Muscle reflex classification of low-back pain

N.P. Reeves a,b, J. Cholewicki b,*, T.E. Milner a

a Department of Kinesiology, Neuromuscular Control Laboratory, Simon Fraser University, Burnaby, BC, Canada V5A 1S6
b Department of Orthopaedics and Rehabilitation, Biomechanics Research Laboratory, Yale University School of Medicine, P.O. Box 208071, New Haven, CT 06520-8071, USA

Received 20 January 2004; received in revised form 29 June 2004; accepted 1 July 2004

Abstract

It has been well documented that low-back pain (LBP) patients have longer muscle response latencies to perturbation than healthy controls. These muscle responses appear to be reflexive and not voluntary in nature, and as a result, might be useful for objectively classifying LBP. The goal of the study was to develop an objective and accurate method for classifying LBP using a sudden load-release protocol. Subjects were divided into two groups: learning group (20 patients and 20 controls), and holdout group (15 patients and 12 controls). Subjects exerted isometric trunk force against a cable in four different directions. Following cable release, the trunk was suddenly displaced eliciting a muscle reflex response. Reflex latencies for muscles switching-on and shutting-off were determined using electromyogram signals from 8 trunk muscles. Independent t tests were performed on the learning group to determine which reflex parameters were to be entered into logistic regression analysis to produce a classification model. The holdout group was used to validate this classification model. The three-parameter model was able to correctly classify 83% of the learning group, and 81% of the holdout group. Using reflex parameters appears to be an accurate and objective method for classifying LBP.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Classification; Reflex; Sudden load-release; Electromyography; Low-back pain

1. Introduction

Low-back pain (LBP) is a common condition affecting a large percentage of the population. It is estimated that between 70% and 85% of the population will experience LBP at some point in their lives [3,4,13,35]. The majority of these cases resolve without medical intervention within the first 6 weeks [35]; however, the minority of cases that progress to become chronic bear a significant cost burden [1,10,33]. Not surprisingly, LBP is one of the most prevalent and costly health problems in Western Society [2], since it comprises 25% of all workplace injuries [38] and accounts for approximately 40% of compensation costs [33].

To address the impact of LBP, companies have established disability management programs to reintegrate workers following an injury in a safe and timely fashion, thus reducing the likelihood of a case becoming a chronic condition. One obstacle in effective disability management is the mistrust between employers and workers. This may be particularly true of back injuries where the cause may be unknown, or where the injury event was trivial (i.e., picking-up a light object). On one side, there are employers and insurance groups who sometimes question the legitimacy of a worker’s injury, and on the other side, there are injured workers who feel pressure from employers and insurance groups to return to work before they are ready. If an objective method was developed that could discriminate those...
with and without LBP, and determine when those with LBP are work-ready, disability decision-making would be less arbitrary and more case-specific.

Starting in the late 1980s, a number of investigators have used spectral electromyography (EMG) to assess fatigue in the paraspinal muscles in order to classify LBP [9,14,19,26,30–32]. Typically, mean or median frequency is used as a fatigue marker during sustained isometric contraction at a specified effort level. It has been claimed that these tests are objective due to the fact that it is difficult to control voluntarily the action potential signal registered with EMG [26]. However, with these protocols, subjects are required to perform maximum voluntary efforts (MVEs) to set the resistance level during the test, and because of inaccuracy in predicting MVE, the power of these tests is diminished and may lead to erroneous results. It could be argued that during MVE tests, motivational factors and/or fear of re-injury can have a significant effect, which in turn could alter spectral parameters during the fatigue trials. In addition, fatigue-based tests require significant exertions both during the estimation of MVE and during the actual test, which can range from 40% to 80% of MVE. This also brings in to question whether performing these tests may be safe for someone who may have diminished tolerance for spinal loading. And finally, a number of researchers cast doubt about the validity of EMG spectral methods applied to the erector spinae muscle group [7,15]. Consequently, an objective and safe method for classifying LBP still remains elusive.

