

# A Roland Morris Disability Questionnaire Target Value to Distinguish between Functional and Dysfunctional States in People with Low Back Pain

Paul W. Stratford, PT, MSc;\* Daniel L. Riddle, PT, PhD, FAPTA†

## ABSTRACT

**Purpose:** To estimate a threshold Roland Morris Disability Questionnaire (RMQ) value that could be used to classify patients with low back pain (LBP) as functional or dysfunctional. **Methods:** In this secondary analysis of data from a study that estimated clinically important RMQ change scores, participants were adults with LBP attending one of three physical therapy clinics. Diagnostic test methodology and a reference standard of goals met were applied to estimate a threshold RMQ value that best distinguished between participants with a functional status and those whose status was dysfunctional. **Results:** Of 143 participants, 104 (73%) met their goals. An RMQ threshold value of 4/24 best distinguished between those who met their goals and those who did not. Sensitivity and specificity for a threshold score of 4 were 94% (95% CI, 88–98) and 69% (95% CI, 52–83), respectively. **Conclusions:** A threshold value of 4 RMQ points provided a reasonably accurate classification of patients. Further research is necessary to cross-validate this estimate and to examine the stability of the estimated value in people with diverse functional demands.

**Key Words:** health status; low back pain; outcome measures.

## RÉSUMÉ

**Objectif :** Estimer une valeur seuil du questionnaire Roland-Morris (QRM) qui pourrait servir à classer les patients qui souffrent de lombalgie dans les catégories de patients « fonctionnels » ou « dysfonctionnels ». **Méthodes :** Dans la présente analyse secondaire de données provenant d'une étude qui faisait l'estimation des cotations de changements importants sur le plan clinique selon le QRM, les participants étaient des adultes atteints de lombalgie, qui fréquentaient l'une de trois cliniques de physiothérapie. On a appliqué la méthodologie de test de diagnostic et une norme de référence des objectifs atteints pour estimer une valeur seuil du QRM qui permettait de distinguer le plus clairement possible les participants qui étaient fonctionnels de ceux qui étaient dysfonctionnels. **Résultats :** Parmi les 143 participants, 104 (73%) ont atteint leurs objectifs. Une valeur seuil du QRM de 4/24 permettait de distinguer le plus clairement possible ceux qui avaient atteint leurs objectifs de ceux qui ne les avaient pas atteints. La sensibilité et la spécificité pour une cotation seuil de 4 étaient de 94% (IC de 95%, 88–98) et de 69% (IC de 95%, 52–83) respectivement. **Conclusions :** Une valeur seuil de 4 points selon le QRM permettait de classer les patients de façon raisonnablement exacte. Il faudra effectuer une recherche approfondie pour faire la contre-validation de cette estimation et pour examiner la stabilité de la valeur estimée chez les personnes ayant diverses demandes fonctionnelles.

Establishing and evaluating measurable goals is a key component of physical therapy practice. Two aspects of a measurable goal are the outcome value and the expected interval for achieving the goal. Measurable goals can be written in terms of change scores or target values. For example, “increase knee flexion by 10° in 2 weeks” is a measurable goal written as a change score, whereas “increase knee flexion to 143° in 4 weeks” is a measurable goal written as a target value. To date, research and application have focused on estimating, reporting, and defining successful outcomes in terms of change scores.

If the threshold change score is met or exceeded, the outcome is considered clinically significant or successful. A limitation of this method is that it is possible for a patient to meet the threshold change score yet still have significant functional limitations. In this article, we estimate the target value for the Roland Morris Disability Questionnaire (RMQ), a measure of functional limitation, in people with low back pain (LBP).<sup>1</sup>

In both clinical practice and clinical research, continuous outcomes are often compressed into dichotomous decisions.<sup>2</sup> Evaluating whether a patient has achieved a

---

From the: \*School of Rehabilitation Science, McMaster University, Hamilton, Ont.; †Departments of Physical Therapy and Orthopaedic Surgery, Virginia Commonwealth University, Richmond, Va.

