Published in final edited form as:

Spine J. 2013 December 1; 13(12): . doi:10.1016/j.spinee.2013.06.021.

Reliability of clinical examination in the diagnosis of neurogenic vs. vascular claudication

Andrew J. Haig, M.D.^{*}, Paul Park, M.D.[#], Peter Henke, M.D.⁺, Karen S.J. Yamakawa, M.S.^{*}, Christy Tomkins, Ph.D.[%], Juan Valdivia, M.D.[#], and Sierra Loar, MPA^{*}

^{*}Department of Physical Medicine and Rehabilitation, The University of Michigan

[#]Department of Neurological Surgery, The University of Michigan

⁺Department of Vascular Surgery, The University of Michigan

[%]Mount Royal University, Calgary, Alberta, Canada

Introduction

Lumbar spinal stenosis is the most common reason for older people to undergo spine surgery. In the United States, the operation rate is both escalating and highly varied across regions (1).

For decades, various researchers, reviewers, and consensus committees have accepted the general concept that lumbar spinal stenosis is a syndrome of clinical presentation associated with radiological evidence. However a number of recent studies have challenged the relationship between imaging results and the clinical syndrome (2, 3, 4, 5). The latter half of the definition, a 'clinical syndrome', has never been defined in a quantifiable or reproducible manner. In fact, it is widely recognized that people who have clinical stenosis may or may not have a number of different presentations, ranging from back pain to leg pain to painless weakness to ataxia, each of which can be caused by other disorders (6).

Neurogenic claudication is the feature most identified with spinal stenosis. Characteristically it is described as back or leg pain that worsens with ambulation or prolonged standing, reaching a point where the sufferer cannot continue without resting, typically with the trunk flexed (6). Animal and human studies have created strong pathophysiological ties between compromise within the spinal canal and claudication (7, 8). A recent review concluded that persons without neurogenic claudication were 1/4th as likely to have spinal stenosis, although it should be noted that the review used a somewhat circuitous reference standard of 'expert opinion with radiographic or anatomic confirmation' (9).

Vascular claudication is a common symptom that presents in a way similar to neurogenic claudication (10, 11). One common test for the cause of vascular claudication, peripheral vascular disease, is the ankle-brachial index (ABI). The ABI is a simple test in which blood pressure cuffs are inflated in the arms and legs and a ratio of blood pressures is created. It is

© 2013 Elsevier Inc. All rights reserved.

Address correspondence to Dr. Haig at: The University of Michigan, 325 E. Eisenhower, Ann Arbor, MI 48108, 734-763-4200, Fax 734-615-1770, andyhaig@umich.edu.

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 citable 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.

highly sensitive and specific to the presence of peripheral vascular disease (12). However, as demonstrated in the Framingham Offspring Study, only about half of people with positive ABI's have *symptoms* of vascular claudication (13). Thus, in a conundrum similar to that encountered by spine physicians, vascular physicians cannot rely on a simple test to prove that claudicatory symptoms are coming from the arteries.

In light of this information, the validity of the clinician's impression that a person has neurogenic claudication has taken on new importance, especially when the clinician is a surgeon who might offer invasive treatment. This paper is designed to examine the reliability of a surgeon's diagnostic impression of neurogenic claudication or vascular claudication based upon an extensive history and physical examination, performed under the most optimal circumstances.

Methods

Subjects

In an ethical review board approved study, three groups of persons aged 55–90 were recruited. The figure shows the recruitment and testing process. A neurogenic claudication group was identified from review of the computerized medical record of neurosurgery and orthopedic surgery spine clinics for persons who were diagnosed with spinal stenosis, had difficulty walking 200 yards (approximately 200 meters), and were offered surgery. The vascular claudication group was identified through review of the records of the vascular surgery, vascular medicine, and vascular diagnostic laboratory for persons who complained of leg pain with ambulation which their doctor attributed to vascular claudication and who also had a positive ankle-brachial index. Asymptomatic volunteers were recruited from postings seeking persons without back pain or vascular disease from the community. Placement in each of these groups will be referred to as the 'recruitment diagnosis'.

