SPINE Volume 34, Number 19, pp 2060–2065
©2009, Lippincott Williams & Wilkins

Quantification of a Meaningful Change in Low Back Functional Impairment

Sue A. Ferguson, PhD,* William S. Marras, PhD,* Deborah L. Burr, PhD,† Stephen Woods, MD,‡ Ehud Mendel, MD,§ and Purnendu Gupta, MD¶

Study Design. Repeated measures study design.

Objective. Determine a meaningful change in low back functional impairment as measured with the lumbar motion monitor.

Summary of Background Data. A quantitative functional performance probability (*P*(n)) measure has been developed and is scored from 0.00 to 1.00. Previous research has shown that a 0.5 cut-off provides excellent sensitivity and specificity for identifying impaired and healthy low back function. However, a meaningful change in the *P*(n) measure has not been defined.

Methods. The lumbar motion monitor was used to repeatedly measure *P*(n) in 3 groups of subjects including (1) asymptomatic, (2) recovering low back pain (LBP) and, (3) nonrecovering LBP. The asymptomatic group had 20 subjects. The recovering and nonrecovering LBP had 18 and 8 subjects, respectively. The asymptomatic group was tested 5 times at 1-week intervals. The 2 LBP groups were tested every 2 weeks for 3 months (6 evaluations).

Results. The *P*(n) in the asymptomatic group did not significantly change over the observed period. On the basis of the variability in the asymptomatic group it was hypothesized that a meaningful change in *P*(n) was 0.14. The defined meaningful change was evaluated in 2 patient with LBP populations. The *P*(n) in the recovered LBP group significantly improved during the 3 month observation period and there was a corresponding reduction of symptoms. In the recovering LBP group the within subject standard deviation was 0.14 and all patients had at least 1 visit to visit change greater than 0.14. Furthermore, 11 of the 18 recovering patients with LBP had a meaningful change between the first 2 visits. In contrast, none of the nonrecovering LBP group had a meaningful change between the first 2 visits.

Conclusion. A meaningful change in *P*(n) was defined as 0.14.

Key words: low back pain, functional performance, meaningful change. **Spine 2009;34:2060–2065**

Low back disorders have been quantified using pain, disability, and physical impairment measures.[1] However, according to Waddell,[1] physical impairment measures provide the only objective measure of low back pain (LBP). An accurate measure of functional impairment may provide a method to quantify the extent of a low back disorder as well as gauge progress during recovery.[2] A good physical impairment measure should be standardized,[1] reproducible, and have good sensitivity and specificity.[3] In addition, physical impairment measures need to have clinically meaningful change values defined in order to help clinicians interpret the physical impairment measure.

A physical impairment measure has been developed to directly quantify the extent of low back impairment.[2,4–6] The technique used the lumbar motion monitor (LMM), which was worn by the patient or subject, to measure trunk kinematics (range of motion, velocity, and acceleration) in 3 dimensions.[7] Over 700 healthy controls and low back disorder patients have participated in studies to develop the database used to quantify the extent of impairment. The results of the functional evaluation produced a functional performance probability (*P*(n)), a score that can vary from 0.0 to 1.0. The *P*(n) score less than 0.5 indicates the person has impaired functional performance for his or her age and gender whereas a score greater than 0.5 indicates the individuals low back function was healthy for his or her age and gender. *P*(n) has been referred to as probability of normal in earlier articles.[2,4–6] The *P*(n) score has been shown to a have a sensitivity of 90% and specificity of 94%.[6] Thus, the *P*(n) distinguished well between healthy and low back disorder patient function. However, a meaningful change in the *P*(n) has not been defined.

