Do various baseline characteristics of transversus abdominis and lumbar multifidus predict clinical outcomes in non-specific low back pain? A systematic review

Arnold Y.L. Wong, Eric C. Parent, Martha Funabashi, Tasha R. Stanton, Greg N. Kawchuk

PII: S0304-3959(13)00378-3
DOI: http://dx.doi.org/10.1016/j.pain.2013.07.010
Reference: PAIN 8900

To appear in: PAIN

Received Date: 26 March 2013
Revised Date: 5 July 2013
Accepted Date: 10 July 2013


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Title: Do various baseline characteristics of transversus abdominis and lumbar multifidus predict clinical outcomes in non-specific low back pain? A systematic review.

Arnold YL Wong, PT, MPhil, Eric C Parent, PT, PhD, Martha Funabashi, PT, MSc, Tasha R Stanton, PT, PhD, Greg N Kawchuk, DC, PhD

1. Department of Physical Therapy, University of Alberta, Edmonton, Alberta, Canada
2. Clinical scientist, Glenrose Rehabilitation Hospital, Edmonton, Alberta, Canada
3. Sansom Institute for Health Research, School of Health Sciences, University of South Australia, Adelaide, South Australia, Australia.

Corresponding author:
Greg Kawchuk DC, PhD
Department of Physical Therapy, University of Alberta,
3-48 Corbett Hall, Edmonton, Alberta, Canada, T6G 2G4
Phone: 780-492-6891
Fax: 780-492-4429
Email: greg.kawchuk@ualberta.ca

Number of text pages (without figures and tables): 38
Number of figures: 1
Number of tables: 5
Number of appendices: 3
Number of legend page: 1
1. Introduction (word count 500)

Approximately 9.2% of the global population is affected by low back pain (LBP) while LBP-related disability is the leading cause of disability in the world [91]. Although certain cases of LBP are ascribed to specific pathology, 90% of sufferers experience LBP of unknown origin or pathology (known as non-specific LBP) [24].

Notwithstanding the lack of consensus regarding the causal relation between deficits in spinal muscles and the onset of LBP, transversus abdominis (TrA) and lumbar multifidus (LM) play important roles in intersegmental spinal control [32,46,53,73,75,100] and may affect the progression and recurrence of LBP. Anatomically, transversus abdominis connects to the lumbar vertebrae through the thoracolumbar fascia forming a corset-like structure encircling the trunk, which controls intra-abdominal pressure and vertebral stiffness [4,31,34,43]. In contrast, lumbar multifidus muscles are deep paraspinal muscles with densely packed short muscle fibers that generate large forces over short distances [20] for intersegmental control [9,100].

Various investigations have demonstrated associations between LBP and the characteristics of TrA/LM [15,18,57]. Research has revealed that patients with acute or chronic LBP have increased fat infiltration and abnormal changes of Type I and II fibers in LM [2,3,42,56,58,70] while patients with unilateral LBP display localized asymmetrical LM atrophy at the painful vertebral level [15,33,70,92]. Functionally, patients with LBP
demonstrate significantly less TrA/LM thickness changes than asymptomatic individuals during voluntary tasks as measured by B-mode ultrasound imaging (USI) [14,18,48,82], and a delayed anticipatory onset of TrA or deep LM fibers as measured by intramuscular electromyography during trunk loading or limb movement [37,38,39,57]. Collectively, it is hypothesized that the aberrant morphometry, histology and activation of these muscles in individuals with acute [16,35,47,75], chronic [61] and recurrent LBP [75] may cause prolonged LBP and its recurrence [32,57].

Recent research suggests that baseline morphometry of TrA/LM may be a treatment effect modifier (a characteristic that predicts who will or will not benefit specifically from a particular treatment) or a prognostic factor for LBP-related clinical outcomes [26,29,85]. Specifically, decreased LM thickness change during contraction as measured by B-mode USI was correlated with the predictors for clinical success with a stabilization exercise program [29]. Poor baseline TrA lateral slide measured with B-mode USI may also predict positive recovery of chronic LBP regardless of the intervention [85]. Taken together, a comprehensive review of evidence for the predictive or treatment modification value of TrA/LM measurements may help clinicians predict LBP prognosis and determine the best treatments for specific patient subgroups. However, to our knowledge, no systematic review on this topic has been conducted.

The primary objective of this review was to summarize the evidence with respect to the
ability of morphometry, histology or activation of TrA/LM to predict clinical outcomes and recurrence of non-specific LBP among untreated or conservatively treated patients. The predictive ability of TrA/LM includes the prognostic value and treatment effect modification.

The secondary objective was to review if baseline features of TrA/LM would have differential prognostic effects or modify treatment effects in subgroups of patients with different (1) chronicity, (2) age, and (3) inception or survival stage.

2. Methods

This review protocol was registered with PROSPERO (CRD42012002703). The methodology and reporting format of this review follows the recommendations and guidelines of the Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) [63] and the Meta-analysis of Observational Studies in Epidemiology (MOOSE) [80].

Relevant articles were systematically identified through MEDLINE, EMBASE, PEDro, SPORTDiscus, CINAHL, and Cochrane Library (from the beginning of each database up to December 2012) using keywords, MeSH and free text words which included: low back pain, LBP, backache, lumbago, lumbalgia, multifidus (truncated), LM, MF, transversus abdominis, TrA, lumbar muscle, erector spinae, stabilizing muscle, predict (truncated), follow-up, cohort,
course, longitudinal, and randomized control (truncated). Appendix I shows the exact search strings utilized. In addition, ClinicalTrial.gov, NIH Clinical Center Clinical Research Studies, and Current Controlled Trials Register were searched to identify relevant ongoing research. The principal investigators of relevant ongoing research, 11 prominent researchers who have published more than 5 articles in this area, as well as the corresponding author of each included article, were contacted to identify any additional studies of potential relevance. The results of the search strategy are shown in Figure 1.

2.2. Selection criteria of studies

Only full reports published in English, Chinese, French or Portuguese and meeting the following criteria were included for analysis (see Appendix II for details).

2.2.1. Design

Longitudinal cohort research is the preferable design for identifying a causal relation for prognostic factors [26,64] while randomized controlled trial design is desirable for evaluating treatment effect modifiers [78]. As such, the eligible study design included: primary longitudinal (both prospective and retrospective) cohort studies and randomized controlled trials. Other eligible study designs included: observational studies, case series with ≥10 subjects involving untreated or conservatively treated patients with nonspecific LBP, as well
as systematic reviews and meta-analyses. Studies from books were excluded as secondary sources of information.

Studies that investigated “inception” or “survival” cohorts were included where inception cohort refers to a group of patients early in the course of nonspecific LBP while survival cohort refers to a group of patients who are at various stages of the LBP development at the time of recruitment [13,41].

2.2.2. Population

Studies that investigated self-ambulatory adults, aged above 18 years, with acute (<6 weeks), subacute (6 to 12 weeks) or chronic (>3months) non-specific LBP were included [28]. Non-specific LBP was defined as pain or discomfort between the 12th rib costal margin and above the gluteal folds, with or without leg pain, where LBP pain is not attributed to specific physical cause or pathology [90,101]. Articles were excluded if they were (1) multiple reports that included duplicated results of identical patient cohorts, or (2) studies involving less than 80% of participants with nonspecific LBP.

2.2.3. Predictors or treatment effect modifiers

It is hypothesized that spinal stability is maintained by the passive (such as disc, ligaments and facets), active (spinal muscles) and neural control (sensory and neuromotor
control) subsystems. Dysfunctions in any of the three subsystems may increase the risk of spinal injury by overloading the other subsystems [67,68,69]. As such, aberrant changes in morphometry, histology or activation of TrA/LM may affect or predict future clinical outcomes or recurrence of LBP. To be included, studies had to investigate the effect of static and dynamic morphometry, histology or activation of TrA/LM in predicting the clinical outcomes of patients with non-specific LBP. Static morphometry was defined as the measurement of architectural characteristics (such as shape, cross-sectional area, volume, length, depth, diameter and pennation angles) of a given muscle at rest [98] whereas dynamic morphometry referred to the measurement of change in architectural features of a muscle during contraction [98]. Histology was defined as the study of microscopic composition of a muscle (such as muscle fiber types).[7] Muscle activation indicated the change in myoelectric signals as measured by electromyography during muscle contraction [8]. Studies that investigated the onset of lateral abdominal muscles (TrA, obliquus internus and obliquus externus) as measured by M-mode USI or tissue Doppler imaging were included because they are valid assessment of deep abdominal muscles function by taking account of intra- and interindividual variability in the onset of the three muscles [59,60,85,88]. In order to identify potential studies that investigated the morphometry, histology or activation of TrA/LM at baseline, our search strategy did not place any restrictions on the type of muscle measurement techniques.
2.2.4. Clinical outcomes

Clinical outcomes included type and duration of LBP symptoms, LBP-related disability, return to work or sports, recurrence of LBP, LBP-related medications and healthcare professional visits.

2.2.5. Intervention

Studies with or without conservative treatments were eligible for review as long as they met the population and study design criteria. Research that solely evaluated the effectiveness of treatments was excluded (i.e. research did not perform prediction or treatment effect modification analyses).