**Correspondence to:** Paul W. Stratford, School of Rehabilitation Science, Institute for Applied Health Sciences, McMaster University, 1400 Main St. W., Hamilton, ON L8S 1C7; stratfor@mcmaster.ca

**Contributors:** All authors designed the study; collected, analyzed, and interpreted the data; drafted or critically revised the article; and approved the final draft.

**Competing interests:** None declared.

*Physiotherapy Canada* 2016; 68(1);29–35; doi:10.3138/ptc.2014-85

target goal value is a popular approach for doing so. In clinical practice, the decision frequently involves determining whether the target value has been achieved to aid in the decision to discharge a patient from active treatment. In clinical research, outcomes are often labelled *successes* or *failures* to avoid a situation in which, despite a statistically significant between-groups difference in mean scores, few or none of the patients in the intervention group achieve a clinically important improvement. The potential for this disconnect stems from the fact that an important between-groups difference is less than an important within-patient improvement.<sup>3</sup> For the RMQ, a between-groups difference of 2 points is considered clinically important, whereas a within-patient change of 4 or 5 points is recognized as the threshold for a clinically important improvement.<sup>4</sup>

Recognizing that a successful outcome is multifaceted, Jacobson and colleagues<sup>5</sup> proposed that a patient must move from a dysfunctional to a functional state and that the change must be statistically reliable. To address the functional–dysfunctional state criterion, Jacobson and colleagues applied diagnostic test methodology to answer the question “Does the level of functioning at posttest suggest that the participant is statistically more likely to be in the functional population than in the dysfunctional population; that is, is the posttest score statistically more likely to be drawn from the functional than the dysfunctional distribution?”<sup>5(p.340)</sup>

More recently, Tubach and colleagues<sup>6</sup> have introduced a variation on this theme referred to as the *Patient Acceptable Symptom State*, defined as “the value beyond which patients can consider themselves well.”<sup>6(p.34)</sup> In addition to meeting the threshold functional score, Jacobson and colleagues<sup>5</sup> proposed a second standard—the reliability change index (RCI)—to minimize the chance that a patient with a pretest score close to the target value will move from a dysfunctional score to a functional score as a result of measurement error. The RCI is calculated as the difference between pretest and posttest scores divided by the standard error of measurement (SEM) times the square root of 2.<sup>7</sup> It is interpreted as a standard normal deviate (*Z* score). Accordingly,  $SEM \times \sqrt{2}$  multiplied by an RCI of 1.65 (where 1.65 is the *Z* value for the 90% CI) would be analogous to a minimal detectable change at the 90% confidence level.

The RMQ is a commonly used patient-reported outcome measure<sup>1</sup> that assesses pain-related functional status; its measurement properties are consistent with or better than those of competing measures.<sup>8</sup> A body of work has estimated threshold change scores for the RMQ using both reliability-based and diagnostic test methodologies.<sup>9–13</sup> Typical estimates of true and important change for the RMQ are around 5 points.<sup>9,14,15</sup>

A challenge of assessing health concepts such as functional status is the lack of a gold standard or error-free reference standard against which the health concept or

outcome of interest can be directly compared. In such situations, a construct validation process is applied, which involves forming theories about the condition or outcome of interest and then testing the extent to which the measure provides results consistent with the theories.<sup>16</sup> We are aware of only one study that has attempted to estimate a recovery score for the RMQ. Using an 11-point global perceived effect reference standard (−5 = vastly worse, 5 = completely recovered) completed by patients, Kamper and colleagues<sup>17</sup> estimated that an RMQ score of  $\leq 2$  best distinguished between patients who considered themselves completely recovered and all other patients. However, they found little difference in diagnostic accuracy between this cut-score and a cut-score of  $\leq 4$ . Given the limited information available concerning a target RMQ value, our study’s purpose was to estimate a threshold target RMQ value that could be used to classify patients with LBP as functional or dysfunctional. In the context of a construct validation design, our theory was that patients who were meeting their treatment goals would be more likely to be in a functional state than those who were not meeting their goals.

