# Accepted Manuscript


Title: Clinical classification criteria for neurogenic claudication caused by lumbar spinal stenosis. The N-CLASS criteria

Author: Stéphane Genevay, Delphine S. Courvoisier, Kika Konstantinou, Francisco M. Kovacs, Marc Marty, James Rainville, Michael Norberg, Jean-François Kaux, Thomas D. Cha, Jeffrey N. Katz, Steven J. Atlas

PII: S1529-9430(17)31052-5
DOI: https://doi.org/doi:10.1016/j.spinee.2017.10.003
Reference: SPINEE 57511

To appear in: *The Spine Journal*

Received date: 30-5-2017
Revised date: 25-8-2017
Accepted date: 2-10-2017

Please cite this article as: Stéphane Genevay, Delphine S. Courvoisier, Kika Konstantinou, Francisco M. Kovacs, Marc Marty, James Rainville, Michael Norberg, Jean-François Kaux, Thomas D. Cha, Jeffrey N. Katz, Steven J. Atlas, Clinical classification criteria for neurogenic claudication caused by lumbar spinal stenosis. The N-CLASS criteria, *The Spine Journal* (2017), https://doi.org/doi:10.1016/j.spinee.2017.10.003.

# Clinical Classification Criteria for Neurogenic Claudication caused by Lumbar Spinal Stenosis. The N-CLASS criteria.

Stéphane Genevay MD
Division of Rheumatology, University Hospitals of Geneva, Switzerland

Delphine S. Courvoisier PhD
Quality of Care Division, University Hospitals of Geneva, Switzerland and Division of Rheumatology, University Hospitals of Geneva, Switzerland

Kika Konstantinou PhD
Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University UK

Francisco M. Kovacs MD
Spanish Back Pain Research Network. Moncloa University Hospital, Madrid, Spain

Marc Marty MD
Department of Rheumatology, Henri-Mondor Hospital, Créteil, France

James Rainville MD
Physical Medicine and Rehabilitation, New England Baptist Hospital, Boston, USA

Michael Norberg MD
Physical Medicine and rehabilitation,University hospital of Lausanne, Switzerland

Jean-François Kaux MD
Physical Medicine and Sport Traumatology Department, University and University Hospital of Liège,, Belgium.

Thomas D. Cha MD
Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, MA, USA

Jeffrey N. Katz MD
Department of Orthopaedic Surgery and Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston MA, USA

Steven J. Atlas MD
Division of General Internal Medicine, Massachusetts General Hospital, Boston MA, USA

**Corresponding author:**

Stéphane Genevay

Division of Rheumatology,

University Hospitals of Geneva,

1211 Geneva, Switzerland

Tel: +4122 372 36 11

Fax: +4122 372 35 30

Email: stephane.genevay@hcuge.ch



Abstract

BACKGROUND CONTEXT

Since imaging findings of lumbar spinal stenosis (LSS) may not be associated with symptoms, clinical classification criteria based on patient symptoms and physical examination findings are needed.

PURPOSE

To develop clinical classification criteria that identify patients with neurogenic claudication (NC) caused by LSS.

STUDY DESIGN

Two stage process. Phase 1: Delphi process; Phase 2: cross-sectional study.

PATIENT SAMPLE

Outpatients recruited from spine clinics in 5 countries.

OUTCOME MEASURE

Items from history and physical examination.

METHODS

Phase 1: A list of potential predictors of NC caused by LSS was based on the available literature and evaluated through a Delphi process involving seventeen spine specialists

(surgeons and non-surgeons) from 8 countries. Phase 2: Nineteen different clinical spine specialists from 5 countries identified patients they classified as having: 1) NC caused by LSS 2) Radicular pain caused by lumbar disc herniation (LDH), or 3) non-specific low back pain (NSLBP) with radiating leg pain. Patients completed survey items and specialists documented examination signs. Coefficients from General Estimating Equation models were used to select predictors, generate a clinical classification score and obtain a receiver operating characteristic (ROC) curve. Conduction of the Delphi process, data management and statistical analysis were partially supported by an unrestricted grant of less than 15000 US dollars from Merck Sharp and Dohme. No fees were allocated to participating spine specialists.

RESULTS

Phase 1 generated a final list of 46 items related to LSS. In phase 2, 209 patients with leg pain caused by LSS (n=63), LDH (n=89) or NSLBP (n=57) were included. Criteria which independently predicted NC ($p<0.05$) were: age over 60; positive 30 second extension test; negative straight leg test; pain in both legs; leg pain relieved by sitting, and leg pain decreased by leaning forward or flexing the spine. A classification score using a weighted set of these criteria was developed. The proposed N-CLASS score ranged from 0 to 19, had an area under the curve of 0.92, and the cutoff (>10/19) to obtain a specificity of >90.0% resulted in a sensitivity of 82.0%.