It has been well documented in the literature that the response latencies of trunk muscles are delayed in LBP patients [11,18,27,28,37]. Radebold et al. [27] used a quick load-release protocol to measure response latencies for trunk muscles and found muscles typically responded within 100 ms following the load-release. This quick response is indicative of a reflex rather than a voluntary response, which suggests that this measure may serve as the basis for an objective classification tool. Although differences have been observed between LBP patients and healthy controls, the use of reflex parameters has not been used to classify LBP. There is some evidence to suggest that with rehabilitation reflex responses can improve [37]. If this is the case, then by monitoring performance in a reflex-based test, timelines for return to work could be established, and the likelihood of durable return to work improved.

In a preliminary study using 15 subjects (eight LBP patients and seven healthy controls), we compared fatigue and reflex parameters to determine which were better for classifying people with and without LBP. From this study, no significant differences ($p < .05$) were found for fatigue parameters (although the initial mean frequency was close to significant at $p = .07$), while a number of significant parameters were found for reflex responses between groups. We concluded that the fatigue parameters were more sensitive to inaccuracies in predicting MVE. Interestingly, LBP patients showed no sign of fatigue indicating that their MVEs may have been under-estimated. Also, these preliminary data showed that reflex parameters were better at classifying LBP than fatigue parameters: classification rates were higher and fewer scores fell near the cut-off point in the reflex-based test than the fatigue tests. In addition, the subjects reported that the reflex-based test was less physically stressful than the fatigue-based test.

Based on these results, a large study was undertaken with a larger population to develop a classification model based on reflex parameters from sudden unloading. A large holdout group was used to cross-validate the model.

2. Methods and materials

2.1. Approach

The current experiment used a sudden load-release method in four directions of isometric trunk exertions to study the muscle response patterns in patients with LBP and healthy controls. EMG signals from eight trunk muscles were recorded for 1 s before and 2 s after the load-release. Based on these signals, the reflex latencies for muscles switching-on and shutting-off following sudden load-release were determined. Data was collected both at Yale and Simon Fraser Universities using a similar experimental apparatus, and combined for the purposes of this study. Before testing, all subjects read and signed a consent form which outlined the testing protocol. The protocol was approved by the institutional review boards at the respective universities.

2.2. Subjects

Sixty-seven people in total participated in the study of which 35 were LBP patients and 32 were healthy controls. Healthy control subjects were defined as persons who had never experienced back pain lasting longer than 3 consecutive days. LBP patients were defined as persons past the acute phase of their condition with no neurological deficits and no structural deformities, genetic spinal disorders or previous spinal surgery. Patients had experienced LBP for periods ranging from 6 months to 35 years, with pain intensity that varied from mild to severe with some pain-free intervals. Patient’s pain was centralized to the back region, and patient’s with radicular pain into the leg were excluded from the study. The consumption of analgesics, mostly nonsteroidal anti-inflammatory, varied from daily use to medication as needed. Most patients were undergoing rehabilitation, typically involving some form of super-
vised exercise therapy. Testing was conducted prior to treatment sessions. The Roland-Morris Disability Questionnaire showed, on average, low scores (5.14 out of 24, SD = 4.32). Eight of the LBP patients were receiving compensation benefits at the time of testing. Compensation patients were screened for non-organic signs of LBP[36], and patients were omitted from the study if they received a positive score on any of the tests. The purpose of this study was to identify individuals with physiologically based impairment. For this reason, we used a clean subject pool free of psychological or non-organic issues. All patients were screened by a healthcare professional before testing to assure that other inclusion criteria were met.

The subject pool was divided into two groups with 40 being assigned to a learning group (Table 1) used to develop the classification model, and 27 to a holdout group (Table 2) to test the accuracy of the classification model. For the learning group, healthy controls were matched as closely as possible for gender, age, weight, and height.

2.3. Task

Subjects were placed in a semi-seated position in the testing apparatus (Fig. 1) for exerting isometric contraction in trunk flexion, extension and lateral bending to the left and right. This apparatus was designed to restrain hip motion while allowing the upper body to move freely in any direction. Consequently, response patterns of trunk muscles were responsible for postural readjustments. Isometric force was applied through a cable attached to a chest harness at approximately T9 and was held with an electromagnet. The release of this electromagnet produced sudden unloading resulting in displacement of the trunk, and initiating reactive responses in the trunk muscles.