2.3. Selection of studies

The selection of studies was divided into two stages (Fig. 1). In the first stage, the identified citations from the databases were stored in RefWorks 2.0 (RefWorks-COS, Maryland, USA) and the duplicated citations were excluded. A research assistant removed the identifiable information (except the title and abstract) of the citations [86]. Two reviewers (AW and MF) screened the titles and abstracts independently and discarded the irrelevant citations using standardized screening forms based on the selection criteria. Each reviewer
filled in the forms and listed the reasons for the inclusion or the exclusion of articles. The screened lists were compared between the two reviewers. Disagreement was resolved by consensus. To minimize the risk of discarding studies incorrectly, articles that were chosen by either reviewer were included for the next stage of the review. Piloting of the study selection process was conducted on the first 100 potential citations. The rationale for inclusion and exclusion were discussed and clarified to ensure the consistency between reviewers. In the second stage, the full-text articles of the potentially eligible studies were retrieved. The research assistant then blacked-out the information regarding authors, institutions and journals. The same screening procedures used at the first stage were then repeated. If selection disagreement persisted following the joint review process, a third reviewer (EP) was consulted for a final decision. The reference lists of the included papers were searched for relevant articles. Forward citation tracking with Scopus and Web of Science was conducted to identify relevant publications that have cited the included articles.

The inter-reviewer reliability at each selection stage was analyzed by percent agreement and Kappa coefficients. Kappa is interpreted as: 0.00 to 0.20 for poor agreement; 0.21 to 0.40 for fair agreement; 0.41 to 0.60 for moderate agreement; 0.61 to 0.80 for good agreement; and 0.81 to 1.00 for almost perfect agreement [54].

2.4. Risk of bias assessment of selected studies
There is no standardized assessment tool for appraising the risk of bias in prognostic studies [5]. Based on the recommendations on evaluation of prognostic and treatment effect modification studies [1,5,25,27,81], a previously published risk of bias assessment tool was modified and adopted [10]. The modifications aimed to evaluate the risk of bias related to confounding variables and treatment effect modifiers. The modified assessment tool covers seven potential bias areas including patient population, follow up of patients, prognostic factors, treatment, outcome measurement, confounding measurement and statistical analysis (Table 1). The maximum score of this checklist is 26. The cut-off point for distinguishing a high quality study is 50% of maximum score [10]. A study with higher scores indicates that the study has lower risk of bias. Since the cut-off point was arbitrary, sensitivity analyses using 60% and 70% cut-off points were conducted to test the robustness of these cut-off points.

Assessment of Multiple Systematic Reviews (AMSTAR) was used to evaluate the risk of bias in the included systematic reviews, if any. AMSTAR was chosen because it has shown high reliability, as well as face, construct and external validity for assessing the risk of bias of systematic reviews [76,77].

The two reviewers conducted the risk of bias assessment independently on the included studies and were blinded to information about authors, institutions and journals. The inter-reviewer reliability for risk of bias assessment was analyzed by percent agreement and...
intraclass correlation coefficient (ICC, model 3,1) of the final score. A consensus meeting was arranged to resolve the differences between both reviewers.

2.5. Data extraction

Two reviewers independently extracted the following information: study design, participants’ characteristics (such as age, gender, type and duration of LBP), sample size, intervention (such as type, dose, duration and follow-up frequency), predictive factors, clinical outcomes (such as functional outcome score, work status and recurrence), statistical analysis (such as Cox regression analysis, log rank test, multivariate Kaplan-Meier survival analysis, and linear/multiple/logistic/Poisson regression models), and results at multiple assessment time points (such as odds ratios, beta-coefficients, risk ratios, positive or negative likelihood ratios, and hazard ratios). If the studies assessed multiple outcomes with different statistical methods, we solely extracted the information that was relevant to our research questions. To ensure accuracy of data extraction, a consensus meeting was held between the reviewers to discuss the extracted data.

2.6. Synthesizing evidence for prognostic factors and treatment effect modifiers

Depending on the study quality, homogeneity of study population (such as age, chronicity of LBP), type of predictive factors, outcome measures, and follow-up time,
meta-analysis was considered for the studies with low risk of bias using a 50% cut-off point. Specifically, meta-analyses for the prognostic/treatment modification effect of each muscle were planned based on the follow-up duration (i.e. immediately after treatment, less than 1-year and more than 1 year). Meta-analysis with reference to the secondary objectives (acute, subacute and chronic LBP, age between 18 to 40 years, 41 to 60 years and above 60 years, and the "inception" or "survival" cohort) would be performed if the relevant data were homogeneous. If meta-analysis was not conducted, qualitative analysis was used to summarize the predictive value of each predictor. The level of evidence for each predictor was assessed using the criteria adopted in previous prognostic research [12]. The level of evidence (strong, moderate, limited, no, and inconclusive evidence) was classified based on the risk of bias of the cohort studies and the consistency of the research findings (Table 2). The results represent the extent of evidence that substantiate a given feature of TrA or LM in predicting the clinical outcomes of untreated or conservatively treated patients with nonspecific LBP. If ≥75% of all the included studies reported a factor that showed a uniform association in the same direction, the evidence was considered consistent [12].

3. Results

3.1 Literature search

The literature search of databases yielded 2,321 potentially relevant articles (Fig. 1).
Two thousand two hundred and twelve articles (2,212) were excluded based on title, abstract and duplication. In the first stage of screening, the percentage agreement between the two reviewers was 96.8% with a Kappa coefficient of 0.69. One hundred fifteen (115) full-text articles were retrieved. Forty-eight articles (48) were identified based on forward tracking, reference lists of selected articles, or suggestions from experts. One hundred and ten articles (110) were excluded after assessing the full text. The major reasons for exclusion were a cross-sectional or diagnostic study design, absence of assessment of morphometry or activation of TrA/LM, and not assessing the ability of TrA or LM morphometry/activation in predicting clinical outcomes. Five articles met the selection criteria [19,22,59,85,102]. The percentage agreement between the two reviewers was 100% (Kappa coefficient = 1.00) in the second stage of screening.

Three included articles investigated TrA [19,59,85] and two included studies examined LM (Table 3) [22,102]. Of these articles, one evaluated the treatment effect modification of TrA dynamic morphometry [19] and the rest investigated the prognostic ability of TrA/LM dynamic morphometry [22,59,85,102]. No included studies assessed the predictive ability of TrA or LM on LBP recurrence.

3.2. Risk of bias assessment and synthesis of evidence

The risk of bias assessment scores is shown in Table 4. The raw agreement for these
scores between two reviewers was 82%. The ICC$_{3,1}$ was 0.76. Full agreement was reached following the consensus meeting. When the cut-off point was set at 50% of the maximum risk of bias score, all the included studies were classified as low risk of bias (Table 4). If the cut-off point was set at 60% of the maximum score, two included articles were classified as high risk of bias (Table 4) [22,85]. When the cut-off point was set at 70% of the maximum score, four included articles were classified as high risk of bias (Table 4) [22,59,85,102]. The major sources of high risk of bias were due to (1) no inclusion of inception cohort, (2) less than 1 year follow-up, (3) no a priori description of potential predictors, (4) no adjustment for confounders, (5) insufficient participants for multiple statistical tests, and (6) selective reporting of results. Regardless of the cut-off points, the conclusions of this review remain unchanged. Since the data obtained from the included studies was clinically heterogeneous (such as different physical tests for TrA and LM, various clinical outcome measures, and inconsistent follow-up duration), meta-analysis was precluded [87].

3.3 Participants

The sample sizes of the included studies ranged from 25 to 87 patients [19,22,59,85,102]. The average age of the participants in these studies ranged from 33.3 to 54.9 years. Four out of five included articles investigated patients with chronic LBP [19,59,85,102]. For the study that did not target patients with chronic LBP, a sample with mixed duration of LBP was
studied (i.e. pain duration ranging from 2 days to 24 years; median duration of 186 days) [22].

In other words, all included studies recruited survival cohorts of patients. The participants of the included studies were recruited from hospitals, physiotherapy and general practice clinics, as well as community-based advertisements [19,22,59,85,102]. None of the studies targeted patients with acute/subacute LBP.

3.4. Physical tests for TrA and LM

Although the review aimed to include studies that assessed baseline characteristics of TrA/LM using various measurement methods, all included studies coincidentally measured TrA/LM using various USI methods, which have published reliability and validity for measuring static and dynamic morphometry of TrA/LM [50,51,60,88,94,101]. Since the included studies did not use electromyography to measure muscle activation, the findings of this review were limited to the investigation of baseline TrA/LM morphometry. Two studies assessed voluntary TrA contraction ratio during abdominal drawing-in maneuver task (ADIM) using B-mode [85] and M-mode [59] USI, respectively. TrA contraction ratio was the ratio between TrA muscle thickness at maximal contraction and TrA muscle thickness at rest [85]. Similarly, automatic TrA contraction ratio during an isometric knee flexion/extension contraction task [19] and TrA lateral slide during ADIM [85] was measured by B-mode USI. TrA lateral slide was quantified by the maximum lateral displacement of v-shaped midline
border of TrA from its resting to contracted positions [85]. The anticipatory onset of lateral abdominal muscles during a contralateral rapid arm movement task was measured by M-mode USI [85] and tissue Doppler imaging [59]. The anticipatory onset of lateral abdominal muscles was expressed as the onset time of the earliest abdominal muscle morphometric change with reference to the onset of the deltoid muscle of the moving arm [85]. The automatic LM contraction during two types of contralateral arm lifting task was measured by B-mode USI [22,102]. The dynamic morphometry of LM during contraction was quantified by the percent change in LM muscle thickness. Specifically, the change in LM thickness (contracted - rest) was expressed as a proportion of the thickness at rest (and multiplied by 100) [102].