The issue of meaningful change or minimal important change in health scores has been in the published data for decades, not only in low back published issues but other health issues.[8,9] Recently, Ostelo *et al*[10] reviewed the published issues on LBP questionnaire score and stated there were no agreed on scientific grounds for determining a minimal important or meaningful change. Crosby *et al*[11] reviewed different statistical approaches for establishing a minimum important change score including anchor and distribution based approaches. The Crosby *et al*[11] review illustrated how easy it was to become bogged down in the statistical minutia and loose sight of the practical issue of determining a change criterion. Furthermore, a statistical change might not indicate a clinically meaningful change in an outcome measure.[12] Thus,

From the *Department of Integrated Systems Engineering, Biodynamics Laboratory, The Ohio State University, Columbus, OH; †College of Public Health and Health Professions, University of Florida, Gainesville, FL; ‡Department of Physical Medicine and Rehabilitation, The Ohio State University, Columbus, OH; §Department of Neurosurgery, The Ohio State University, Columbus, OH; and ¶Section of Orthopeadic Surgery and Rehabilitation Medicine, University of Chicago, Chicago, IL.

Acknowledgment date: November 16, 2008. Revision date: March 9, 2009. Acceptance date: March 12, 2009.

The device(s)/drug(s) that is/are the subject of this manuscript is/are exempt from FDA or corresponding national regulations.

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Sue A. Ferguson, PhD, Department of Integrated Systems Engineering, Biodynamics Laboratory, The Ohio State University, 210 Baker Systems, 1971 Neil Avenue, Columbus, OH 43210; E-mail: ferguson.4@osu.edu

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Table 1 Characteristics of Each Sample Population

	Asymptomatic	Nonrecovering	Recovering
Sample size	20	8	18
Age	22.6 (5.3)	44.6 (12.6)	31.7 (9.8)
% Males	50%	62%	66%
No. subject with previous history of low back pain	NA	4	12
Time since last episode	—	2.9 yr	1.97 yr
No. patients filing worker's compensation	—	2	2
No. patients with litigation	—	0	1
Initial McGill pain questionnaire score	—	2.87 (1.36)	2.22 (0.94)
Final McGill pain questionnaire score	—	1.56 (1.50)	0 (0)

the goal of this study was to identify a meaningful change score that can be used to identify meaningful functional improvement in future studies.

Three objectives were developed to achieve the goal of this article. The first objective was to determine the reproducibility of the $P(n)$ score in a healthy population. Second, determine a meaningful change in $P(n)$ based on the expected variability of the healthy population. The third objective was to evaluate the definition of meaningful change in $P(n)$ in a patient population. These objectives should establish a clinically meaningful change in $P(n)$ that can be used in future studies as well as by clinicians to define improvement or degradation in low back functional performance.

■ Materials and Methods

Two experiments were conducted to achieve the objectives of the study. First, an asymptomatic repeatability study was conducted. The participant's low back function was tested 5 times at 1 week intervals. The inclusion criterion for the asymptomatic study was no history of LBP. Second, study was conducted on a patient with LBP. Acute patients with LBP were tested every 2 weeks for 3 months for a total of 6 visits. Acute LBP was defined as an on set of symptoms within the past 3 months. Inclusion criterion for patients with LBP was local LBP symptoms. Exclusion criterion was radicular pain symptoms below the knee without motor or sensory loss. Patients were retrospectively placed into 2 groups either recovering or nonrecovering at the end of the study.

Subjects

In the asymptomatic study 20 subjects participated, 10 males and 10 females. In the LBP study there were 26 patients. The inclusion criterion for the recovering group in the study was that patients be recovered on $P(n)$ (>0.5) and symptoms. The inclusion criterion for the nonrecovering group was that patients have $P(n)$ less than 0.5 and have symptoms at 3 months. Eight patients were retrospectively placed in the nonrecovering group and 18 were placed in the recovered. The exclusion criterion for patients was recovery of only 1 outcome measure. All participants signed a university consent form before participation. Descriptive characteristics for the patient and asymptomatic study participants have been listed in Table 1.

Equipment

The LMM was used to evaluate trunk kinematics. The LMM measured position, velocity, and acceleration in all 3 planes of the body and has been previously validated.[7] The LMM was attached to the person using orthoplast. A laptop computer was used for data collection and storage as well as the display of the twisting position.

Experimental Design

Both the asymptomatic study and patient with LBP study were repeated measure designs. The asymptomatic controls were tested 5 times at 1-week intervals and at the same time of day. The patients with LBP were tested every 2 weeks for 3 months for a total of 6 evaluations.