Through review of medical record and interviews, potential subjects were excluded from the study if they had previous back surgery, cardiopulmonary precautions for ambulation testing, or risk factors for neuromuscular disease (diabetes, alcohol more than 12 drinks per week, personal or family history of neuromuscular disease, or history of previous significant focal lower limb nerve injury). Because a later phase of the study involved imaging and electrodiagnostic testing, contraindications to MRI scanning (metal, obesity, unmanageable claustrophobia), relative contraindications or technical issues related to EMG testing (previous lumbar surgery, coumadinization, severe immune disorder, extreme obesity, 3+ or greater pitting edema, or implanted electrodes such as defibrillators), were exclusions. Among the groups with the recruitment diagnoses of neurogenic or vascular claudication, anyone who had another disorder that, in the investigator's opinion, might be limiting ambulation more than the vascular or neurogenic claudication, were eliminated. This provision included persons with severe osteoarthritis or other unforeseen diagnoses. Subjects were required to be competent and willing to travel to the testing appointments at their own expense. All subjects were compensated for their efforts.

Testing protocol

Subjects completed the lengthy standard university spine program questionnaire. This includes basic demographic, medical, social, family, and spine history, and an extensive review of systems. Standardized measures analyzed in the current study include a 10 cm visual analog scale (VAS) for pain (14), the Pain Disability Index (15), the McGill Pain Scale (16), and the Quebec Back Pain Disability Scale (17). Other validated scales administered included the Swiss Stenosis Questionnaire (18), the Walking Impairment Questionnaire (19), the Multidimensional Pain Inventory (20), and the SF-36 (21).

All subjects underwent ABI testing per the protocol described by McDermott (2001), using a hand-held doppler probe (Nicolet Vascular Pocket Dop II, Golden, Colo). The ABI was done to ensure that neurogenic and asymptomatic volunteers did not have coincident vascular disease and to ensure that the clinically performed ABI's on the vascular patients were reproducibly abnormal. Those whose ABI results did not match their recruitment diagnosis were eliminated. Specifically, neurogenic and normal subjects were eliminated for a positive ABI, and vascular were eliminated for a negative study

Each subject underwent a history and physical examination by one of three faculty spine neurosurgeons and a single vascular surgeon. The history included a required review of the extensive questionnaires. Afterwards the clinicians were also asked to obtain a spine symptom history as they would in clinic, with the exception that they were not allowed to ask about diagnostic tests, treatments, or the names or specialties of physician consultants. They were allowed to ask about other medical problems and symptoms, however aside from the questionnaire, no specific questions were required. During the physical examination the surgeons were required to observe for spinal deformity, hair loss, varicosities, and pain provocation with testing of range of motion and palpation. In addition they performed a detailed lower limb test for strength and deep tendon reflexes, plantar reflex, sensation at the great toe, nerve tension signs, capillary refill and dorsalis pedis pulse, and observed for pain reproduction on hip range of motion. The physicians were not constrained by this extensive history and physical examination, and were encouraged to ask questions and perform maneuvers as needed to clarify their impression.

Masking was achieved by first having a research assistant coach the subjects to provide truthful information about all aspects of their complaints and medical history, but to not provide any information about specialists, tests, or treatments related to their claudication complaint. The surgeons were instructed not to seek such information. The research assistant was present in the room to enforce this rule and observe any violations. After testing the surgeon rated any potential unmasking events.

The main outcome from examination was a single bottom-line choice of neurogenic claudication, vascular claudication, both, or neither. In addition the clinicians' rated certainty and severity of each of these diagnoses on a 100 mm visual analog scale. To ensure that no other confounding diagnoses could be detected, the clinicians were required to rate the possibility of three other causes for leg pain; venous insufficiency, painful limb arthritis, and polyneuropathy on 100mm VAS scales; and were asked to write in any other pertinent diagnoses.

Statistical analysis

After importing the cleaned data from an Excel spreadsheet, PASW Statistics 18 (SPSS Inc., Chicago, Illinois) was used for checking duplicate records and analyzing data. Chi-square tests were performed to examine the relationship between categorical variables. One-way analysis of variance (ANOVA) was used to compare group mean differences, with a Tukey HSD adjustment for multiple pair-wise comparisons between group means. Cohen's Kappa test was used to examine the strength of association between the opinions on diagnostic and clinical findings of different specialists, (e.g. neurosurgeon and vascular surgeon). Kappa statistics are appropriate for looking at associations between categorical variables that have the same number of categories. Significant results indicate that the two variables of interest are somehow associated. We adopted Fleiss's (22) interpretations of Kappa 0.75 as excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor. In all the analyses, a p-value of < 0.05 was considered as statistically significant.

