The validity of a clinical test for the diagnosis of lumbar spinal stenosis

Robert Dobbs, Stephen May, Philip Hope

PII: S1356-689X(16)30357-5
DOI: 10.1016/j.math.2016.05.332
Reference: YMATH 1860

To appear in: Manual Therapy

Received Date: 11 January 2016
Revised Date: 23 May 2016
Accepted Date: 25 May 2016


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 page

The validity of a clinical test for the diagnosis of lumbar spinal stenosis

Robert Dobbs¹, Stephen May²*, Philip Hope¹

¹ Physiotherapy Department, Chesterfield Royal Hospital, Chesterfield, Derbyshire

² 38 Collegiate Crescent, Faculty of Health and Wellbeing, Sheffield Hallam University, Sheffield, S10 2BP

s.may@shu.ac.uk

0114 2252370

*Corresponding author
The validity of a clinical test for the diagnosis of lumbar spinal stenosis

Abstract

Background: The diagnosis and management of acquired lumbar spinal stenosis (ALSS) is an area of growing interest with an increase in its prevalence and detection in the older population.

Objectives: To investigate the diagnostic accuracy of a modified extension test (MExT) for diagnosing ALSS in subjects aged fifty or over.

Methods: Symptomatic response of the bi-component MExT was evaluated and compared against MRI findings in 30 subjects. Estimates of sensitivity, specificity, likelihood ratios and post-test probabilities were all calculated, and the capability of the test to discriminate between grade and location of stenosis was also appraised.

Results: MExT sensitivity was high at 92% (95% CI, 72-99%) leading to a significant negative likelihood ratio at -LR 0.2 (95% CI, 0.03-1.36); conversely, specificity was low at 40% (95% CI, 7-82%) with only a small positive likelihood ratio of +LR 1.53 (95% CI, 0.74-3.16). All correlations between the MExT and concurrent grade, or location of stenosis appeared weak and insignificant.

Conclusions: The MExT was found to demonstrate acceptable criterion validity in relation to ruling-out a diagnosis when a negative result was observed; however, further validation appears warranted.

Words 184

Keywords: Lumbar, Spinal stenosis, Extension Test, Diagnosis, Validity
The validity of a clinical test for the diagnosis of lumbar spinal stenosis

1. INTRODUCTION

World epidemiological projections indicate that the number of people within the population aged sixty-five or over will have increased from sixteen to twenty five per cent by the year 2040 (Population Reference Bureau, 2010; Maloney-Backstrom et al., 2011). In response to such demands, health research agendas have increasingly shifted focus towards the management of degenerative conditions proposing direct efficiency savings through enhanced examination and treatment rigour (IAGG, 2007). One such condition gaining attention is acquired lumbar spinal stenosis (ALSS). Point prevalence estimates suggest degenerative lumbar conditions may afflict up to twenty per cent of the older-aged population (Keller et al., 2003; Lyle et al., 2005). Of those individuals referred to an orthopaedic spinal specialist, approximately fourteen per cent are found to exhibit severe stenotic change requiring surgical decompression (Aalto et al., 2006; Slatis et al., 2007; Watters et al., 2008). As such, ALSS has fast become the leading cause of spinal surgery in this population (Mannion et al., 2010; Steurer 2010), whilst its deleterious influence on locomotor capability and psychosocial aspects of health renders it a precursor for falls, depression and cardiovascular disease (Middleton and Fish, 2009; Kim et al., 2011).

The clinical syndrome of ALSS develops as a consequence of incremental damage to spinal tissues. Narrowing of the intervertebral disc reduces the distance between adjacent vertebrae, whilst an alteration to biomechanical force initiates arthritic change and ligamentous laxity (Fredrickson et al., 2001; Vo et al., 2005; Papadakis et al., 2011). Subsequent thickening of the ligamentum flavum, pedicles or vertebral facets diminish the space available for neural and vascular structures, with ischaemic change ensuing within the spinal canal (central stenosis), nerve root canal, or intervertebral foramina (foraminal stenosis) (Vo et al., 2005). Sufferers of ALSS present with symptoms of lower extremity pain or paraesthesia, occurring with or without back pain especially with positions of
extension, which limit walking capacity secondary to neurogenic claudication (Bal et al. 2006; Malmivaara et al., 2007; Ogikubo et al., 2007).

