BMJ Open

Validity of increasing the number of motor control tests within a test battery for discrimination of low back pain conditions in people attending a physiotherapy clinic: a case–control study

Christoph Biele,[1] Dirk Möller,[1] Harry von Piekartz,[1] Toby Hall,[2] Nikolaus Ballenberger[1]

¹Social Sciences, Osnabrück University of Applied Sciences, Osnabrueck, Germany
²School of Physiotherapy and Exercise Science, Curtin University, Perth, Western Australia, Australia

Correspondence to
Dr Nikolaus Ballenberger;
n.ballenberger@hs-osnabrueck.de

ABSTRACT

Objectives To develop a time-efficient motor control (MC) test battery while maximising diagnostic accuracy of both a two-level and three-level classification system for patients with non-specific low back pain (LBP).

Design Case–control study.

Setting Four private physiotherapy practices in northern Germany.

Participants Consecutive males and females presenting to a physiotherapy clinic with non-specific LBP (n=65) were compared with 66 healthy-matched controls.

Primary outcome measures Accuracy (sensitivity, specificity, Youden index, positive/negative likelihood ratio, area under the curve (AUC)) of a clinically driven consensus-based test battery including the ideal number of test items as well as threshold values and most accurate items.

Results For both the two and three-level categorisation system, the ideal number of test items was 10. With increasing number of failed tests, the probability of having LBP increases. The overall discrimination potential for the two-level categorisation system of the test is good (AUC=0.85) with an optimal cut-off of three failed tests. The overall discrimination potential of the three-level categorisation system is fair (volume under the surface=0.52). The optimal cut-off for the 10-item test battery for categorisation into none, mild/moderate and severe MC impairment is three and six failed tests, respectively.

Conclusion A 10-item test battery is recommended for both the two-level (impairment or not) and three-level (none, mild, moderate/severe) categorisation of patients with non-specific LBP.

BACKGROUND

Low back pain (LBP) is a major health problem. The real-world prevalence and incidence of LBP ranges in routinely collected data from 1.4% to 20.0% and 0.02% to 7.0%.[1] In 2013, the overall costs of LBP in the USA

Strengths and limitations of this study

► The sample was large and comprised consecutively recruited subjects with non-specific low back pain (LBP) attending private physiotherapy clinics at multiple sites, indicating high generalisability.

► Subjects with LBP at all stages (acute, subacute or chronic) were matched for age and gender with healthy asymptomatic people.

► A clinically centred motor control test battery to identify impairments in control in people with LBP was developed by consensus from a panel of clinical experts.

► Statistical procedures were applied to determine reliability and diagnostic accuracy in categorising motor control ability at two levels (deficient or not) and three levels (not, mildly/moderately, and severely deficient).

► A limitation was the unequal distribution of the study sample characteristics in subgroups, with limited numbers presenting with subacute pain.

were estimated to be US$119–238 billion.[2] LBP can be labelled as specific or non-specific based on imaging, as well as acute, subacute or chronic based on the time with pain. However, based on these groupings, no conclusions can be drawn with respect to appropriate physical therapy interventions. In particular in the acute stage, when LBP has been present for 6 weeks therapy regimens have not been shown to be effective in aiding recovery. According to Fritz _et al_ the outcome in a group of patients with acute LBP did not improve significantly when physical therapy was applied.[3] One explanation for the failure of current management of LBP is the focus on biomedical approaches, which is associated

with an exponential increase in healthcare costs with a concurrent increase in disability and chronicity.[4]

It has also been suggested that understanding heterogeneity in LBP may help develop more effective management strategies[4] through individualised patient-centred care. Altered lumbopelvic motor control (MC) may be one subgroup or component of LBP,[4] [5] potentially explaining heterogeneity which is suggested to manifest in 70% of people with chronic non-specific LBP.[6] Identification of MC dysfunction is determined by motor control tests (MCT). Integration of such tests in treatment strategies seems to be promising.[4–9] Altogether, approximately 30 MCTs have been described in six different positions.

Despite the use of MCT in clinical practice, there is no standardised regimen to apply these tests and classify patients with chronic LBP (CLBP), and there is no consensus on how to re-educate impairments identified. In addition, the psychometric properties of selected single MCTs diverge substantially. For example, the inter-rater reliability of the 'Bent knee fall out test' ranges from 0.38 to 0.95.[10–12]

Test batteries of MC vary in the number and nature of the included tests and the psychometric properties of different test batteries diverge substantially.[5] [11] [12] For example, one approach applies up to 18 tests,[13] while another only six.[9] Hence, it is unclear how many and which tests are needed to identify dysfunction of MC accurately and time efficiently at the same time. Additionally, the accuracy of MCT batteries has only been tested in people with or without LBP. Even though it has been shown in a three-group comparison that patients with CLBP demonstrated more failed tests, the accuracy and validity of tests in more than two groups (eg, different levels of impairment, severity and chronicity) has not been evaluated in detail.[14]

The aim of this study was to identify an MCT battery that is, on the one hand, time efficient and, on the other hand, obtains a maximum of diagnostic accuracy in discrimination between subjects with and without non-specific LBP. In this context, specific research questions were to (1) assess the optimal number of MCT items including thresholds for categorisation, (2) identify the combination of single tests which serve best for categorisation, and (3) evaluate the appropriateness of a three-level categorisation (none, mild, moderate/severe MC impairment).