Results

The process of screening is documented in Table 1. Among persons contacted, proportionally fewer vascular recruits agreed or qualified to participate. Anticoagulation was a frequent reason for exclusion. Table 2 describes the characteristics of the three populations included in the final analysis. A one-way ANOVA was used to test for differences in demographical, pain and disability variables among the three groups (neurogenic claudication, vascular claudication and asymptomatic). Significant differences between the groups were found for age, McGill total score, Pain Disability Index, VAS pain average, and Quebec Back Pain Disability Scale. The asymptomatic volunteers were significantly younger (Tukey post-hoc comparison $M = 61.7$ years, 95% CI [59.6, 63.8]) than the vascular group. ($M = 68.9$ years, 95% CI [63.1, 74.6], $p = 0.024$). As expected they also had lower pain and disability scores than the other groups, all $p < 0.01$. The neurogenic claudication group had significantly higher Pain Disability Index ($p = 0.014$) and VAS pain average ($p = 0.021$) than the vascular claudication group.

Study surgeon agreement with recruitment diagnosis was excellent (22), with Kappa's of 0.761 and 0.803 for the neurosurgeons and vascular surgeon respectively ($p < 0.001$). The neurosurgeons and vascular surgeon disagreed with recruitment diagnosis in 14% and 12% of cases, however, and 6 cases which the vascular surgeon felt had 'both' diagnoses were excluded from analysis. Five of these were recruited as neurogenic and one as vascular claudication. (Table 3) The two masked surgeons who examined an individual agreed on diagnosis in 70/84 (83%) cases, with a Kappa value of 0.717 ($p < 0.001$) (Table 4). This analysis did not include the 6 cases rated as 'both' by the vascular surgeon, which would raise the potential rate of disagreement from 17% to 22%.

A number of aspects of the patient history related to the surgeon's diagnosis. There were significant group mean differences in VAS pain; persons whom the vascular surgeon and one of the three neurosurgeons agreed had neurogenic claudication had higher pain than the agreed-upon vascular cases (post-hoc Tukey analysis $M = 5.7$, 95% CI [4.8, 6.5] versus $M = 2.6$, 95% CI [0.0, 5.2], $p = 0.014$). The asymptomatic subjects were not included in this analysis (Table 5). In contrast, Table 6 demonstrates that there was poor inter-rater reliability for most parts of the physical examination.

Discussion

This study found that the clinical impression of neurogenic claudication and vascular claudication is a valid construct, reproducible between the presenting diagnosis made by the treating surgeon, one masked vascular and three masked spine neurosurgeon at our institution. However, in this very carefully selected population of severely involved persons and completely asymptomatic persons with no other confounding disorders, there were also a number of disagreements between clinicians. This raises a concern that the discriminant value of the clinical impression of claudication may not be adequate when the stakes are high, as is the case when spine surgery is considered.

The first descriptions of neurogenic claudication and vascular claudication are attributed to giants in the field of medicine, DeJérine (1911) and Charcot (1858) cited by Gilfillan more than half a century ago (11). Since then the possibility of confusion between neurogenic and vascular claudication has been a common topic of discussion and teaching. However we could find no rigorous research comparing the two. Dodge (23) found that, of 172 patients with symptoms of claudication and lumbar spinal stenosis 'proved' by now disputed myelography or computed tomography criteria, nine had peripheral vascular disease identified with ultrasonography and arteriography. Fasih (24) reviewed vascular surgical

clinic records to show that, of 1070 patients, 33 were diagnosed with a non-vascular problem, including 28 who were thought to have spinal stenosis, again with debatable evidence.

Despite this work, the intersection between vascular and neurogenic claudication had not been adequately explored prior to this study. We found that the masked study surgeons demonstrated good or substantial agreement, matching each other in 83% of cases. Factors that relate to agreement can be explored to understand the relative value of different aspects of the clinical encounter in making a final diagnosis.

Table 5 shows a number of strong relationships regarding the patient report of pain and disability. In contrast, poor, slight, and fair levels of agreement were found for four pertinent components of the physical examination. This suggests that these physical examination findings were not the factors used by surgeons in coming to agreement on diagnosis, even where they may have added to certainty. The sample size was not sufficient to explore various specific aspects of the physical examination or the extensive history and patient questionnaires, so it is quite possible that some specific factors are highly diagnostic. It is also possible that a more careful orientation to the examination, such as training to ensure similar skill in performing these basic maneuvers or additional training to ensure similar scoring might have increased the inter-rater reliability. Nevertheless, these findings suggest that listening to the patient's symptoms -and the patient's ability to provide a clear and rational description of their symptoms - are critical aspects of the diagnosis.