1.1 Diagnostic methods

However no widely accepted diagnostic criterion for ALSS currently exist (Goh et al., 2004; Haig et al., 2006; Genevay et al., 2010; Genevay and Atlas, 2010). Recent reviews have cited heterogeneity and unsatisfactory research standards, which prevent conclusions on the performance of diagnostic tests being drawn (De Graaf, 2006; Genevay and Atlas, 2010). Moreover, appraisal of these tests employed found many to be grounded in anecdotal evidence rather than robust measurement (Haig et al., 2006; Browne and Roberts, 2008; Deyo et al., 2009). For this reason supplementary analysis of all diagnostic procedures is needed. A recent review reported postural effects on symptoms to have the most useful diagnostic criteria, whereas clinical tests were less consistently useful (de Schepper et al. 2013).

With a distinctive clinical presentation initial diagnosis of ALSS is made in relation to patient report alone; however, clinical prediction rules based solely on patient report have only been done on small populations (Sugioka et al., 2008). Accurate findings from physical testing would help to validate patient report, however selection of these tests has proven problematic. Neurological examination is commonly observed to be normal (Bassewitz and Herkowitz, 2001; Genevay and Atlas, 2010), thus the most pertinent physical findings typically relate to symptomatic change in accordance with lumbar movement. Although both treadmill and bike tests have been postulated as useful when differentiating ALSS from claudication arising from peripheral vascular disease (Fritz et al., 1997; Tenhula et al., 2000; Deen et al., 2000; Yukawa et al., 2002), research findings appear inconsistent (Dong and Porter, 1989; Moon et al., 2005).

Thus postural tests gauging the symptomatic response to lumbar extension appears pertinent to the initial physical examination. Studies have exhibited the extension movement to narrow spinal space by up to twenty per cent (Panjabi et al., 1983; Inufusa et al., 1996; Fritz et al., 1998), with percentages substantially
increasing when coupled with degenerative changes (Schroenstrom and Willen, 2001; Westergaard et al., 2009; Kishner et al., 2010). Regardless of a strong theoretical foundation, only one study by Katz et al., (1995) is frequently cited by other authors when appraising validity (Fritz et al., 1998; Lurie, 2005; Vo et al., 2005; Westergaard et al., 2009). Evaluating a range of physical components this study observed a small positive likelihood ratio of 1.6 with thirty-seconds of lumbar extension and provocation of at least thigh pain. Positive findings on coupled quadrant movements of extension and side-flexion is recognised as 'Kemps sign' (Jenis and Howard, 2000), and is stated to occur frequently in foraminal stenosis (Watanabe et al., 2010; Eguchi et al., 2010). Thus further evaluation of 'Kemps sign' or combined extension and side-flexion movement seems warranted, as it would be a simple and straightforward test to employ in the examination of those with suspected ALSS, because of age and findings from the history.

1.2 Aims

The aim of this study was to prospectively evaluate the validity of the ‘Modified Extension Test’ (MExT) an extension-based test proposed by the authors. This test gauges the symptomatic response to lumbar movement, and is postulated to support the diagnosis of ALSS whether located centrally or foraminaly. Moreover, no consistent information currently exists regarding whether symptoms occur more rapidly following the commencement of extension in those with severe stenotic change. Such discernments would support the identification of those individuals more likely to require surgical consideration (Atlas et al., 1996; Amundsen et al., 2000; Benoist, 2002; Slatis, 2007; Watters et al., 2008), and hence, was selected for sub-analysis despite the inconsistencies that there was such a link in previous literature (Wang et al., 2008; Haig and Tomkins, 2010; Kishner et al., 2010).