METHODS

Study design

This trial was designed as a case–control study with cases being consecutively recruited.

Patient and public involvement

Neither patients nor the public were involved in the design or planning of the study.

Study sample

Subjects were recruited from four different private physiotherapy practices in Germany. Consecutive patients referred to the physiotherapy clinics with a doctor's diagnosis of non-specific LBP were assessed and consequently enrolled if they met eligibility criteria. Participants were subdivided into acute (<6 weeks), subacute (6–12 weeks) and chronic (>12 weeks) pain. This procedure continued until the desired sample size was achieved. Subjects were required to provide a written consent prior to their participation. Healthy controls were volunteers who did not have any back pain within the previous 3 months prior to the testing and were comparable in age and gender. These subjects came from the same institutions as symptomatic subjects, however they were not receiving treatment for LBP. Participants were randomly allocated to each rater and were required to complete an 11-item MCT battery. A sample size of 66 subjects were required in each group. This was based on an area under the curve (AUC) of 0.64 (medium effect size), with type I error of 5% and type II error of 20%. Additionally, a 5% dropout rate was taken into account.

A medical doctor made the diagnosis of non-specific LBP according to a differential diagnostic flow chart.[15] Subjects were also required to suffer LBP within the last 3 months, to be at least 18 years of age and to have sufficient understanding of German language. Healthy control subjects had to be at least 18 years of age and to have sufficient understanding of German language.

MCT battery

A clinically centred MCT battery was developed by a consensus of clinical experts working with patients with LBP who had more than 15 years of clinical experience and were qualified as International Federation of Orthopaedic Manipulative Physical Therapists (IFOMPT) orthopaedic manual therapy practitioners. Requirements for the composition of the MCT battery were to have tests that challenge through movement of both the upper and lower extremities MC of the lumbopelvic region, in different positions and through open and closed kinetic chain. Furthermore, no more than 11 tests were allowed, as the MCT battery is aimed at clinical use where time efficiency is an important issue. The 11 tests are listed in table 1.

The subjects were tested in individual treatment rooms, performing the whole MCT battery in one session. All raters were blinded to any patient data including the presence of LBP. The tests were demonstrated and explained to the patients by the rater. If a patient needed more than three attempts to perform the test movement correctly, the test was considered as failed. Each single test was rated as fail (−) or pass (+). The total number of failed test results was summed as a summary score.

Nine raters were educated for 4 hours on the execution and evaluation of the MCTs by a physiotherapist who is a clinical expert in this field. Finally, raters were tested for agreement on 2 February 2017 in a pilot study of four subjects. Each subject was tested by one of the nine raters randomly chosen. The other raters were in attendance, observing but not allowed to communicate with other

Table 1 The 11-item motor control test battery: included tests are shown and briefly described in online supplementary appendix 1

No	Test name	Acronym	Position
1	Forward bend	Fb	Standing
2	Return from forward bend	rFb	Standing
3	Sitting forward lean	SFL	Sitting
4	Sitting knee extension	SKE	Sitting
5	Pelvic tilt	PT	Standing
6	One-leg stance	OLS	Standing
7	Side bending	LAT-F	Standing
8	Rocking forward	Rf	Four-point kneeling
9	Prone knee flexion	PKF	Prone
10	Hip abduction/lateral rotation	BKFO	Supine
11	Shoulder flexion L spine against wall	ShFw	Standing

raters. Raters filled in the test questionnaire according to their own individual rating. Analysis of the data for the pilot test revealed substantial reliability for the interpretation of the test among raters (weighted kappa=0.74).

Statistical analysis

Baseline characteristics and demographic data were compared between the two groups by χ^2 tests, and independent t-tests for categorical and interval data, respectively. In the case of non-normally distributed data Mann-Whitney U test was applied.

Diagnostic accuracy is defined in terms of clinical (known-group) validity which is demonstrated when a test or questionnaire can discriminate between two groups assumed to differ on a variable of interest.[16] For the present study, clinical validity was evaluated based on the status of LBP. According to Luomajoki et al,[14] it can be assumed that subjects with LBP would show more positive test results than those without. And within patients with LBP it might be assumed that subacute and chronic patients would present more positive test results than acute patients. Whether a test battery consisting of more items increases accuracy of discrimination between LBP and no LBP was also evaluated. In order to evaluate the accuracy of the test battery depending on the number of single tests, parameters of diagnostic criteria for the summary scores of the 2, 4, 6, 8, 10 and 11-item solutions of the test battery were calculated.

