

CLINICAL CLASSIFICATION CRITERIA FOR NEUROGENIC CLAUDICATION CAUSED BY LUMBAR SPINAL STENOSIS. THE N-CLASS CRITERIA

Stéphane Genevay, MD^a, Delphine S. Courvoisier, PhD^{a,b}, Kika Konstantinou, PhD^c, Francisco M. Kovacs, MD^d, Marc Marty, MD^e, James Rainville, MD^f, Michael Norberg, MD^g, Jean-François Kaux, MD^h, Thomas D. Cha, MDⁱ, Jeffrey N. Katz, MD^j, Steven J. Atlas, MD^k

- a. *Division of Rheumatology, University Hospitals of Geneva, 4 rue Gabrielle-Perret-Gentil, 1205, Geneva, Switzerland*
- b. *Quality of Care Division, University Hospitals of Geneva, Rue Gabrielle Perret-Gentil, 1205, Geneva, Switzerland*
- c. *Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG, UK*
- d. *Unidad de la Espalda Kovacs Hospital Universitario HLA-Moncloa Avda de Valladolid 81, 28008, Madrid, Spain*
- e. *Department of Rheumatology, Henri-Mondor Hospital, 51 Ave du MI de Lattre de Tassigny, Créteil, France*
- f. *The Spine Center, New England Baptist Hospital, 125 Parter Hill, MA 02120, Boston, USA*
- g. *Physical Medicine and Rehabilitation, University hospital of Lausanne, Av Pierre-Decker 4, Switzerland*
- h. *Physical Medicine and Sport Traumatology Department, University and University Hospital of Liège, Av de l'hôpital, B35, 4000 Liège, Belgium*
- i. *Department of Orthopaedic Surgery, Massachusetts General Hospital, 55 Fruit St., Yawkey 3A, MA 02114, Boston, USA*
- j. *Orthopedic and Arthritis Center for Outcome Research, Brigham and Women's Hospital, 75 Francis Street, BTM 5016, MA 02115, Boston, USA*
- k. *Division of General Internal Medicine, Massachusetts General Hospital, 50 Stanifort Street, Room 966, MA 02114, Boston, USA*

KEYWORDS : Cauda equina; Classification criteria; Diagnostic process; Low back pain; Lumbar spinal stenosis; Neurogenic claudication; Radicular pain

ABSTRACT

Background Context : Because 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 : The objective of this study was to develop clinical classification criteria that identify patients with neurogenic claudication (NC) caused by LSS.

Study Design : This study is a two-stage process that includes Phase 1, the Delphi process, and Phase 2, the cross-sectional study.

Patient Sample : Outpatients were recruited from spine clinics in five countries.

Outcome Measure : The outcome measure includes items from the patients' history and physical examination.

Methods : In Phase 1, a list of potential predictors of NC caused by LSS was based on the available literature and was evaluated through a Delphi process involving 17 spine specialists (surgeons and non-surgeons) from eight countries. In Phase 2, 19 different clinical spine specialists from five 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. The patients completed survey items and the specialists documented the examination signs. Coefficients from general estimating equation models were used to select predictors, to generate a clinical classification score, and to obtain a receiver operating characteristic curve. Conduction of the Delphi process, data management, and statistical analysis were partially supported by an unrestricted grant of less than 15,000 US dollars from Merck Sharp & 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 that independently predicted NC ($p < .05$) were age over 60 years, 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 and 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%.

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

Introduction

Neurogenic claudication (NC), also called pseudoclaudication [1], [2], is the cardinal symptom caused by lumbar spinal stenosis (LSS) [3]. Lumbar spinal stenosis represents a degenerative process involving the narrowing of the spinal canal around the nerve roots of the cauda equina within the dural sac caused by facet joint osteoarthritis, hypertrophic 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 caused by LSS [6]. On its own, the size of the spinal canal is not a valid diagnostic criterion; because there is no agreement on what defines “normal” and “stenotic,” stenotic images can be seen in asymptomatic subjects, and there is a limited correlation between anatomical 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 to establish homogenous groups of patients for clinical or population studies, which is essential in multicenter studies and contributes to the generalizability of results [9], [11], [12], [13], [14]. For other musculoskeletal diseases (eg, rheumatoid arthritis, spondyloarthropathy), the 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 patients with LBP with leg pain and neurologic signs from other categories of patients with LBP 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.

Methods

The present study was designed according to rules defined by Fries 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 the 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 (Supplementary 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 seven-point scale from 1 (useless) to 7 (very useful).