CONCLUSION.

Clinical criteria independently associated with neurogenic claudication due to LSS were identified. Use of these symptom and physical variables as a classification score for clinical research could improve homogeneity among enrolled patients.

Keywords: Lumbar spinal stenosis; neurogenic claudication; classification criteria

## INTRODUCTION

Neurogenic claudication (NC), also called pseudoclaudication,[1, 2] is the cardinal symptom caused by lumbar spinal canal stenosis (LSS).[3] LSS represents a degenerative process involving the narrowing of the spinal canal around the nerve roots of the cauda equina within the dural sac due to facet joint osteoarthritis, hypertropic thickening and bulging of the ligamentum flavum, and bulging of the intervertebral disc.[4] Since the first descriptions of the relationship between symptoms of NC and radiographic images demonstrating LSS almost 70 years ago,[1, 5] hundreds of scientific contributions have been published including randomized controlled trials and clinical practice guidelines. A key limitation of the existing literature is the heterogeneity of eligibility criteria for identifying patients with symptoms due to LSS.[6] On its own, the size of the spinal canal is not a valid diagnostic criterion, since there is no agreement on what defines "normal" and "stenotic", stenotic images can be seen in asymptomatic subjects, and there is limited correlation between anatomic findings and symptoms.[7, 8] Consequently, eligibility criteria vary across studies and limit their generalizability, compromising attempts to compare results.[9] These limitations have been recognized in proposals to develop consensus criteria to define and classify patients with symptomatic LSS.[3, 10]

In the absence of specific biomarkers, the use of classification criteria is a key step to identify patients with a specific disease and establish homogenous groups of patients for clinical or population studies, which is essential in multicenter studies and contributes to generalizability of results. [9,11-14] For other musculoskeletal diseases (e.g. rheumatoid arthritis, spondyloarthropathy), widespread adoption of classification criteria has been a key factor spurring advances in diagnosis and treatment.[11] In the field of low back pain (LBP), prior attempts to differentiate LBP patients with leg pain and neurological signs from other

categories of LBP patients failed to define any specific diagnostic criteria for these categories.[12]

In view of the large economic burden related to the management of LBP syndromes including LSS,[13] there is a clear need to develop validated clinical classification criteria for research and clinical purposes.[9] During a workshop at the 11th Forum for Primary Care Research in Low Back Pain, a multidisciplinary, international study to develop classification criteria for LBP related leg symptoms was conceived.

## METHODOLOGY

This study was designed according to rules defined by Fries et al. for constructing classification criteria, [14] and focused on NC caused by LSS and radicular pain caused by lumbar disc herniation (LDH). Here, we report on the development and validation of clinical classification criteria for NC caused by a LSS. Criteria for LDH have been previously reported. [15]

### *Selection of potential items*

The first phase of the study began with a literature review to identify and collate items, followed by a Delphi process as proposed by Lin et al.[16] A list of patient-reported symptoms and clinical signs considered useful in diagnosing NC caused by LSS or radicular pain caused by LDH was generated from a structured literature review with additional items suggested by a multidisciplinary group of 17 spine specialists (Appendix A).

### *Delphi Process*

A Delphi consensus process consisting of rounds of expert review was then conducted to reduce the list of items to those deemed potentially important for the diagnosis of each syndrome.[17] For each round, the specialists rated the usefulness of each criterion "to differentiate patients with NC caused by LSS from all others" on a 7 point scale from 1 (useless) to 7 (very useful).

For round 1, items were excluded if they had a mean score <3, had a rating of 1 by more than 25% of the spine specialists, or more than 50% answering "don't know". For round 2, retained items were re-scored. Items were selected for the clinical phase if the mean score was ≥ 4.5 and the difference between the two rounds was ≤ 1. Additional rounds were planned until all items were either included or excluded.

*Clinical Study*

In phase 2, 19 different spine specialists (surgeon and non-surgeon), working in English and French speaking countries (Appendix A), screened patients presenting at their clinics with back-related leg pain, for study eligibility. Since the original item list was created in English, a version of the items in the case report form was translated into French following rules defined for cross-cultural adaptation and validation.[18]

Patients were included if the expert diagnosed the patient with NC caused by LSS. As a comparison group, patient with radicular pain caused by LDH, or non-specific LBP with referred leg pain were also recruited.[19] Since symptoms of NC are related to central stenosis, patients with only foraminal stenosis were not included. Patients with specific causes of back pain (e.g infection) were excluded. Additional exclusion criteria were: patients younger than 18 years old, LBP without any leg pain, leg pain not related to a spinal problem,

unable to read or understand study site's native language, or declining study participation. As recommended for the development of classification criteria, participating spine specialists were asked to enroll similar numbers of patients from each of the three diagnostic groups.[20] Neither patients nor specialists received any form of compensation for their participation in this study.