3.5. Adjustment of confounders

Three included studies adjusted for some confounders in the respective data analyses [19,85,102]. Ferreira and co-workers adjusted for the baseline values of a numerical pain rating scale, the Roland Morris disability index, and a patient-specific functional scale in the respective regression models [19]. Similarly, Zielinski and colleagues performed regression analyses after adjusting for the Oswestry disability index and numeric pain rating scores at baseline [102]. Unsgaard-Tondel and colleagues adjusted for age, gender, body mass index, pain duration, initial Fear-Avoidance Beliefs Questionnaire physical score and initial pain
intensity in their linear and logistic regression models [85]. Two studies did not adjust for any confounders [22,59].

3.6. Clinical outcome measures

All studies used both a self-reported pain rating scale and at least one functional assessment scale as clinical outcome measures [19,22,59,85,102] but in contrast, Unsgaard-Tondel and colleagues investigated only the predictive ability of baseline dynamic morphometry of TrA/LM on pain intensity [85]. The Modified Oswestry disability index and the Roland Morris Disability questionnaire were the commonly adopted functional assessment scales. One study also included a patient-specific functional scale [19].

3.7. Intervention and follow-up

The included studies adopted different conservative treatments including sling exercise [85], motor control exercise [19,59,85,102], general exercise [19,85], spinal mobilization [19] and spinal manipulative therapy [22]. Although no limitation was placed on the type of conservative treatment evaluated, none of the included articles investigated pharmaceutical intervention or the predictive ability of TrA/LM on untreated patients.

The follow-up time of the included studies ranged from 1 week to 1 year. Four included studies showed less than 80% attrition when they reassessed the participants immediately.
after the treatment [19,22,59,102]. Although the attrition in the study involving 1-year follow up was 79.8%, their linear and logistic regression analyses used data from 56.8% to 62.4% of the total number of participants recruited [85].

3.8. Statistical analysis

Linear regression [59,85,102], logistic regression [19,85] and correlation [22,59] were commonly used in the included studies to quantify the predictive value of TrA and LM dynamic morphometric variables. However, the report of statistical results varied among studies. While some studies reported odds ratios or interaction effects together with the corresponding confidence intervals [19,85], others simply reported a p-value or “non-significant” [22,59,102]. Most studies estimated the relation between predictors and clinical outcomes based on significance testing of the corresponding absolute values in outcome scales [19,22,59,102] whereas one study used clinically important pain reduction for analysis [85].

3.9. Static morphometry and histology of TrA or LM in predicting clinical outcomes of LBP

The literature search did not identify any study that investigated the static morphometry or histology of TrA or LM in predicting clinical outcomes or recurrence of LBP in patients with non-specific LBP. As such, there was no evidence to support the predictive value of
TrA/LM static morphometry or histology on clinical outcomes of nonspecific LBP.

3.10. Baseline dynamic morphometry of TrA or LM as a prognostic factor for clinical outcomes of LBP

Table 3 summarizes the data extracted from the included studies that investigated the dynamic morphometry of TrA/LM at baseline in predicting clinical outcomes of LBP. Depending on the types of physical tests, the predictive value of TrA/LM varies. Unsgaard-Tondel and co-workers found that better baseline TrA lateral slide during ADIM in patients with chronic LBP was associated with higher pain intensity at 1-year following various types of exercise therapy (beta coefficient = 0.13) [85]. They also noted that better baseline TrA lateral slide was more likely to have less clinically important pain improvement (odds ratio = 0.76) for at 1-year follow-up regardless of the type of exercise treatment [85]. However, the baseline TrA contraction ratio (contracted TrA thickness/relaxed TrA thickness) during ADIM was not correlated with the improvement in clinical outcomes (pain intensity or LBP-related disability) of patients irrespective of the type of exercise intervention [59,85].

Conflicting predictive results were also noted in baseline percent LM thickness change. Fritz and colleagues found that baseline percent thickness change of LM (during a contralateral arm lifting task) was not associated with LBP-related disability following 1-week spinal manipulative therapy [22]. A 6-week motor control exercise study also
revealed that baseline percent LM thickness change could not predict post-treatment pain intensity [102]. However, the same research found that a larger percent LM thickness change during a resisted contralateral arm lifting task at baseline was a weak, but significant predictor of improved post-treatment LBP-related disability in a subgroup of patients who were unlikely to improve with stabilization exercise ($r = 0.38$, $p = 0.013$) [102].

Overall, depending on the types of physical tests, there was limited evidence about whether baseline TrA dynamic morphometric variables were prognostic factors for LBP-related disability or pain intensity in patients with chronic LBP following various exercise interventions (Table 5). Similarly, there was limited evidence in support of no relation between baseline LM dynamic morphometry and post-treatment LBP intensity but there was conflicting evidence for a relation between baseline LM dynamic morphometry and post-treatment LBP-related disability (Table 5).

### 3.11. Baseline dynamic morphometry of TrA or LM as a treatment effect modifier for clinical outcomes of LBP

One study reported that baseline dynamic morphometry of TrA was a treatment effect modifier [19]. Ferreira and co-workers found that patients with poorer baseline TrA thickness contraction ratio receiving a course of motor control exercise yielded better pain reduction than those receiving a general exercise regimen at 8-week follow-up (Table 3) [19]. No study
has investigated the treatment effect modification ability of LM at baseline.

Collectively, as a result of limited publication on this topic, limited evidence corroborated that poorer automatic TrA contraction at baseline was a treatment effect modifier favoring motor control exercise over general exercise (Table 5). There was no evidence regarding the treatment effect modification of baseline LM dynamic morphometry.

3.12. Baseline anticipatory onset of lateral abdominal muscles or LM morphometric change in predicting clinical outcomes of LBP

Two studies reported that baseline anticipatory (feedforward) onset timing of lateral abdominal muscles morphometric change could not predict short- or long-term clinical outcomes (LBP-related disability or pain intensity) of individuals with chronic LBP following various types of exercise interventions (Table 3) [59,85]. No study has investigated the predictive value of the anticipatory onset of LM morphometric change on the clinical outcomes of patients with non-specific LBP.

There was limited evidence against the role of anticipatory onset of lateral abdominal muscles morphometric change in predicting the short- or long-term clinical outcomes of patients with chronic LBP following various exercise regimens (Table 5) [59,85]. There was no evidence concerning the value of feedforward onset of LM morphometric change in predicting the clinical outcomes of patients with non-specific LBP.
3.13. Effect of chronicity, age, and inception/survival cohorts on predictive ability of TrA/LM

Since the majority of the participants in the included studies should be classified as patients with chronic LBP, the value of morphometry of TrA/LM in predicting clinical outcomes of patients with acute/subacute LBP remains unknown. Similarly, as the mean ages of the participants in the included studies were fairly homogenous, subgroup analysis based on different age groups was precluded. No included study recruited “inception” patient cohorts. As such, no subgroup analysis of “inception” and “survival” cohorts was performed.

4. Discussion (word count 1499)

The primary objective of this review was to summarize the evidence regarding the ability of morphometry, histology or activation of TrA/LM to predict clinical outcomes and LBP recurrence among untreated or conservatively treated patients. There was limited evidence that baseline TrA contraction ratio and anticipatory onset of lateral abdominal muscles morphometric change were not related to the short- or long-term clinical outcomes of patients with chronic non-specific LBP following various exercise interventions. Depending on the type of physical test, there was conflicting evidence regarding the dynamic morphometry of TrA/LM in predicting LBP-related disability or pain reduction in patients with chronic non-specific LBP after various conservative treatments. No identified research
investigated the baseline static morphometry of TrA/LM in predicting clinical outcomes of patients with non-specific LBP. Similarly, there has yet to be a study that examines the value of baseline TrA/LM characteristics in predicting LBP recurrence. Interestingly, limited evidence showed that poorer baseline TrA contraction ratio was a treatment effect modifier favoring motor control exercise over general exercise.

No subgroup analysis (based on age, chronicity, and “inception” or “survival” cohorts) was performed in this review given the absence of relevant subgroup classification and analysis in the primary studies.