## METHODS

This study was a secondary analysis of data gathered during a previously reported investigation.<sup>14</sup> The institutional review board exempted this study from the requirement to obtain informed consent because the data were obtained as part of routine care at these sites. All data were de-identified before analysis. The purpose of the original study was to estimate the minimal clinically important change for RMQ scores and to determine the extent to which the change estimates were dependent on baseline scores. The present study, however, focused on identifying the final discharge score that best differentiates functional from dysfunctional patients, as categorized by the goals-met reference standard, with the goal of fully exploring a threshold for interpreting discharge RMQ scores.

### Participants

Patients were eligible for inclusion if they were referred by physicians between November 1993 and December 1995 to any of three physical therapy clinics for treatment of LBP. Patients were excluded if they were referred with other conditions in addition to LBP or if, in the therapist’s judgment, they had been diagnosed with other problems that might adversely affect their disability. The study sample was one of convenience and included all patients who fulfilled the eligibility criteria and who provided both admission and discharge RMQ scores.

### Protocol

Patients completed the RMQ immediately before their initial examination and immediately after their final visit; they were blinded to their initial RMQ responses when

completing their discharge RMQ. On the basis of information obtained at the initial assessment, therapists in collaboration with patients developed a treatment plan and set goals. At the final visit, the therapist specified the reason for discharge: (1) patient achieved all treatment goals, (2) loss of insurance coverage, (3) referral to another practitioner, (4) patient did not appear for appointment, and (5) other. Therapists were unaware of the RMQ discharge score when specifying the reason for discharge.

### Goal setting and evaluation

All study sites were part of the same private practice corporation and applied a similar approach to history taking and assessment. Using a collaborative process, patients and therapists jointly set treatment goals. Therapists were required to set at least one goal, and if a patient's employment status was affected by LBP, achievement of the goal was conditional on his or her return to the pre-injury work level. Patients were asked to identify functional activities affected by their LBP; goals were patient specific and included a spectrum of activity levels. Patients indicated to their therapists whether or not they had achieved their goals. All goals had to be met for a patient to be classified as "goals met."

### Roland Morris Disability Questionnaire

The RMQ is a 24-item patient-reported outcome measure that inquires about pain-related disability resulting from LBP.<sup>1</sup> Items are scored 0 if left blank or 1 if endorsed, for a total RMQ score ranging from 0 to 24; higher scores represent higher levels of pain-related disability. Typical RMQ test-retest estimates are in the range of 0.79 to 0.88 points for relative reliability (intra-class correlation) and 1.7 to 2.0 points for absolute reliability (SEM).<sup>18,19</sup> The threshold for important change has been estimated to be approximately 5 RMQ points.<sup>15,19,20</sup>

### Data analysis

We analyzed the data using STATA version 13.1 (StataCorp LP, College Station, TX). Patient characteristics and RMQ scores were summarized as quartiles for continuous variables and as proportions or counts for categorical variables.

We applied receiver operating characteristic (ROC) curve analysis to identify the RMQ score that most accurately classified patients as having met or not met their treatment goals.<sup>21</sup> An ROC curve plots sensitivity (the number of patients correctly identified by the RMQ as having met their goals divided by the number of patients who truly met their goals) against 1 – specificity (number of patients correctly identified by the RMQ as not having met their goals divided by number of patients who truly did not meet their goals). The area under the ROC curve, which can take any value from 0 to 1, quantifies the measure's accuracy: The closer the area is to 1, the

greater the accuracy. An area of 0.50 indicates that the measure does no better than chance at classifying patients as having met their goals or not. We estimated the cut-point that best classified patients as having met or not met their goals as the RMQ score that jointly maximized sensitivity and specificity, then applied sensitivity and specificity values obtained for the cut-point score to estimate the chance that a patient's goals were met by combining the pre-RMQ chance of a patient's achieving his or her goals with sensitivity and specificity information (see Appendix for an example). We use the term *information gain* to denote the difference between the pre-RMQ chance of labeling a patient's outcome as goals met and the post-RMQ chance of doing so.