Self-Report LBP. Self-report of LBP status was measured using the McGill Pain Questionnaire.[13]

Functional Performance. The functional performance protocol required the participant to control their twisting position while they flexed and extended their trunk as fast as they could comfortably.[2,4–6,14] The controlled twisting positions were 0°, 15° clockwise and counterclockwise and 30° clockwise and counterclockwise. The order of clockwise and counterclockwise was counterbalanced by subject.

Procedure

All patients were screened at the first visit by an orthopedic surgeon (P.G.). Once approved for the study, the participants completed the functional performance evaluation and questionnaire. This procedure was repeated at each visit. In the case of asymptomatic group no questionnaire was completed. The LMM testing procedure required less than 30 minutes.

Data Analysis

Questionnaire. The McGill Pain Questionnaire present pain intensity score was scored to measure symptoms on a 0 to 5 scale.[13]

Functional Performance Probability. The kinematic measures from the LMM were calculated with previously reported validated techniques.[7] The output included range of motion, velocity, and acceleration for all 3 planes. The data were normalized by age and gender. The $P(n)$ calculation used 6 variables including ability to perform tasks, twisting range of motion, sagittal range of motion at zero, sagittal extension velocity at zero, sagittal extension acceleration at zero, and lateral range of motion at zero. The model generates a probability of the functional performance being in the asymptomatic group.

Statistical Analysis

Means and standard deviation were calculated across visits using SAS version 9.1. Further analysis of variance using the SAS proc mixed with a randomized blocks design was performed to examine significance of trial as well as quantify the within and between subject variance. Individual changes scores were evaluated for each subject by group.

■ Results

The asymptomatic population had an average $P(n)$ above 0.9 at every visit as shown in Figure 1A. The average score ranged from 0.913 to 0.944 and the standard deviation range from 0.053 to 0.099. The average of all

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

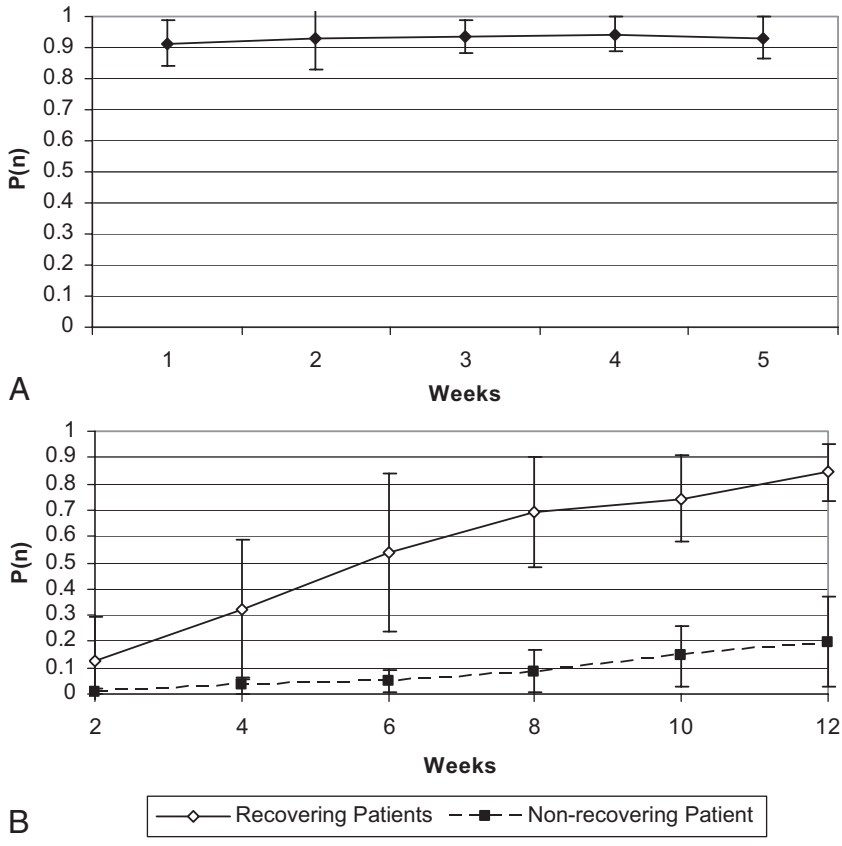

Figure 1. **A**, *P*(n) in the asymptomatic population across time. **B**, *P*(n) in patients with low back pain across time.