4.1. Characteristics of the included studies

4.1.1 Sample size

The sample sizes of the included studies were relatively small. Theoretically, each candidate predictor in a multi-variable regression model requires at least 10 participants to ensure the precision of estimation and to prevent erroneous associations [11,55]. Four times more participants are required if a study aims to estimate the effect of a treatment effect modifier [6]. However, four included studies did not meet these criteria [22,59,85,102]. Conducting regression analysis with suboptimal sample size might reduce the statistical power to identify statistically significant associations [71].
4.1.2. Physical tests

The dynamic morphometry of a muscle measured with B-mode USI is affected by many factors. Although muscle activation may change muscle dimensions, many factors affect the magnitude of the dimensional change. Specifically, dynamic morphometry of a muscle in 2-dimensional ultrasound images is influenced by the resting state of the muscle, the extensibility of the musculotendinous unit, the type of contraction, the competing force from nearby muscles, the out-of-plane muscle motion, the orientation of the ultrasound transducer, and the operator’s experience [97]. As such, the reported correlation coefficient between the magnitude of change in abdominal muscle thickness and the corresponding electromyography ranged from 0.14 to 0.93 [8,36,45,62,96]. Given that four included studies used B-mode USI to measure dynamic morphometry of TrA/LM without reporting their operators’ qualification [19,22,85,102], their findings should be interpreted with caution [95]. Future studies should use electromyography or novel measurement methods [40,66] to measure the activation of trunk muscles [99].

While ADIM was commonly used to test TrA activity [59,85], the performance of ADIM relies on the examiner’s instruction and the participant’s capability to perform the desirable action [52]. Inconsistent ADIM performance might explain the absence of correlation between baseline TrA contraction ratio and LBP intensity at 9-week or 1-year follow-up [59,85]. Instead of using ADIM, assessment of automatic TrA activity using an active straight
leg raise test [52] or an isometric knee flexion/extension task [18] is preferable because these tests eliminate the voluntary component of ADIM.

The great variability in onset timing of lateral abdominal muscles in individuals with [23,89] and without LBP [60,88] may account for the lack of relation between baseline anticipatory onset of these muscles and future clinical outcomes of LBP. Future research should investigate whether the predictive ability of feedforward activation of TrA exists in certain patient subgroups using intramuscular electromyography.

4.1.3. Intervention and follow-up

The treatment intensity of the included studies varied from 2 sessions over 1 week to 9 weeks of once-weekly treatment [19,22,59,85,102]. These results might not be generalizable to treatments of different frequencies and durations. Additionally, some treatments (such as general exercise) in the included studies might improve clinical outcomes through the alternations of self-belief about disability or exercise [44,49,74,93]. Therefore, baseline TrA/LM characteristics might be less likely to predict the clinical outcomes of patients receiving these treatments.

Four included studies investigated the association between predictors and the immediate post-treatment clinical outcomes [19,22,59,102]. Although this might be the best stage to estimate the predictive value of various baseline predictors, it would be imperative to
investigate their predictive ability at a longer-term follow-up.

4.1.4. Adjustment of confounders

We determined that if our evidence synthesis discarded the studies that did not adjust for confounders [22,59], our conclusions would remain. Psychological factors, which have been identified as independent predictors of LBP outcomes [21,59,72], were not considered in most of the included studies [19,22,59,102]. Analyzing data without adjusting for psychological factors may affect estimates of the predictive ability of TrA/LM [83,84].

4.1.5. Statistical analysis

Four included studies [19,22,59,102] analyzed the relation between predictors and clinical outcomes using the absolute values in outcome scales. One study used clinically important pain reduction for analysis [85]; a measure argued to have greater clinical significance although it requires a larger sample size to obtain a significant result.

Two studies used multiple statistical tests [85,102], which increase the risk of false positive findings. Zielinski and coworkers used eight separate linear regression models to investigate whether baseline LM percent thickness change could predict the post-exercise clinical outcomes in 25 patients [102]. Similarly, Unsgaard-Tondel and colleagues used multiple statistical tests and observed only baseline TrA lateral slide was related to the LBP
intensity of patients with chronic LBP at 1-year follow up [85]. Future studies should replicate these experiments with larger sample sizes and a priori hypotheses to minimize the risk of accepting false positive results.

4.2. Strengths of this review

This review has multiple strengths. To improve the coverage of relevant articles, we identified candidate publications in four languages through a systematic search, and contacted the corresponding authors of the included articles and prominent researchers in the area. To enhance the accuracy and consistency of data extraction, we used standardized forms for screening and data extraction, and conducted a pilot exercise. Further, this review protocol was registered in PROSPERO to improve the credibility.

Our risk of bias assessment tool has incorporated the recommended criteria for evaluating prognostic studies and treatment effect modification studies [5,10,25,27]. To provide a comprehensive overview of the assessment results, we reported specific items in detail. We also conducted two sensitivity analyses using 60% and 70% cut-off points to test the robustness of the assessment. The results showed that different cut-off points would not change the conclusions of this review.

4.3. Limitations
Like other systematic reviews, the present review may be subject to publication bias where positive studies are selectively published [1,17]. To optimize the chance of including publications with positive and negative results, a comprehensive search strategy was used. Although a funnel plot was not used in this review given the heterogeneity and small number of included studies [79], the results of our included studies only marginally support or even reject the value of baseline features of TrA/LM in predicting clinical outcomes of patients with LBP. Therefore, our review is unlikely to be affected by publication bias.

4.3. Implications

This review was conducted based on the hypothesis that aberrant changes in morphometry or activation of TrA/LM may affect the progression or recurrence of LBP [32,65]. However, different factors may affect our findings. Firstly, the deficits in TrA/LM may be the consequence of LBP rather than the cause of LBP [30]. This premise is supported by the fact that experimental pain causes changes in anticipatory onset of TrA [35] and dynamic morphometry of TrA/LM [47]. Hence, baseline deficits of TrA/LM may not predict LBP prognosis. Secondly, all included studies recruited participants with chronic non-specific LBP whose prognosis may be affected by multiple factors (such as duration of pain and coping strategy) [93]. Therefore, our results might have been different if the included studies had involved patients with acute/subacute LBP. Further, given the
heterogeneity of non-specific LBP, baseline characteristics of TrA/LM may only predict clinical outcomes of certain patient subgroups. Thirdly, four [19,59,85,102] out of the five included studies [19,22,59,85,102] reported the post-treatment changes in TrA/LM characteristics. However, regardless of the presence or absence of post-treatment TrA/LM characteristic changes, there were only weak associations between baseline TrA/LM characteristics and future clinical outcomes. The weak relations might be attributed to the small sample size or the heterogeneity of survival patient cohorts. Finally, since B-mode USI cannot measure muscle activation, our findings might have been different if the included studies had used electromyography. Overall, despite the conflicting evidence regarding baseline TrA/LM deficits in predicting clinical outcomes of patients with LBP, some longitudinal studies substantiate the deficits of TrA/LM as risk factors for the development of LBP or lower limb injury in healthy individuals (Appendix III).

Given the above, high quality large-scale longitudinal studies with long-term follow-up and adjustment for confounders are needed to clarify the potential role of TrA/LM in predicting the prognosis or recurrence of LBP. Future studies should also investigate whether the post-treatment changes in TrA/LM characteristics accompany the improvement of clinical outcomes.
Summary of this study (word count 23)

There is conflicting evidence regarding the ability of baseline morphometry of TrA/LM to predict clinical outcomes of conservatively treated patients with non-specific LBP.
Acknowledgements

The authors would like to thank librarian Jeanette Buckingham for the design of literature search. We also thank Lydia Dani for assisting the procedures of removing the duplicated citations and the identifiable information of the potential citations. Arnold Wong is supported by the Alberta Innovates-Health Solutions Graduate Studentship and the Golden Key Graduate Scholar Award. Greg Kawchuk is supported by the Canadian Research Chair Program. There are no conflicts of interest.
References


Figure Legend
Figure 1. A flow diagram of the literature search.
Search Strategy

(Number of articles obtained from each electronic database till Dec 2012)

MEDLINE (n = 584)  EMBASE (n = 921)
CINAHL (n = 327)  Cochrane library (n = 102)
PEDro (n = 87)  SPORTDiscus (n = 125)
Clinical trial.gov + Clinical Center Clinical Research Studies + Current Controlled Trials Register (n = 175)
Articles from contacted researchers or authors (n = 4)

Potentially relevant articles identified and screened for retrieval (n = 2325)

Excluded due to duplication (n = 461)
Excluded based on title and abstract (n = 1751)

Potentially appropriate articles for retrieval (n = 113)

Reasons for exclusion (n = 110)
1. Results related to superficial trunk muscles
2. Diagnostic, cross-sectional or case study design
3. Clinical prediction rule derivation only
4. Irrelevant to studying the prognostic/treatment modification effect of transversus abdominis or multifidus
5. Studying risk factors in healthy individuals only
6. Studying diagnosis other than non-specific low back pain
7. Related to temporal changes of function of multifidus or transversus abdominis and clinical outcomes

Initially included articles (n = 3)

Finalized included articles (n = 5)

Potential reference from selected articles (n = 12)
Citation tracking by Scopus/Web of Science (n = 32)
Table 1. Risk of bias assessment tool for prospective cohort studies (modified from Chorti et al., 2009) [9]