Approval was obtained from the ethical committee of the Geneva University Hospitals, Geneva, Switzerland and additional approvals were obtained from each participating institution.

Consecutive patients who sought care at the participating back pain clinics and met inclusion criteria were invited to join the study. Informed consent was obtained prior to enrollment. During the same visit, enrolled patients completed a questionnaire and the spine specialist provided information on symptoms and findings obtained from the physical examination (definitions of clinical tests were provided). Spine specialists categorized patients into one of the 3 diagnosis groups using any relevant test or procedure felt necessary as part of routine practice (including advances imaging, MRI or CT scans and electrodiagnostic studies), and rated their degree of confidence with the diagnosis on a visual analog scale from 0 (not confident at all) to 10 (extremely confident). Cases in which the confidence level was below 7 were excluded from the analysis in order to retain groups of patients that fit unambiguously into one of the three categories. The clinicians were not aware that patients diagnosed with less than 7 confidence scores would be excluded.

*Statistical analysis*

To select the best set of clinical criteria for neurogenic claudication caused by LSS, all items identified by the Delphi method were included in analyses and expert clinical diagnosis

served as the "gold standard". We first used univariable multiple logistic regression with generalized estimating equation (GEE) models (logit link) with neurogenic claudication caused by LSS versus the combined LDH and NSLBP groups as the outcome and each criterion as the predictor. The GEE model with an exchangeable correlation matrix was used to account for the multiple study spine specialists. Items were included in the multivariable models if the univariable p-value was <0.1 but excluded if selected in 10 patients or fewer. We then ran two multivariable GEE models, one with all the patient-reported criteria, and the other with all physical examination criteria. All items with p-value $p<0.1$ were introduced in a subsequent multivariable model and we chose the model based on the lowest value of the quasi-information criteria (M1) logit link, exchangeable correlation matrix. Sensitivity analyses were then performed to attempt to simplify the models while maintaining sensitivity and specificity. To test the appropriateness of model selection, we also used the least absolute shrinkage and selection operator (LASSO) method and compared the criteria retained using this statistical model selection with the sequential model selection described above.

Based on the coefficients of the final GEE model, we assigned a weight to each criterion retained, and established the "Neurogenic CLAudication caused by lumbar Spinal Stenosis" (N-CLASS) classification criteria set. The psychometric quality of the N-CLASS was assessed using receiver operating characteristic (ROC curve) and area under the curve (AUC). To determine the score cutoff, we aimed at obtaining a specificity of at least 90%, thus creating a classification score that includes few false positive (i.e., patients considered as having a N-CLASS score above the cutoff, but diagnosed by the gold standard as not having NC caused by LSS). This score and its cutoff were then used to compute sensitivity, specificity, with their respective 95% confidence intervals. All analyses were done using R

v3.2.3, with libraries geepack for the GEE analysis, MESS for the quasi-information criterion, and the glmnet library for the LASSO model selection.

*Sample size calculation*

Assuming at least 10 patients per variable are needed for analyses using logistic regression and a total of 10 criteria in the final model, the required sample size was 100 patients. However, because patients recruited by the same expert are not independent, we multiplied this sample size by a design effect, assuming an intraclass correlation of 0.05 [21] and an average number of patients per physician (cluster size) of 15 [21]. This led to a final sample size of 170 (100 x [1+(15-1)x0.05]). Assuming similar numbers of recruited patients per diagnosis (60 patients with radicular pain caused by LDH, 60 patients with neurogenic claudication caused by LSS, and 50 patients with non-specific LBP), this sample size allowed for estimating a 95% confidence interval around a sensitivity of 80% with a half-interval of 10.1% (i.e., 69.9% and 90.1%), and a specificity of 80% with a half-interval of 7.5% (72.5% and 87.5%).

## RESULTS

*Delphi process*

The literature review and items identified by the group of spine specialists resulted in a list of 236 potential items for spine-related leg pain symptoms and physical examination findings. Out of the 236 items, 96 were associated with neurogenic claudication caused by LSS while the others were associated with radicular pain due to LDH. In the 1$^{st}$ round, 3 of the 96 items were excluded, all based on mean scores <3, leaving 93 items. In the 2$^{nd}$ round, 47 items were excluded. Of the 46 remaining items, 22 were patient-reported symptoms and 24 were physician-reported findings. As all items had a stable evaluation (≤1 point difference between

rounds on the usefulness scale), the Delphi process ended. In a similar manner, items associated with radicular pain due to a LDH were identified.[15]

*Clinical study*

Among 213 enrolled patients (average 10.7 patients enrolled per expert), 4 were excluded as the spine specialists rated their confidence with diagnosis to be <7. The remaining 209 patients included 63 with neurogenic claudication caused by LSS, 89 with radicular pain caused by LDH, and 57 with NSLBP with referred leg pain (Table 1). Tests employed by the spine specialists as part of the diagnostic evaluation included MRI or CT scan for 203/209 patients and EMG for 25/209 patients.