The study needs to be interpreted in light of its methodology. Research on inter-rater reliability of the history and physical examination requires some level of confidence that the clinicians are sufficiently trained and experienced. These faculty surgeons were quite familiar with the diseases and with the history and physical examination components that they were required to perform. The framework within which the surgeons decided on the diagnosis is also important. Their agreement came only after a requirement to review a patient questionnaire that includes more data and more standardized testing than any clinical surgical encounter that we know of, and their physical examination requirements may also have been well beyond the usual practice of a surgeon. The question they knew they would be asked had only three possible answers, in great contrast to the differential diagnosis generated in the real world. Both of these factors drive up the inter-rater reliability compared to other situations, so any extrapolation from the current study would need to assume that a similar construct was applied.

Variations on this work can result in different information. A study that included more surgeons, surgeons with more diverse training, experience, or practice patterns, might show different results. There might be value in studying reliability between other types of clinicians such as physiatrists, neurologists, therapists, or primary care physicians. These persons may even see the syndrome differently than our small number of surgeons. A study that included persons with disorders such as mechanical back pain would better represent the more complex and equivocal decision-making in the community. There is value in performing each of these studies, but the agreement between surgeons is perhaps the clinically most important issue regarding reproducibility of the diagnosis, since surgeon's diagnostic impressions may drive a choice of invasive procedures.

The current study was designed to examine the question of reliability in the most optimal and clinically obvious situation. There is no gold standard for the diagnosis of stenosis. For stenosis subjects the offer of surgery by faculty surgeons represents a high level of certainty and severity that exceeds the typical definition of stenosis, which specifies only criteria of imaging evidence and clinical impression. To our knowledge no study overtly uses this as a

criterion for diagnostic certainty or severity. However this criterion means that the study population more closely matches major trials such as the SPORT study, where patients must be at a stage where they could reasonably be randomized or selected for surgery. (25)

Statistical agreement is reassuring to scientists, who can now use the clinical impression as a valid metric for inclusion in research studies. A caveat is that the impression must be based on the exhaustive information gathered in this study. In clinical care an acceptable level of certainty is more complex. It involves consideration of the risks and benefits of whatever procedure is contemplated. For example, the eight subjects who were offered surgery for stenosis by surgeons who had seen their imaging tests, but who were seen as asymptomatic volunteers by our expert neurosurgeon may have wished for more certainty.

While surgery may be highly effective for persons who have failed aggressive conservative intervention, it cannot be taken lightly. A meta-analysis of surgical complications from spinal stenosis surgery (26) noted a 12% complication rate, including 1% chance of deep infection, 2.8% chance of deep venous thrombosis, and 0.32% chance of perioperative death. Re-operation rates are quoted at 6% to 18% over periods of 1–4 years (27).

One hopeful answer is electrodiagnostic testing. Two studies have shown nearly 100% specificity of a certain electrodiagnostic protocol for clinically apparent stenosis and, depending on the severity of presentation, moderate to excellent sensitivity for the clinical impression of spinal stenosis, while detecting common neuromuscular diseases that can confound the diagnostic impression (2). In cases where diagnostic certainty is important—including persons who have anything but a classical clinical diagnosis and are being considered for surgery, this electrodiagnostic protocol is indicated.

Finally, there is the question of terminology. Lumbar spinal stenosis is often called a syndrome, but in fact it is a disease. A disease is defined by certain pathophysiological changes. In the case of lumbar stenosis the disease is defined by pathophysiological changes to the nerve roots of the lumbar spine that occur as a result of insufficient space. A syndrome is a constellation of findings that some group of experts agrees upon. The people who define a syndrome often hope that this set of rules is specific enough that others can use them with good reproducibility. They frequently seek out evidence that the syndrome approximates the disease, so that they can design research and clinical interventions that impact the disease.

For generations various clinicians, scientists, and policymakers have used the term ‘syndrome’ to describe stenosis. Initially the association of that word with stenosis was likely an attempt to warn physicians away from treating radiological findings without clinical findings. However the definitions of the syndrome have continued to be vague to the point of uselessness. Typical definitions found in practice guidelines and study inclusion criteria consist of the clinician’s impression based on (unspecified) history criteria plus radiological evidence based on no specific measurement. To meet this vague standard the current study does include persons who couldn’t walk a certain distance, had MRI scans that their surgeon thought showed stenosis, and did not have other diseases to explain the walking problem. The ‘offer to operate’ was imposed so that there would be little question that the non-study clinician felt that there was a disease of consequence.