We provide a vignette illustrating a 50% pre-RMQ chance for patients meeting their goals in the Discussion with elaboration in the Appendix. Previous research has shown that misclassification errors (i.e., labeling patients as having met their goals when in fact they have not, or vice versa) are minimized when the pretest chance of an outcome is 50%.<sup>22</sup> In the context of writing a measurable goal for discharge, the best timeframe for an individual patient to achieve his or her goal would be when 50% of patients with similar characteristics would be expected to achieve their goals. Moreover, typical recovery curves, such as the one provided by Gurcay and colleagues<sup>23</sup> that we refer to in the Discussion section, provide the average change trajectory for patients. Taken literally, this would be the 50th percentile value, which translates into a 50% chance that any patient sharing the characteristics of the sample would achieve the target value within the specified timeframe.

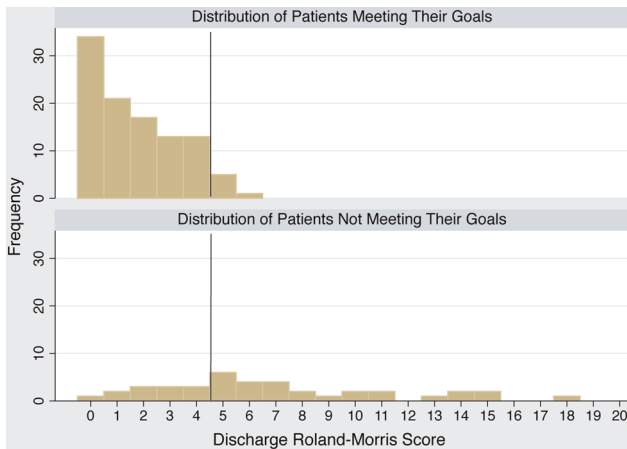
## RESULTS

The original study by Riddle and colleagues<sup>14</sup> described the characteristics of our participant sample in detail. In brief, 143 patients provided admission and discharge RMQ scores. The median duration of treatment was 30 (1st, 3rd quartiles: 19, 150) days, and the sample's median age was 39 (1st, 3rd quartiles: 31, 48) years; 60 patients (42%) were male, and 7 (5%) were receiving workers' compensation. Table 1 and Figure 1 report summary RMQ scores.

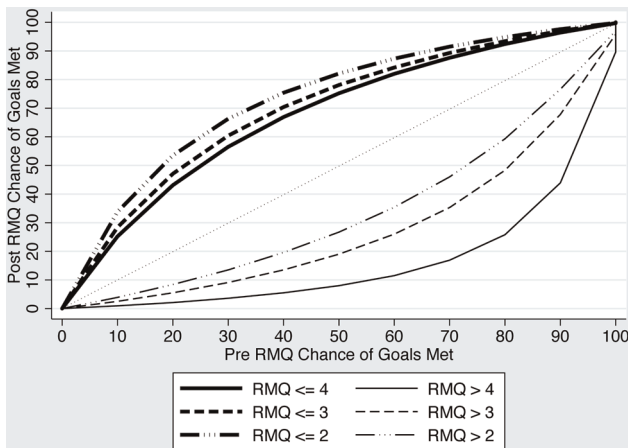
**Table 1** Summary Roland Morris Scores

	Scores at the 1st, 2nd (median), and 3rd quartiles	
	Goals met (n = 104)	Goals not met (n = 39)
Admission	4, 8*, 13	6, 11*, 16
Discharge	0, 1*, 3	4, 6*, 10
Change	3, 6*, 12	1, 3*, 7

\*Median (2nd quartile).



**Figure 1** Distribution of discharge Roland Morris Disability Questionnaire scores by goals-met status.



**Figure 2** Information gain curves for Roland Morris Disability Questionnaire (RMQ) cut-point scores.

Of the 143 patients who took part in this study, 104 (73%) met their treatment goals. The median RMQ discharge score for patients who met their goals was 1 (1st, 3rd quartiles: 0, 3), versus 6 (1st, 3rd quartiles: 4, 10) for those who did not meet their goals. The area under the ROC curve was 0.88 (95% CI, 0.82–0.95), and the RMQ goals-met threshold value was 4 or less. Sensitivity and specificity for the threshold value were 94% (95% CI, 88–98) and 69% (95% CI, 52–83), respectively.