2. METHODS

2.1 Ethical approval

This was a prospective blinded concurrent-criterion related validity design to establish MExT sensitivity, specificity and predictive capability in diagnosing
ALSS. Ethical approval was granted by the Nottingham Two proportionate review sub-committee, with site specific permission and insurance obtained in conjunction with the local CRH NHS trust research and development team. Institutional clearance was further established on behalf of Sheffield Hallam University Health and Care Ethics Committee. All patients provided signed, informed consent prior to participation.

2.2 Sample

Subjects aged fifty or over were recruited from Chesterfield Royal Hospital (CRH) over a six-month period. Consecutive patients attending either the Extended Scope Practitioner (ESP) or musculoskeletal (MSK) physiotherapy clinic were screened, and where deemed eligible were invited to participate. Recruitment was based on presenting symptoms, with inclusion criterion requiring subjective report of unilateral or bilateral pain or paraesthesia radiating below the gluteal fold. In accordance with guidelines for MRI scan referral, symptoms were required to have been present for a minimum of six weeks (Bussieres et al 2008). Patients who had a lack of mental capacity to provide informed consent, or were unable to understand English, or there was a pre-existing diagnosis of peripheral vascular disease, or of Ankylosing Spondylitis were excluded. A STARD flowchart is provided in Figure 1.

In the absence of a defined ‘true’ diagnostic gold standard, Magnetic Resonance Imaging (MRI) was selected as the reference test upon which to evaluate the index test’s (MExT) validity. Whilst concern has been documented regarding the relative inconsistency between clinical symptoms and degree of observed radiographic change (Wang et al., 2008; Haig and Tomkins, 2010; Kishner et al., 2010), clinical guidelines recommend its use as a primary tool in ALSS diagnosis (Watters et al., 2008). MRI has demonstrated both significant positive likelihood ratios, (between 8.1 and 16.2), and negative likelihood ratios (between 0.3-0.19) establishing its usefulness for ruling in or out a diagnosis (Fritz et al., 1998). Moreover, the reliability of image interpretation for ALSS has been established with Lurie et al., (2008) observing Kappa scores of (0.82) and (0.83) for inter-rater and intra-rater reliability respectively. The definitions used are in tables 1 and 2.
Table 1: Definitions of central ALSS severity based upon cross-sectional area of the canal and presence of cerebrospinal fluid.

<table>
<thead>
<tr>
<th>Severity</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe ALSS</td>
<td>&gt;50% narrowing with absence of cerebrospinal fluid</td>
</tr>
<tr>
<td>Moderate ALSS</td>
<td>&gt;50% narrowing with more nerve root than cerebrospinal fluid</td>
</tr>
<tr>
<td>Mild ALSS</td>
<td>&lt;50% narrowing with cerebrospinal fluid equal to nerve root</td>
</tr>
<tr>
<td>No ALSS</td>
<td>&lt;50% narrowing with more cerebrospinal fluid than nerve root</td>
</tr>
</tbody>
</table>

Table 2: Definitions of foraminal ALSS severity based upon degree of nerve root epidural fat.

<table>
<thead>
<tr>
<th>Severity</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe ALSS</td>
<td>Compression with total ablation of fat around the nerve root</td>
</tr>
<tr>
<td>Moderate ALSS</td>
<td>Compression with some distortion of fat around the nerve root</td>
</tr>
<tr>
<td>Mild ALSS</td>
<td>Foraminal encroachment, without fat compromise</td>
</tr>
<tr>
<td>No ALSS</td>
<td>No foraminal encroachment, without fat compromise</td>
</tr>
</tbody>
</table>

ALSS = acquired lumbar spinal stenosis

2.3 Procedure

The investigation was undertaken at CRH with subjects assessed by an ESP of more than twenty years’ experience. The primary researcher obtained informed consent before key demographic and clinical findings were recorded. Prior to MExT testing a neurological examination was carried out that included dermatomes, myotomes, reflexes, and straight leg raise test as described by Petty (2006).