One might wish for a more specific definition of the syndrome. No group of experts has agreed upon a viable or reproducible one. In fact, when researchers have attempted to make the radiological criteria more specific—by defining certain diameters or areas of the thecal sac or spinal canal, or even the radiologist’s gestalt impression—the syndrome lost any meaningful relationship to the disease. (3, 4, 28, 29) Regarding the clinical impression of claudication, the current study shows that; after reviewing this study’s questionnaires, asking

clarifying questions, and performing the stated physical examination; a surgeon's impression that a person has neurogenic claudication, holds together fairly well, but not perfectly. While words like neurogenic claudication are often discussed in guidelines and other documents, no strict rules for clinically labeling a patient as having neurogenic claudication have evolved. This protocol may help.

Neurogenic claudication, however, doesn't seem to be a sine-qua-non for the clinical side of stenosis. No guideline to our knowledge excludes persons if they do not have neurogenic claudication. Such a definition would exclude findings such as static neurological deficit or even back pain that some experts believe reflect the pathophysiology of stenosis. A recent study took an innovative approach to illustrating other variables that are actually considered by clinicians. (30) In a recursive on-line survey of 97 Physical Medicine and Rehabilitation physicians, Sandella found "Leg pain while walking" (66%), "must sit down or bend" (66%) and "flex forward while walking" (49%) were the most commonly selected questions. "Normal foot pulses" (19%), "back pain" (16%) "leg pain" (15%), "relief with rest" (14%), and "sensory deficits" (12%) were of intermediate value, while "problems with balance", "haven fallen recently", and "the sacroiliac joint is not the main pain generator" were all chosen less than 5% of the time. Statistically significant ($P<.05$) change in certainty ceased after 5 questions at 86.2% certainty. This is yet another method of seeking consensus before agreeing upon a syndrome.

It is entirely possible that the expanding and varied use of surgery for stenosis can be traced to vague and never-validated radiological criteria combined with scarce evidence that any particular clinical presentation is specific or reliable. There are potential solutions. Faced with an even more daunting lack of anatomical proof the American Psychological Association has gone far in advancing diagnosis and treatment of mental health problems through its rigorous Diagnostic and Statistical Manual (30). The Diagnostic and Statistical Manual is explicit about inclusion and exclusion from certain psychiatric syndromes. While it is revised as more clinical knowledge evolves, and people may well have diseases but not qualify under the manual's rules, the consensus on measurable and reproducible criteria has served that field well.

Perhaps spine policy makers can make a decision that will allow spine care to emulate this success. Regardless, until research shows that some well-defined syndrome precisely matches the disease, educators, researchers, and policymakers must be careful to provide specific framework for the complaints that they diagnose, research, and teach. More important, individual clinicians must recognize the vulnerability of their judgment as they propose treatments that have real risk and potential benefit.

Acknowledgments

Siera Goodnight, Derek Wood, and Danielle Sandella assisted in testing and recruitment of subjects.

The project described was supported by Award Number R01HD059259 from the Eunice Kennedy Shriver National Institute Of Child Health & Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute Of Child Health & Human Development or the National Institutes of Health.

References

1. Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI. United States trends in lumbar fusion surgery for degenerative conditions. *Spine*. 2005; 30:1441–1445. [PubMed: 15959375]
2. Haig AJ, Tomkins C. Diagnosis and Treatment of Lumbar Spinal Stenosis. *JAMA*. 2010; 303:71–72. [PubMed: 20051574]