For Round 1, items were excluded if they had a mean score of <3, had a rating of 1 by more than 25% of the spine specialists, or had more than 50% answering “don't know.” For Round 2, retained items were rescored. 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 (surgeons and non-surgeons), working in English- and French-speaking countries (Supplementary Appendix A), screened patients presenting at their clinics with back-related leg pain for study eligibility. Because 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, patients with radicular pain caused by LDH, or non-specific LBP with referred leg pain were also recruited [19]. Because symptoms of NC are related to central stenosis, patients with only foraminal stenosis were not included. Patients with specific causes of back pain (eg, 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, inability to read or to understand the study site's native language, or declining from 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 the present 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 before 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 three diagnosis groups using any relevant test or procedure felt necessary as part of routine practice (including advanced imaging, magnetic resonance imaging or computed tomography 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 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 non-specific LBP 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 were 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 a p-value of $<.1$ 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 previously.

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 the receiver operating characteristic curve and the 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 positives (ie, patients considered as having an 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 the sensitivity and the specificity, with their respective 95% confidence intervals. All analyses were done using R v3.2.3 (R Foundation for Statistical Computing, Vienna, Austria), 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 the 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 ($100 \times [1 + (15 - 1) \times 0.05]$). 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% (ie, 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, whereas the others were associated with radicular pain caused by LDH. In the first round, 3 of the 96 items were excluded, all based on mean scores of <3, leaving 93 items. In the second 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 caused by an LDH were identified [15].

CLINICAL STUDY

Among 213 enrolled patients (average of 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 non-specific LBP with referred leg pain (Table 1). Tests employed by the spine specialists as part of the diagnostic evaluation included magnetic resonance imaging or computed tomography scan for 203 of the 209 patients and electromyography for 25 of the 209 patients.

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)	<.001
Sex (female)	30 (49.2)	63 (45.0)	.69
Duration of back pain (y)	7.6 (9.4)	5.8 (8.1)	.18
Duration of leg pain (y)	3.8 (5.1)	2.4 (5.3)	.06
Worst pain location			.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			.48
French speaking	30 (47.6)	79 (54.1)	
English speaking	33 (52.4)	67 (45.9)	
Physician specialty			.83
Surgeon	35 (55.6)	80 (54.8)	
Rheumatologist	10 (15.9)	28 (19.2)	
Rehabilitation spec.	18 (28.6)	38 (26.0)	

LSS, lumbar spinal stenosis; LDH, lumbar disc herniation; NSLBP, non-specific low back pain; spec., specialist.
 Data are presented as n (%).

The statistical analysis included items thought to be related to LSS and those thought to be related to LDH (Supplementary Appendix B). Overlapping items (eg, worse pain when sneezing, coughing, or staining) were combined to create single variables; items reported by 10 patients or fewer were dropped from further analyses; and duplicate items (ie, items associated with both clinical entities) were discarded, yielding a final count of 75 items. In the 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 $p < .1$ were included in a subsequent multivariate analysis leading to the identification of seven items (Table 2), four 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 three physician-reported items: negative straight leg raise [SLR] test, positive 30-second extension test, and difficulty in squatting caused by 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), and the cutoff value to obtain a specificity of $\geq 90\%$ resulted in a sensitivity of 81.7%. Removing the item “difficulty in squatting caused by 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.

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	<.001	—
Age >60 y	1.761	5.82	<.001	1.8
30-s Extension test	1.768	5.86	.007	1.8
Negative SLR-60	0.899	2.46	.04	0.9
Difficulty in squatting caused by weakness	0.382	1.46	.44	0.4
Patient reports pain in both legs	1.583	4.87	<.001	1.6
Patient reports pain relieved by sitting	1.714	5.55	<.001	1.7
Patient reports decreased pain when leaning forward or flexing the spine	1.257	3.52	.03	1.3

GEE, generalized estimating equation; LSS, lumbar spinal stenosis; OR, odds ratio; SLR, straight leg raise test; Negative SLR-60 test, SLR test is negative in the absence of pain below 60° of hip flexion.

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

Straight leg raise test*	Typical leg pain produced when the leg is being raised, considered negative in the absence of pain below 60° of hip flexion
30-s Extension test*	Typical leg symptoms (eg, pain, paresthesia, and weakness) are reproduced during active spine extension performed in standing position for 30 s [22]
Difficulty in squatting caused by weakness	Patient not able to squat or unable to rise from a squatting position because of muscle weakness

*Clinical tests included in the Neurogenic Claudication Caused by Lumbar Spinal Stenosis (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 neurogenic claudication caused by lumbar spinal stenosis; SLR, straight leg raise test.