Prior to MExT commencement the order in which the constituent test component would be administered was pre-determined through random allocation. The randomisation process comprised a blinded independent therapist selecting from thirty coloured balls concealed within a box, (fifteen blue and fifteen red). When a blue ball was selected component one, (extension), was to be undertaken first;
conversely, when a red ball was selected component two, (extension/side-flexion), was the initial test component.

Each subject received the MExT (Figure 2), whilst an independent rater timed each component up to the sixty second cut-off point. The result of the MExT was documented against each component, and when positive, the associated time for symptoms to occur was recorded. A positive test was deemed to be when pain below the gluteal fold was exacerbated or produced with one or both components of the test; exacerbation or production of back pain alone did not count. The participant was subsequently monitored for ten minutes and advised on what measures should be undertaken in the event of adverse effects. In order to limit partial verification bias each participant was referred for an MRI scan irrespective of MExT result (Pewsner et al., 2004; Whiting et al., 2004). The subsequent lumbar MRI scan was undertaken on a 1.5 Tesla, Siemens, Avanto scanner within a week of the index test, before being analysed by a consultant radiologist blinded to clinical findings and index test result. The results from both the index and reference tests were returned to the principal investigator and evaluated, whilst each subject received an ESP follow-up appointment to discuss further management options.
Figure 2. Index Test ‘MExT’ Procedure

Initial set up

Standing position

The participant adopts the neutral standing position, with the therapist situated to the side of the individual. The therapist assumes a comfortable step stance posture using the medial aspect of their knee (closest to the participant) to maintain the knee in extension and provide support to the participant (see limb placement image below). The therapist then assumes an appropriate hold upon which to facilitate the test. The participant is asked to place their arm closest to the therapist across their chest with their hand close to the contralateral shoulder. The therapist places one hand on the lower lumbar spine and the other is located across the participant’s arm. The adopted position allows appropriate comfort for participant and therapist. Prior to test commencement, the participant is advised that they should inform the therapist if any symptomatic change occurs.

Limb Placement
Component 1

The therapist asks for the timer to commence on the command ‘Go’ as they gently manoeuvre the participant into their full available extension, remaining in that position for up to one minute. The procedure is terminated prior to one minute if either the leg symptoms have been reproduced (a positive test), or if the participant reports excessive lumbar spine discomfort to continue, (an inconclusive test). If the participant fails to report reproduction of their leg symptoms during this time period the test is deemed negative. The therapist terminates the test with the command ‘stop’ before the participant is returned to the neutral position and provided with a one minute rest to allow symptoms to return to normal prior to initiating stage three.

Component 2

The participant is once again advised to inform the therapist should any symptomatic change occur. The therapist asks for the timer to commence as the participant is manoeuvred into full extension, and then side-flexion towards the symptomatic side. On the occasion that symptoms are bilateral the examiner selects only the most symptomatic side to test, with results only determined positive if symptoms are reproduced on this same side. If symptoms are reproduced on the contralateral side the test is deemed negative. The procedure is terminated in relation to the predefined criteria (as in component one) with all findings documented. The researcher asks the participant to rest for up to ten minutes following the procedure, taking the opportunity to answer any additional questions the participant may have and to ensure no immediate adverse effects have occurred. The physiotherapist only performs the test once to mirror clinical practice, whilst additionally reducing the likelihood of developing post-test soreness.
2.4 Pilot study

Preceding commencement of the trial, a pilot study of five subjects was completed with results not included in the final analysis. Study procedures were appraised and adapted accordingly, with inter-rater reliability of both stopwatch recordings and MRI interpretation found to be above the limit set for acceptable reliability >0.75 (Streiner and Norman, 2003). The minimal detectable change was at a 95% confidence level.