3. Haig AJ, Geisser ME, Tong HC, et al. Electromyographic and magnetic resonance imaging to predict lumbar stenosis, low-back pain, and no back symptoms. *J Bone Joint Surg Am.* 2007; 89:358–366. [PubMed: 17272451]
4. Kalichman L, Cole R, Kim DH, et al. Spinal stenosis prevalence and association with symptoms: the Framingham Study. *Spine J.* 2009; 9:545–550. [PubMed: 19398386]
5. Yagci I, Gunduz OH, Ekin G, Diracoglu D, Us O, Akyuz G. The utility of lumbar paraspinal mapping in the diagnosis of lumbar spinal stenosis. *Am J Phys Med Rehabil.* 2009; 88:843–851. [PubMed: 19661776]
6. North American Spine Society. Evidence Based Clinical Guidelines for Multidisciplinary Spine Care: Diagnosis and Treatment of Degenerative Lumbar Spinal Stenosis. Burr Ridge, IL: North American Spine Society; 2007.
7. Sekiguchi M, Kikuchi S, Myers RR. Experimental spinal stenosis: relationship between degree of cauda equina compression, neuropathology, and pain. *Spine.* May 15.2004 29:1105–1111. [PubMed: 15131438]
8. Kobayashi S, Uchida K, Takeno K, Baba H, Suzuki Y, Hayakawa K, Yoshizawa H. Imaging of cauda equina edema in lumbar canal stenosis by using gadolinium-enhanced MR imaging: experimental constriction injury. *AJNR Am J Neuroradiol.* 2006; 27:346–353. [PubMed: 16484408]
9. Suri P, Rainville J, Kalichman L, Katz JN. Does this older adult with lower extremity pain have the clinical syndrome of lumbar spinal stenosis? *JAMA.* 2010; 304:2628–2636. [PubMed: 21156951]
10. Hirsch AT, Haskal ZJ, Hertzner NR, Bakal CW, Creager MA, Halperin JL, Hiratzka LF, Murphy WR, Olin JW, Puschett JB, Rosenfield KA, Sacks D, Stanley JC, Taylor LM Jr, White CJ, White J, White RA, Antman EM, Smith SC Jr, Adams CD, Anderson JL, Faxon DP, Fuster V, Gibbons RJ, Hunt SA, Jacobs AK, Nishimura R, Ornato JP, Page RL, Riegel B. American Association for Vascular Surgery. Society for Vascular Surgery. Society for Cardiovascular Angiography and Interventions. Society for Vascular Medicine and Biology. Society of Interventional Radiology. ACC/AHA Task Force on Practice Guidelines Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease. American Association of Cardiovascular and Pulmonary Rehabilitation. National Heart, Lung, and Blood Institute. Society for Vascular Nursing. TransAtlantic Inter-Society Consensus. Vascular Disease Foundation. ACC/AHA 2005 Practice Guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease): endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. *Circulation.* 2006; 113:e463–e654. [PubMed: 16549646]
11. Gilfillan RS, Berry JD. Diagnosis and surgical treatment of arteriosclerosis. *Geriatrics.* 1954; 9:409–420. [PubMed: 13191545]
12. Bernstein EF, Fronek A. Current status of non-invasive tests in the diagnosis of peripheral arterial disease. *Surg Clin North Am.* 1982; 62:473–487. [PubMed: 7089825]
13. Murabito JM, Evans JC, Nieto K, Larson MG, Levy D, Wilson PW. Prevalence and clinical correlates of peripheral arterial disease in the Framingham Offspring Study. *Am Heart J.* 2002; 143:961–965. [PubMed: 12075249]
14. Jensen MP, Karoly P, Braver S. The measurement of clinical pain severity: a comparison of six methods. *Pain.* 1986; 27:117–126. [PubMed: 3785962]
15. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. *Pain.* 1990; 40:171–182. [PubMed: 2308763]
16. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain.* 1975; 1:277–279. [PubMed: 1235985]
17. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement properties. *Spine.* 1995; 20:341–352. [PubMed: 7732471]

18. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossell AH, Katz JN. Measurement properties of a self-administered outcome measure in spinal stenosis. *Spine*. 1996; 21:796–803. [PubMed: 8779009]
19. Coyne KS, Margolis MK, Gilchrist KA, Grandy SP, Hiatt WR, Ratchford A, Revicki DA, Weintraub WS, Regensteiner JG. Evaluating effects of method of administration on Walking Impairment Questionnaire. *J Vasc Surg*. 2003; 38:296–304. [PubMed: 12891111]
20. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain*. 1985; 23:345–356. [PubMed: 4088697]
21. Ware JE, Sherbourne CD. The MOS 36-item Short Form Health Survey (MOS SF-36): I. Conceptual framework and item selection. *Medical Care*. 1992; 30:473–481. [PubMed: 1593914]
22. Fleiss, JL.; Levin, B.; Paik, MC. Statistical methods for rates and proportions. 3rd ed.. Hoboken: John Wiley & Sons; 2003.
23. Dodge LD, Bohlman HH, Rhodes RS. Concurrent lumbar spinal stenosis and peripheral vascular disease. A report of nine patients. *Clin Orthop Relat Res*. 1988; 230:141–148. [PubMed: 3365886]
24. Fasih T, Pickin M, Cuschieri RJ. Non-vascular claudication: a clinical conundrum. *Int J Clin Pract*. 2003; 57:87–89. [PubMed: 12661788]
25. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical Versus Nonoperative Treatment for Lumbar Spinal Stenosis Four-Year Results of the Spine Patient Outcomes Research Trial. *Spine*. 2010; 35(14):1329–1338. [PubMed: 20453723]
26. Turner JA, Ersek M, Herron L, et al. Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature. *Spine*. 1992; 17:1–8. [PubMed: 1531550]
27. Katz JN, Lipson SJ, Chang Izzc, et al. Seven to 10 year outcome of decompressive surgery for lumbar spinal stenosis. *Spine*. 1996; 21:92–98. [PubMed: 9122770]
28. Boden S, Davis DO, Dina TS, et al. Abnormal magnetic resonance scans of the lumbar spine in asymptomatic subjects. *J Bone Joint Surg Am*. 1990; 72:403–408. [PubMed: 2312537]
29. Haig AJ, Tong HC, Yamakawa KS, et al. Spinal stenosis, back pain, or no symptoms at all? a masked study comparing radiologic and electrodiagnostic diagnoses to the clinical impression. *Arch Phys Med Rehabil*. 2006; 87(7):897–903. [PubMed: 16813774]
30. Sandella D, Haig AJ, Tomkins C, Yamakawa KSJ. An Innovative Physiatrist Survey to Quantify the Recursive Logical Process of Diagnosing Lumbar Spinal Stenosis. *AAPMR*. 2012 Accepted, PM&R October 2012.
31. American Psychiatric Association. Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition (DSM- IV). American Psychiatric Association; 2000. ISBN 978-0-89042-027-0