The selection of optimal items for each item solution was achieved by applying an algorithm offered in the subselect package.[17] Given a predefined number of items/variables the algorithm identifies the items/variables out of various items/variables by maximising discriminant criteria like

Roy's root with the purpose of maximal discrimination of given subgroups, here no LBP versus LBP, by the identified items/variables. Once the optimal items for each solution were selected the number of failed single items/tests was summed up for each subject to achieve the summary score. Based on the summary score, diagnostic criteria were calculated for the potential of correctly classifying no LBP and LBP. Therefore, the receiver operating characteristic (ROC) statistics including the optimal cut-off by means of the Youden index, AUC, sensitivity and specificity were calculated. The closer these criteria are to one the better the diagnostic accuracy. The Youden index takes both sensitivity and specificity into account and ranges from 0 (correct classification only due to chance) to 1 (100 % correct classification). The following criteria have been used to interpret the AUC: excellent discrimination (AUC=0.90–1.0); good discrimination (AUC=0.80–0.90); fair discrimination (AUC=0.70–0.80); poor discrimination (AUC=0.60–0.70); and discrimination no better than chance (AUC≤0.50).[17] Furthermore, effect size according to Cohen's D was computed (<0.2 no effect, ≥0.2 small effect, ≥0.5 moderate effect, ≥0.8 large effect).

Additionally, for each item solution it was aimed to identify cut-offs for a three-class classification into no MC problem, mild/moderate MC problem and severe MC. Therefore, the summary score of each item solution against the status of having no LBP was compared, acute LBP and subacute/chronic LBP. The diagnostic criteria for the accuracy of correctly classifying the three groups were the generalised Youden index and the volume under the curve. Due to the presence of three cut-offs both the generalised Youden index needs to be applied instead of the classical Youden index and the volume under the curve needs to be used instead of the AUC. The generalised Youden index reflects the accuracy of correctly classifying all three subgroups (probability of correctly identifying no LBP, mild/moderate LBP and severe LBP) by means of the summary score of MC. Interpretation of the generalised Youden index is identical to the original Youden index ranging from 0 to 1 representing no and maximal accuracy, respectively. The volume under the surface (VUS) is computed due to the fact that the three probabilities span a three-dimensional space. Interpretation of the VUS is similar to AUC with the value 1 representing maximum accuracy. In contrast to the AUC, where 0.5 reflects accuracy due to chance, a VUS value of 0.167 means accuracy due to chance. The following criteria have been used to interpret the VUS in accordance with AUC: excellent discrimination (VUS=1.00 to 0.84); good discrimination (VUS=0.84 to 0.66); fair discrimination (VUS=0.66 to 0.50); poor discrimination (VUS=0.50 to 0.33); and discrimination no better than chance (VUS≤0.17). Furthermore, for the optimal item solution the probabilities of belonging to a clinical group were calculated based on the MC summary score for both the two-class (no LBP vs LBP) and the three-class (no LBP vs acute vs subacute/chronic LBP) classification based on

Table 2 Diagnostic criteria for the 2, 4, 6, 8, 10 and 11-item solutions for the two-class categorisation

Items, n	AUC	Sensitivity	Specificity	Youden	LR+	LR−	Effect size	Cut-off >	Test items
2	0.72	0.60	0.80	0.40	3.00	0.50	0.97	1	2, 9
4	0.80	0.65	0.83	0.48	3.82	0.42	1.23	2	2, 4, 9, 10
6	0.83	0.60	0.89	0.49	5.45	0.45	1.40	3	2, 4, 5, 7, 9, 10
8	0.85	0.90	0.63	0.53	2.43	0.16	1.45	3	2, 4, 5, 4, 6, 7, 8, 10
10	0.85	0.75	0.82	0.57	3.40	0.20	1.45	3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
11	0.85	0.77	0.80	0.57	3.70	0.19	1.45	4	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Forward bend (1), return from forward bend (2), sitting forward lean (3), sitting knee extension (4), pelvic tilt (5), one-leg stance (6), side bending (7), rocking forward (8), prone knee flexion (9), hip abduction/lateral rotation (10), shoulder flexion resting L spine on wall (11).
AUC, area under the curve; LR, likelihood ratio.

the predicted probabilities of a multinomial regression model.