1. Patient Population
   a. Case definition
      Operational definition of cases including exclusion criteria  2
      Operational definition of cases but no exclusion criteria  1
      No explicit definition of cases or cannot tell  0
   b. Source population
      Clear description of the source population  1
      Unclear or no description of the source population  0
   c. Representativeness
      Patients representative of clinical practice  2
      Patients unlikely to be representative of clinical practice  1
      Cannot determine  0
   d. Patient selection
      Inception cohort (defined in relationship to onset of symptoms)  2
      Survival cohort, including a subset of the sample with an acute episode/ < 3 months (which is analyzed separately)  1
      Survival cohort; unable to define subsets within the cohort or cannot tell  0
   e. Participants
      Clinical and demographic characteristics described  2
      Insufficient description of participants characteristics  1
      No explicit description of participants characteristics  0

2. Study attrition/Follow up
   a. Follow-up (extent and length)
      Follow-up of ≥ 80% of total sample to at least 1 year as planned  3
      Follow-up of ≥ 80% of total sample for less than 1 year or patients followed for varying lengths of time within 1 year  2
      Follow-up < 80% of total sample completed as planned  1
      Unclear  0
   b. Reasons for loss to follow-up and reporting
      Reasons were provided with descriptive characteristics of this group/ in case of no attrition  2
      Reasons were provided without descriptive characteristics of this group  1
      Unclear  0

3. Predictors (such as prognostic factors/treatment effect modifiers)
   A priori clear definition or description of and rationale for potential predictor(s)  2
   including information on standardization or using reliable and validated measurements
   A clear definition or description of and rationale for predictor(s) but insufficient detail on standardization or validation of measurements, or the predictor(s) was/were not specified in advance.
   Inadequate description of/rationale for potential prognostic factors, or a post-hoc factor  0
4. Treatment details
   Description and standardization and/or randomization of treatment (or description of control for confounding activities for study without treatment)*
   Description of treatment but no standardization or randomization (or description of control for confounding activities for study without treatment)
   No information on the treatment provided*

5. Outcome measures
   Blinded outcome criteria appropriate to the research questions with reports of standardized or valid measurements
   Unblinded outcome criteria appropriate to the research questions
   No explicit outcome criteria (e.g. patient significantly improved)

6. Confounding variables
   Important confounding variables are clearly defined and measured with valid and reliable measurements
   Confounding variables are not defined or considered

7. Analysis
   a. Statistical analysis
      Adjusted proportions provided or appropriate multivariate techniques used to adjust for other prognostic factors*
      Crude proportions but data stratified or presented in a manner which would allow for analysis of subsets
      Crude proportions for at least one response (e.g. remission and/or recurrence) or inadequate sample size
   b. Selective reporting of results
      All the results were presented
      Only some results were presented
   c. Effect size
      Effect sizes was given or sufficient information to calculate effect sizes (e.g. odds ratios, relative risk, correlations, likelihood ratios, significance of the interaction tests)
      Not reported
   d. Standard errors and/or confidence intervals for the estimates
      Standard errors and/or confidence intervals for the estimates was given or sufficient information was given to calculate them
      Not reported

* Score 0 for a treatment effect modifier if there is no randomization
# > 10 events per independent variable per prognostic factor; > 40 events per independent variable for treatment effect modifier; more participants are needed as the number and the correlation of covariates increase.
Table 2. Level of evidence (adopted from Cornelius et al., 2010) [11].

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Consistent results (≥ 80%) from at least two high quality studies</td>
</tr>
<tr>
<td>Moderate</td>
<td>One high quality study and consistent findings (≥ 80%) in one or more low quality studies</td>
</tr>
<tr>
<td>Limited</td>
<td>Findings in one high quality cohort or consistent results (≥ 80%) among low quality studies</td>
</tr>
<tr>
<td>No</td>
<td>No study identified</td>
</tr>
<tr>
<td>Conflicting</td>
<td>Inconsistent results irrespective of study quality</td>
</tr>
</tbody>
</table>
Table 3. Data extraction of included studies reporting the functions of transversus abdominis, lumbar multifidus or lateral abdominal muscles of lumbar spine and the prediction of clinical outcomes in patients with low back pain

<table>
<thead>
<tr>
<th>Studies</th>
<th>Study population; average age and standard deviation; follow-up</th>
<th>Intervention</th>
<th>Muscle (prognostic factors/treatment modifiers)</th>
<th>Outcome measure</th>
<th>Confounder adjustment</th>
<th>Relations between baseline muscle function and future clinical outcomes (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferreira et al., 2010</td>
<td>Patients with chronic non-specific low back pain recruited from physical therapy departments at three teaching hospitals (n=34); 45.4 ± 17.3 years for motor control group, 54.9 ± 11.3 years for general exercise group, 45.4 ± 17.7 years for spinal manipulation group; 8 weeks</td>
<td>A randomized controlled trial. Each group received 12 sessions over 8 weeks: (1) Motor control exercises: training function of specific deep muscles of lumbar region, coordination of deep trunk muscles and pelvic floor muscles with diaphragmatic respiration pattern, incorporation of deep trunk muscles contraction during functional tasks, home exercises (2) General exercises: strengthening and stretching of the main muscles of the body, cardiovascular fitness, education, home exercises (3) Spinal manipulative therapy: joint mobilization but not thrust manipulation techniques to the spine or pelvis</td>
<td>Percentage thickness change of TrA at baseline as measured by B-mode USI during an isometric knee flexion/extension task (Treatment effect modifiers)</td>
<td>A) Global impression of recovery</td>
<td>Corresponding baseline of clinical outcomes</td>
<td>Linear regression to compare the effect of motor control exercise (vs general exercise) on clinical outcomes using baseline percentage thickness change of TrA as an independent variable A) interaction effect*: -19.9 (-45.3 to 5.5); p=0.116 B) interaction effect*: 16.1 (-34.0 to 66.2); p=0.506 C) interaction effect*: -13.7 (-67.2 to 39.9); p=0.597 D) interaction effect*: 18.1 (1.2 to 34.9); p=0.037</td>
</tr>
<tr>
<td>Fritz et al., 2011</td>
<td>Patients with low back pain with or without leg symptoms recruited from physical therapy clinics and community-based advertisements (n=50); 33.3 ± 12.9 years; 1 week</td>
<td>An uncontrolled study. Each patient received two sessions of spinal manipulative therapy 3-4 days apart.</td>
<td>Baseline LM percent thickness change as measured by B-mode USI during CALT in prone (A prognostic factor)</td>
<td>Modified ODI</td>
<td>No adjustment for any confounders</td>
<td>Zero-order correlation between immediate post-spinal manipulation change in LM percent thickness change in the first session and percentage ODI improvement in the first week r= -0.013 (non-significant)</td>
</tr>
</tbody>
</table>
Mannion et al., 2012 [61]

Patients with chronic non-specific low back pain with or without referred pain recruited from rheumatology, orthopedics and neurology of participating hospitals (n=37); 44 ± 12.3 years; 9 weeks

An uncontrolled study. Each patient received 1 treatment session per week for 9 weeks.

Motor control exercise: training function of specific deep muscles of lumbar region, coordination of deep trunk muscles and pelvic floor muscles with diaphragmatic respiration pattern, incorporation of deep trunk muscles contraction during functional tasks, education, home exercises

Lateral abdominal muscle function at baseline:
1) voluntary TrA-CR during ADIM as measured M-mode USI
2) Anticipatory onset of lateral abdominal muscles (TrA, OI, OE) during rapid arm movement as measured by TDI

RMDQ:

No adjustment for any confounders

Unsgaard-Tondel et al., 2012 [85]

Participants with chronic non-specific low back pain recruited from general practitioner or physiotherapists and by advertising among staff at a local hospital (n=87); 40.2 ± 11.2 years for patients who experienced clinically important pain reduction and 41.2 ± 11 years for patients who did not have clinically important pain reduction; 1 year

A randomized controlled trial. Each group received weekly sessions over 8 weeks:

1) Motor control exercises (40 mins each): training function of TrA, LM and pelvic floor muscles, contraction of deep trunk muscles and pelvic floor muscles, incorporation of deep trunk muscles contraction during functional tasks, home exercises

2) General exercises (1 hour): general trunk strengthening and stretching of main muscles of the body, cardiovascular fitness, education, home exercises

3) Sling exercises (40 mins each): performing different progressive back exercises in slings that hung down from the ceiling. Participants needed to maintain a neutral spine

Baseline abdominal muscle function:
(i) TrA lateral slide during ADIM as measured by B-mode USI
(ii) TrA-CR10 during ADIM as measured by B-mode USI
(iii) OI-CR10 during ADIM as measured by B-mode USI
(iv) Anticipatory onset of abdominal muscles during rapid arm movements as measured by M-

NPRS

In linear regression model A, age, initial fear avoidance for physical activity, gender, body mass index, initial pain intensity and pain duration were adjusted for.

In linear regression model B, age, initial pain level and pain duration were adjusted for.

In logistic regression model C, age, initial fear avoidance for

The regression results were obtained from the pooled sample of all treatment groups.

A) linear regression of pain at 1-year follow-up in relation to:
Baseline (i) β = 0.13 (0.00 to 0.26); R²=0.036
Baseline (ii) β = -0.03 (-0.13 to 0.09); R²=0.002
Baseline (iii) β = 0.00 (-0.22 to 0.22); R²=0.000
Baseline (iv) β = 0.00 (-0.03 to 0.03); R²=0.001

B) linear regression of pain at 1-year follow-up in relation to both baseline and change in (i), (iii) and (iv), the baseline variables (i, iii and iv) explained 35% of total variance of pain at 1-year follow-up. P value for this model was not reported.