Figure.
Subject recruitment and testing

Table 1

Subject recruitment and classification

	Total	Vascular impression and +ABI	Offered surgery for stenosis	Asymptomatic volunteer	Statistic X ² / F stat
Cases screened	403	128	168	107	
Preliminarily qualified	124	21	54	49	23.885
% of screened	30.8%	16.4%	32.1%	45.8%	<0.001**
Contacted and consented	98	15	46	37	2.333
% of qualified	79.0%	71.4%	85.2%	75.5%	0.311
Neurosurgeon exam	94	13	45	36	4.027
% of qualified	75.8%	61.9%	83.3%	73.5%	0.134
Vascular surgeon exam	96	15	44	37	1.043
% of qualified	77.4%	71.4%	81.5%	75.5%	0.594
Study ABI	94	14	44	36	2.051
% of qualified	75.8%	66.7%	81.5%	73.5%	0.359
Both surgeons and ABI done	90	12	43	35	3.896
% of qualified	72.6%	57.1%	79.6%	71.4%	0.143

* Significant at the 0.05 level (2-tailed).

** Significant at the 0.01 level (2-tailed).

† Statistical testing not performed due to expected count less than 5 in 6 cells.

ABI=Ankle Brachial Index, a highly sensitive and specific test for peripheral artery disease.

Table 2

Descriptors of the final study population

	Total	Vascular impression and +ABI	Offered surgery for stenosis	Asymptomatic volunteer	Statistic
In the final population	90	12	43	35	
Age, years	64.7 ±8.3	68.9 ±9.0	66.0 ±8.9	61.7 ±6.1	4.588 0.013*
Sex, % male	49 (54.4%)	8 (66.7%)	26 (60.5%)	15 (42.9%)	3.246 0.197
Race					-/-
Black	7 (7.8%)	1 (8.3%)	4 (9.3%)	2 (5.7%)	
White	79 (87.8%)	11 (91.7%)	39 (90.7%)	29 (82.9%)	
Asian	4 (4.4%)	0 (0.0%)	0 (0.0%)	4 (11.4%)	
BMI (N=68)	30.7±6.1	33.1±9.7	31.7±4.8	29.2±6.3	1.813 0.171
McGill Total	14.5±15.4	17.8±14.6	23.6±14.8	2.3±5.0	31.299 <0.001**
Pain disability index	18.5±18.3	17.3±17.0	30.6±15.6	4.9±10.6	31.625 <0.001**
VAS pain average this week	3.0±2.9	3.0±2.6	5.0±2.3	0.6±1.4	42.429 <0.001**
Quebec Back Pain Disability Scale	29.6±25.3	30.3±15.8	46.3±21.3	7.6±12.1	43.145 <0.001**

* Significant at the 0.05 level (2-tailed).

** Significant at the 0.01 level (2-tailed).

-/- Statistical testing not performed due to expected count less than 5 in 6 cells.

ABI= Ankle Brachial Index, a highly sensitive and specific test for peripheral artery disease.

Table 3

Study surgeon agreement with the patient's unmasked clinical surgeon's diagnosis.