SLR is positive if typical leg pain is produced between 0 and 60°.

Figure. Receiver operating characteristic curve of the score obtained using the full model and the neurogenic claudication caused by lumbar spinal stenosis score. Se, sensitivity; Spe, specificity.

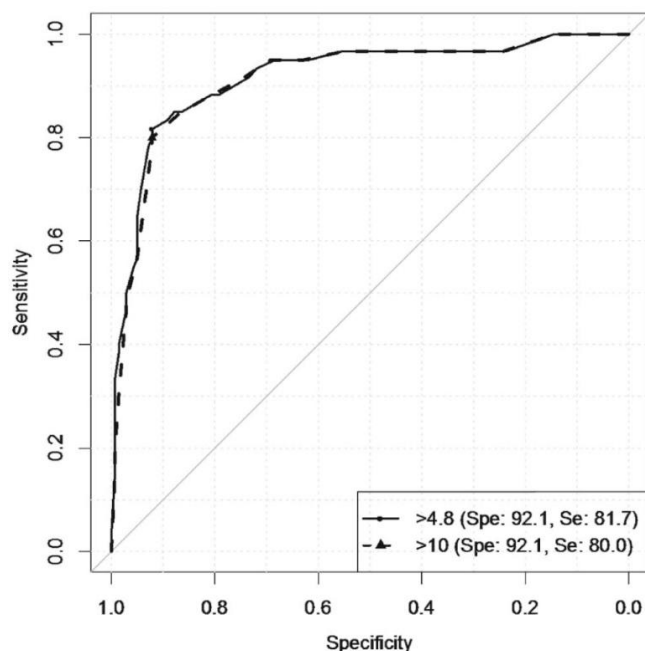


Table 5. Results of the final (S1) GEE model to predict diagnosis of neurogenic claudication caused by LSS

	Estimate	OR	p	Score
Intercept	-5.728	0.00	<.001	—
Age >60 y	1.874	6.51	<.001	1.9
Positive 30-s extension test	1.853	5.38	.003	1.9
Negative SLR-60 test	1.016	2.76	.03	1.0
Patient reports pain in both legs	1.535	4.64	<.001	1.5
Patient reports leg pain relieved by sitting	1.602	4.96	<.001	1.6
Patient reports leg pain decreased by leaning forward or flexing the spine	1.544	4.68	.01	1.5

GEE, generalized estimating equation; LSS, lumbar spinal stenosis; OR, odds ratio; SLR, straight leg raise test; Negative SLR-60 test, SLR test is negative in the absence of pain below 60° of hip flexion.

Items retained in the final model demonstrated fractional weights that varied twofold (see the respective scores in 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 N-CLASS, a weight of 4 is attributed for age >60 years and 30-second extension test, 3 for all patient-reported criteria (ie, feeling pain in both legs, leg pain relieved by sitting, and pain decreased by leaning forward or flexing the spine), and 2 for the 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 cutoff of >10 provided a sensitivity of 80.0% (95% confidence interval: 67.7%–89.2%) and a specificity of 92.1% (95% confidence interval: 86.4%–96.0%) (simplified model 3, Table 4).

Table 6. Neurogenic claudication caused by lumbar spinal stenosis score (simplified weighted score)

Age >60 y	4
Positive 30-s 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 is positive if leg pain is produced below 60°.

The patient is classified as having neurogenic claudication caused by lumbar spinal stenosis if the total score (ranging from 0 to 19) is 11 or more; specificity 92.1%, sensitivity 80.0%.

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 [23]. The use of classification criteria is advocated and promoted for classifying conditions that lack highly specific biomarkers [20], [23], [24]. The present 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 six items, four symptoms and two 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 the physical examination findings. SLR is a typical clinical finding in radicular pain caused by 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 patients with NC [22], [25].

COMPARISON WITH THE EXISTING LITERATURE

Several studies have sought to develop diagnostic criteria to classify patients with NC caused by LSS [22], [26], [27], [28], [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 the criteria 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 [23]. 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 (ie, classification vs. diagnostic).

Whereas two items included in N-CLASS have been reported in one of the five studies on diagnostic criteria (ie, “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 [22], [27], [28] and leg pain relieved by sitting [22], [26], [27], [28]). Overall, at least 50% of the criteria of N-CLASS are included in four of the five 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 (ie, classification vs. diagnostic criteria, see previous discussion). Second, in the Tomkins-Lane study, the Delphi process was performed on a small, preselected 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 (eg, “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 a significant difference among our three groups; see Supplementary Appendix B).