Each subject performed three trials at a predetermined force level of 65 N for males and 40 N for females. These force magnitudes were established in a preliminary study, and are approximately 20% of the maximal isometric exertion averaged for flexion, extension, and lateral bending. An oscilloscope was used to provide visual feedback to the subject. The force was displayed as a line which moved up and down while the target was displayed as a stationary line. Subjects were asked to keep the force output line steady at the target and were asked not to anticipate the release. The time of release was randomly varied after the target force level was reached.

2.4. Data collection and processing

Eight channels of EMG were recorded using bipolar, Ag–AgCl, disposable electrodes with a centre-to-centre
spacing of 30 mm (at Yale), and bipolar, stainless steel electrodes with an active circuit and centre-to-centre spacing of 13 mm (at SFU). The electrodes were positioned over the following muscles on each side of the body: rectus abdominis (3 cm lateral to the umbilicus), external oblique (approximately 15 cm lateral to the umbilicus), thoracic erector spinae (5 cm lateral to T9 spinous process), and lumbar erector spinae (3 cm lateral to L5 spinous process). All EMG signals were band-pass filtered between 20 and 250 Hz (at Yale) and 20–500 Hz (at SFU), differentially amplified, and then A/D converted at a sample rate of 1600 Hz (at Yale) and 2000 Hz (at SFU). Initially, 12 channels of EMG were collected. The four additional channels recorded the activity of latissimus dorsi, and internal oblique muscles bilaterally for the Yale subject pool. These muscles were later removed from the Yale data and were not measured for the SFU subject pool. Due to electrode position for these muscles, there were some concerns with artefact entering the signal from contact with the pads and harness. Analysis of the Yale data indicated that the latissimus dorsi muscle group played a minor role in postural adjustments following load-release as indicated by lower EMG activity. Discriminant analysis on the Yale data indicated that the internal oblique muscle group activity was not significantly different between healthy controls and patients ($p = .158$).

To facilitate the detection of switching-on and shutting-off of muscles from EMG signals (Fig. 2), a computer algorithm was developed. When rectified mean EMG increased by 1.5 standard deviations, muscle switching-on was detected. When rectified mean EMG decreased by 1.5 standard deviations, muscle shutting-off was detected. Time between this detection and a force drop represented the reflex latency. A decrease of 50% (at Yale) and 5% (at SFU) in mean force prior to the release was used as the threshold. At SFU, the increased strength of the electromagnet resulted in a slower rate of force decline; consequently, the force threshold appeared to be consistent between the Yale and SFU data sets. Statistical analysis revealed that there were no significant differences between mean reflex latencies at Yale and SFU ($p = .793$).

To clarify, agonistic muscles were defined as muscles that were active before the load-release and were expected to shut-off. Antagonistic muscles were quiet before the load-release and were expected to switch-on. In flexion, flexors acted as the agonists (FOFF) group and extensors as the antagonists (FON) group. In extension, extensors acted as the agonists (EOFF) group and flexors as the antagonists (EON) group. In lateral bending to the left, ipsilateral muscles (left side) acted as the agonists (LOFF) group and contralateral muscles (right side) acted as the antagonists (LON) group, and vice versa for lateral bending to the right (ROFF and RON, respectively).

**2.5. Data analysis**

The average reflex latency of each muscle for the three trials was determined. These reflex latencies were grouped and averaged into switching-on and shutting-off times for each of the four directions. EON, FON, LON, and RON correspond to the average reflex latency for muscles switching-on in the four respective directions. EOFF, FOFF, LOFF, and ROFF correspond to the average reflex latency for muscles shutting-off in the four respective directions.