Figure 2 shows information gain when an RMQ threshold score of 4 or less is applied. The diagonal dotted line represents no information gain (i.e., the post-RMQ chance equals the pre-RMQ chance). The solid curve above the diagonal represents the chance that a patient's goals will have been met given an RMQ score  $\leq 4$ ; the solid curve below the diagonal shows the chance of a patient's goals' being met given an RMQ score  $> 4$ . For a patient whose pre-RMQ chance of goals met is 50% and whose RMQ score is  $\leq 4$ , the post-RMQ chance of goals

met is 75%, which represents an information gain of 25%. For a similar patient with a pre-RMQ chance of goals met of 50% and a RMQ score of  $> 4$ , the post-RMQ chance of goals met is 8%, which represents an information gain of 42%. Also shown in Figure 2 for the purpose of comparison are information curves for cut-points of 2 and 3. These curves reveal a modest increase in accuracy when a cut-point of  $\leq 2$  or  $\leq 3$  is applied compared with when a cut-point of  $\leq 4$  is applied. However, a substantial decrease in accuracy is noted when a score of  $> 2$  or  $> 3$  is applied, compared with a score of  $> 4$ .

## DISCUSSION

We applied a goals-met reference standard as the construct for a functional outcome and found that a threshold discharge score of 4 best distinguished patients who achieved their goals from those who did not. Given a 50% chance of achieving a successful outcome before administering the RMQ, a subsequent RMQ score of  $\leq 4$  suggests a 75% chance that the patient has achieved a successful outcome, and a score of  $> 4$  indicates a 92% chance that the patient has not yet done so.

Although several reports have suggested RMQ cut-off scores for success of 2 to 4, only Kamper and colleagues used experimental methodology to arrive at a cut-off value.<sup>17,24,25</sup> Applying diagnostic test methodology and a reference standard of complete recovery, they determined that the diagnostic odds ratio was maximized for a RMQ cut-off score of 2. An important difference between their study and ours is that Kamper and colleagues' reference standard for success was complete recovery. Only 8% of their sample was classified as completely recovered, compared with the 73% deemed to have met their goals in our study. Conceivably, it is possible for patients to meet their goals or to be functional without being completely recovered, and this may translate into a higher RMQ cut-point score. The apparent difference between the two studies' findings may also be explained by sampling variability and the fact that both studies provide estimates of the same population value.

Addressing a somewhat different but related question, Denis and colleagues<sup>26</sup> reported that an RMQ cut-point score of 2 best distinguished between nurses working with LBP and those off work because of LBP. The area under the ROC curve was 0.94, and the associated sensitivity and specificity values for this threshold were 92% and 83%, respectively. In addition to sampling variability and the obvious design difference between our study and that of Denis and colleagues, another possible explanation for the difference in threshold value estimates is that the demands of nursing require a higher level of function (i.e., lower RMQ score) than the average physical demands of the patients in our study.

Our results and those of Kamper and colleagues<sup>17</sup> and Denis and colleagues<sup>26</sup> could be viewed as three examples of construct validation applied to establishing a

functional-state RMQ score. Each study applied a difference reference standard; when no gold standard exists for the outcome of interest, support for a result or outcome score is strengthened by the extent to which different reference standards or research designs yield similar results.<sup>16</sup> Collectively, the three studies suggest that a functional state score falls in the range of 2–4 RMQ points. Moreover, the differences among studies are consistent with the notion that people whose occupations or activities of daily living are more demanding are likely to require lower RMQ scores to achieve a functional state.

Our findings are applicable to both researchers and practitioners. Researchers can use the information from our study in two ways. First, researchers seeking to categorize patients' recovery as success or failure could complement the existing practice of defining success in terms of a change score with a second standard for a target value. This two-pronged approach would have implications for a study sample's eligibility criteria. For example, it is not uncommon for investigators to apply a minimum RMQ value of 4 or 5 points as a criterion for study eligibility.<sup>27–29</sup> We suspect that the rationale for this—not always declared—is that patients must have the potential for a true or clinically important change and that a change of 4 or 5 points is consistent with the minimal detectable change for the RMQ.<sup>18,30</sup> If one considers both the target value for success and the minimal improvement for success, however, a minimum eligibility score of 8 or 9 RMQ points might be more appropriate (i.e., 4-point target value plus a change of 4 or 5 points).