To be of value for basic science, epidemiologic, or clinical research, clinical criteria must have a good specificity [11], [24]. Prior studies of diagnostic criteria reported specificity that was lower (ie, 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), meaning that it would only misclassify 9% of the subjects, identical to Konno et al. [27], but with a much better specificity (92.1% vs. 72%).

STRENGTHS AND LIMITATIONS

The present study was conducted following current recommendations for the development of classification criteria [14]. Face validity is likely to be good, because the items included in N-CLASS are commonly reported in the literature. The diversity of spine specialists involved in the present study suggests good content validity. The inclusion of a heterogeneous population of patients with back-related leg pain also supports good construct validity. The present study also has several limitations. The gold standard used for the 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 (ie, 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

the present study were different from those involved in the Delphi phase. Nevertheless, experts' initial clinical suspicion may have influenced anamnesis and the interpretation of patients' symptoms and findings from physical examination.

Clinicians' skills may also influence the accuracy of the score, because 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 that are routinely taught to medical students. Future studies should also be performed to confirm the validity of the N-CLASS in an independent population.

Conclusions

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

ACKNOWLEDGMENTS

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 thank MSD for their financial support.

SUPPLEMENTARY MATERIAL

Supplementary material related to this article can be found at

<https://doi.org/10.1016/j.spinee.2017.10.003>.

References

- [1] Verbiest H. A radicular syndrome from developmental narrowing of the lumbar vertebral canal. *J Bone Joint Surg Br* 1954;36-B:230–7.
- [2] Ehni G. Spondylotic cauda equina radiculopathy. *Tex Med* 1965;61:746–52.
- [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;24:253–65.
- [5] Van Gelderen C. Ein orthotisches (lordotisches) kaudasyndrom. *Acta Psychiatr Neurol* 1948;23:57–68.
- [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;13:1826–34.
- [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;21:783–8.

- [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* 2010;35:803–11.
- [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* 2013;38:E469–81.
- [11] Aggarwal R, Ringold S, Khanna D, et al. Distinctions between diagnostic and classification criteria? *Arthritis Care Res (Hoboken)* 2015;67:891–7.
- [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. *Ann Rheum Dis* 2014;73:968–74.
- [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). *Spine J* 2017;17.
- [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;18:659–70.
- [17] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32:1008–15.
- [18] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
- [19] Bogduk N. On the definitions and physiology of back pain, referred pain, and radicular pain. *Pain* 2009;147:17–19.
- [20] Singh JA, Solomon DH, Dougados M, et al. Development of classification and response criteria for rheumatic diseases. *Arthritis Rheum* 2006;55:348–52.
- [21] van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *J Clin Epidemiol* 2012;65:1212–18.
- [22] Katz JN, Dalgas M, Stucki G, et al. Degenerative lumbar spinal stenosis. Diagnostic value of the history and physical examination. *Arthritis Rheum* 1995;38:1236–41.
- [23] June RR, Aggarwal R. The use and abuse of diagnostic/classification criteria. *Best Pract Res Clin Rheumatol* 2014;28:921–34.
- [24] Johnson SR, Goek ON, Singh-Grewal D, et al. Classification criteria in rheumatic diseases: a review of methodologic properties. *Arthritis Rheum* 2007;57:1119–33.
- [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;16:170–8.
- [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;16:1951–7.
- [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. *Fam Pract* 2008;25:237–44.
- [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* 2016;41:1239–46.

Appendix A List of Spine Specialists

Age	Gender	Country	Specialty	Clinical care for LBP, years	Practice setting	Involved in research on LBP, years	Nature of research
DELPHI STUDY							
53	Male	France	Rheumatologist	15	hospital	15	observational studies ; clinical trials ; others
50	Male	USA	Internal Medicine	23	primary care	20	observational studies ; systematic reviews
44	Male	Australia	Physiotherapy	21	primary care	10	clinical trials ; systematic reviews
45	Female	UK	Physiotherapist	15	primary care ; hospital	6	observational studies ; clinical trials ; systematic reviews
60	Male	Switzerland	Rheumatology	30	hospital	25	observational studies ; clinical trials
54	Male	Canada	clinical epidemiology /chiropractic	28	hospital	10	observational studies ; systematic reviews
54	Male	Switzerland	Rheumatology	10	hospital	5	observational studies ; clinical trials ; systematic reviews ; others
46	Male	Spain	back pain	27	primary care	24	observational studies ; clinical trials ; systematic reviews
53	Male	France	Rheumatologist	15	hospital	15	observational studies ; clinical trials ; systematic reviews ; others
47	Male	USA	Internal Medicine	12	primary care ; hospital	12	observational studies ; clinical trials ; systematic reviews ; others
50	Male	united states	general internal medicine	23	primary care	20	observational studies ; systematic reviews
39	Male	Netherlands	Epidemiology, physiotherapy	12	primary care	8	observational studies ; clinical trials ; systematic reviews
35	Male	USA	Physical & Rehabilitation Medicine	4	hospital	4	observational studies ; systematic reviews
52	Female	Australia	Rheumatology	20	hospital	15	observational studies ; others
50	Male	Canada	Epidemiology	0	N/A	25	observational studies ; clinical trials ;