The statistical analysis included items thought to be related to LSS and those thought to be related to LDH (Appendix B). Overlapping items (e.g. worse pain when sneezing, coughing or staining) were combined to create single variables, items reported by ten patients or fewer were dropped from further analyses, and duplicate items (i.e. items associated with both clinical entities) were discarded, yielding a final count of 75 items. In univariable analysis, 37 of 75 criteria were significantly associated with a diagnosis of neurogenic claudication caused by LSS including 17 patient-reported and 20 physician-reported items.

Multivariate analysis was conducted separately for patient-reported items and physician-reported items. Items with $p < 0.1$ were included in a subsequent multivariate analysis leading to the identification of 7 items (Table 2), 4 patient-reported items (age over 60, bilateral leg pain, leg pain relieved by sitting, and leg pain decreased by flexing the spine or leaning forward) and 3 physician-reported items, negative straight leg raise [SLR] test, positive 30 second extension test, and difficulty in squatting due to weakness. The definition of these

clinical items is provided in Table 3. The score derived from this model (M1, Table 4) had an AUC of 0.92 (Figure 1), and the cutoff value to obtain a specificity of ≥90% resulted in a sensitivity of 81.7%. Removing the item, "difficult in squatting due to weakness", resulted in a negligible reduction in AUC, sensitivity and specificity (Table 4). However, removing the SLR test had a strong negative impact on sensitivity. The model without the squat exam item was then considered as the final model (Table 5). The Lasso model selection method retained the same six items as being the most predictive of neurogenic claudication caused by LSS. Thus, this sensitivity analysis confirmed the results of the sequential method using univariable and multivariable analyses.

Items retained in the final model demonstrated fractional weights that varied two-fold (see the respective scores, Table 5). To translate these weights into an easy to use scoring method, the score of each item was multiplied by 2 and rounded to the nearest integer. Hence, in the criteria set for "Neurogenic Claudication caused by Lumbar Spinal Stenosis" ( N-CLASS) a weight of 4 is attributed for age >60 and 30 second extension test, 3 for all patient-reported criteria (i.e., feeling pain in both legs, leg pain relieved by sitting and pain decreased by leaning forward or flexing the spine) and 2 for negative SLR test (Table 6). A patient was classified as having neurogenic claudication caused by LSS if the total score (ranging from 0 to 19) was 11 or more. This cut off of >10 provided a sensitivity of 80.0% [95%CI: 67.7% – 89.2%] and a specificity of 92.1% [95%CI: 86.4% – 96.0%] (Table 4, simplified model 3).

## DISCUSSION

Classification criteria are defined as a set of disease characteristics used to group individuals into a well-defined homogenous population with similar clinical disease features [22]. Their use is advocated and promoted for classifying conditions which lack highly specific

biomarkers. [20, 22, 23] This study was conducted by a multidisciplinary international team of spine specialists, using a modified Delphi process for item generation and a clinical validation study to produce a set of clinical classification criteria for NC caused by LSS. It identified a final set of 6 items, 4 symptoms and 2 physical examination findings, that independently predicted NC caused by LSS. Using coefficients from the final regression model, a classification criteria set was developed; patients with a score >10 in the N-CLASS score have a 90% chance of having NC caused by LSS.

Given the limitations of physical examination findings in the evaluation of patients with suspected NC caused by LSS, most of the final items included were patient-reported variables (bilateral leg pain, leg pain relieved by sitting and leg pain decreased by flexing the spine or leaning forward in addition to patient age). [4] Two items were derived from physical examination findings. SLR is a typical clinical finding in radicular pain due to LDH and its absence in LSS is well recognized [4]. The 30 second extension test is less well known but was reported to have some specificity in identifying patient with NC.[24, 25]

*Comparison with the existing literature*

Several studies have sought to develop diagnostic criteria to classify patients with NC caused by LSS, [24, 26-29] but we are not aware of studies that have used a Delphi process to identify potential items and then develop and validate classification criteria in a clinical study.