5 standard deviations was 0.07. Doubling the standard deviation of 0.07 to 0.14 would incorporate 95% of all the data. It is hypothesized that a change of 0.14 or more in the *P*(n) represents a clinically meaningful change and would correctly identify 95% of cases as a meaningful change.

The analysis of variance using randomized blocks with a fixed effect for trial and random subject effect indicated no significant difference across trials for the asymptomatic group (*P* value, 0.3797). The within and between subject variance was 0.0024 and 0.0027, respectively. The within subject standard deviation was approximately 0.05. The residual plots from the model exhibited a random distribution about zero characteristic of a normal distribution. Thus, the asymptomatic subjects were very stable in their performance from week to week.

The baseline average *P*(n) in the recovered patient group was 0.125 with a standard deviation of 0.17, as shown in Figure 1B. The recovered patient group had a final *P*(n) of 0.844 with a standard deviation of 0.11. The average change in *P*(n) between trials was 0.14.

The analysis of variance using randomized blocks with a fixed effect for recovered patients showed a significant trial effect with a *P* value of 0.0001 indicating statistical meaningful improvement between observation points. The within subject variance was 0.0197 and the between subject variance was 0.0255. The within subject standard deviation was 0.14 for recovered patients.

The nonrecovering LBP group had a baseline average *P*(n) of 0.010 with a standard deviation of 0.009 and a final average *P*(n) score of 0.199 with a standard deviation of 0.174, as illustrated in Figure 1B. The *P*(n) scores in the nonrecovering LBP group did show some improvement however, it does not reach the 0.5 necessary to be considered recovered. Also, the self reported symptoms listed in Table 1 indicated that these patients were experiencing mild symptoms after 3 months.

The randomized blocks analysis of variance also showed significant improvement between trials in the nonrecovering LBP group. The within subject variance was 0.0064 and the between subject variance was 0.0024. The within subject standard deviation was approximately 0.08.

Figures 2A-C examined the individual change scores instead of group means. In the asymptomatic group 90% of the population had no change score greater than 0.14. In the nonrecovering group 5 of 8 or 62% had no change scores greater than 0.14. Predominantly, these meaningful changes occurred between the last 2 visits. The recovered patients on the other hand all had at least one change score greater than 0.14 and 50% of the group had 2 change scores greater than 0.14.

Table 2 examined the number of meaningful changes from visit to visit by group as well as the effect sizes from visit to visit by group. The most striking difference in the table was between the first 2 visits for the recovering and nonrecovering LBP groups. The recovering group had 11