One person performed each component of the test while two other people simultaneously did the stopwatch timing. The intra-class correlation coefficient (ICC) was 0.99 whilst the standard error of measurement (SEM) was 1.72 and the minimal detectable change was 4.76. Moreover, MRI interpretation evaluation was based upon agreement of stenotic grade, with a kappa coefficient value of 0.83 (95% CI 0.48, 1.0).

2.5 Data Analysis

Data was analysed using SPSS (Statistical Package for the Social Sciences) version 16.0. Construction of a two by two contingency table allowed for initial identification of test validity, with point-estimates of sensitivity and specificity ascertained using a calculator alongside 95% confidence intervals (Bland, 2000). Sensitivity and specificity are measures of test performance, discriminating between subjects with or without a specific disease. In conjunction with statistical and clinical recommendation acceptable levels were duly fixed at > 80% (Streiner and Norman, 2003; Riegelman, 2005).

Positive and negative likelihood ratios were further calculated alongside 95% confidence intervals to provide primary analysis of test validity. Likelihood ratios estimate the extent to which an individual with ALSS is more likely to test positive or negative when compared to someone without the disease. Evaluation utilising likelihood ratios is preferable to that of the more commonly used positive and negative predictive values, as likelihood ratios are not dependent upon disease prevalence (McGee, 2002). As such, a positive likelihood ratio of > 2.0 was perceived to highlight an important increase in ALSS likelihood (McGee, 2002),
conversely, a negative likelihood ratio of < 0.5 was adjudged to propose a reduced ALSS likelihood. Post-test probabilities of a subject having ALSS in relation to a specific MExT result were established from the likelihood ratios, and confirmed in association with Fagan’s nomogram (Akobeng, 2007; Campo et al., 2010).

Sub-analysis of the relationship between symptom onset following MExT commencement and the grading of ALSS whether mild, moderate or severe was undertaken using Spearman’s rank correlation coefficient. Similarly, the Phi coefficient was utilised to establish whether an association existed between positive findings on a specific MExT component and a diagnosis of central or foraminal ALSS. Significant findings were deemed to be \( p < 0.05 \).

3. FINDINGS

3.1 Demographics

Two-hundred and seventy-six patients with various upper and lower quadrant symptoms were assessed between January and July in the defined clinics. Thirty-one subjects were eligible for recruitment based on the inclusion criteria described above. Only one subject was subsequently withdrawn from the study after declining an MRI scan. The remaining thirty subjects, seventeen males and thirteen females (mean age 64 years, S.D. 6.9 years) continued through to analysis. Both index and reference tests were available for all thirty subjects, with no index or reference tests determined inconclusive. There were no lasting adverse events from performing either the index or reference tests. The characteristics of the sample are given in table 3.

Table 3: Sample Characteristics and Clinical findings

<table>
<thead>
<tr>
<th>Subject characteristics / Clinical presentation</th>
<th>Number within sample (( n = 30 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. male (%)</td>
<td>17 (53)</td>
</tr>
<tr>
<td>Age years, (S.D)</td>
<td>64 (6.9)</td>
</tr>
<tr>
<td>Lower limb pain or paraesthesia below the knee (%)</td>
<td>23 (77)</td>
</tr>
<tr>
<td>Bilateral lower limb symptoms (%)</td>
<td>10 (33)</td>
</tr>
<tr>
<td>Dermatome loss(^a) (%)</td>
<td>10 (33)</td>
</tr>
</tbody>
</table>
In relation to index and reference tests, eighty-seven per cent of the sample tested positive to at least one component of the MExT. In accordance with reference test findings sample prevalence of ALSS was calculated as eighty-three per cent, with thirty per cent found to have central stenosis, thirty-three per cent foraminal stenosis, and seventeen per cent both central and foraminal forms. Three per cent were positive for stenosis at the lateral recess with ten per cent additionally found to have an associated degenerative spondylolisthesis at either the L4/5 or L5/S1 level.