Clinical diagnostic criteria are different from classification criteria. Diagnostic criteria are designed to help clinicians to detect and diagnose patients suffering from a given condition. [11] A high sensitivity is expected as they are meant to be broadly inclusive and avoid leaving subjects with that condition undiagnosed. In contrast, the emphasis for classification criteria is on obtaining a high specificity to ensure that all patients diagnosed with a condition actually have symptoms attributable to it.[22] Sensitivity and specificity being on a continuum and

having an inverse relationship, there will inevitably be some difference in the items retained in the respective criteria set (i.e. classification vs. diagnostic).

While two items included in N-CLASS have been reported in one of the five studies on diagnostic criteria (i.e. “pain in both legs” included in Cook et al. and “negative SLR” in Konno et al), the others have all been reported several times (the maximum being four times for age [24, 27, 28] and leg pain relieved by sitting.[24, 26-28] Overall, at least 50% of the criteria of N-CLASS are included in 4 of the 5 studies on diagnostic criteria. However, in the most recently published study, only 1 out of 10 items were included in N-CLASS, despite most of these items being in our study.[29] Although both studies used a panel of spine experts to select items, several methodological differences between the studies may be relevant. First, the purposes were different (i.e. classification vs. diagnostic criteria, see above). Second, in the Tomkins-Lane study, the Delphi process was performed on a small, pre-selected group of items and not from a comprehensive group of items derived from literature review and expert opinion. Finally, their criteria have not yet been tested in clinical practice. Most of the items included in their criteria were tested in the clinical phase of our study and were not discriminant (e.g. “leg pain brought on by walking” was reported by 82% in the control group, “leg pain increased with walking” in 89% and “absence of abnormal foot pulse” in 79%, without significant difference among our three groups, see Appendix B).

To be of value for basic science, epidemiological or clinical research, clinical criteria must have a good specificity.[11, 23] Prior studies of diagnostic criteria reported specificity that was lower (i.e. 80% or less) [27, 28] than the 92% reported in the present study. Interestingly, the high specificity of the N-CLASS was obtained while keeping the sensitivity above 80%. By comparison, in the study by Cook et al, specificity greater than 90% would result in a sensitivity of 6%. Accuracy reflects the discriminant ability of a test, combining both sensitivity and specificity. Previous studies report accuracy between 0.8 and 0.92.[27, 28] N-

CLASS has an accuracy of 0.91 (Figure 1), meaning that it would only misclassify 9% of subjects, identical to Konno et al [27] but with a much better specificity (92.1% versus 72%).

*Strengths and limitations*

This study was conducted following current recommendations for the development of classification criteria.[14] Face validity is likely to be good, since the items included in N-CLASS are commonly reported in the literature. The diversity of spine specialists involved in this study suggests good content validity. The inclusion of a heterogeneous population of patients with back-related leg pain also supports good construct validity. This study also has several limitations. The gold standard used for diagnosis of NC caused by LSS was diagnosis by experts. Although this is the recommended practice for diseases for which no validated biomarkers are available, and diagnosis was based on best clinical practice (i.e., consistency of symptoms, signs and results from appropriate imaging and other diagnostic tests), it carries the intrinsic risk of circular reasoning. We tried to minimize this risk by ensuring that experts involved in the clinical phase of this study, were different from those involved in the Delphi phase. Nevertheless, experts' initial clinical suspicion may have influenced anamnesis, and interpretation of patients' symptoms and findings from physical examination.

Clinicians' skills may also influence the accuracy of the score, since the latter depends on the accuracy of data gathered during the clinical encounter. The N-CLASS score may not be as accurate if the clinician performing the evaluation is not skilled in examining patients with spine symptoms. However, both the SLR test and the 30 second extension are simple tests, which are routinely taught to medical students. Future studies should also be performed to confirm the validity of the N-CLASS in an independent population.

## CONCLUSION

This international multidisciplinary study is the first to propose classification criteria for NC due to LSS. When designing future research studies on LSS, use of N-CLASS score could improve the homogeneity of the studied populations and increase the quality of study comparisons and data pooling.

AKNOWLEDGMENT

We express our gratitude to all spine specialists who participated in the Delphi process and in the recruitment of patients as well as the patients who kindly participated. We also wish to thank MSD for their financial support

CONLFICT OF INTEREST

This study received financial support from an unconditional scientific grant from MSD. MSD had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. Publication of this study was not contingent upon approval from the study sponsor. No fees were allocated to participating spine specialists.