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

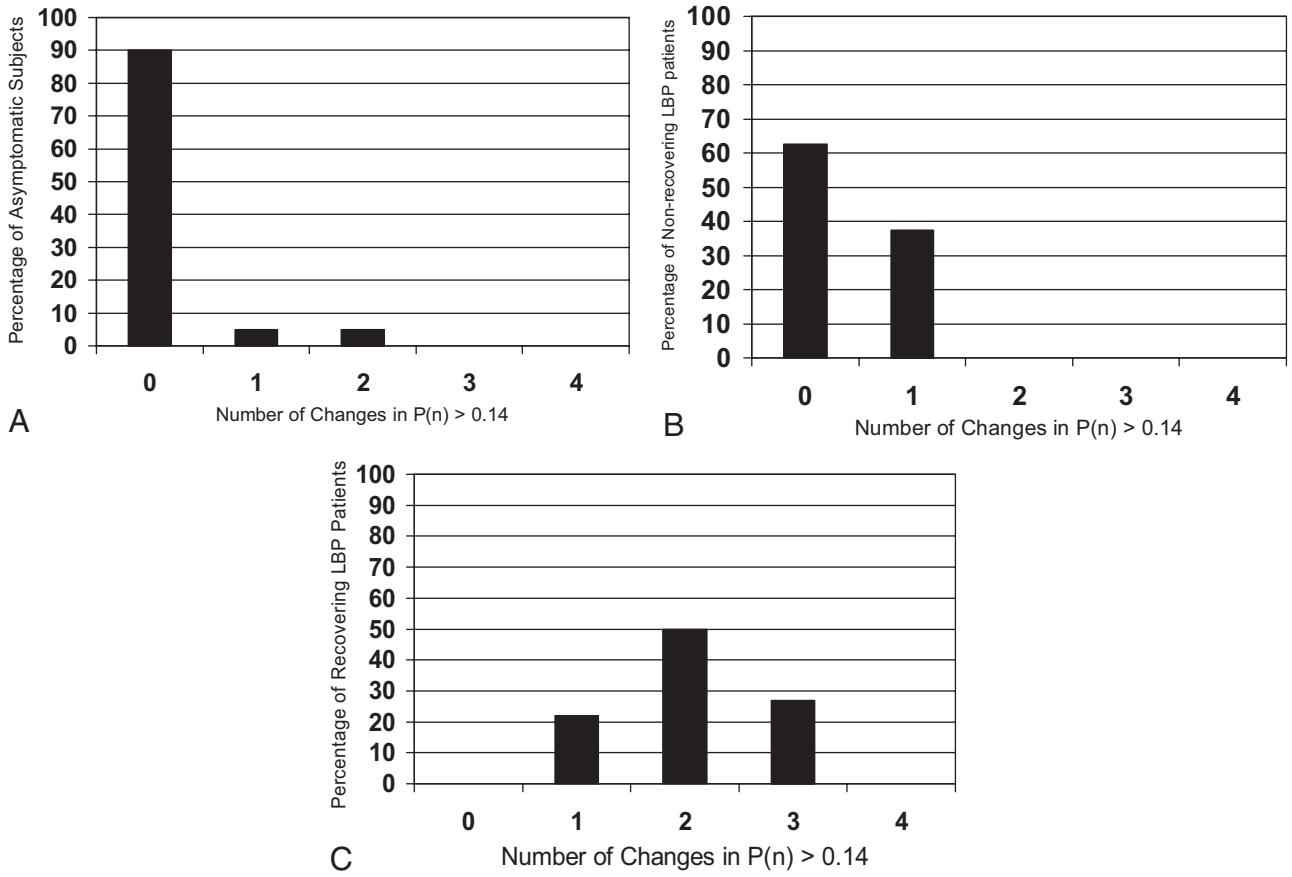

Figure 2. **A**, percentage of asymptomatic subjects by the number of changes in *P*(n) greater than 0.14. **B**, percentage of nonrecovering LBP patients by the number of changes in *P*(n) greater than 0.14. **C**, percentage of recovered LBP patients by the number of changes in *P*(n) greater than 0.14.

of the 18 subjects with a meaningful change whereas the nonrecovering group had none of the 8 subjects with a meaningful change. The nonrecovering group had all the meaningful changes between visit 4 and 6 whereas the recovering group had 23 of the 37 meaningful changes occur in the first 3 visits. Table 2 showed that in the asymptomatic group 2 subjects had a meaningful change between visits 1 and 2 and 1 subject had a meaningful change between visits 2 and 3. Even with these individual cases the effect sizes between visit 1 and 2 as well as visits 2 and 3 were quite small 0.02 and 0.01, respectively. In all cases the *P*(n) score was greater than 0.5 for the asymptomatic group.

■ Discussion

The *P*(n) measure did not significantly change across visits in the healthy population. This indicates that the *P*(n) measure was reproducible in asymptomatic subjects. Thus, the first objective of the study to show that the *P*(n) measure was reproducible has been accomplished. Furthermore, *P*(n) meets all the criteria of a good physical measure,[1,3] it has been standardized, reproducible, has high sensitivity and specificity.[6]

The second objective of the study was to define a meaningful change in *P*(n). The meaningful change was set at 0.14 based on doubling the average standard deviation in the asymptomatic population. This technique of averaging the standard deviation across subjects and trial combines the within and between subject variation. Doubling the average standard deviation in the asymptomatic population includes 95% of the population based on a normal distribution. There have been no agreed on methods for determining a meaningful change.[10] The strength of defining a meaningful change