RESULTS

All participants gave their informed consent. Mean baseline characteristics in the no LBP group (n=66) were: age 41.3 years (SD: 13.3 years), body mass index (BMI) $25.5 \, kg/m^2$ (SD: $5.3 \, kg/m^2$), gender 53.0% female, physical activity (of at least moderate intensity with slightly elevated heart rate or breathing such as riding the bike or gardening) 197.4 min/week (SD: 248.9 min/week) and absolute oswestry disability index (ODI) score (out of 50) 2.1 (SD: 3.47).

Mean baseline characteristics in the LBP group (n=65) were: age 46.4 years (SD: 12.7 years), BMI $26.9 \, kg/m^2$ (SD: $5.7 \, kg/m^2$), gender 64.6% female, physical activity 98.8 min/week (SD: 151.0 min/week) and absolute ODI score (out of 50) 10.6 (SD: 7.4). Age and physical activity were significantly different between groups.

Two-class categorisation

In table 2 the diagnostic criteria are presented for the 2, 4, 6, 8, 10 and 11-item solutions for the two-class categorisation in which MC is compared with LBP status. With each additional two test items at least one diagnostic criterion (AUC, Youden index, effect size) increases including up to the 10-item solution. The 11-item solution, however, does not provide any gain in accuracy.

As an all item solution results in no additional increase of diagnostic accuracy, the 10-item solution is considered the optimal solution for detecting LBP in correlation with MC deficiency. The overall discrimination potential of the test is good (AUC>0.8). The optimal cut-off for the 10-item solution for classification into LBP/no LBP is 3. Therefore, at least four items need to be considered as failed in order to be classified as LBP. The optimal identified items/tests are forward bend (1), return from forward bend (2), sitting forward lean (3), sitting knee extension (4), pelvic tilt (5), one-leg stance (6), side bending (7), rocking forward (8), prone knee flexion (9) and hip abduction/lateral rotation.

Figure 1 shows the resulting ROC curves for 2, 6 and 10-item solutions. With an increase of items, the AUC increases substantially.

In figure 2 the distribution of the summary scores (10-item solution) of the MCT battery against LBP status is
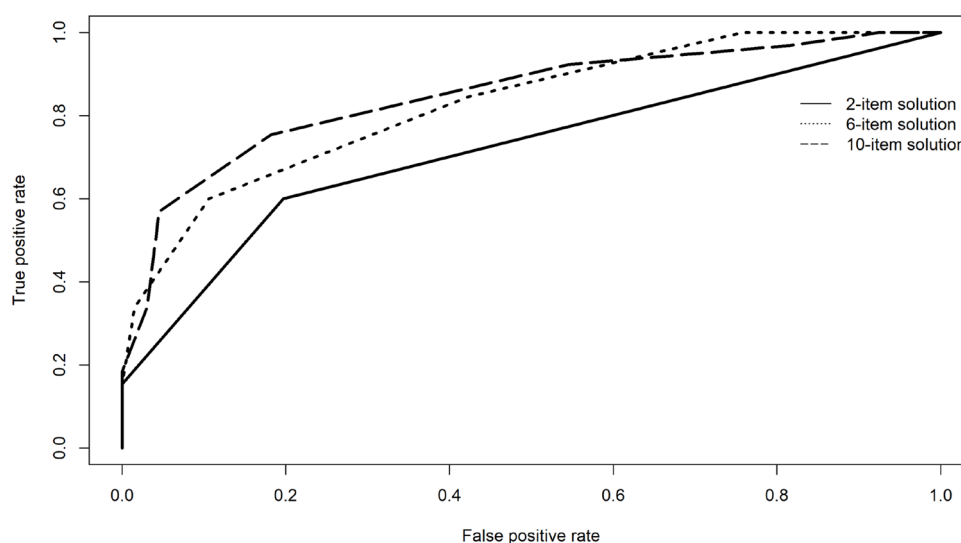