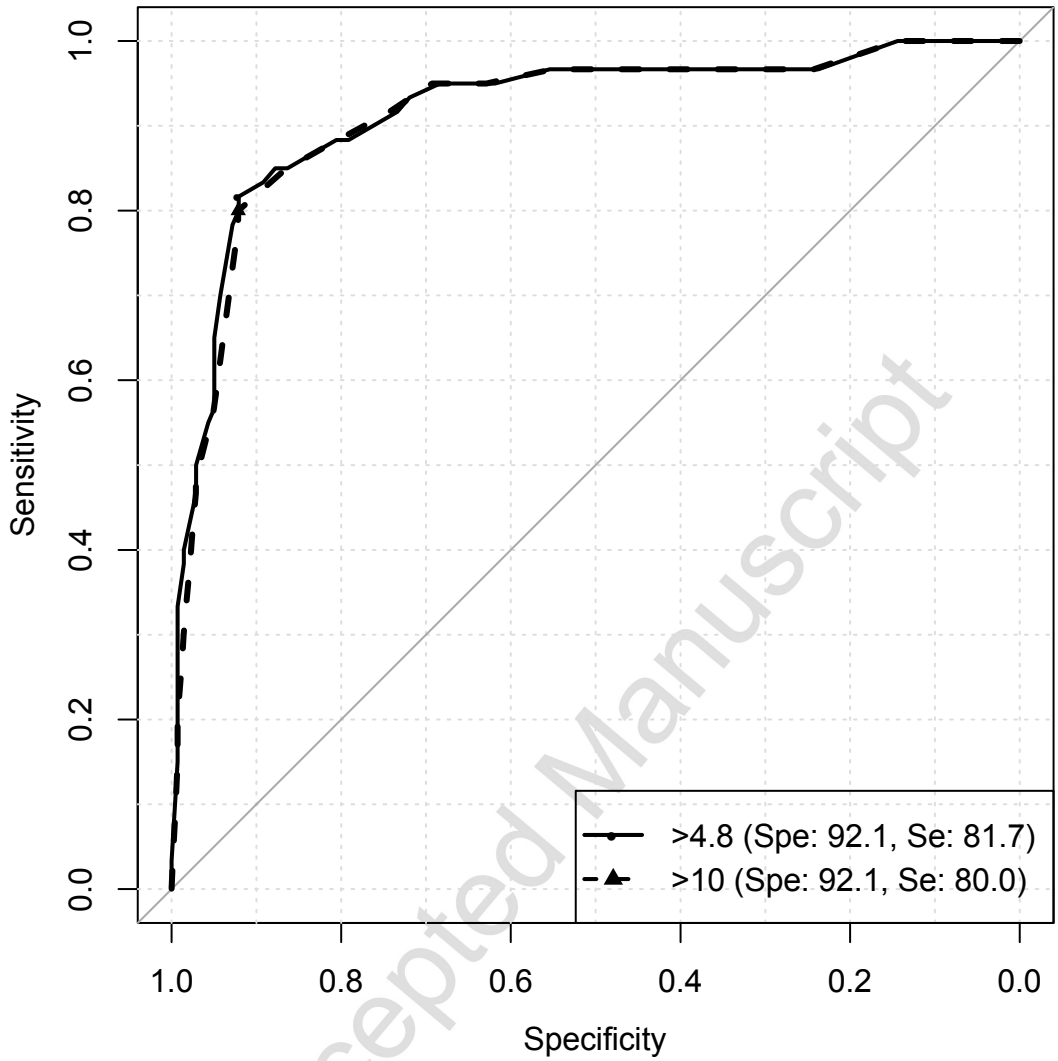
REFERENCES:

1. Verbiest H. A radicular syndrome from developmental narrowing of the lumbar vertebral canal. J Bone Joint Surg Br. 1954;36-B(2):230-7.
2. Ehni G. Spondylotic cauda equina radiculopathy. Tex Med. 1965;61(10):746-52.
3. Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. BMJ. 2016;352.
4. Genevay S, Atlas SJ. Lumbar spinal stenosis. Best Pract Res Clin Rheumatol. 2010;24(2):253-65.
5. Van Gelderen C. Ein orthotisches (lordotisches) Kaudasyndrom. Acta Psychiatr Neurol. 1948;23(1-2):57-68.
6. Haig AJ, Park P, Henke PK, et al. Reliability of the clinical examination in the diagnosis of neurogenic versus vascular claudication. Spine J. 2013;13(12):1826-34.