### 3.2 Sensitivity, specificity and likelihood ratios

Sensitivity, specificity, and positive and negative likelihood ratios are reported in table 4. Pre and post-test probabilities were calculated from pre and post-test odds and confirmed against Fagan’s nomogram. Whilst the pre-test probability of a subject within the sample having ALSS was 83% the post-test probability marginally increased to 88% when a positive MExT was observed. Likewise post-test probabilities based on the negative likelihood ratio observed a decrease in the likelihood of ALSS to 49% when a negative result was detected.

Sub-analysis of correlations between grade of stenosis and symptom onset following MExT commencement (Table 5), failed to demonstrate a significant correlation for either central or foraminal elements in conjunction with Spearman’s correlation coefficient. In relation to central stenosis a negative correlation of $R = -0.147$ was observed with symptom onset more rapid in severe stenosis, but this correlation was not statistically significant ($p = 0.684$). Similar outcomes were observed in relation to foraminal stenosis with a negative correlation of $R = -0.211$, which was not statistically significant ($p = 0.418$).

<table>
<thead>
<tr>
<th>Myotome loss(^a) (%)</th>
<th>1 (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflex loss(^b) (%)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Positive SLR(^c) (%)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Intermittent Claudication (%)</td>
<td>25 (82)</td>
</tr>
</tbody>
</table>

\(a\) = areas of sensory abnormality; \(b\) = weakness in muscle compared to the other side; \(c\) = diminished or absent; \(d\) = reduced range compared to the other side
Consideration of MExT selectivity in relation to stenosis location failed to identify notable correlations between the matched components (Table 6). Findings of central stenosis and a positive finding on the on the extension component of the test observed a statistically non-significant negligible association ($r = 0.11$, $p = 0.54$). There was a statistically non-significant association identified between the extension / lateral flexion component of the test and foraminal stenosis ($r = 0.09$, $p = 0.62$). The extension component of the test and foraminal stenosis had an association of ($r = 0.24$, $p = 0.19$) whilst the extension / lateral flexion component of the test and central stenosis had an association of ($r = 0.15$, $p = 0.41$), but neither was significant. The lack of statistical significance could be due to small sample size and lack of adequate statistical power.

**Table 4: MRI compared to MExT (in participant numbers)**

<table>
<thead>
<tr>
<th></th>
<th>MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ve scan findings</td>
</tr>
<tr>
<td>MExT</td>
<td></td>
</tr>
<tr>
<td>+ve test</td>
<td>23</td>
</tr>
<tr>
<td>-ve test</td>
<td>2</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>92% (72-99%)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td></td>
</tr>
<tr>
<td>+ LR (95% CI)</td>
<td>1.53 (0.74-3.16)</td>
</tr>
<tr>
<td>- LR (95% CI)</td>
<td>0.2 (0.003-1.36)</td>
</tr>
</tbody>
</table>

+ve = one or both components were provocative of symptoms; -ve = neither component provoked symptoms; CI = confidence intervals; LR = likelihood ratios

**Table 5: Analysis of time to onset of symptoms (in seconds), and correlation with stenosis severity**

<table>
<thead>
<tr>
<th>Stenosis grade (N)</th>
<th>Mean time to symptom onset when positive on component 1 for central stenosis</th>
<th>Mean time to symptom onset when positive on component 2 for central stenosis</th>
<th>Mean time to symptom onset when positive on component 1 for foraminal stenosis</th>
<th>Mean time to symptom onset when positive on component 2 for foraminal stenosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foraminal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central stenosis</td>
<td>Foraminal stenosis</td>
<td>Central and foraminal stenosis</td>
<td>No stenosis</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>MExT extension positive / lateral flexion negative</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MExt extension negative / lateral flexion positive</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MExT extension / lateral flexion positive</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>MExT extension / lateral flexion negative</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