C) logistic regression of clinically important reduction in pain (NPRS ≥
position during various sling exercises to activate both deep and superficial trunk muscles, home exercises

mode USI (Prognostic factors)

physical activity, gender, body mass index, and pain duration were adjusted for. In logistic regression model D, only age and pain duration were adjusted for.

2) at 1-year follow-up in relation to:
Baseline (i) OR = 0.76 (0.62 to 0.93)
Baseline (ii) OR = 1.05, (0.92 to 1.20)
Baseline (iii) OR = 0.86 (0.65 to 1.13)
Baseline (iv) OR = 0.99 (0.96 to 1.03)

D) logistic regression of clinically important reduction in pain at 1-year follow-up in relation to both baseline and change in (i), (iii) and (iv), only baseline (i) was statistically significantly associated with improved pain level, OR = 0.75 (0.57 to 0.98)

Zielinski et al., 2012 [98]

Participants with chronic low back pain with or without recurrences recruited from local physical therapy and physician clinics and/or from self-referral; for CPR-eligible patients (n =11) 34.3 ± 10.8 CPR-ineligible patients (n = 14) 42.1 ± 10 years for; 6 weeks

One arm of a randomized controlled study. Each participants received 1 treatment session per week over 6 weeks

Motor control exercise: training function of specific deep muscles of lumbar region, strengthening trunk flexors, extensors, and oblique muscles, education on proper body biomechanics and spinal protection during activities of daily living, home exercises

Percent thickness change of LM thickness as measured by B-mode USI during a resisted or non-resisted CALT in prone (A prognostic factor)

Modified ODI; NPRS

Participants’ initial ODI and NPRS scores.

- Linear regression of post-treatment change of ODI in relation to baseline percent change of LM thickness in CPR-eligible patients
  r = 0.36, p=0.06 (no resistance task)
r = 0.14, p=0.39 (resistance task)
In CPR-ineligible patients
  r = 0.014, p=0.65 (no resistance task)
r = 0.38, p=0.013 (resistance task)
- Linear regression of post-treatment change of NPRS in relation to baseline percent change of LM thickness in CPR-eligible patients
  r = 0.094, p=0.89 (no resistance task)
r = 0.13, p=0.81 (resistance task)
In CPR-ineligible patients
  r = 0.065, p=0.94 (no resistance task)
r = 0.094, p=0.08 (resistance task)

*Interaction effect indicates whether the post-treatment pain intensity following motor control exercise (in contrast to general exercise) is affected by the participant’s percent thickness change of TrA at baseline. A positive interaction effect means that lower baseline percent thickness change of TrA has a better pain reduction if the participant receives motor control exercise rather than general exercise.

ADIM = abdominal drawing-in maneuver; B-mode = brightness mode; CALT = contralateral arm lifting task; CPR = clinical prediction rule; CPR-eligible patients = patients who either had lumbar spine hypomobility and/or had any three of four clinical features identified by the clinical prediction rule for
stabilization exercise success; CPR-ineligible patients = patients who neither had lumbar spine hypomobility nor had any three of four clinical features identified by the clinical prediction rule for stabilization exercise success; LM = lumbar multifidus; M-mode = motion mode; NPRS = numeral pain rating scale; ODI = Oswestry disability index; OE = obliquus externus; OI = obliquus internus; OI-CR10 = oblique internus thickness contraction ratio measured as: (thickness of oblique internus thickness at maximum contraction/oblique internus resting thickness) x 10; OR = odds ratio; RMDQ = Roland Morris disability questionnaire; TDI = tissue Doppler imaging; TrA = transversus abdominis; TrA-CR = transversus abdominis thickness contraction ratio measured as: (thickness of transversus abdominis at maximum contraction/transversus abdominis resting thickness); TrA-CR10 = transversus abdominis thickness contraction ratio measured as: (thickness of transversus abdominis at maximum contraction/transversus abdominis resting thickness) x 10; USI = ultrasound imaging.
Table 4. Risk of bias assessment scores of the included studies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Case definition</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1b. Source population</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1c. Representativeness</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1d. Patient selection</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1e. Participants</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2a. Follow-up</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2b. Reasons for attrition</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3. Prognostic factors/Modifiers</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4. Treatment details</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5. Outcome measures</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6. Confounders</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7a. Analysis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7b. Selective reporting of results</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7c. Effect size</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7d. Standard errors/confidence intervals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total scores</td>
<td>19</td>
<td>15</td>
<td>16</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>Risk of bias (cut-off at 50%)*</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Risk of bias (cut-off at 60%)*</td>
<td>low</td>
<td>high</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Risk of bias (cut-off at 70%)*</td>
<td>low</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>

*Cut-off score at 50%: 13; Cut-off score at 60%: 16; Cut-off score at 70%: 18
Table 5. Level of evidence for conclusions on the neuromuscular function of transversus abdominis or lumbar multifidus as a treatment effect modifier and prognostic factor with study quality appraisal cut-off point at 50%

<table>
<thead>
<tr>
<th>Prognostic factors or treatment moderators</th>
<th>Clinical outcomes</th>
<th>Number of cohort studies with positive or no results and risk of bias of the studies (high/low)</th>
<th>Level of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td>Factors identified</td>
<td>Positive results</td>
<td>No results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>TrA</td>
<td>Percentage thickness change of TrA during an isometric knee flexion/extension task*</td>
<td>RMDQ</td>
<td>Patient-specific functional scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Global perceived recovery</td>
<td>VAS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMDQ</td>
<td>(short-term)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPRS</td>
<td>(long-term)</td>
</tr>
<tr>
<td>Lateral abdominal muscles</td>
<td>Anticipatory onset of lateral abdominal muscles prior to arm elevation</td>
<td>RMDQ</td>
<td>Patient-specific functional scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPRS</td>
<td>(short-term)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPRS</td>
<td>(long-term)</td>
</tr>
<tr>
<td>LM</td>
<td>LM percent thickness change during CALT+</td>
<td>Modified ODI</td>
<td>1[98]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPRS</td>
<td>1[98]</td>
</tr>
</tbody>
</table>

* Percentage thickness change of TrA measured as change in muscle thickness as percentage of resting thickness
+ Lumbar multifidus percent thickness change expressed as change in muscle thickness as percentage of resting thickness

No result means that the result did not support the factor as a treatment modifier or a prognostic factor
ADIM = abdominal drawing-in maneuver; CALT = contralateral arm lifting task; conflicting = conflicting level of evidence; high = cohort study with high risk of bias; limited = limited level of evidence; low = cohort study with low risk of bias; LM = lumbar multifidus; inconclusive = contradictory findings regarding the factor as a treatment modifier or prognostic factor; NPRS = numeric pain rating scale; ODI = Oswestry disability index; PGRS = Pain graphic rating scale; positive = results supported the factor as a treatment modifier or a prognostic factor; RMDQ = Roland Morris disability questionnaire; strong = strong level of evidence; TrA-CR = transversus abdominis thickness contraction ratio measured as: (thickness of transversus abdominis at maximum contraction/transversus abdominis resting thickness); TrA-CR10 = transversus abdominis thickness contraction ratio measured as: (thickness of transversus abdominis at maximum contraction/transversus abdominis resting thickness) x 10.
Appendix I. Keywords for literature search and search strategies

<table>
<thead>
<tr>
<th>Related items</th>
<th>Key Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>Explode low back pain (MeSH), lower back (MeSH), LBP, lumbar pain (MeSH), lumbar pain (MeSH), lumbago (MeSH), lumbalgia (MeSH), backache (MeSH), explode sciatica (MeSH)</td>
</tr>
<tr>
<td>Lumbar multifidus and Transversus abdominis</td>
<td>Lumbar multifidus (MeSH), SM, LM, transversus abdominis (MeSH), TrA, lumbar muscle (MeSH), back muscle (MeSH), muscle activity, core muscle, erector spinae (MeSH), stabilizing musc* (MeSH), muscular (MeSH)</td>
</tr>
<tr>
<td>Study designs</td>
<td>Explode cohort studies (MeSH), incidence (MeSH), explode mortality (MeSH), follow-up studies (MeSH), mortality (MeSH), prognos*, predict*, course, explode randomized control*, RCT, course</td>
</tr>
</tbody>
</table>