An independent $t$ test was used to determine if reflex latencies were statistically different ($p < .05$) between patients and healthy controls. Parameters reaching significance were then entered into logistic regression analysis to determine which parameters should be included in the classification model. From the regression analysis, a prediction equation was formulated and applied to the holdout group to validate the model. To test the robustness of classification, four new randomly assigned groups were developed and the accuracy for classification for these groups was determined. The average score of the five groups was used to establish classification accuracy.
3. Results

In general, patients tended to have longer reflex latencies than healthy controls and showed more variability as indicated by mean reflex latencies and standard deviations between groups (see Table 3).

Results from the independent *t* tests identified five parameters that were significantly different (*p* < .05) between the patients and controls (Table 3). All of these parameters were entered into binary logistic regression analysis using a forward stepwise method. This analysis yielded the following prediction equation incorporating three parameters:

\[ y = \exp(29 - 247 \times FON - 149 \times EOFF - 34 \times FOFF), \]

\[ \text{probability} = \frac{y}{y + 1}. \]

Re-entering the learning group into the classification model resulted in 86.7% of patients and 84.2% of healthy controls being correctly classified with the overall model accuracy of 85.3% (six missing data points). For the holdout group, the model accurately classified 81.8% of patients and 76.9% of healthy controls with the overall model accuracy being 79.2% (five missing data points). Missing cases are the subjects for whom the algorithm could not identify reflex latencies for one or more of the parameters used for classification. Consequently, no classification score could be obtained for those individuals. As depicted by the classification plots, the majority of the cases scored at either end of the range making the model more categorically distinct (Figs. 3 and 4).

The four randomly assigned groups used to develop new classification models produced similar results for classification accuracy. Averaging the scores from all five groups resulted in 83% of the learning group and 81% of the holdout group being correctly classified (Table 4).

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Shuttering-off (ms)</th>
<th>Switching-on (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agonist</td>
<td>Antagonist</td>
</tr>
<tr>
<td>Patients</td>
<td>Healthy controls</td>
<td>Patients</td>
</tr>
<tr>
<td>Healthy controls</td>
<td></td>
<td>Healthy controls</td>
</tr>
<tr>
<td>Extension (E)</td>
<td>65 (18)</td>
<td>87 (28)</td>
</tr>
<tr>
<td></td>
<td><em>p</em> &lt; .001</td>
<td><em>p</em> = .094</td>
</tr>
<tr>
<td>Flexion (F)</td>
<td>99 (67)</td>
<td>82 (15)</td>
</tr>
<tr>
<td></td>
<td><em>p</em> = .028</td>
<td><em>p</em> &lt; .001</td>
</tr>
<tr>
<td>Lateral bending left (L)</td>
<td>81 (38)</td>
<td>87 (15)</td>
</tr>
<tr>
<td></td>
<td><em>p</em> = .019</td>
<td><em>p</em> = .237</td>
</tr>
<tr>
<td>Lateral bending right (R)</td>
<td>62 (41)</td>
<td>91 (32)</td>
</tr>
<tr>
<td></td>
<td><em>p</em> = .125</td>
<td><em>p</em> = .044</td>
</tr>
</tbody>
</table>

Fig. 3. Classification plot for learning group set #1. “P” refers to patients and “C” refers to healthy controls.

![Classification plot for learning group set #1](image)

Fig. 4. Classification plot for holdout group set #1. “P” refers to patients and “C” refers to healthy controls.

![Classification plot for holdout group set #1](image)