Table 2 Number of Meaningful Changes Between Each Visit by Group

Population	Visit 1–2	Visit 2–3	Visit 3–4	Visit 4–5	Visit 5–6
No. subjects with change >0.14 in the asymptomatic group (N = 20)	2	1	0	0	NA
No. subjects with change >0.14 in the nonrecovering group (N = 8)	0	0	0	1	2
No. subjects with change >0.14 in the recovering group (N = 18)	11	12	7	3	4
Effect size in asymptomatic group	0.02	0.01	0.01	0.01	NA
Effect size in nonrecovering group	0.02	0.02	0.04	0.06	0.05
Effect size for recovering group	0.20	0.22	0.15	0.05	0.10

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

in $P(n)$ at 0.14 is that the risk of falsely indicating a meaningful change when none has decreased. The disadvantage of defining a meaningful change at 0.14 is that the chance of Type II error (false negative) might be increased, however, our analysis of the patient populations indicated this was not likely.

The defined meaningful change of 0.14 in $P(n)$ was evaluated with 2 patient populations. First, in a group of recovering patients with recovery defined retrospectively, patients showed significant $P(n)$ improvement over time. The average $P(n)$ change between visits in this group was 0.14. The within subject standard deviation between visits was 0.14. Examining the visit to visit changes illustrated in Figure 2B showed that all of the recovering patients had at least 1 visit to visit change greater than the defined meaningful change of 0.14. Clinically, these patients all had local LBP at their initial visit and all were symptom free at their final evaluation. This shows that the hypothesized meaningful change in functional performance probability corresponds to a reduction in LBP symptoms.

Second, in the nonrecovering LBP group there was a significant improvement in $P(n)$ during the 3 months of recovery. However, the average change score between visits was 0.038, nearly 4 times smaller than the proposed meaningful change of 0.14. On an individual basis 5 of the 8 patients had no individual change score greater than 0.14, and the remaining 3 patients had only 1 change score greater than 0.14, threshold of meaningful change. Furthermore, the timing of these meaningful changes occurred after the fourth evaluation as shown in Table 2. Clinically the symptoms in this group measured with the McGill Pain Questionnaire improved from 2.87 to 1.56 indicating the group had mild to discomforting pain symptoms at the end of the 3 months of evaluations. Thus, those patients with continued pain symptoms at the end of 3 months tend not to have meaningful change in $P(n)$.

Beaton[15] discussed factors that affect the amount of change quantified in the responsiveness of a measure. Included in the factors listed by Beaton[15] was the patient population studied. This study illustrated the difference in within subject variation from the 3 different population's asymptomatic, recovering LBP, and nonrecovering LBP. Furthermore, this change in the within subject variability of the population may influence the definition of a meaningful change. The 0.14 definition of a meaningful change is equal to the within subject standard deviation in the recovered patients with LBP. The within subject standard deviation in asymptomatic group and nonrecovering LBP group were both smaller than the meaningful change definition. The recovering LBP group also had clinical improvement of being symptom free. Thus, it has been hypothesized that the defined 0.14 meaningful change might be conservative but truly represents a meaningful clinical change in low back function.

This was the first study to examine a meaningful change in an objective quantitative functional performance measure. There have been several studies examining meaningful change in LBP questionnaire scores. Ostelo *et al*[10] reviewed the published data on change scores using several low back questionnaires including the Oswestry Disability Index, Roland Morris Disability Questionnaire, and Quebec Pain Disability Questionnaire. The results of the review showed that questionnaires with scores ranging from 0 to 100 had a meaningful absolute change score between 10 to 20 points. The $P(n)$ score in the current study also has a range of 100 points only on a difference scale from 0.00 to 1.00. Thus, the range of $P(n)$ scores was similar to the questionnaires and the defined meaningful change score of 0.14 is within of absolute change scores from the questionnaires reviewed by Ostelo *et al*.[10]

Ostelo *et al*[10] has pointed out that there are not agreed on scientific grounds or empirical methods for determining a meaningful change in a health outcome. Several researcher have indicated that it is easy to get caught up in the specifics of the statistical methods for defining a minimal important change.[8,9,11,12] The goal of the study was to define a meaningful $P(n)$ change that is indicative of a significant change in function. Future studies are needed to confirm the findings of this study which defines a meaningful change in $P(n)$ at 0.14. In addition, functional performance probability already has a cut-off of 0.5 below which performance is impaired and above which performance is considered healthy. The meaningful change score defined in the study may be used alone or in conjunction with 0.5 anchor based cut-off. If we can identify individuals with deteriorating functional performance from 0.85 to 0.70 it might be possible to prevent a costly low back injury, while an individuals low back function does not fall into the impaired category.