Figure 1 Receiver operating characteristic (ROC) curves for 2, 6 and 10-item solutions.
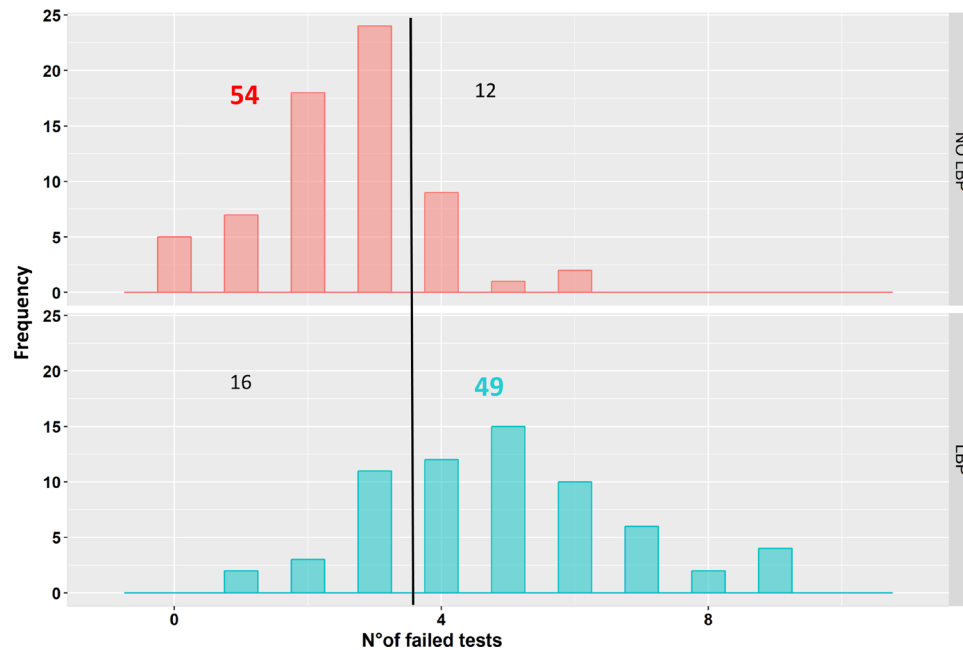
Figure 2 Distribution of motor control summary scores including categorisation numbers is depicted. LBP, low back pain.

depicted including the optimal cut-off of three tests. Out of 66 subjects without LPB 56 are classified correctly and out of 65 subjects with LPB 49 are classified correctly (specificity=82%, sensitivity=82%) .

Figure 3 plots the probability of having no LBP versus LPB according to the number of failed items based on the 10-item solution. For instance, with exactly five failed MCTs the predicted probability of having LPB is approximately 80%.

Three-class categorisation

In table 3 the diagnostic criteria are presented for the 2, 4, 6, 8 and 10 and all item solutions for the

three-class categorisation in which the MC summary score is compared with the status of no LPB, acute LBP and subacute/chronic LBP. With each increase of two items all diagnostic criteria (VUS, Youden index, effect size) increase including the eight-item solution.

When adding two more items to the eight-item solution (10-item solution) there is only a gain in the Youden index from 0.31 to 0.40. An all item solution results in no additional increase of diagnostic accuracy. Hence, also for the three-class categorisation the 10-item solution is the optimal solution for detecting a deficiency in MC. The overall discrimination potential of the test is fair
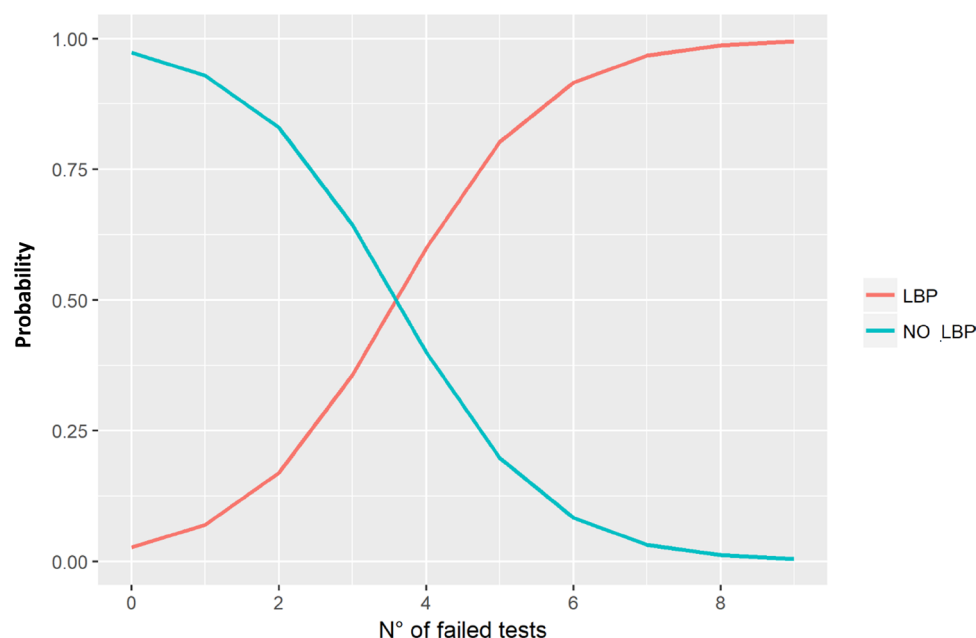

Figure 3 Predicted probability of having LBP/no low back pain depending on number of failed tests. LBP, low back pain.

Table 3 Diagnostic criteria for the 2, 4, 6, 8, 10 and 11-item solutions for the three-class categorisation

Items, n	VUS	Sensitivity (chronic status)	Sensitivity (acute status)	Specificity	Youden	Effect size	Cut-off ≤, ≥
2	0.41	0.80	0.47	0.23	0.25	1.02	0, 2
4	0.45	0.36	0.38	0.83	0.29	1.30	1, 3
6	0.47	0.42	0.53	0.57	0.27	1.50	1, 4
8	0.52	0.74	0.44	0.63	0.31	1.56	2, 4
10	0.52	0.48	0.50	0.82	0.40	1.56	3, 6
11	0.52	0.52	0.47	0.80	0.40	1.56	3, 6

VUS, volume under the surface.