	Clinical Diagnosis as Recruited				Statistic	
		Vascular impression and +ABI	Offered surgery for stenosis	Asymptomatic volunteer	Kappa value	P
Masked Examiner Impression	Neurosurgeon diagnosis N=90	n=12	n=43	n=35	0.761	<0.001
	Vascular claudication	9	1	0		
	Neurogenic claudication	1	34	1		
	Asymptomatic	2	8	34		
	Vascular Surgeon Diagnosis N=84	n=11	n=38	n=35	0.803	<0.001
	Artery disease	9	1	0		
	Spinal stenosis	0	32	2		
	Asymptomatic	2	5	33		

Table 4

Agreement between masked study surgeons on diagnosis

		Vascular surgeon diagnosis			Statistic	
		Vascular claudication n=10	Neurogenic claudication n=34	Asymptomatic n=40	Kappa value	P
Total N=84 Neurosurgeon diagnosis	Vascular claudication	8	1	0	0.717	< 0.001
	Neurogenic claudication	1	26	4		
	Asymptomatic	1	7	36		

Table 5

Patient pain and function as related to disagreement between clinicians.

Total n=54	Agreed neurogenic claudication n=26	Agreed vascular claudication n=8	Disagreed vascular vs. neurogenic* + either vs. 'neither'**, n=20	F stat	P
Patient information					
VAS pain average this week	5.7 ± 2.0	2.6± 2.5	4.1±2.5	5.185	0.009
PDI total score	32.7±14.7	18.0±19.0	26.4±16.9	2.680	0.079
McGill total	26.2±15.2	18.6±17.1	18.2±13.3	1.909	0.159
Quebec Back Pain Disability Scale	51.4±18.1	30.8±15.1	40.0±23.2	3.143	0.053

* An examination of the nature of the 20 disagreements between the two clinicians shows that they disagreed in 2 subjects on the type of claudication, e.g. neurogenic vs. vascular claudication.

** Disagreed either vs. 'neither' includes the following:

Six persons whom the vascular surgeon thought had both types of claudication (either) whom the neurosurgeon thought were neurogenic (5) and vascular (1) subjects. Four persons whom the vascular surgeon felt were asymptomatic ('neither') but the neurosurgeon thought had neurogenic claudication. Eight persons whom the neurosurgeon felt were asymptomatic ('neither') but who were felt to be neurogenic (7) and asymptomatic (1) subjects by the vascular surgeon.

Table 6

Reliability of the clinical examination between the two observers for the population as a whole and for persons thought to have neurogenic claudication.

Agreement on clinical examinations	Neurological and vascular examiners agreement on diagnosis			
	Total Population Agreed + Disagreed N=90 Number (%)	Total Population Agreed N=70 Number (%)	Total Population Disagreed N=20 Number (%)	Agreed Neurogenic Claudication N=26 Number (%)
Strength⁽¹⁾				
Agreed deficit	4 (4.4%)	3 (4.3%)	1 (5.0%)	3 (11.5%)
Agreed no deficit	71 (78.9%)	58 (82.9%)	13 (65.0%)	16 (61.5%)
Disagreed deficit vs. no deficit	15 (16.7%)	9 (12.9%)	6 (30.0%)	7 (26.9%)
Kappa	0.254*	0.331*	0.063	0.295
Reflex⁽²⁾				
Agreed abnormal	11 (12.2%)	9 (12.9%)	2 (10.0%)	8 (30.8%)
Agreed normal	47 (52.2%)	36 (51.4%)	11 (55.0%)	7 (26.9%)
Disagreed abnormal vs. normal	32 (35.5%)	25 (35.7%)	7 (35.0%)	11 (42.3%)
Kappa	0.220*	0.228*	0.186	0.210
Great toe sensation				
Agreed abnormal	0 (0.0%)	0 (0.0%)	3 (15.0%)	2 (7.7%)
Agreed normal	81 (90.0%)	64 (91.4%)	17 (85.0%)	24 (92.3%)
Disagreed abnormal vs. normal	9 (10.0%)	6 (8.6%)	-	-
Kappa	-0.036	-0.040	**	**
Standing lumbar tenderness (N=89, 70, 19, 26)				
Agreed tender (abnormal)	2 (2.2%)	1 (1.4%)	1 (5.3%)	1 (3.8%)
Agreed normal	75 (84.3%)	61 (87.1%)	14 (73.7%)	17 (65.4%)
Disagreed abnormal vs. normal	12 (13.5%)	8 (11.4%)	4 (21.1%)	8 (30.7%)
Kappa	0.187	0.141	0.269	0.019

* Kappa significant at < 0.05.

** Statistic not computed due to at least one variable in a 2-way table is a constant.

⁽¹⁾ Strength: Right and left toe dorsiflexion and plantarflexion, ankle dorsiflexion and plantar flexion, knee extension, and hip flexion and internal rotation.

⁽²⁾ Reflex: Right and left patellar and Achilles.