**Medline search strategies:**

1. (lumbo* or lumbar).mp. [mp=title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]
2. exp Back/ or back.mp. or exp Lumbar Vertebrae/ or exp Lumbosacral Region/ or exp Coccyx/ or coccyx.mp. or torso.mp. or exp Torso/
3. 1 or 2
4. muscle, skeletal.mp. or exp Muscle, Skeletal/ or muscul*.mp. or exp Muscular Atrophy/
5. 3 and 4
6. (paraspinal or multifid* or back muscle or back extensor*).mp.
7. (LM or MF).mp. [mp=title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]
8. 3 and 7
9. 5 or 6 or 8
10. exp Abdominal Muscles/ or transversus abdominis.mp. or transverse abdominis.mp. or rectus abdominis.mp. or exp Rectus Abdominis/ or TrA.mp. or external oblique*.mp. or internal oblique*.mp.
11. (core or (stabilization or stabilizing)).mp.
12. 5 and 11
13. 10 or 12
14. back pain.mp. or exp Back Pain/ or exp Spinal Diseases/ or lumbalgia.mp. or exp Low Back Pain/ or LBP.mp. or sciatica.mp. or exp Sciatica/ or (backache or lumbago or dorsalgia or lumbar pain).mp.
15. prognosis.mp. or exp Prognosis/ or incidence.mp. or exp Incidence/ or morbidity.mp. or exp Morbidity/ or exp Risk Factors/ or (predict* or course or prognos*).mp.
16. exp Follow-Up Studies/ or followup.mp. or follow-up.mp. or exp Cohort Studies/ or cohort.mp. or exp Longitudinal Studies/ or longitudinal.mp. or exp Prospective Studies/ or prospective.mp. or exp Retrospective Studies/ or retrospective.mp.
17. 15 and 16
18. 9 and 14 and 17
19. limit 18 to ("young adult (19 to 24 years)" or "adult (19 to 44 years)" or "young adult and adult (19-24 and 19-44)" or "middle age (45 to 64 years)" or "middle aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)") and (Chinese or English or French or Portuguese))

20. 13 and 14 and 17

21. limit 20 to ("young adult (19 to 24 years)" or "adult (19 to 44 years)" or "young adult and adult (19-24 and 19-44)" or "middle age (45 to 64 years)" or "middle aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)") and (Chinese or English or French or Portuguese))

EMBASE search strategies
1. lumbar.mp. or exp lumbar vertebra/ or exp lumbar spine/ or exp back/ or exp lumbosacral spine/ or lumbo*.mp. or lumbosacral.mp. or ilium.mp. or exp iliac bone/ or coccyx.mp. or exp coccygeal bone/ or exp trunk/ or trunk.mp. or torso.mp.

2. muscle atrophy/ or muscle contraction/ or muscle.mp. or exp muscle/ or exp muscle function/ or muscul*.mp. or exp musculoskeletal system/

3. 1 and 2

4. paraspinal.mp. or exp back muscle/ or back extensor*.mp. or multifid*.mp.

5. (LM or MF).mp.

6. 2 and 5

7. 3 or 4 or 6

8. exp rectus abdominis muscle/ or exp abdominal wall musculature/ or transverse abdominis.mp. or transversus abdominis.mp. or rectus abdominis.mp. or TrA.mp. or external oblique*.mp. or internal oblique*.mp.

9. (core or (stabilization or stabilizing)).mp. or exp spine stabilization/

10. 2 and 9

11. 8 or 10

12. back pain.mp. or exp backache/ or lumbago.mp. or exp low back pain/ or LBP.mp. or (dorsalgia or lumbalgia).mp. or sciatica.mp. or exp ischialgia/

13. survival analysis.mp. or exp survival/ or exp mortality/ or mortality.mp. or morbidity.mp. or exp morbidity/ or follow-up.mp. or followup.mp. or exp follow up/ or exp longitudinal study/ or longitudinal.mp. or exp cohort analysis/ or cohort.mp. or risk factor.mp. or exp risk factor/ or exp prediction/ or predict*.mp. or progno*.mp. or course.mp. or exp prognosis/

14. (follow-up or followup).mp. or exp follow up/ or exp longitudinal study/ or longitudinal.mp. or exp cohort analysis/ or cohort.mp. or prospective.mp. or exp prospective study/ or exp retrospective study/ or retrospective.mp.

15. morbidity.mp. or exp morbidity/ or risk factor.mp. or exp risk factor/ or exp prediction/ or predict*.mp. or progno*.mp. or course.mp. or exp prognosis/ or exp incidence/ or incidence.mp.

16. 14 and 15

17. 7 and 12 and 16

18. limit 17 to ((chinese or english or french or Portuguese) and (adult <18 to 64 years> or aged <65+ years>))

19. 11 and 12 and 16

20. limit 19 to ((chinese or english or french or Portuguese) and (adult <18 to 64 years> or
aged <65+ years>)

CINAHL search strategies
1. MM "Lumbar Vertebrae" OR Lumbar OR back OR MM "Back" OR trunk OR MM "Torso" OR spine OR MM "Spine+" OR MM "Coccyx" OR lumbosacral OR MM "Lumbosacral Plexus+
2. MM "Musculoskeletal System+" OR MM "Muscles+" OR MM "Muscle, Skeletal+" OR muscle*
3. S1 and S2
4. (paraspinal) and (S2)
5. (MM "Multifidus Muscles") OR "multifid*"
6. MF OR back extensor* OR LM
7. S3 or S4 or S5 or S6
8. ( (MM "Rectus Abdominis Muscles") OR "transversus abdominis" ) OR MM "Abdominal Muscles+" OR abdominal muscle*
9. (core) and (S2)
10. (stabiliz*) and (S2)
11. transverse abdominis OR TrA OR internal obliqu* OR external obliqu*
12. S8 or S9 or S10 or S11
13. ( (MM "Sciatica") OR "sciatica" ) OR MM "Low Back Pain" OR MM "Pelvic Pain+" OR MM "Chronic Pain" OR MM "Back Pain+" OR lumbar pain OR back pain OR LBP OR dorsalgia OR lumbago OR lumbalgia OR backache
14. "course" OR ( "Predict*" OR ( ((MM "Risk Factors+") OR "risk factor" ) ) OR ( ((MM "Morbidity+") OR "morbidity" OR (MM "Comorbidity") ) OR ( (MM "Prognosis+") OR "prognos*" ) OR ( (MM "Incidence") OR "Incidence" ) ) )
15. ( (MH "Randomized Controlled Trials") OR "randomized controlled" ) OR ((MM "Concurrent Prospective Studies") OR "cohort" OR (MM "Nonconcurrent Prospective Studies") ) OR longitudinal OR (( "follow-up" OR (MM "Prospective Studies+") ) OR followup ) OR longitudinal study OR ( (MM "Retrospective Design") OR "retrospective" ) OR "cohort" OR cohort study )
16. S14 and S15
17. S7 and S13 and S16
18. S7 and S13 and S16 Limiters - Age Groups: All Adult
19. S12 and S13 and S16
20. S12 and S13 and S16 Limiters - Age Groups: All Adult

Sport Discus search strategies
1. lumbar OR back OR trunk OR torso OR spine OR lumbosacral OR Lumbar vertebrae OR coccy* OR iliac OR ilium
2. Musculoskeletal OR muscle*
3. S1 and S2
4. MF
5. ( back extensor* OR multifid* OR LM ) OR ( paraspinal AND ( Musculoskeletal OR muscle* ) )
6. S3 or S4 or S5
7. internal obliqu* OR external obliqu* OR TrA OR ((rectus abdominis OR transversus abdominis OR transverse abdominis) OR abdominal muscle* OR (abdomen AND (Musculoskeletal OR muscle*))) OR (core AND (Musculoskeletal OR muscle*)) OR (stabiliz* AND (Musculoskeletal OR muscle*))

8. backache OR sciatica OR LBP OR dorsalgia OR lumbago OR lumbalgia

9. pain AND (lumbar OR back OR trunk OR torso OR spine OR lumbosacral OR Lumbar vertebrae OR coccy* OR iliac OR ilium)

10. S8 or S9

11. retrospective OR prospective OR followup OR follow-up OR follow up OR longitudinal

12. (predict* OR morbidity OR comorbidity OR prognos* OR risk factor* OR) OR course

13. S11 and S12

14. S6 and S10 and S13

15. S7 and S10 and S13

Other search strategies for PEDro and Cochrane Library:
(Low back pain OR LBP OR lumbago OR lumbalgia OR backache) and (lumbar multifidus OR LM OR transversus abdominis OR TrA OR SM) and (prognos* OR predict* OR longitudinal OR cohort OR course OR survival analysis OR follow-up)
Appendix II. Inclusion and exclusion criteria

Inclusion Criteria

Type of study design: Systematic review/meta-analysis of cohort studies or follow-up studies; validation of clinical decision rule; follow-up of untreated control groups in randomized controlled trials; retrospective cohort studies; observational studies; case series with > 10 subjects; non-duplicated information from multiple reports of the same patient cohorts

Articles: English, Chinese or Portuguese language; non-English, non-Chinese and non-Portuguese language studies in English-language review articles; inclusion of morphology measurement or muscle activity of LM or TrA, and at least one clinical outcome of LBP.