53	Female	Canada	Rehabilitation Medicine	4	primary care	25	systematic reviews ; others observational studies
44	Male	Canada	Sports medicine, occupational medicine	11	primary care	5	observational studies ; systematic reviews

CLINICAL STUDY

50	Male	Switzerland	Physical & Rehabilitation Medicine	15	hospital	13	observational studies
44	Female	France	Rheumatology	17	hospital	13	observational studies ; clinical trials
57	Female	France	Rheumatology	25	hospital	20	observational studies ; clinical trials
54	Female	Belgium	Rheumatology	22	hospital	0	
35	Male	Belgium	Physical & Rehabilitation Medicine	10	hospital	2	observational studies
46	Male	Switzerland	Orthopaedic Spine Surgeon	10	hospital	10	observational studies ; clinical trials
35	Female	France	Rheumatology	9	hospital	0	
35	Female	France	Rheumatology	8	hospital	0	
38	Male	USA	Orthopaedic Spine Surgeon	5	hospital	5	
37	Male	USA	Physical & Rehabilitation Medicine	5	hospital	5	observational studies ; clinical trials
52	Male	USA	Physical & Rehabilitation Medicine	25	hospital	22	observational studies ; clinical trials; systematic review
41	Male	UK	Neurosurgeon	10	hospital	10	observational studies ; clinical trials
57	Male	UK	Neurosurgeon	29	hospital	22	observational studies ; clinical trials
NA	Female	UK	Neurosurgeon	NA	hospital	NA	
NA	Male	UK	Neurosurgeon	NA	hospital	NA	
43	Male	UK	Orthopaedic Spine Surgeon	12	hospital	10	observational studies ; clinical trials
50	Male	UK	Orthopaedic Spine Surgery	15	hospital	18	observational studies ; clinical trials
45	Male	UK	Orthopaedic Spine Surgeon	15	hospital	5	observational studies ; clinical trials
NA	Male	UK	Orthopaedic Spine Surgeon				

NA : not available

Appendix B Items selected at the end of the Delphi and their frequency of report in the clinical phase

Criterion	Reference Condition (RP vs. NC) *	NC caused by LSS (N=63)	Other diagnostic (N=146)	p
<i>Physician documented symptoms</i>				
Age > 60 y.o.	NC	44	48	<0.001
Past history of low back pain	RP	51	128	0.29
Past history suggesting radiculopathy	RP	32	69	0.78
Pain radiates to ankle	RP	37	110	0.02
Any monoradicular pain	RP	34	99	0.07
Monoradicular pain L3 (anterior thigh +/- knee)	RP	5	7	0.88
Monoradicular pain L4 (anterior thigh + knee +/- tibial medial)	RP	8	12	0.72
Monoradicular pain L5 (lateral leg, anterior ankle, dorsum foot, inner side of the foot, great toe)	RP	16	40	0.90
Monoradicular pain S1 (posterior leg, heel, outside foot and sole, little toe)	RP	8	48	<0.001
Increased back pain with Valsalva	RP	9	41	0.05
Increased leg pain with Valsalva	RP	15	52	0.13
History of increased pain during extension	NC	32	42	<0.001
Urinary incontinence	RP/NC	3	0	0.04
Urinary retention	RP/NC	0	1	1.00
Perianal numbness	RP/NC	2	2	0.75
<i>Physician documented exam</i>				
Increased pain during extension	NC	39	45	<0.001
Buttock/leg symptoms following 30" of lumbar extension	NC	36	37	<0.001
Disturbed squat	RP/NC	16	13	<0.001
Disturbed walking on heels	RP/NC	10	20	0.82
Disturbed walking on toes	RP/NC	8	13	0.54
Unable to do 10 toe raise	RP/NC	14	20	0.15
Weakness after walking	NC	5	4	0.16
Lasegue sign	RP	13	64	<0.001
SLR , any angle	RP	15	66	0.01
SLR with 60° angle (as estimated by the clinician)	RP	8	60	<0.001
Crossed SLR	RP	2	14	0.19
Flip test	RP	5	41	<0.001
Femoral traction	RP	7	10	0.43
Hypoesthesia in a dermatome	RP	20	42	0.85
Perianal sensory	RP/NC	0	1	1.00
Decreased rectal tone	RP/NC	1	0	0.66
Leg muscle atrophy	RP/NC	6	7	0.32
Peripheral pulses	NC	50	114	0.89