Figure 1B illustrated that the recovered patients with LBP on average had a functional performance probability of 0.84 indicating there was some residual loss in performance. Ferguson *et al*[14] found that based on functional performance probability in a group of acute patients with LBP only 68% of the population was recovered. Schiottz-Christensen *et al*[16] in a 1-year prospective study found that 2% of acute patients with LBP were still on sick leave at a 1-year follow-up, however, 45% of patients continued to have LBP symptoms. Henschke *et al*[17] found that 72% of participants had fully recovered after 12 months. Measuring the functional performance probability over time would provide a direct objective measurement of functional impairment throughout the recovery process. It would also quantify the degree of degradation from a relapse or exacerbation.

Limitation

A limitation of the study was the modest sample sizes in the 3 groups. In addition, the starting point of impairment was not considered in defining these functional changes only the differences. None the less, we think that

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

these functional findings provided compelling evidence for defining a meaningful change in functional status.

■ Key Points

- An objective quantitative low back functional performance (*P*(n)) measure has been developed.
- Based on the variability in an asymptomatic group, a meaningful change in *P*(n) has been defined as 0.14.
- In a group of recovering patients with LBP, all patients had at least 1 visit to visit change greater than 0.14.

References

1. Waddell G. *The Back Pain Revolution*. Edinburgh, United Kingdom: Churchill Livingstone; 1998.
2. Marras WS, Parnianpour M, Ferguson SA, et al. The classification of anatomic- and symptom-based low back disorders using motion measure models. *Spine* 1995;20:2531–46.
3. Andersson GBJ. Sensitivity, specificity, and predictive value a general issue in the screening for disease and in the interpretation of diagnostic studies in spinal disorders. In: Frymoyer J, ed. *The Adult Spine: Principles and Practice*. New York, NY: Raven Press; 1991:277–87.
4. Marras WS, Parnianpour M, Ferguson SA, et al. Quantification and classification of low back disorders based on trunk motion. *Eur J Phys Med Rehabil* 1993;3:218–35.
5. Marras WS, Parnianpour M, Kim J, et al. A normal database of dynamic trunk motion during repetitive trunk flexion and extension as function of task asymmetry, age and gender. *IEEE* 1994;2:137–46.
6. Marras WS, Ferguson SA, Gupta P, et al. The quantification of low back disorder using motion measures methodology and validation. *Spine* 1999; 24:2091–100.
7. Marras WS, Fathallah F, Miller R, et al. Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics. *Int J Ind Ergon* 1992;9:75–87.
8. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996;49:1215–9.
9. Wells GA, Tugwell P, Kraag GR, et al. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557–60.
10. Ostelo R, Deyo R, 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;33:90–4.
11. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health related quality of life. *J Clin Epidemiol* 2003;56:395–407.
12. Wright JG. The minimal important difference: who's to say what is important? *J Clin Epidemiol* 1996;49:1221–2.
13. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975;1:277–99.
14. Ferguson SA, Marras WS, Gupta P. Longitudinal quantitative measures of the natural course of low back pain recovery. *Spine* 2000;25:1950–6.
15. Beaton D. Understanding the relevance of measured change through studies of responsiveness. *Spine* 2000;25:3192–9.
16. Schiottz-Crhistensen B, Nielsen GL, Hansen VK, et al. Long-term prognosis of acute low back pain in patients seen in general practice: a 1-year prospective follow-up study. *Fam Pract* 1999;16:223–32.
17. Henschke N, Maher CG, Refshauge KM, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ* 2008;337:a171.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.