4. Discussion

The primary goal of the study was to determine if muscle reflex parameters could accurately classify LBP. Clear group differences were observed in reflex responses to sudden load-release: patients showed longer latencies and more variability than healthy controls. This was particularly true for parameters dominated by responses of the erector spine muscle group (i.e., FON and EOFF, Table 3). Therefore, using these reflex parameters appears to be an accurate method for classifying LBP. From a safety perspective, only low-level exertions (20% MVE) are required for this type of test which is substantially lower than that of fatigue-based protocols. No subjects experienced any discomfort during testing or had their injury aggravated suggesting that the test is safe. In addition, this classification protocol is not as susceptible to the inaccuracies of predicting effort level than fatigue-based protocols.
The unloading reflex is a common phenomenon found in everyday activities. During sudden unloading, agonist muscles experience a brief period of inhibition, while concurrently, antagonist muscles activation levels are increased. This activation pattern minimizes the perturbation by altering the net joint torque. The neural pathways for control of trunk agonist and antagonist muscles are not clearly understood. A number of receptors may provide feedback to respond to an unloading perturbation. Muscle spindle density is high in rotatores, which is a small muscle spanning one or two segments of the spine [6]. It is believed that this muscle acts as a kinesthetic sensor to monitor trunk position [24]. Human supraspinal and interspinal ligaments were found to be well innervated with Ruffini corpuscles receptors that are believed to be sensitive to both fast movement and static motion depending on their firing threshold [12]. Within the facet joints of the spine, mechanoreceptors were identified that are believed to provide feedback regarding joint position, tension, and pressure [20]. Also, in the intervertebral disc, receptors resembling golgi tendon organs were found in the deeper layers of the annulus fibrosus, and pacinian-like receptors were identified in the superficial annulus layers [21] suggesting that the discs can act as more than just a shock absorption system [29]. Likely, convergent afferent input from various receptors contributes to the divergent reflex response observed during unloading.

It is possible that trauma to the spine might damage receptors or nerve fibers which innervate them and may produce “dead zones” in sensory feedback or at least alter their sensitivity. This is supported by a number of studies that have documented decreased trunk proprioception [8,22,25], increased postural sway, and decreased postural control with LBP [16,17,23,28]. In terms of the load-release protocol, such dead zones may be responsible for the delays in reactive responses. Possibly greater than normal displacement is then required before a reflex response is elicited.

The reflex latencies found in the present study matched closely with those found in earlier work [18,27,28,37], and is indicative of reflex responses and not voluntary movement. Onset times are slightly longer than those reported by Tani et al. [34] and Zedka et al. [39], and may be explained in part by the experimental protocol. These two experiments used a tapping protocol to elicit a stretch reflex, whereas in the present study, a trunk perturbation was used to initiate reflex responses. The inertial mass of the trunk may have resulted in a slight delay in muscle stretching which in turn would result in longer latencies for muscle response. During testing, subjects with less upper body mass tended to be displaced more than heavier subjects as result of inertial differences. If the sensory receptors mediating the reflex responses are affected by length and velocity changes in muscle, this difference in inertial mass may result in a decrease in model accuracy. Perhaps a model that normalizes effort to acceleration following release or upper body weight instead of setting effort to a predetermined force level may produce better classification.

To test if the protocol was sensitive to changes in trunk stiffness that could be modulated volitionally, reflex latencies were compared between coactivation and no coactivation conditions in preliminary tests on control subjects. From the results of this study, it appears that the erector muscle group is not affected significantly by coactivation, and in turn suggests that the model is not significantly affected by changes in trunk stiffness.

There is considerable redundancy in force generators in the spine [5] which allows for a variety of recruitment patterns to achieve the same kinematic output. For this reason, the average muscle response to the load-release protocol was used for classification. The accuracy of the model may have been improved if smaller groups of muscles or individual muscle responses were used for classification, but this improvement in accuracy would come at the expense of general validity of the model. The high prediction accuracy in the holdout group suggests that the model presented in this paper should be equally accurate when applied to the general population. However, with this said, the authors feel that the model accuracy could be improved if customized to specific populations. Proportionately more males than females were misclassified (seven out nine were males) suggesting the protocol may be gender sensitive. Unfortunately due to the small study population, it was not possible to generate two gender specific classification models. In addition, both