Subjects: self-ambulatory adult patients (>18 years of age) with acute (< 6 weeks), subacute (6 to 12 weeks) or chronic (> 12 weeks) nonspecific LBP (Hayden et al., 2005)

Personal characteristics: human subjects of any gender, BMI and ethnicity

Causes of pain: idiopathic; traumatic without vertebral fracture, degenerative; experimentally induced

Setting: in-patient; outpatient clinics; laboratories

Treatment: non-surgical treatment with explicit description of therapeutic dosage (e.g. drugs, physical therapy modalities or manual therapy), or studies without intervention

Follow-up: at least one follow-up after the baseline measurements (with reported timing for follow-up)

Prognostic or treatment effect modifiers:
Morphology or neuromuscular function of LM or TrA

Type of outcome measures:
At least one of the following LBP clinical outcome measures: Quebec Back Pain Disability scale; Roland Morris disability questionnaire (RMDQ); Oswestry Disability Index (ODI); Japanese orthopedic association score; McGill pain questionnaire; Patient-Specific Functional Scale; pain score (e.g. visual analog scale, numerical pain rating scale); subjective improvement of symptoms (global rating of change); return-to-work/school; days off work/school; recurrent rate; quality of life (e.g. short-form 36, short-form 12, EQ-5D, etc.); subjective complaints

LM or TrA measures: magnetic resonance imaging; computed tomography; ultrasound imaging; electromyography (intramuscular or surface)

Measurement period: multiple assessments over different time periods in prospective or retrospective studies

Statistical analysis: odds ratio; risk ratio; linear or logistic regression; multiple regression; cross tabulations; chi-squared test; Kaplan-Meier curves; cox regression; ANOVA; Wilcoxon rank sum test; Kruskal Wallis test

Exclusion criteria

Designs: single cross-sectional studies; multiple reports that included duplicated results of same patient case series (only included the study with the largest series)

Medical conditions: pregnancy related back pain; LBP patients from sources outside the back (non-spinal LBP); fibromyalgia or myofascial pain syndrome in >20% of the overall sample size; spondylolisthesis; spondylolysis; stenosis; disc herniation;
back surgery; cancer; neural disease (e.g. multiple sclerosis or Guillain Barre Syndrome); acute major trauma; spinal cord injury; congenital lumbar/sacral condition; spasticity resulting from spinal cord injury or neural diseases; cauda equina syndrome; infection; solely osteoporosis; vertebral compression fracture; rheumatoid arthritis; scoliosis; stroke; cerebral palsy

Articles: papers written in non-English, non-Chinese, non-French and non-Portuguese; letters; studies in books
Appendix III. Longitudinal cohort studies reporting the ability of neuromuscular functions or morphology of transversus abdominis, abdominal muscles or lumbar multifidus in asymptomatic participants in predicting the risk of future low back pain or lower limb injuries

<table>
<thead>
<tr>
<th>Studies</th>
<th>Study population and follow-up</th>
<th>Muscle function (risk factor)</th>
<th>Outcome measure</th>
<th>Confounders adjustment</th>
<th>Results (adjusted for confounders)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hides et al., 2001 [33]</td>
<td>Pain free individuals recovered from acute LBP following 4-week of medical management or 4-week motor control exercises; 1 year and 3 years</td>
<td>CSA asymmetry of LM in Motor control exercise group (MC): 0.7% ± 2.5% Medical management group (MM): 16.8% ± 9.3%</td>
<td>Recurrence episode Severity Treatment sought</td>
<td>No adjustment for any confounders</td>
<td>At year 1: MM were 12.4 times more likely to experience recurrence of LBP than MC (P&lt;0.001) Likelihood of recurrence in MC: 25% Likelihood of recurrence in MM: 80% “as severe as” the original episode in MC: 33% “as severe as” the original episode in MM: 56% Treatment sought by MC: 15% Treatment sought by MM: 42% At year 2-3: MM were 9.3 times more likely to experience recurrence of LBP than MC (P&lt;0.01) Likelihood of recurrence in MC: 28.6% Likelihood of recurrence in MM: 76.9% “as severe as” the original episode in MC: 14% “as severe as” the original episode in MM: 17% Treatment sought by MC: 20% Treatment sought by MM: 25%</td>
</tr>
<tr>
<td>Moseley, 2004 [67]</td>
<td>Patients with neck pain of more than 4 months and less than 1-year duration, with or without radiating shoulder and/or arm pain recruited from a physiotherapy clinic (n=50). Age-matched controls were recruited by an advertisement on a notice board (n=50); 2 years</td>
<td>Baseline deep abdominal muscle contraction during ADIM as categorized by PBU: ≥ 4mmHg (normal) 2-4mmHg (uncertain) &lt; 2mmHg (abnormal)</td>
<td>Occurrence of chronic or recurrent LBP at 2-year telephone follow-up</td>
<td>No adjustment for any confounders</td>
<td>Logistic regression for developing chronic or recurrent LBP at 2 years (DV) as predicted by ADIM category (IV) category: β = 0.562 and Exp(β) = 1.755; p=0.02 Patients who reported LBP at 2-year follow-up (Fisher’s exact test, p&lt;0.03): 14% of patients with normal ADIM category 25% of patients with uncertain ADIM category 74% of patients with abnormal ADIM category reported LBP at 2 years Controls who reported LBP at 2-year follow-up (Fisher’s exact test, p&lt;0.04): 11% of controls with normal ADIM category 14% of controls with uncertain ADIM category 75% of controls with abnormal ADIM category</td>
</tr>
<tr>
<td>Roussel et al., 2009 [75]</td>
<td>Students in a full-time professional dance program (n=32); every 2 weeks during a 6-month follow-up</td>
<td>Ability of deep abdominal muscles control as measured by PBU (in mmHg) using:</td>
<td>Occurrence of low back or lower limb injuries</td>
<td>No adjustment for any confounders</td>
<td>A stepwise conditional logistic regression with: DV = development of injuries to lumbar spine or lower limbs; IV = KLAT (Exp(β) = 0.588, p = 0.015) IV = SB (Exp(β) = 8.782, p = 0.029)</td>
</tr>
<tr>
<td>Study</td>
<td>Population</td>
<td>Methodology</td>
<td>Outcome Measures</td>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>-------------</td>
<td>------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Paalanee et al., 2011 ([68])</td>
<td>Adolescents responded to a postal questionnaire (n=337); 3 years (from 18 to 21 years old)</td>
<td>CSA and fat infiltration of LM as measured by radiologist on MRI images</td>
<td>Pain clusters classification based on LBP intensity, LBP duration, pain medication usage, frequency of physician visits, sports restriction and LBP episode: (1) No pain (2) Minor pain (3) Moderate pain (4) Recent onset pain (5) Always painful</td>
<td>No adjustment for any confounders</td>
<td></td>
</tr>
<tr>
<td>Hides et al., 2011 [32]</td>
<td>Professional male Australian Rules Football players (n=34); 5 months during preseason period</td>
<td>MRI measurement of CSA of LM; Change of CSA of trunk during ADIM using MRI (at rest and during contraction) (i) TrA lateral slide (ii) TrA contraction thickness ratio (iii) OI contraction thickness ratio (iv) first onset of abdominal muscles</td>
<td>Incidence of injury that prevent an athlete from completing a full training session. Severity of HGT injury that caused missing training sessions (A) 0 (B) 1-3 (C) &gt;4</td>
<td>Adjusted for age, height and weight</td>
<td>Difference in CSA of LM between players with different severity of injuries: (A) &gt; (C) at L3 to L5 level (P&lt;0.01) (B) did not differ from (C) at L3 to L5 level (p&gt;0.05) CSA of LM did not change as a result of the injury (P&gt;0.05) Baseline LM CSA had high specificity in predicting grade (C) injury versus grade (A) or (B) injury Baseline LM CSA at L5 level had high specificity (96.4) and sensitivity (83.3) in predicting grade (C) injury</td>
</tr>
</tbody>
</table>

ADIM = abdominal drawing-in maneuver; CPR = clinical prediction rule; CSA = cross-sectional area; DV = dependent variable; HGT = hip-groin-thigh; IV = independent variable; LM = lumbar multifidus; MRI = magnetic resonance imaging; PBU = pressure biofeedback unit; TrA = transversus abdominis.
Abstract

Although individual reports suggest that baseline morphometry or activity of transversus abdominis or lumbar multifidus predict clinical outcome of low back pain (LBP), a related systematic review is unavailable. Therefore, this review summarized evidence regarding the predictive value of these muscular characteristics. Candidate publications were identified from six electronic medical databases. After review, 5 cohort studies were included. Although this review intended to encompass studies using different muscle assessment methods, all included studies coincidentally used ultrasound imaging. No research investigated the relation between static morphometry and clinical outcomes. Evidence synthesis showed limited evidence supporting poor baseline transversus abdominis contraction thickness ratio as a treatment effect modifier favoring motor control exercise. Limited evidence supported that high baseline transversus abdominis lateral slide was associated with higher pain intensity following various exercise interventions at 1-year follow-up. However, there was limited evidence for the absence of relation between the contraction thickness ratio of transversus abdominis or anticipatory onset of lateral abdominal muscles at baseline and the short- or long-term LBP intensity after exercise interventions. There was conflicting evidence for a relation between baseline percent thickness change of lumbar multifidus during contraction and the clinical outcomes of patients following various conservative treatments. Given study heterogeneity, the small number of included studies, and the inability of conventional grey-scale B-mode ultrasound imaging to measure muscle activity, our findings should be interpreted with caution. Further large-scale prospective studies, employing the appropriate technology (i.e. electromyography to assess muscle activity), should be conducted to investigate the predictive value of morphometry or activity of these muscles with respect to LBP-related outcomes measures.