7. Ishimoto Y, Yoshimura N, Muraki S, et al. Associations between radiographic lumbar spinal stenosis and clinical symptoms in the general population: the Wakayama Spine Study. Osteoarthritis Cartilage. 2013;21(6):783-8.
8. Steurer J, Roner S, Gnannt R, Hodler J. Quantitative radiologic criteria for the diagnosis of lumbar spinal stenosis: a systematic literature review. BMC Musculoskelet Disord. 2011;12:175.
9. Genevay S, Atlas SJ, Katz JN. Variation in eligibility criteria from studies of radiculopathy due to a herniated disc and of neurogenic claudication due to lumbar spinal stenosis: a structured literature review. Spine (Phila Pa 1976). 2010;35(7):803-11.
10. de Schepper EI, Overdevest GM, Suri P, et al. Diagnosis of lumbar spinal stenosis: an updated systematic review of the accuracy of diagnostic tests. Spine (Phila Pa 1976). 2013;38(8):E469-81.
11. Aggarwal R, Ringold S, Khanna D, et al. Distinctions Between Diagnostic and Classification Criteria? Arthritis Care Res (Hoboken). 2015;67(7):891-7.
12. Stynes S, Konstantinou K, Dunn KM. Classification of patients with low back-related leg pain: a systematic review. BMC Musculoskelet Disord. 2016;17:226.
13. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Annals of the rheumatic diseases. 2014;73(6):968-74.
14. Fries JF. Methodology of validation of criteria for SLE. Scand J Rheumatol Suppl. 1987;65:25-30.
15. Genevay S, Courvoisier DS, Konstantinou K, et al. Development of clinical classification criteria for radicular pain caused by lumbar disc herniation: the RAPIDH criteria (RAdicular PaIn caused by Disc Herniation). The Spine Journal. 2017;Available online 5 May 2017.
16. Lin ML, Lin WT, Huang RY, et al. Pulsed radiofrequency inhibited activation of spinal mitogen-activated protein kinases and ameliorated early neuropathic pain in rats. Eur J Pain. 2014;18(5):659-70.
17. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32(4):1008-15.
18. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186-91.
19. Bogduk N. On the definitions and physiology of back pain, referred pain, and radicular pain. Pain. 2009;147(1-3):17-9.
20. Singh JA, Solomon DH, Dougados M, et al. Development of classification and response criteria for rheumatic diseases. Arthritis Rheum. 2006;55(3):348-52.
21. van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! J Clin Epidemiol. 2012;65(11):1212-8.
22. June RR, Aggarwal R. The use and abuse of diagnostic/classification criteria. Best Pract Res Clin Rheumatol. 2014;28(6):921-34.
23. Johnson SR, Goek ON, Singh-Grewal D, et al. Classification criteria in rheumatic diseases: a review of methodologic properties. Arthritis Rheum. 2007;57(7):1119-33.
24. Katz JN, Dalgas M, Stucki G, et al. Degenerative lumbar spinal stenosis. Diagnostic value of the history and physical examination. Arthritis Rheum. 1995;38(9):1236-41.
25. Takahashi N, Kikuchi S, Yabuki S, Otani K, Konno S. Diagnostic value of the lumbar extension-loading test in patients with lumbar spinal stenosis: a cross-sectional study. BMC Musculoskelet Disord. 2014;15:259.
26. Cook C, Brown C, Michael K, et al. The clinical value of a cluster of patient history and observational findings as a diagnostic support tool for lumbar spine stenosis. Physiother Res Int. 2011;16(3):170-8.
27. Konno S, Hayashino Y, Fukuhara S, et al. Development of a clinical diagnosis support tool to identify patients with lumbar spinal stenosis. Eur Spine J. 2007;16(11):1951-7.
28. Sugioka T, Hayashino Y, Konno S, Kikuchi S, Fukuhara S. Predictive value of self-reported patient information for the identification of lumbar spinal stenosis. Family practice. 2008;25(4):237-44.

29. Tomkins-Lane C, Melloh M, Lurie J, et al. Consensus on the Clinical Diagnosis of Lumbar Spinal Stenosis: Results of an International Delphi Study. Spine (Phila Pa 1976). 2016.

1 Figure 1: ROC curve of the score obtained using the full model and the N-CLASS score.




2

3

1 Table 1: Patients' characteristics

| Characteristics reported as means (standard deviation), unless otherwise specified) | Neurogenic claudication caused by LSS N=63 | Other (Radicular pain caused by LDH or NSLBP with referred leg pain) N=146 | *p* |
|---|---|---|---|
| Age | 64.8 (14.3) | 47.9 (14.9) | <0.001 |
| Sex (female) | 30 (49.2%) | 63 (45.0%) | 0.69 |
| Duration of back pain (years) | 7.6 (9.4) | 5.8 (8.1) | 0.18 |
| Duration of leg pain (years) | 3.8 (5.1) | 2.4 (5.3) | 0.06 |
| Worst pain location | | | 0.47 |
| Back | 10 (16.9%) | 33 (24.3%) | |
| Leg | 27 (45.8%) | 61 (44.9%) | |
| Both equally | 22 (37.2%) | 42 (30.9%) | |
| Country of residence | | | 0.48 |
| French-speaking | 30 (47.6%) | 79 (54.1%) | |
| English-speaking | 33 (52.4%) | 67 (45.9%) | |
| Physician specialty | | | 0.83 |
| Surgeon | 35 (55.6%) | 80 (54.8%) | |
| Rheumatologist | 10 (15.9%) | 28 (19.2%) | |
| Rehabilitation spec. | 18 (28.6%) | 38 (26.0%) | |