Table 4
Classification scores for four randomly assigned groups

<table>
<thead>
<tr>
<th>Equations</th>
<th>Learning group accuracy</th>
<th>Holdout group accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( y = \exp(19 - 184 \times \text{FON} - 78 \times \text{EOFF} - 23 \times \text{FOFF}) )</td>
<td>80% (5 missing)</td>
<td>83% (4 missing)</td>
</tr>
<tr>
<td>2 ( y = \exp(24 - 229 \times \text{FON} - 113 \times \text{EOFF} - 14 \times \text{FOFF}) )</td>
<td>78% (3 missing)</td>
<td>81% (6 missing)</td>
</tr>
<tr>
<td>3 ( y = \exp(25 - 242 \times \text{FON} - 107 \times \text{EOFF} - 30 \times \text{FOFF}) )</td>
<td>83% (4 missing)</td>
<td>77% (5 missing)</td>
</tr>
<tr>
<td>4 ( y = \exp(26 - 264 \times \text{FON} - 92 \times \text{EOFF} - 29 \times \text{FOFF}) )</td>
<td>89% (5 missing)</td>
<td>83% (4 missing)</td>
</tr>
<tr>
<td>Average</td>
<td>83%</td>
<td>81%</td>
</tr>
</tbody>
</table>
false positive and false negative misclassified subject tended to be older than correctly classified subject (45 years versus 36.5 years, respectively) suggesting that the classification is age sensitive. Group means indicate that there were no significant differences in height and weight of those misclassified from those who were not.

It should be mentioned that this type of classification was intended for individuals with mechanical LBP and not for individuals with non-organic issues. For this reason, a clean subject pool, free of non-organic signs, was used in the development of the classification model. Classification indicating LBP suggests that there is a physiological impairment in terms of delayed reflex responses. It has been shown that rehabilitation efforts can correct this impairment [37]. Presumably, reflex parameters should not be affected by non-organic signs, so subjects with non-organic signs would not be classified as having physiological symptoms and could, therefore, be prescribed different treatment, perhaps more psychologically based.

Finally, it should be mentioned that the algorithm for detecting muscle switching-on and shutting-off sometimes required manual correction. A number of different protocols were tested to improve accuracy as well as reduce the time spent processing data. However, because of signal artefacts, it is difficult to completely automate this process. Consequently, there is some subjectivity in determining whether a change in the signal was a result of a myoelectric event or from an external source. This is particularly true for the abdominal muscles which tend to have lower signal-to-noise ratio due to increased subcutaneous fat. However, signals generated from the erector spinae muscle group appeared to be free of artefacts making the process of determining reflex latencies more objective for parameters derived from these muscle groups.

Acknowledgments

This work was made possible by research grants from the National Institute of Arthritis and Musculoskeletal and Skin Disease (NIH Grant 5R01 AR46844), the Research Secretariat at the Workers’ Compensation Board of British Columbia, and the Whitaker Foundation.

References


N. Peter Reeves received a B.Sc. (Kinesiology) from the University of Waterloo (1994), and M.Sc (Kinesiology) from Simon Fraser University (2002). Currently, he is a Ph.D. student (Biomedical Engineering) at Yale University. He is interested in understanding how the back works and becomes injured in order to develop effective interventions. Prior to starting his Master’s, Peter worked for a number of years as an ergonomic consultant developing and managing industrial injury prevention and management programs.

Jacek Cholewicki is a spinal biomechanist and Associate Professor of Orthopaedics and Rehabilitation at Yale University School of Medicine in New Haven, Connecticut. He holds cross-appointments to the departments of Mechanical Engineering and Biomedical Engineering where he teaches a graduate and undergraduate course in biomechanics. He received his Bachelor, Masters, and Ph.D. degrees in Kinesiology from the University of Waterloo in Canada in 1986, 1990, and 1994, respectively. He has been at Yale since 1993, initially as a Postdoctoral Research Associate. His prior research addressed the issues of lumbar and cervical spine function, spine injury mechanisms, tissue loading, and biomechanical modeling using both in vivo and in vitro experimental approaches. His current research interests include motor control of spine stability and the effectiveness of abdominal belts.

Theodore Milner received his B.Sc. (Honors) in Physics and his M.Sc. and Ph.D. in Physiology from the University of Alberta. From 1984-86 he was a post-doctoral fellow in Brain and Cognitive Science at MIT. He accepted a faculty position in the Institute of Biomedical Engineering at the Université de Montréal in 1986. In 1992, he moved to the School of Kinesiology at Simon Fraser University where he is currently a full professor. His research interests include adaptive control of limb mechanics, human robotics, electromyography and neurophysiology of human movement.