Practitioners could apply our results to assist in framing measurable goals and in approximating the confidence in a clinical decision when the cut-point score is applied. The following clinical vignette illustrates how information from our study could be integrated with existing evidence to develop measurable goals.

#### **Application vignette**

##### ***Patient history.***

Mr. Smith is a 37-year-old man who presents with an RMQ score of 15/24.

##### ***Complementary literature-based evidence.***

True change for the RMQ has been estimated to be 4–5 points.<sup>18,19</sup> Gurcay and colleagues<sup>23</sup> have described the following change profile for patients similar to Mr. Smith (values abstracted from a figure): Initial assessment, 15/24; 1-week follow-up, 9/24; 2-week follow-up, 5/24; 3-week follow-up, 4/24; 4-week follow-up, 2/24; 8-week follow-up, 1/24.

##### ***Measurable goals.***

*Change or progress goal example:* “To decrease Mr. Smith's RMQ score by 5 or more points in 1 week.”

*Target goal example:* “To decrease Mr. Smith's RMQ score to 4 or fewer points in 3 weeks.”

This value could also represent a target for discharge from active treatment. The 3-week interval is based on Gurcay and colleagues' recovery data, which show the average interval for achieving a RMQ score of 4.

Although substantial effort has been devoted to reporting point estimates to guide clinical decisions, little attention has been directed toward determining with how much confidence a practitioner can apply the proposed value. One approach is to examine information gain, by which we mean the extent to which confidence in a clinical decision increases as a result of applying and interpreting the value of a measure. For example, suppose a practitioner scheduled Mr. Smith's reassessment at a point when the typical patient with similar characteristics had a 50% chance of crossing into a functional state. Now suppose that Mr. Smith's previous RMQ score was 9, and today it is 4. Applying the results of our study, we could be approximately 75% certain that Mr. Smith had moved into a functional state. In this example, the information gain is 25% (i.e., from 50% to 75%).

#### **Limitations**

Our study has several limitations. First, we must acknowledge that what constitutes a functional state is likely unique to each person; this study used patient-specific goals, but no overt attempt was made to link these goals to functionality. Second, we did not collect patient-specific goals but only the therapists' dichotomous assessments of whether all goals were met or not met for each patient. Collecting and reporting all patient-specific goals would have better characterized the type of functional activities that were judged to be problematic. This would be particularly informative given the work of Denis and colleagues,<sup>26</sup> which suggests that lower RMQ scores may be required for people performing more demanding activities. Third, we do not know the extent to which diagnostic severity was equally distributed between groups. However, it is evident that patients in the goals-not-met group had higher RMQ scores at admission (see Table 1). A fourth limitation is the relatively small sample size of patients who did not meet their goals.

#### **CONCLUSIONS**

Our goal was to estimate a threshold target RMQ value that could be used by researchers and practitioners to augment the classification of patients as functional or dysfunctional. We found that a threshold value of 4 RMQ points provided a reasonably accurate classification of patients. We believe that further inquiry is necessary to cross-validate our estimate and to examine the stability of the estimated value in patients with diverse demands in terms of activities of daily living. An important aspect of this subsequent research would be to standardize the goal-setting process and classification of success (goals met).

## KEY MESSAGES

### What is already known on this topic

Substantial information exists on estimates of true and important change scores for the Roland Morris Disability Questionnaire (RMQ). These estimates are typically in the range of 4–5 change points. In addition, one study has estimated an RMQ threshold score of 2 points for complete recovery.

### What this study adds

This study applies the framework of Jacobson and colleagues,<sup>5</sup> which considers two criteria for success: first, that a patient's follow-up score is statistically more likely to be drawn from the functional rather than the dysfunctional distribution; second, that the change in a patient's score is reliable. Using a goals-met reference standard, our study suggests that an RMQ cut-off of 4 provides a reasonable value for distinguishing between functional and dysfunctional states.