Table 2: GEE model with logit link and exchangeable correlation matrix to predict neurogenic claudication caused by LSS

| | Estimate | OR | p | score |
|---|---|---|---|---|
| Intercept | -4.514 | 0.00 | <0.001 | -- |
| Age > 60 | 1.761 | 5.82 | <0.001 | 1.8 |
| 30 second extension test | 1.768 | 5.86 | 0.007 | 1.8 |
| Negative SLR-60 | 0.899 | 2.46 | 0.04 | 0.9 |
| Difficulty in squatting due to weakness | 0.382 | 1.46 | 0.44 | 0.4 |
| Patient reports pain in both legs | 1.583 | 4.87 | <0.001 | 1.6 |
| Patient reports pain relieved by sitting | 1.714 | 5.55 | <0.001 | 1.7 |
| Patient reports decreased pain when leaning forward or flexing the spine | 1.257 | 3.52 | 0.03 | 1.3 |

SLR-60: straight leg raise test is positive if leg pain is produced below 60°.

Table 3. Description of clinical tests retain from the multivariable analysis

| | |
|---|---|
| straight leg raise test* | Typical leg pain produced when leg is being raised. Considered negative in the absence of pain below 60° of hip flexion. |
| 30 second extension test* | Typical leg symptoms (e.g pain, paresthesia, weakness) are reproduced during active spine extension performed in standing position for 30 seconds [24] |
| difficulty in squatting due to weakness | Patient not able to squat or unable to raise from a squatting position because of muscle weakness |

* Clinical tests included in the N-CLASS criteria

Table 4: Sensitivity analysis of full and simplified prediction models, with each model's estimation of AUC, sensitivity and specificity.

| | AUC | Thres | Sensitivity | Specificity |
|---|---|---|---|---|
| M1: full model (see Table 2) | 0.917 | >4.8 | 0.817 | 0.921 |
| Simplified model 1: removed "difficulty to squat" | 0.914 | >5 | 0.800 | 0.921 |
| Simplified model 2: "removed negative SLR" | 0.901 | >4.9 | 0.733 | 0.916 |
| Simplified model 3: simplified weighted model, including "negative SLR" but not "difficulty to squat" (see Table 6) | 0.914 | >10 | 0.800 | 0.9214 |

M1: Full model including all variables. Thres: Threshold for a "positive" score, indicative of a strong suspicion of NC due to LSS. SLR: Straight Leg Raise test. SLR is positive if typical leg pain is produced between 0 and 60°.

Table 5: Results of the final (S1) GEE model to predict diagnosis of neurogenic claudication caused by LSS

| | Estimate | OR | p | score |
|---|---|---|---|---|
| Intercept | -5.728 | 0.00 | <0.001 | -- |
| Age > 60 | 1.874 | 6.51 | <0.001 | 1.9 |
| Positive 30 seconds extension test | 1.853 | 5.38 | 0.003 | 1.9 |
| Negative SLR-60 test | 1.016 | 2.76 | 0.03 | 1.0 |
| Patient reports pain in both legs | 1.535 | 4.64 | <0.001 | 1.5 |
| Patient reports leg pain relieved by sitting | 1.602 | 4.96 | <0.001 | 1.6 |
| Patient reports leg pain decreased by leaning forward or flexing the spine | 1.544 | 4.68 | 0.01 | 1.5 |

SLR-60: straight leg raise test; test is said positive if leg pain is reproduced below 60° of passive hip flexion.

Table 6: N-CLASS (Neurogenic CLAudication caused by lumbar Spinal Stenosis) score (simplified weighted score).

| | |
|---|---|
| Age > 60 | 4 |
| Positive 30 seconds extension test | 4 |
| Patient reports pain in both legs | 3 |
| Patient reports leg pain relieved by sitting | 3 |
| Patient reports leg pain decreased by leaning forward or flexing the spine | 3 |
| Negative SLR-60 test | 2 |

SLR-60: straight leg raise test; test is said positive if leg pain is reproduced below 60° of passive hip flexion.

The patient is classified as having Neurogenic Claudication caused by LSS if the total score (ranging from 0 to 19) is 11 or more. Specificity 92.1%, sensitivity 80.0%.