin DC, Deyo RA, Battie M, Street J, Barlow W. A comparison...