(VUS>0.8). The optimal cut-off of the 10-item solution for categorisation into no MC, mild/moderate MC and severe MC is 3 and 6, respectively. Therefore, at least four items need to be considered as failed in order to be classified as mildly/moderately MC deficient and with six or more failed items persons are classified into severely MC deficient.

In figure 4 the distribution of the summary scores of the MCT battery against LBP status is depicted including the optimal cut-off of three and six failed tests for the 10-item solution.

The diagonal entries show the correctly classified subjects. Out of 68 persons without LBP, 54 show maximally three failed items which corresponds to probability of 82%. Among 34 people with acute LBP, 17 present with a summary score of exactly 4 or 5 (correct classification 50%). Additionally, 15 out of 31 patients failed at least six times which corresponds to 48% having a correct classification. This results in an overall classification accuracy of 60%. Under our null hypothesis, an overall classification accuracy of 33.3% was expected. This corresponds

to a generalised Youden index of 0.4 and a VUS of 0.51 (expected VUS under the null hypothesis was 0.167).

Figure 5 plots the probability of having no LPB, acute or subacute/chronic LBP depending on the number of failed items based on the 10-item solution for the three-class classification. For example, with exactly 10 failed items the predicted probability of having subacute/chronic lLBP is approximately 87%, the probability of having acute LBP is approximately 13% and the probability of having no LBP is approximately 0%.

DISCUSSION

This study sought to identify a time-efficient MCT battery with maximised diagnostic accuracy for both a two-class and a three-class solution for patients with non-specific LBP. For both the two and the three-class categorisation the ideal number of MCT items is 10. With increasing number of failed tests, the probability of having LBP increases. The overall discrimination potential for the two-class categorisation of the MCT is good (AUC>0.8),
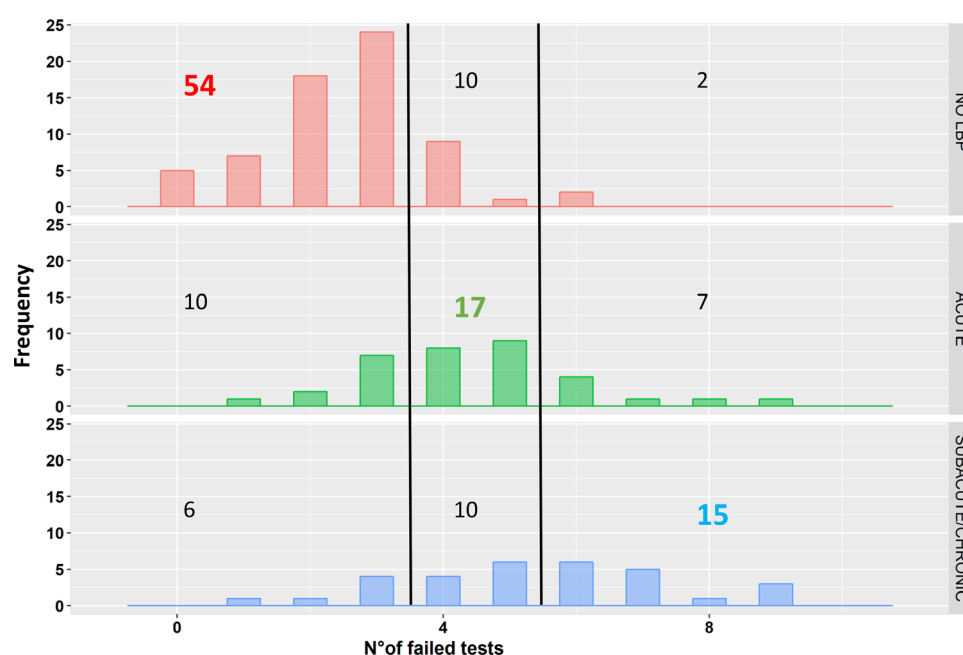
Figure 4 Distribution of motor control (MC) summary scores including classification numbers is depicted. LBP, low back pain.
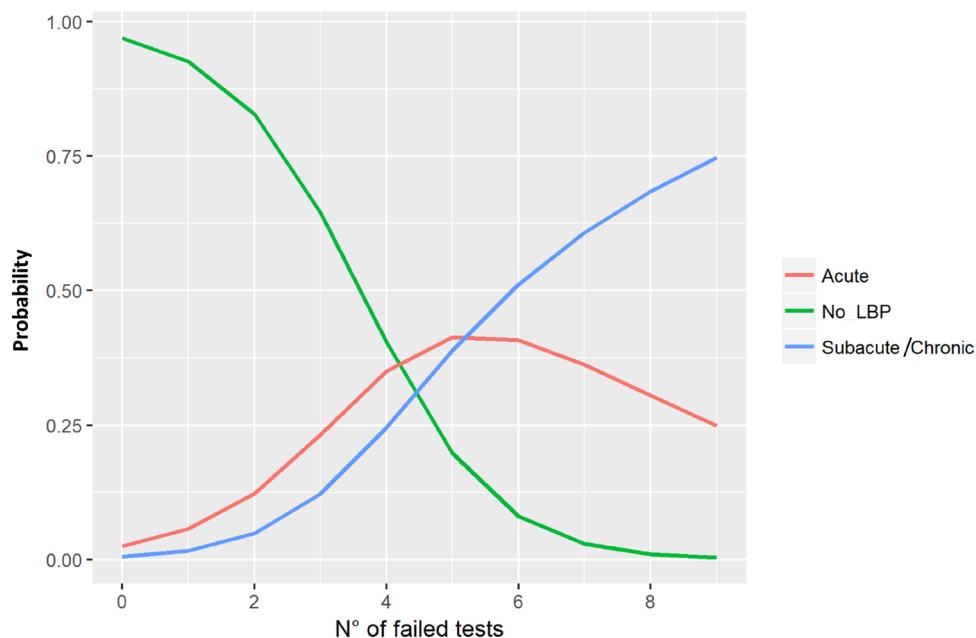
Figure 5 Predicted probability of having acute, subacute/chronic LBP, no low back pain depending on number of failed tests. LBP, low back pain.