## REFERENCES

- 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*. 1983;8(2):141–4. <http://dx.doi.org/10.1097/00007632-198303000-00004>. Medline:6222486
- Froud R, Eldridge S, Lall R, et al. Estimating the number needed to treat from continuous outcomes in randomised controlled trials: methodological challenges and worked example using data from the UK Back Pain Exercise and Manipulation (BEAM) trial. *BMC Med Res Methodol*. 2009;9(1):35. <http://dx.doi.org/10.1186/1471-2288-9-35>. Medline:19519911
- Goldsmith CH, Boers M, Bombardier C, et al.; OMERACT Committee. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. *J Rheumatol*. 1993;20(3):561–5. Medline:8478874
- Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine*. 2000;25(24):3115–24. <http://dx.doi.org/10.1097/00007632-200012150-00006>. Medline:11124727
- Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behav Ther*. 1984;15(4):336–52. [http://dx.doi.org/10.1016/S0005-7894\(84\)80002-7](http://dx.doi.org/10.1016/S0005-7894(84)80002-7).
- Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis*. 2005;64(1):34–7. <http://dx.doi.org/10.1136/ard.2004.023028>. Medline:15130902
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12–9. <http://dx.doi.org/10.1037/0022-006X.59.1.12>. Medline:2002127
- Newman AN, Stratford PW, Letts L, et al. A systematic review of head-to-head comparison studies of the Roland-Morris and Oswestry measures' abilities to assess change. *Physiother Can*. 2013;65(2):160–6. <http://dx.doi.org/10.3138/ptc.2012-12>. Medline:24403680
- Stratford PW, Binkley JM, Riddle DL, et al. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther*. 1998;78(11):1186–96. Medline:9806623
- Beurskens AJHM, de Vet HCW, Köke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain*. 1996;65(1):71–6. [http://dx.doi.org/10.1016/0304-3959\(95\)00149-2](http://dx.doi.org/10.1016/0304-3959(95)00149-2). Medline:8826492
- Coelho RA, Siqueira FB, Ferreira PH, et al. Responsiveness of the Brazilian-Portuguese version of the Oswestry Disability Index in subjects with low back pain. *Eur Spine J*. 2008;17(8):1101–6. <http://dx.doi.org/10.1007/s00586-008-0690-1>. Medline:18512083
- Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther*. 2002;82(1):8–24. Medline:11784274
- Grotle M, Brox JI, Vøllestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine*. 2004;29(21):E492–501. <http://dx.doi.org/10.1097/01.brs.0000143664.02702.0b>. Medline:15507789
- Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. *Phys Ther*. 1998;78(11):1197–207. Medline:9806624
- Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine*. 2008;33(1):90–4. <http://dx.doi.org/10.1097/BRS.0b013e31815e3a10>. Medline:18165753
- Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. New York: Oxford University Press; 2008. <http://dx.doi.org/10.1093/acprof:oso/9780199231881.001.0001>.
- Kamper SJ, Maher CG, Herbert RD, et al. How little pain and disability do patients with low back pain have to experience to feel that they have recovered? *Eur Spine J*. 2010;19(9):1495–501. <http://dx.doi.org/10.1007/s00586-010-1366-1>. Medline:20229120
- Ostelo RW, de Vet HC, Knol DL, et al. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol*. 2004;57(3):268–76. <http://dx.doi.org/10.1016/j.jclinepi.2003.09.005>. Medline:15066687
- Stratford PW, Finch E, Solomon P, et al. Using the Roland-Morris Questionnaire to make decisions about individual patients. *Physiother Can*. 1996;48(2):107–10.
- Monticone M, Baiardi P, Ferrari S, et al. Development of the Italian version of the Oswestry Disability Index (ODI-I): a cross-cultural adaptation, reliability, and validity study. *Spine*. 2009;34(19):2090–5. <http://dx.doi.org/10.1097/BRS.0b013e3181aa1e6b>. Medline:19730216
- Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis*. 1986;39(11):897–906. [http://dx.doi.org/10.1016/0021-9681\(86\)90038-X](http://dx.doi.org/10.1016/0021-9681(86)90038-X). Medline:2947907
- Stratford PW. Diagnosing patient change: impact of reassessment interval. *Physiother Can*. 2000;52(3):225–8.
- Gurcay E, Bal A, Eksioğlu E, et al. Acute low back pain: clinical course and prognostic factors. *Disabil Rehabil*. 2009;31(10):840–5. <http://dx.doi.org/10.1080/09638280802355163>. Medline:19736660
- Grotle M, Brox JI, Glomsrød B, et al. Prognostic factors in first-time care seekers due to acute low back pain. *Eur J Pain*. 2007;11(3):290–8. <http://dx.doi.org/10.1016/j.ejpain.2006.03.004>. Medline:16677837
- Jensen TS, Albert HB, Sorensen JS, et al. Magnetic resonance imaging findings as predictors of clinical outcome in patients with sciatica receiving active conservative treatment. *J Manipulative Physiol Ther*. 2007;30(2):98–108. <http://dx.doi.org/10.1016/j.jmpt.2006.12.004>. Medline:17320730
- Denis S, Shannon HS, Wessel J, et al. Association of low back pain, impairment, disability and work limitations in nurses. *J Occup Rehabil*. 2007;17(2):213–26. <http://dx.doi.org/10.1007/s10926-007-9065-4>. Medline:17252204
- Saner J, Kool J, de Bie RA, et al. Movement control exercise versus general exercise to reduce disability in patients with low back pain