4. **DISCUSSION**

To the authors' knowledge, the current investigation is the first to propose and evaluate the validity of the MExT, a bi-component movement test, postulated to support the diagnosis of ALSS through physical examination of individuals who were 50 years or older, and had unilateral or bilateral symptoms in the lower extremity radiating distal to the gluteal fold. The current evaluation observed the MExT to demonstrate high sensitivity (92%) with a clinically significant negative likelihood ratio (-LR 0.2). In accordance with negative post-test probabilities, the
likelihood of a subject within our cohort having ALSS reduced from 83% to 49% when a negative result was observed. Such findings would support the prospective validity of the MExT to assist in ruling-out a diagnosis when a negative result is observed. In relation to the test’s proficiency to assist in ruling-in a diagnosis, the converse appeared true. Specificity point estimates were low (40%), and positive likelihood ratios were only small, (+LR 1.53). Neither value achieved the desired level for clinical acceptability, (>80% or > 2.0 respectively) with only a five per cent increase in subject post-test probability when a positive result was observed.

Analysis of whether a relationship existed between stenosis grade and symptom onset following MExT commencement failed to demonstrate a significant correlation for either central or foraminal elements. Correlations were statistically non-significant irrespective of whether stenotic types were stratified by the theoretical preferential component or calculated in relation to the component of quickest onset of symptoms. Analysis of alternate MExT components in relation to stenotic location additionally failed to identify a clear association. Whilst ‘Kemps sign’ has been endorsed as an important clinical observation for foraminal stenosis (Lyle et al., 2005; Watanabe et al., 2010), findings from this current study would alternatively promote the evaluation of symptomatic response to both extension alone and extension with side-flexion, irrespective of the stenotic location suspected. As such, inclusion of both movement components would allow for greater test sensitivity than any one movement alone.

As identified in an updated systematic review of the accuracy of diagnostic tests (de Schepper et al. 2013) items from the history had the highest diagnostic value: radiating leg pain worse on standing, absence of pain when sitting, easing of pain when bending, and a wide-based gait. MRI was the most promising imaging test; and about 50 clinical tests were also examined. These included abnormal gait, weakness, reflexes, a combination of signs and symptoms, a questionnaire, pain drawings, and walking on a treadmill. On the whole clinical tests had either good levels of sensitivity or good levels of specificity, but not both; similar to the findings in this study.
Whilst the current cohort was consistent with other studies of this nature in regards to age and gender (Katz et al., 1995; Lyle et al., 2005; Konno et al., 2007); appraisal of post-test validity against previous extension tests varied. Katz et al. (1995) reported a small positive likelihood ratio (+LR 1.6) on ‘thigh pain with thirty seconds extension’; however, point estimates were notably different to the current study with reduced sensitivity (51%), and greater specificity (69%). The increase in test sensitivity may have related to longer test duration. Subjects in the MExT maintained the extended posture for up to thirty seconds longer than previous studies. Whilst test modifications may assist identification of ALSS with greater frequency than traditional extensions tests, such alterations may additionally account for the noted drop in specificity.

In relation to alternative clinical diagnostic tools the MExT observed a comparative capability to rule out ALSS diagnosis. Sensitivity estimates and negative likelihood ratios were similar to those of the prediction-rule grounded, ‘ALSS diagnostic support tool’ (Konno et al., 2007). However, in relation to overall properties the MExT appeared subordinate, failing to achieve the same level for specificity (72 %) or positive likelihood ratios (+LR 3.31). Likewise greater discriminative capability to rule-in a diagnosis was observed during a two-stage treadmill test with all positive likelihood ratios exceeding those of the MExT (+LR >2.5) (Fritz et al., 1997).