with an optimal cut-off of three tests. Therefore, at least four items need to be considered as failed in order to be classified as MC deficient.

The overall discrimination potential of the three-class categorisation is fair (VUS>0.5). The optimal cut-off of the 10-item solution for categorisation into no, mild/moderate and severe MC deficiency is three and six MCTs, respectively. Therefore, at least four MCT items need to be considered as failed in order to be classified as mildly/moderately MC deficient and with six or more failed items persons are classified as severely MC deficient.

The optimal identified items/tests for both classifications are: forward bend (1), return from forward bend (2), sitting forward lean (3), sitting knee extension (4), pelvic tilt (5), one-leg stance (6), side bending (7), rocking forward (8), prone knee flexion (9), and hip abduction/lateral rotation (10).

The discrimination potential for the two-class solution is substantially better than for the three-class solution. Especially since the sensitivity within chronic and acute patients is low (0.48 and 0.5) indicating a high misclassification rate mainly between mild/moderate and severe MC deficiency. The two-class solution can be recommended for use both in the clinical and scientific context with emphasis on ruling out MC deficiency as indicated by the stronger negative likelihood ratio as compared with the positive likelihood ratio (0.2 vs 3.4). Hence, the probability of having no MC deficiency (negative test result) in subjects with LBP is 0.2 times the probability of having no MC deficiency (negative test result) in healthy subjects (ruling out). In contrast, the probability of having MC deficiency (positive test result) in subjects with LBP is 3.4 times the probability of having MC deficiency (positive test result) in healthy subjects (ruling in).

Even though not as accurate as the two-class categorisation, the three-class categorisation can still be recommended for clinical use. While distinction between groups with four or five failed MCTs is poor, the distinction with six or more failed tests is much more accurate. For instance, with 10 failed tests the predicted probability of having subacute/chronic LBP is 87% and only 13% of having acute LBP.

Our results are strongly in line with other research. Previously it has been demonstrated that a six-item test battery could discriminate people with LBP from healthy subjects with an effect size of 1.18.[14] Furthermore, and similar to our results, it was also shown that patients with CLBP had more failed tests than people with acute or subacute LBP and healthy subjects and that discrimination between the subacute and chronic state was poor. Our test results based on the six-item battery are very similar to previous report.[14] However, it was demonstrated that increasing the test battery from six to 10 items increases diagnostic accuracy as indicated by the Youden index increasing from 0.49 to 0.57 which leads to the recommendation of augmenting the number of MCT items to 10. This is in line with other research[18] where more tests are recommended in order to yield more possibilities of personalisation of individual therapy.

This option is also warranted by the clinically driven composition of our MCT battery that contains tasks in open and closed kinetic chain with both upper and lower extremity movements impacting on MC of the lumbopelvic region in different positions such as standing, sitting, four-point kneeling and lying down. Furthermore, the findings of this study represent a clear enhancement to the existing body of knowledge as a conclusion can be drawn with respect to an optimal

number of MCTs to include in the MCT battery. This holds true for a two and three-class categorisation which has not been subject to investigation yet in this field of research.