- and movement control impairment: a randomised controlled trial. *BMC Musculoskelet Disord.* 2011;12(1):207. <http://dx.doi.org/10.1186/1471-2474-12-207>. Medline:21943318
28. Lehtola V, Luomajoki H, Leinonen V, et al. Efficacy of movement control exercises versus general exercises on recurrent sub-acute nonspecific low back pain in a sub-group of patients with movement control dysfunction: protocol of a randomized controlled trial. *BMC Musculoskelet Disord.* 2012;13(1):55. <http://dx.doi.org/10.1186/1471-2474-13-55>. Medline:22494776
29. Dascanio V, Birks Y, Clark L, et al. Randomized cohort trial was shown to be feasible for evaluating treatments in low back pain. *J Clin Epidemiol.* 2014;67(8):940–6. <http://dx.doi.org/10.1016/j.jclinepi.2014.04.004>. Medline:24836758
30. Stratford PW, Binkley J, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther.* 1996;76(4):359–65, discussion 366–8. Medline:8606899

## APPENDIX

### Sample Calculations for a Patient to Have Met His or Her Goals for a Pre-RMQ Chance of 50%

Step 1. Convert a pre-RMQ chance of 50% to a pre-RMQ odds.

Pre RMQ odds = 50:50 or 1:1

Step 2. Convert sensitivity and specific values of 94% and 69% to likelihood ratios.

Likelihood ratio given a RMQ score  $\leq 4$  = [sensitivity / (1 – specificity)].

Likelihood ratio given a RMQ score  $\leq 4$  =  $[0.94 / (1 - 0.69)] = 3.03$ .

Likelihood ratio given a RMQ score  $> 4$  = [(1 – sensitivity) / specificity].

Likelihood ratio given a RMQ score  $> 4$  =  $[(1 - 0.94) / 0.69] = 0.09$ .

Step 3. Combine likelihood ratios with pre-RMQ chance of meeting goal (50% or 1:1 in this example) to calculate post-test odds of meeting goal.

Given a RMQ score  $\leq 4$ , the post-RMQ odds =  $1 \times 3.03$  or 3.03.

Given a RMQ score  $> 4$ , the post-RMQ odds =  $1 \times 0.09$  or 0.09.

Step 4. Convert the post-RMQ odds to the chance a patient's goals have been met.

Post-RMQ chance = post-RMQ odds / (post-RMQ odds + 1)

Post-RMQ chance of 3.03 =  $3.03 / 4.03 = 0.75$  or 75%

Post-RMQ chance of 0.09 =  $0.09 / 1.09 = 0.08$  or 8%