Quadriceps weakness				<0.001
Unilateral	RP	5	6	
Bilateral	NC	5	0	
Normal		50	136	
Quadriceps weakness correlates with pain distribution	RP	6	8	0.59
Ankle dorsiflexion weakness				<0.001
Unilateral	RP	18	1	
Bilateral	NC	1	4	
Normal		125	57	
Ankle dorsiflexion weakness correlates with pain distribution	RP	3	13	0.01
Ankle eversion weakness				0.40
Unilateral	RP	2	10	
Bilateral	NC	1	1	
Normal		58	132	
Ankle eversion weakness correlates with pain distribution	RP	2	5	0.26
Hallucis weakness				0.13
Unilateral	RP	4	18	
Bilateral	NC	3	2	
Normal		55	124	
Hallucis weakness correlates with pain distribution	RP	4	10	0.11
Plantar weakness				<0.001
Unilateral	RP	0	11	
Bilateral	NC	2	0	
Normal		59	132	
Plantar weakness correlates with pain distribution	RP	1	5	0.10
Patellar reflex				<0.001
Unilateral	RP	45	128	
Bilateral	NC	7	14	
Normal		10	2	
Patellar reflex correlates with pain distribution	RP	4	8	0.25
Ankle reflex				<0.001
Unilateral	RP	45	113	
Bilateral	NC	2	22	
Normal		15	15	
Ankle reflex correlates with pain distribution	RP	4	17	0.00
<i>Patient reported symptoms</i>				
Worse pain in				0.46
Back		10	33	

Leg(s)	RP	27	61	
Both		22	42	
Pain in both legs	NC	41	26	<0.001
Pain below knee	RP	47	115	0.63
Numbness in leg	NC	36	93	0.39
Pins and needles in leg	RP	28	95	0.01
Worse pain when coughing	RP	18	49	0.54
Worse pain when sneezing	RP	19	54	0.43
Worse pain when straining	RP	31	61	0.42
Worse pain when sitting	RP	18	95	<0.001
Pain brought by walking	NC	51	85	<0.001
Pain increase with walking	NC	54	89	<0.001
Pain stops you from walking	NC	52	81	<0.001
Incline forward while walking	NC	42	50	<0.001
Leg symptoms when cycling	NC	7	17	1
Weakness in legs while walking	NC	45	77	<0.001
Pain decreased by sitting	NC	9	86	<0.001
Pain relieved by sitting	NC	40	25	<0.001
Pain decreased by stooping	NC	25	42	0.15
Pain relieved by stooping	NC	17	17	0.01
Pain decreased by leaning forward	NC	48	65	<0.001
Pain relieved by leaning forward	NC	27	29	<0.001
Pain decreased by flexing	NC	30	39	<0.001
Pain relieved by flexing	NC	15	20	0.10
Pain relieved by standing still	NC	20	25	0.02
Perianal numbness	RP/NC	6	6	0.22
Difficulty urinating	RP/NC	3	7	1
Urinary incontinence	RP/NC	11	6	<0.001
Bowel incontinence	RP/NC	5	2	0.03
Combined variables				
Femoral traction test or Lasegue sign	RP	18	69	0.02
Femoral traction test or SLR	RP	20	72	0.03
Femoral traction test or SLR 60°	RP	14	67	<0.001
Femoral traction or flip test	RP	11	49	0.03
Bilateral motor weakness	NC	44	112	<0.001
Bilateral decreased reflex	NC	16	8	<0.001
Leg pain increased with Valsalva	RP	48	94	0.11
Symptoms of Cauda Equina	RP/NC	4	2	0.07
Reference condition: RP = Radicular pain caused by lumbar disc herniation; NC = neurogenic claudication caused by lumbar spinal stenosis.				
DH: disc herniation. SLR: straight leg raise test				