4.1 Limitations

The findings of this study should be considered in relation to a number of study limitations. Through recruitment of only a small sample size (n = 30), the evaluation had an increased likelihood of type II error. The small sample size and lack of statistical power may have had the effect of producing non-significant results. Confidence intervals around the primary data appeared wide, and hence, estimations of test validity must be interpreted with caution. Prevalence of ALSS within the cohort was high as a result of the inclusion criterion employed. Specificity estimates may have suffered due to inadequate evaluation against subjects with an alternative lumbar, neurogenic or vascular pathology. Recruitment methods through secondary care may have inadvertently biased the sample towards more advanced pathology. A large proportion of the cohort was
observed to exhibit moderate to severe ALSS. As such, it was not possible to fully evaluate the MExT against more mild variants limiting its external validity to primary care settings. Furthermore in such settings there would be a higher degree of diagnostic uncertainty, but in such a context tests could have the greatest ability to influence clinical decision making if they perform well. Whilst MRI was selected as the reference standard due to its objectivity, recognised discriminative capability and consistent clinical application, it must be acknowledged that inconsistency between clinical symptoms and degree of radiographic change exists. Another limitation was the lack of assessment of the reliability between testers on decision making regarding whether there was a positive or negative response to the tests.

4.2  Clinical and research implications

In relation to clinical practice it is not currently possible to support or refute the MExT’s diagnostic validity due to the limitations discussed. Whilst preliminary estimates suggest the MExT may demonstrate part criterion validity, further evaluation is required. As such, the current evaluation should act as a feasibility study upon which to base further investigation into extension test validity. Future studies should aim to reproduce test procedures within larger cohorts across a range of primary and secondary care settings. Replication of current findings would support the value of the MExT in ruling out a diagnosis of ALSS, and hence, would form a valuable part of the physical examination process.

Until such validation has been completed, clinician’s should continue to evaluate all available subjective and objective findings in order to support or refute a potential diagnosis. Moreover, in clinical practice clinician’s should continue to use subjective report of radicular symptoms and the impact upon the individual’s functionality as the pre-requisite for onward referral.

5.  CONCLUSION

Preliminary examination of a modified extension test’s validity to diagnose lumbar spinal stenosis demonstrated conflicting results. Whilst the test demonstrated a
clinically significant capability to rule-out diagnosis, concurrent capability to rule-in diagnosis was unacceptable. All aspects of the test were not capable of determining the severity of stenosis or of differentiating between central and foraminal types. Supplementary research on larger cohorts is required to fully evaluate test validity.

Word count 4,231 including figures

References


Population Reference Bureau


Acknowledgements
The authors would like to thank Pieter Meiring and other staff at Chesterfield Royal Hospital for facilitating the study, and the first author’s family for on-going encouragement throughout the process.

Conflict of interest
The authors declare no conflict of interest.

Funding
No funding was received for this study.
Figure 1. STARD Flow Diagram

Potentially eligible patients
n = 276

Excluded patients
Reasons n = 1 - declined MRI

Eligible patients
n = 31

MeXT test
n = 30

Positive MeXT
n = 26

No MRI
n = 0

MRI
n = 26

Inconclusive result
n = 0

ALSS Present
n = 23

ALSS absent
n = 3

ALSS Present
n = 2

ALSS absent
n = 2

ALSS Present
n = 0

ALSS absent
n = 0

Negative MeXT
n = 4

No MRI
n = 0

MRI
n = 4

Inconclusive result
n = 0

No MRI
n = 0

Inconclusive result
n = 0

ALSS Present
n = 0

ALSS absent
n = 0

Exclusion reason (N = 245):
Upper quadrant symptoms (109)
≤ 50 years (60)
No referred symptoms (76)
No informed consent (0)
Peripheral vascular disease (0)
Ankylosing Spondylitis (0)

Eligible patients
n = 31

Potentially eligible patients
n = 276
Highlights

- Spinal stenosis is likely to be more commonly encountered in clinical practice.
- The validity of a simple clinical test was compared to diagnosis by MRI.
- The test had a reasonable sensitivity of 92%, but a poor specificity of 40%.
- The test may be useful to rule out a diagnosis with a negative test result.