The strengths of our study include the consecutively recruited large sample of subjects arising from different practices indicating high generalisability. However, the subgroups were unbalanced resulting in small number of subjects in the subacute class which did not allow the formation of a separate subgroup with subacute pain. Furthermore, age and physical activity were significantly differently distributed between groups. However, these variables were not statistically associated with the outcome of the test battery and hence, did not confound our results. Additionally, the reliability of single items was not investigated. Only rater reliability of the total test battery was assessed after the training programme in the pilot phase. However, reliability was assessed in previous studies indicating acceptable values.[10–12] Instead of clinical preselection of single tests a statistical-based selection which might have increased diagnostic accuracy might have been applied. However, this might have resulted in a pool of single tests that is quite different from the one agreed by clinical consensus.

Further research might address points as follows: the MCT battery should be validated in an external and independent study population including larger sample size especially with respect to subgroups such as subacute pain. Furthermore, it would be of interest to assess the ability of incorporating the MCT battery in clinical classification systems for patients with LBP and whether this improves ability of subgrouping. Finally, it needs to be evaluated if treatment according to this MCT battery improves the clinical outcome of patients with LBP.

In conclusion, the use of a 10-item MCT battery is recommend for both the two-class and a three-class categorisation of patients with non-specific LBP into MC deficient and not deficient or mildly/moderately, severely and not MC deficient. The MCT battery should include the items forward bend, return from forward bend, sitting forward lean, sitting knee extension, pelvic tilt, one-leg stance, side bending, rocking forward, prone knee flexion and hip abduction/lateral rotation. The two versions show good and fair discrimination potential.

REFERENCES

1. Fatoye F, Gebrye T, Odeyemi I. Real-World incidence and prevalence of low back pain using routinely collected data. *Rheumatol Int* 2019;39:619–26.
2. Ma VY, Chan L, Carruthers KJ. Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Arch Phys Med Rehabil* 2014;95:986–95.
3. Fritz JM, Magel JS, McFadden M, *et al*. Early physical therapy vs usual care in patients with recent-onset low back pain: a randomized clinical trial. *JAMA* 2015;314:1459–67.
4. O' Sullivan P, Dankaerts W, O' Sullivan K. Movement System Impairment Syndromes of the Low Back. In: Jull G, Moore A, Falla D, eds. *Grieve's Modern Musculoskeletal Physiotherapy*. 4th edn. Edinburgh: Elsevier, 2015: 465–9.
5. O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005;10:242–55.
6. O'Sullivan PB. Lumbar segmental 'instability': clinical presentation and specific stabilizing exercise management. *Man Ther* 2000;5:2–12.
7. Van Dillen LR, Sahrmann SA, Norton BJ, *et al*. Movement system impairment-based categories for low back pain: stage 1 validation. *J Orthop Sports Phys Ther* 2003;33:126–42.
8. Hodges P. The role of motor control training. In: Jull G, Moore A, Falla D, eds. *Grieve's Modern Musculoskeletal Physiotherapy*. 4th edn. Edinburgh: Elsevier, 2015: 482–7.
9. Luomajoki HA, Bonet Beltran MB, Careddu S, *et al*. Effectiveness of movement control exercise on patients with non-specific low back pain and movement control impairment: a systematic review and meta-analysis. *Musculoskelet Sci Pract* 2018;36:1–11.
10. Van Dillen LR, Sahrmann SA, Norton BJ, *et al*. Reliability of physical examination items used for classification of patients with low back pain. *Phys Ther* 1998;78:979–88.
11. Enoch F, Kjaer P, Elkjaer A, *et al*. Inter-Examiner reproducibility of tests for lumbar motor control. *BMC Musculoskelet Disord* 2011;12:114.
12. Luomajoki H, Kool J, de Bruin ED, *et al*. Reliability of movement control tests in the lumbar spine. *BMC Musculoskelet Disord* 2007;8:90.
13. Sahrmann S, Van Dillen L. Movement System Impairment Syndromes of the Low Back. In: Jull G, Moore A, Falla D, eds. *Grieve's Modern Musculoskeletal Physiotherapy*. 4th edn. Edinburgh: Elsevier, 2015: 474–81.
14. Luomajoki H, Kool J, de Bruin ED, *et al*. Movement control tests of the low back; evaluation of the difference between patients with low back pain and healthy controls. *BMC Musculoskelet Disord* 2008;9:1–12.
15. Waddell G. *The back pain revolution*. 2nd edn. Edinburgh; New York: Churchill Livingstone, 2004.
16. Michalos AC. *Encyclopedia of quality of life and well-being research*. New York: Springer, 2014.
17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
18. Saragiotto BT, Maher CG, Yamato TP, *et al*. Motor control exercise for chronic non-specific low-back pain. *Cochrane Database Syst Rev* 2016;21.